

Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2013 New J. Phys. 15 123033

(<http://iopscience.iop.org/1367-2630/15/12/123033>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 141.14.132.32

This content was downloaded on 13/01/2014 at 08:00

Please note that [terms and conditions apply](#).

Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar

Igor Ying Zhang^{1,4}, Xinguo Ren^{1,2}, Patrick Rinke¹, Volker Blum^{1,3}
and Matthias Scheffler¹

¹ Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6,
D-14195 Berlin, Germany

² Key Laboratory of Quantum Information, University of Science and
Technology of China, Hefei, 230026 Anhui, People's Republic of China

³ Mechanical Engineering and Materials Science Department and Center for
Materials Genomics, Duke University, Durham, NC 27708, USA

E-mail: zhang@fhi-berlin.mpg.de

New Journal of Physics **15** (2013) 123033 (31pp)

Received 4 September 2013, revised 19 November 2013

Accepted for publication 22 November 2013

Published 19 December 2013

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/15/12/123033

Abstract. We present a series of numerically tabulated atom-centered orbital (NAO) basis sets with valence-correlation consistency (VCC), termed NAO-VCC- nZ . Here the index ' nZ ' refers to the number of basis functions used for the valence shell with $n = 2, 3, 4, 5$. These basis sets are constructed analogous to Dunning's cc-pV nZ , but utilize the more flexible shape of NAOs. Moreover, an additional group of (sp) basis functions, called enhanced minimal basis, is established in NAO-VCC- nZ , increasing the contribution of the s and p functions to achieve the valence-correlation consistency. NAO-VCC- nZ basis sets are generated by minimizing frozen-core random-phase approximation (RPA) total energies of individual atoms from H to Ar. We demonstrate that NAO-VCC- nZ basis sets are suitable for converging electronic total-energy calculations based on valence-only (frozen-core) correlation methods which contain explicit sums over unoccupied states (e.g. the RPA or second-order Møller–Plesset

⁴ Author to whom any correspondence should be addressed.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

perturbation theory). The basis set incompleteness error, including the basis set superposition error, can be gradually reduced with the increase of the index ‘ n ’, and can be removed using two-point extrapolation schemes.

S Online supplementary data available from stacks.iop.org/NJP/15/123033/mmedia

Contents

1. Introduction	2
2. Basis set extrapolation	6
3. Numerical basis set with valence-correlation consistency	7
3.1. Construction	7
3.2. Basis set convergence for total energy	12
3.3. Basis set convergence for energy differences	15
3.4. Role of the confining potential	17
4. Results and discussion	18
4.1. Homonuclear diatomic molecules from H ₂ to Ar ₂ , HDM18	18
4.2. Reaction energies, G2RC	20
4.3. Isomerization energies, ISO34	22
4.4. Non-covalent interactions, S22	23
5. Conclusions and outlook	24
Appendix. Accuracy of numerical integrations	25
References	29

1. Introduction

Density-functional theory (DFT) is currently the most widely used method for first-principles electronic-structure calculations. The random-phase approximation (RPA) [1–5] and methods related to it [1, 6–15] are state-of-the-art density-functional approximations (DFAs) that address some notorious deficiencies of conventional DFAs in, for example, describing non-covalent interactions [11, 13, 14, 16, 17], reaction barrier heights, [11, 13, 18], surface adsorbates [4, 19] and the f-electron metal cerium [20].

Within the adiabatic-connection fluctuation–dissipation theorem, the RPA correlation energy E_c^{RPA} is expressed as [11]

$$E_c^{\text{RPA}} = \int_0^\infty \frac{du}{2\pi} \text{Tr}[\ln(1 - \chi_0(iu)v) + \chi_0(iu)v]. \quad (1)$$

Here $v = 1/|r - r'|$ is the bare Coulomb interaction kernel. $\chi_0(\mathbf{r}, \mathbf{r}'; iu)$ is the Kohn–Sham (KS) independent-particle density response in imaginary frequency

$$\chi_0(\mathbf{r}, \mathbf{r}'; iu) = \sum_\sigma \sum_{n,m} \frac{(f_{n,\sigma} - f_{m,\sigma})}{\epsilon_{n,\sigma} - \epsilon_{m,\sigma} - iu} \phi_{n,\sigma}^*(\mathbf{r}) \phi_{m,\sigma}^*(\mathbf{r}') \phi_{m,\sigma}(\mathbf{r}) \phi_{n,\sigma}(\mathbf{r}'), \quad (2)$$

where $\phi_{n,\sigma}$ and $\epsilon_{n,\sigma}$ are the KS eigenfunctions and eigenenergies with Fermi occupation numbers $f_{n,\sigma}$. Unfortunately, the RPA suffers heavily from its well-known slow convergence

with basis set size [7, 12, 21–23], a trait which is common to all explicit-correlation methods (e.g. second-order Møller–Plesset perturbation theory (MP2), or the coupled-cluster methods). In physical terms, this slow convergence is due to the difficulty of using smooth orbital product expansions to describe the electron–electron Coulomb cusps in real space [24, 25]. In this work, we assess the performance of different types of atom-centered-orbital basis sets in RPA calculations, which include the widely used analytical Gaussian-type orbital (GTO) and the flexible numeric atom-centered orbital (NAO) basis functions.

Over the years, several methods have been pursued that aim to address the slow convergence of the correlation energy, including:

- Explicitly correlated wave function methods (i.e. R12 and F12) [24, 26–28], the transcorrelated method [29, 30], diffusion quantum Monte Carlo [31–33], density functional perturbation theory together with plane-wave basis sets [34, 35] and Nakatsuji’s local Schrödinger equation method [36]. All these methods attempt to accelerate the convergence behavior directly by addressing the electron–electron cusp.
- Complete-basis-set (CBS) extrapolation schemes, which are currently the most popular method to eliminate the basis set incompleteness error in quantum chemistry. These CBS extrapolation schemes are based on standardized basis sets. These basis sets provide a step-wise systematic convergence of the basis set incompleteness error, which enables an analytic extrapolation to the CBS limit particularly for the terms containing unoccupied-state sums [6, 7, 21, 22, 37–41].
- CBS extrapolation in plane-wave basis sets. Plane-wave basis sets are another popular choice for electronic-structure calculations, because they provide an intrinsically and systematically improvable resolution of v in real or reciprocal space by increasing the momentum cutoff parameters. Plane waves are particularly suitable when combined with ‘pseudoization’ strategies that remove the atomic nuclei and core electrons from the explicit parts of the calculation. The plane-wave basis set error convergence of the RPA correlation energy has been investigated by Harl *et al* [16] and Björkman *et al* [42]. A recent development by Shepherd *et al* [40] suggests that, by introducing a new kind of energy cutoff based on the momentum transfer vector, it is possible to obtain accurate extrapolated correlation energies with lower computational cost. In the closely related GW method [43–46] the completeness relation is used to collapse the sum over unoccupied states [47, 48].

The most widely used basis sets in quantum chemistry for CBS extrapolations are Dunning’s (augmented) polarized valence-correlation consistent GTO basis sets (aug)-cc-pVnZ [49–51], with $n = 2, 3, 4, 5$ and higher. In early 2001, Furche investigated the basis set dependence of the frozen-core RPA method for the atomization energy of N_2 using the cc-pVnZ basis sets [22]. Comprehensive studies were carried out in recent years by Eshuis *et al* [21], reporting accurate benchmark data of the frozen-core RPA method, including various covalent and non-covalent interactions using the two-point extrapolation scheme from quadruple- ζ (4Z) and quintuple- ζ (5Z) together with Dunning’s (aug)-cc-pVnZ basis sets. Fabiano *et al* [37] investigated the basis set convergence behavior of all-electron RPA calculations together with Dunning’s correlation consistent GTO basis sets augmented with core and core–valence basis functions (VnZcv with $n = 4, 5, 6$ and 7) for H, C, N, O, F and Ne. It was found that the all-electron RPA method suffers from serious basis set incompleteness errors. Neither the extrapolation from 4Z and 5Z nor bare septuple- ζ (7Z) calculations without extrapolations suffice to provide

converged results for the all-electron RPA method. It was suggested that accurate results can be achieved by extrapolating from 5Z and sextuple- ζ (6Z), if globally optimized two-point extrapolation schemes are employed. These observations reveal that the valence and core correlation contributions converge differently in RPA calculations, indicating that the basis set parts accounting for core and valence correlation can be constructed separately [52, 53].

We here explore the use of NAO basis functions of the form

$$\varphi_i(\mathbf{r}) = \frac{u_i(r)}{r} Y_{lm}(\Omega) \quad (3)$$

as a recipe to obtain an efficient, systematic ‘correlation consistent’ convergence of the unoccupied-state sums in equation (2) for RPA, MP2 and related explicitly correlated approaches. Here, $u_i(r)$ is a radial function and $Y_{lm}(\Omega)$ denotes the real parts ($m = 0, \dots, l$) and imaginary parts ($m = -l, \dots, -1$) of complex spherical harmonics. Analytic prescriptions such as Slater-type orbitals (STOs) [54, 55] or GTOs (including the Dunning basis sets) [49–51, 53] are obviously consistent with this shape, but the radial function $u_i(r)$ may also be chosen independently of any analytical restriction [56–62].

Previous investigations [57, 58] have shown that CBS total energies can be obtained with NAO basis sets for conventional DFAs and HF. For instance, efficient, hierarchical basis sets for DFT calculations for elements 1–102 were developed in 2009 within the Fritz Haber Institute ‘*ab initio* molecular simulations’ (FHI-aims) electronic-structure package [57], FHI-aims-2009 for short. The radial functions of these basis sets are organized in groups (so-called ‘tiers’) of functions in addition to a minimal basis of the occupied orbitals (core and valence) of spherically symmetric free atoms. The additional ‘tiers’ ($n = 1–4$) were found by a step-wise minimization of the LDA total energies of symmetric dimers for all elements from light to heavy. The FHI-aims-2009 basis sets were found to be accurate and transferable for the *total energies* of all local, semilocal and hybrid DFAs and the HF method [57, 63].

However, the unoccupied state sums of equation (2) present a different problem: Obtaining systematic convergence for a spectrum of unoccupied states, the number of which is not bounded by the number of electrons in the system but rather by the spatial resolution needed to approximate the effect of the electron–electron cusp in a systematic way. Ren *et al* [63] have investigated the convergence properties of the FHI-aims-2009 basis sets for explicit-correlation methods such as RPA and MP2. In essence, the convergence of energy differences (e.g. binding energies) is satisfactory, but only if a counterpoise (CP) correction scheme is applied.

In this paper, we develop a new sequence of NAO basis sets with valence-correlation consistency for light elements (H–Ar), in order to extend the reach of this prescription to absolute total energies for explicitly correlated methods and to CBS extrapolation schemes for cases where a direct CP correction is not practical (e.g., energy differences between very different conformations of a single molecule). As outlined below and in the [appendix](#), using NAOs for this purpose offers both physical and numerical advantages in (i) the radial function shape near the nucleus, (ii) their tails toward large distances and (iii) the fact that their spatial extent can be controlled by a single criterion (smooth cutoff distance, which can be large) in a practical calculation. We stress the fact that these new NAO basis sets are not intended to supersede the FHI-aims-2009 basis sets, but instead serve as an independent complement for methods which require a direct treatment of the electron–electron cusp by way of unoccupied-state sums.

Dunning analyzed the contribution of basis functions to the correlation energy on the theoretical level of the frozen-core configuration interaction method with single and double

excitations (CISD) [49]. He found that the first d-type GTO function has the largest effect on the correlation energy, but the second d-type GTO function has approximately the same effect as adding the first f-type GTO function. Likewise, the third d-type, second f-type and first g-type GTO function compose the third grouping. Based on these observations, Dunning generated a series of basis sets by introducing batches of basis functions that give a (approximately) consistent amount of correlation energy from large to small. These GTO basis sets are called ‘correlation consistent’. In this work, we are interested in whether this strategy also works for the NAO basis functions.

Correlation consistent basis sets can provide energies and other ground-state and excited-state properties that converge smoothly toward the CBS limit, especially for explicit-correlation methods, such as RPA, MP2 and CCSD(T). This observation has resulted in many empirical extrapolation formulae that can be used to estimate the CBS limit [64–66]. Using the He ground state energy, Hill [67] found that the energy increments due to partial waves of a given angular momentum number l in a CI calculation are proportional to $(l + \frac{1}{2})^{-4}$, which immediately yields an inverse relation between the correlation energy and the highest angular momentum l_{\max} in a specific basis set [68–70]:

$$E_{l_{\max}}^{\text{cor}} = E_{\infty}^{\text{cor}} + A/(l_{\max} + 1)^3 + O[(l_{\max} + 1)^{-4}]. \quad (4)$$

This relation can be used to extrapolate the energy for atoms, but can only be used approximately for molecules. More importantly, Klopper *et al* [70] argued that this relation is not fulfilled until the function spaces with lower l than l_{\max} are saturated. Dunning’s ‘correlation consistent’ strategy introduces a group of s and p basis functions on top of the minimal basis (named the (sp) correlation set) to facilitate the convergence of the valence correlation energy. Generally, these basis functions mainly affect the description of the valence electrons, and are used to define the index ‘ n ’ of Dunning’s basis sets, i.e. the cc-pVnZ basis sets contain n s (and n p) basis functions for the valence shell (1 from the minimal basis, and $n - 1$ from the (sp) correlation set). The s and p function spaces in cc-pVnZ basis sets have been saturated further in several subsequent studies [50, 52, 53, 71]. Core-correlation consistency requires augmented basis functions with much more compact radial shape, as, e.g. in cc-pCVnZ [53] or cc-pwCVnZ [52]. The calculation of electronic affinities and other properties associated with anions requires augmented diffuse basis functions, i.e. aug-cc-pVnZ [50]. Ahlrichs and co-workers proposed a family of def2 basis sets [71], that have become one of the default choices in the TURBOMOLE program (see footnote 7) [71, 72]. For the first- and second-row elements, the def2 basis sets use the same polarization sets as Dunning’s cc-pVnZ basis sets, but the former generally contains more s and p functions than the latter. These s and p functions were determined by minimizing the HF energies of each element. Distinct from these works, in this paper we investigate the influence of saturating s and p function spaces for valence-correlation consistency. We propose an additional group of sp functions, named enhanced minimal basis, which greatly improves the valence correlation convergence (VCC) behavior of our ‘correlation consistent’ basis sets (see section 3 for details).

Dunning’s ‘correlation consistent’ GTO basis sets can be incorporated into the NAO framework of FHI-aims [57] and other NAO codes without any additional implementation effort. However, our investigations show that these GTO basis sets require more expensive real-space integration grids to achieve a numerical integration accuracy that is comparable to that of the NAO functions in FHI-aims. As shown in the [appendix](#), this is in fact not only just a numerical but also a physical problem of any contracted GTO function with very high exponents

near the nucleus. Regardless of the integration method used, the actual integrand in expressions such as

$$\int d^3r \varphi_i(\mathbf{r}) \hat{H}_{\text{scf}} \varphi_i(\mathbf{r}) \quad (5)$$

(\hat{H}_{scf} denotes the self-consistent field Hamiltonian of standard or generalized Kohn–Sham theory and φ_i denotes a basis function) may become unphysical very close to the nucleus due to the inclusion of very sharp primitive GTOs. In contrast, NAO basis functions do not have this problem as one can use the exact occupied orbitals of free atoms as the minimal basis. Similar numerical integration difficulties also occur for GTOs far from the nucleus, where the analytical tails of diffuse GTOs decay slower than those of the NAO functions. Here, it proves extremely convenient that NAO functions can be rigorously localized by a confining potential [57]. Even if the confinement radius is large, the relevant integration volume is still under direct and unambiguous control (see the [appendix](#) for a detailed discussion).

In this work, we demonstrate the advantage of NAO functions for compact basis sets with valence-correlation consistency. As a first step, we focus on the valence electrons and ground-state total energies only. The description of core and core–valence correlation requires significantly different shapes of basis functions [52, 53] and will be discussed in future work.

2. Basis set extrapolation

As mentioned above, the slow convergence behavior of RPA, MP2 and CCSD(T) total energies with basis set size is often addressed by basis set extrapolation schemes [64–66]. Since we face the same underlying problem (the two-electron cusp), we show below that our valence-correlation consistent NAO total energies can be successfully extrapolated as well. Therefore, we first introduce the two extrapolation strategies to be used in this work.

The first one is the simplest but popular two-point extrapolation scheme [69]

$$\begin{aligned} E[n] &= E[\infty] + A/n^3, \\ E[\infty] &= \frac{E[n_1]n_1^3 - E[n_2]n_2^3}{n_1^3 - n_2^3}, \end{aligned} \quad (6)$$

where ‘ n_1 ’ and ‘ n_2 ’ are the indices of the valence-correlation consistent basis sets. This $1/n^3$ formula was originally proposed for the correlation energy [68, 69] but was also used directly for the total energy [38, 73]. Taking frozen-core CCSD(T) atomization energies of 51 small molecules, Feller *et al* [38] suggested that it is not necessary to extrapolate the correlation energies separately, because the error in the rest is significantly smaller than the error in the correlation component. Considering that the converged total energies can be easily obtained with NAO basis sets for conventional DFAs and HF [57], this two-point extrapolation is used only for total energies. We denote this scheme CBS-TE. Eshuis and Furche [21] stated that they use this two-point formula to extrapolate frozen-core RPA correlation energies. However, they did not say explicitly how to determine the CBS limit of the remaining components. With the same (aug)-cc-pVnZ basis sets, we can reproduce their results using the CBS-TE scheme.

In order to fully exploit the power of correlation consistent basis sets for frozen-core RPA calculations, we also consider another popular two-point extrapolation scheme with two global

optimized parameters α and d , [68, 69] denoted CBS-OPT:

$$\begin{aligned} E[n] &= E[\infty] + A/(n+d)^\alpha, \\ E[\infty] &= \frac{E[n_1](n_1+d)^\alpha - E[n_2](n_2+d)^\alpha}{(n_1+d)^\alpha - (n_2+d)^\alpha}. \end{aligned} \quad (7)$$

In previous investigations [37, 38, 41, 68, 69], the parameters α and d were separately optimized for $n_1 = 2$, $n_2 = 3$ (CBS-OPT[23]), $n_1 = 3$, $n_2 = 4$ (CBS-OPT[34]) and $n_1 = 4$, $n_2 = 5$ (CBS-OPT[45]). However, in this work, we prefer to use the same global parameters for each level of extrapolation, since a sequence of correlation consistent basis sets should provide consistent convergence for the basis set incompleteness error. We will determine the global parameters for (aug)-cc-pVnZ and our new NAO basis set in section 4.1.

3. Numerical basis set with valence-correlation consistency

3.1. Construction

The sequence of basis sets, constructed in this work, are composed of a varying number of primitive NAO basis functions $\{\varphi_i(\mathbf{r})\}$ in the form of equation (3). These functions with spherical harmonics include the radial terms $\{u_i(r)\}$ generated by the exact solutions of radial Schrödinger equations for different radial potentials $\{v_i(r)\}$ within a confining potential $v_{\text{cut}}(r)$:

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r) \right] u_i(r) = \epsilon_i u_i(r), \quad (8)$$

where l denotes the angular momentum quantum number related to the spherical harmonics $Y_{lm}(\Omega)$. In this work, we choose the default $v_{\text{cut}}(r)$ used in FHI-aims [57] to confine the radial extent, which is turned on smoothly at a distance from the nucleus (called the onset radius r_{onset}), and reaches infinity at $r_{\text{onset}} + w$ ($w = 2.0 \text{ \AA}$ throughout this work). While the confining potential $v_{\text{cut}}(r)$ could also be used to provide extra basis flexibility at a given basis size [59, 74, 75], we use it as a purely technical quantity, which should not impact any physical results. Because we only focus on the 18 light elements from H to Ar, our experience in this work (see section 3.4 below) suggests that $r_{\text{onset}} = 4 \text{ \AA}$ suffices to provide well-converged results for most of these elements. However, the basis set optimization for the metal elements (Li, Na, Mg and Al) is apt to include more diffuse basis functions, which exhibit a slower convergence with respect to the confining potential. Therefore, a slightly larger value ($r_{\text{onset}} = 5 \text{ \AA}$) is required for these metal elements.

The FHI-aims-2009 basis sets take the self-consistent free-atom radial potential $v_i(r) \equiv v_{\text{free atom}}(r)$ as default to generate the radial functions $\{u_i(r)\}$ used in the minimal basis, which consists of the occupied orbitals from a non-spinpolarized, spherically symmetric free atom calculation. This minimal basis naturally captures the wave function oscillations near the nucleus not only in bare atoms but also in bonded structures [57] (because the nuclear Coulomb singularity here dominates the potential). We therefore select this minimal basis as the starting point of our new basis set. As mentioned above, this minimal basis is less demanding on the radial grids than the GTO minimal basis of Dunning's correlation consistent basis set (see also the appendix).

In addition to the minimal basis, our new basis sets include three other subsets named polarization set, (sp) correlation set and (sp) enhanced minimal basis, where (sp) indicates that

Table 1. The character of the radial functions included in the NAO-VCC- nZ ($n = 2, 3, 4, 5$) basis sets for the first-row elements. In brackets: number of radial functions of each angular momentum channel. In addition, each basis set includes the minimal basis set of numerically computed radial functions of occupied free atoms.

	(sp) Set		
	Enhanced ^a	Correlation	Polarization set ^b
NAO-VCC-2Z	(2s 1p)	(1s 1p)	(1d)
NAO-VCC-3Z	(2s 1p)	(2s 2p)	(2d 1f)
NAO-VCC-4Z	(2s 1p)	(3s 3p)	(3d 2f 1g)
NAO-VCC-5Z	(2s 1p)	(4s 4p)	(4d 3f 2g 1h)

^a Enhanced minimal basis: (1s) for H and He, (2s1p) for Li–Ne and (3s2p) for Na–Ar.

^b Second-row elements have the same polarization set. However, for H and He, one polarization function is removed in each polarization group for each l , i.e. *none* for 2Z, (1d) for 3Z, (2d 1f) for 4Z, and (3d 2f 1g) for 5Z.

only s and p type orbitals are used. We will establish them in the following paragraphs. All these subsets employ only radial functions $u_i(r)$ constructed from hydrogen-like atoms. We use the notation ‘H(l, z_i)’ to define each of these hydrogen (‘H’)-like functions unambiguously in the context of equation (8). Here, l is the chosen angular momentum channel. The radial functions are generated by hydrogen-like radial potentials $v_i(r) = z_i/r$ (see equation (8)). For a unique definition of $u_i(r)$, it remains to specify the principal quantum number n . n is always chosen to be the minimal one, $n = l + 1$, leading to nodeless radial functions similar to Slater-type orbitals [54, 55]. Each radial function thus defined leads to $2l+1$ basis functions via equation (3). z_i acts as an *a priori* free parameter that controls the extent of the resulting radial functions $u_i(r)$, analogous to the role of the exponent in a GTO. By a suitable choice of z_i for individual subsets of radial functions $\{H(l, z_i)\}$, we can emulate the even-tempered relation [76]

$$z_i = \alpha\beta^{i-1} \quad i = 1, \dots, N_{\text{orb}} \quad (9)$$

to expand the nuclear charges $\{z_i\}$ of hydrogen-like orbitals which belong to the same subset of N_{orb} individual radial functions. (α, β) are the optimization parameters used to expand the z_i within individual subsets $\{H(l, z_i)\}$. The number of such optimization parameters increases when new basis functions with new l are introduced in each subset.

Table 1 summarizes which subsets of radial functions and how many individual radial functions per angular momentum channel l are used to define our new series of basis sets. We follow the concept of *valence-correlation consistency* proposed by Dunning [49] to construct the polarization set by adding numeric hydrogen-like orbitals with increasing l as the basis set quality increases. We call our NAO basis sets with valence-correlation consistency ‘NAO-VCC- nZ ’ ($n = 2, 3, 4, 5$). The index ‘ n ’ equals the highest angular momentum number l_{max} in the specific basis sets for the first- and second-row elements. Taking the first-row elements, the last column in table 1 (‘polarization set’) lists the polarization functions for each index of the basis set. NAO-VCC-2Z utilizes only one d-type hydrogen-like orbital H(d, z_i) in its polarization set, labeled as (1d) (i.e. $N_{\text{orb}} = 1, l = 2$ in this case). However, the size of the polarization set increases quickly. At the highest quality level (NAO-VCC-5Z), we then have 4 H(d, z_i), 3 H(f, z_i),

2 $H(g, z_i)$ and 1 $H(h, z_i)$ hydrogen-like orbitals (4d 3f 2g 1h). We note here that the second-row elements share the form of the polarization set with the first-row elements. However, for H and He, the index ‘ n ’ is not l_{\max} but $l_{\max} - 1$, which again emphasizes the empirical nature of the extrapolation schemes used for the CBS limit.

In order to guarantee VCC, the continued saturation in (sp) basis function space is equally important. According to Dunning’s strategy, we therefore also build (sp) correlation sets of increasing size. As discussed in the introduction, the index ‘ n ’ of Dunning’s (aug)-cc-pVnZ sets was defined as the basis functions used for the valence shell, counted by the ‘ $n - 1$ ’ s (or p) functions in the (sp) correlation sets plus the other one from the minimal basis. This definition of ‘ n ’ is also valid for our new NAO-VCC- n Z basis sets. Like in the polarization set, we group basis functions with the same angular momentum l in the (sp) correlation set, e.g. (4s) or (4p) in NAO-VCC-5Z, following the even-tempered relation. The parameters (α , β) are different for each element, and optimized separately for double- ζ (2Z) to quintuple- ζ (5Z), respectively.

Further investigation in this work reveals that it is beneficial to introduce additional s- and p-type basis functions on top of the (sp) correlation set. We address this point in more detail in the next subsection, observing that these (sp) type functions are especially important to guarantee VCC in small basis sets. The result is that our basis sets generally include what we call a ‘(sp) enhanced part’ of the basis set (see table 1), containing 1 $H(s, z_i)$ for H and He (1s); 2 $H(s, z_i)$ and 1 $H(p, z_i)$ for the first-row elements (2s1p); and 3 $H(s, z_i)$ and 2 $H(p, z_i)$ for the second-row elements (3s2p). Because the basis sizes are the same as those of the minimal basis, we also call this (sp) enhanced set ‘enhanced minimal basis’. We again employ the even-tempered relation to expand the basis functions with the same angular momentum for the enhanced minimal basis. The corresponding parameters are optimized for each level of our basis sets. Concerning the optimization procedure, we first introduce and optimize the (sp) correlation set based on the minimal basis. Then we add the (sp) enhanced set and optimize parameters (α, β) for that, and finally the same for the polarization set.

Regarding the optimization target, we select the total energies of spherically symmetric free atoms from H to Ar in the frozen-core RPA based on the orbitals from a self-consistent Perdew–Burke–Ernzerhof [77] (PBE) DFT calculation. At the DFT-PBE level, all orbitals (core and valence) are self-consistently determined. Frozen-core RPA here implies that the index i in equation (2) only runs over occupied valence orbitals. We denote this RPA procedure RPA@PBE in the following. Unless otherwise stated, frozen-core RPA@PBE calculations are carried out for the rest of the discussion.

The parameters (α, β) are optimized to minimize the RPA@PBE total energies for each element. The optimization is carried out with a Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [78] based on three-point fitted numerical gradients. The corresponding iterations are terminated if the energy deviation in each BFGS step falls below 10^{-6} eV. In order to reach the global minimum as best as we can, a set of initial guesses is generated globally in well-chosen parameter windows with tabulated grids, i.e. $0.0 < \alpha < 50.0$ and $1.0 < \beta < 3.0$.

As an illustration, table 2 reports the optimized parameters for the carbon atom. In the development of correlation consistent GTO basis sets [49], it was found that the minimal exponent of GTOs with the same angular momentum in the polarization set decreases. Meanwhile, the maximal exponents increase with increasing index. The equivalent parameter in hydrogen-like orbitals is the nuclear charge $\{z_i\}$. The aforementioned observation also partially holds for the NAO basis sets, where the maximal nuclear charges z_{\max} for the same angular momentum always grow monotonically from NAO-VCC-2Z to 5Z. Taking $H(d, z_{\max})$

Table 2. Optimized parameters (α , β) and expanded nuclear charges of hydrogen-like orbitals for the (sp) correlation set, the polarization set, and the enhanced minimal basis in the NAO-VCC- nZ ($X = 2, 3, 4, 5$) basis sets for the carbon atom. The NAO-VCC- nZ basis sets employ the same minimal basis for all indices ‘ n ’. The minimal basis is not shown in the table.

NAO-VCC-2Z			NAO-VCC-3Z		
(sp) Correlation	(α, β)	{ z_i }	(sp) Correlation	(α, β)	{ z_i }
(1s)	(1.827,1.000)	1.827	(2s)	(1.807,1.804)	1.807,3.260
(1p)	(2.322,1.000)	2.322	(2p)	(1.974,1.604)	1.974,3.166
Polarization			Polarization		
(1d)	(5.952,1.000)	5.952	(2d)	(5.891,1.134)	5.891,6.677
			(1f)	(10.150,1.000)	10.150
(sp) Enhanced			(sp) Enhanced		
(2s)	(2.354,1.177)	2.354,2.986	(2s)	(6.773,2.046)	6.773,13.859
(1p)	(3.188,1.000)	3.188	(1p)	(3.647,1.000)	3.647
NAO-VCC-4Z			NAO-VCC-5Z		
(sp) Correlation	(α, β)	{ z_i }	(sp) Correlation	(α, β)	{ z_i }
(3s)	(1.596,2.024)	1.567,3.230,6.538	(4s)	(1.623,1.989)	1.623,3.227,6.419,12.768
(3p)	(1.887,2.416)	1.887,4.558,11.011	(4p)	(2.417,1.698)	2.417,4.104,6.969,11.834
Polarization			Polarization		
(3d)	(5.231,1.463)	5.231,7.652,11.196	(4d)	(5.111,1.343)	5.111,6.861,9.211,12.366
(2f)	(10.645,1.015)	10.645,10.800	(3f)	(10.518,1.094)	10.518,11.504,12.583
(1g)	(15.741,1.000)	15.741	(2g)	(16.246,1.034)	16.246,16.792
			(1g)	(22.554,1.000)	22.554
(sp) Enhanced			(sp) Enhanced		
(2s)	(13.753,1.880)	13.753,25.858	(2s)	(15.611,1.900)	15.611,29.662
(1p)	(10.266,1.000)	10.266	(1p)	(10.714,1.000)	10.714

for example, z_{\max} increases from 5.952 (NAO-VCC-2Z), 6.677 (3Z), 11.196 (4Z), to 12.366 (5Z). While the minimal nuclear charges z_{\min} are going down simultaneously for H(d, z_{\min}), this trend is not completely fulfilled for H(f, z_{\min}) and H(g, z_{\min}).

The enhanced minimal basis in this work is originally designed to saturate the (sp) basis functions on top of the (sp) correlation set. Table 2 indicates that the optimization is generally apt to produce more compact basis functions with larger nuclear charges, which demonstrates the importance of saturating the (sp) inner space for describing valence correlation. For the other elements up to the second row, the NAO-VCC- nZ basis sets are listed in the supporting information⁵.

Figure 1 plots the averaged basis sizes of (aug)-cc-pVnZ, NAO-VCC- nZ and FHI-aims-2009. For most of these basis sets, the average counts 18 elements from H to Ar. However, the averaged basis size of aug-cc-pV5Z only covers 14 elements, because the basis sets are unavailable for the four metal elements (Li, Be, Na and Mg). In FHI-aims-2009, no tier-4 basis set was optimized for H, He, Li, Ne and Ar, because the tier-3 basis sets are sufficient to provide

⁵ The supplementary information lists the NAO-VCC- nZ basis sets for all elements of the periodic table up to the second row (see the supplementary data, available from stacks.iop.org/NJP/15/123033/mmedia).

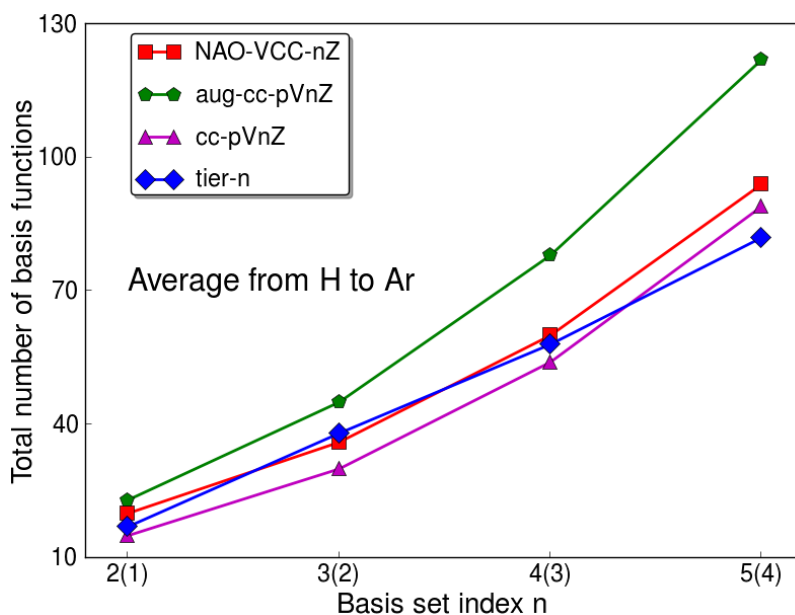


Figure 1. The averaged basis sizes of four atom-centered orbital basis sets from H to Ar. Tier- n denotes different tiers of the FHI-aims-2009 basis sets. On the x -axis, $n = 2, 3, 4, 5$ we list the indices of (aug)-cc-pV n Z and NAO-VCC- n Z, while the number $n = 1, 2, 3, 4$ given in parentheses are those of FHI-aims-2009 tier- n .

converged LDA total energies for symmetric dimers of these five elements. The averaged basis size of tier-4, therefore, takes the remaining 13 elements into account. Unless otherwise stated, the smaller basis sets will be used automatically if ‘(aug)-cc-pV5Z’ and ‘tier-4’ are unavailable. As shown in figure 1, the basis sizes of NAO-VCC- n Z are slightly larger than those of cc-pV n Z due to the (sp) enhanced minimal basis. Both are significantly more compact than the aug-cc-pV n Z sequence, which augments cc-pV n Z with a set of diffuse functions. The basis size of tier-3 is still similar to that of NAO-VCC-4Z and cc-pV4Z, but one level up tier-4 becomes the smallest among the four atom-centered orbital basis sets.

With the aid of the resolution-of-identity (RI) technique, the most expensive step of the RPA method is the construction of the independent-particle response function. The number of required operations is proportional to $N_{\text{aux}}^2 N_{\text{occ}} N_{\text{vir}}$, where N_{aux} corresponds to the number of auxiliary basis functions used to expand the response function, and (N_{occ} and N_{vir}) are the numbers of the occupied and virtual single-particle states, respectively. The memory scaling is dominated by the storage of three-center integrals, which is proportional to $N_{\text{aux}} N_{\text{bs}}^2$, where N_{bs} is the basis size of a specific basis set (see [11, 12] for a detailed discussion). For a given system, for which N_{occ} is then a constant, both $N_{\text{vir}} = N_{\text{bs}} - N_{\text{occ}}$ and $N_{\text{aux}} \ll N_{\text{occ}}(N_{\text{bs}} - N_{\text{occ}})$ are proportional to the basis size N_{bs} , resulting in a cubic scaling of the RPA method in terms of N_{bs} ($O(N_{\text{bs}}^3)$) for both time and memory. Our discussion of figure 1 demonstrates that our new NAO-VCC- n Z basis sets are more compact (i.e. smaller) than the Gaussian orbital aug-cc-pV n Z sequence while retaining the accuracy of aug-cc-pV n Z (see section 4). The NAO-VCC sets therefore facilitate a significant computational speed-up that is comparable to the one of switching from aug-cc-pV n Z to cc-pV n Z.

Table 3. RPA@PBE absolute energies of the carbon atom calculated with various basis sets (in eV).

Carbon	2Z	3Z	4Z	5Z	6Z	CBS-TE[45] ^a	CBS-TE[56] ^b
NCC ^c	-1029.5673	-1030.1646	-1030.4294	-1030.5685		-1030.7145	
ACC ^c	-1029.2036	-1030.1261	-1030.4769	-1030.6031	-1030.6527	-1030.7355	-1030.7208
CC ^c	-1029.0505	-1030.0515	-1030.4373	-1030.5780	-1030.6387	-1030.7255	-1030.7221

^a Two-point total-energy extrapolation ($1/n^3$) from quadrupole- ζ (4Z) to quintuple- ζ (5Z) basis sets.

^b Two-point total-energy extrapolation ($1/n^3$) from quintuple- ζ (5Z) to sextuple- ζ (6Z) basis sets.

^c NCC is the abbreviation for NAO-VCC- nZ , ACC for aug-cc-pV nZ and CC for cc-pV nZ .

3.2. Basis set convergence for total energy

As a particular example, we first examine the performance of NAO-VCC- nZ for the carbon atom shown in table 3. The calculated RPA@PBE total energies become lower monotonically for all three sequences of basis sets with valence-correlation consistency. However, comparing to the extrapolated values, the basis set incompleteness errors in quintuple- ζ (5Z) are still larger than 120 meV. Table 3 reveals that, while diffuse functions help to notably lower the total energies for finite basis sizes, the contribution seems to be negligible when extrapolating to the CBS limit. The CBS value extrapolated from NAO-VCC-4Z and 5Z (CBS-TE[45]) is -1030.7145 eV, which is about 21 meV higher than the extrapolated value from aug-cc-pV4Z and 5Z. However, the discrepancy reduces to only 6 meV by taking the extrapolated value from aug-cc-pV5Z and 6Z (CBS-TE[56]) as the reference, validating the power of NAO-VCC- nZ basis sets to describe valence correlation for explicit-correlation methods.

We further show the convergence behavior of four types of atom-centered orbital basis sets for several first- and second-row elements. In figure 2, the basis set errors of RPA@PBE ground-state total energies are plotted for increasing basis size. Figure 3 presents the same data logarithmically to better illustrate the convergence behavior. The cc-pV nZ basis sets show a consistent convergence for all elements examined here. Adding a set of diffuse functions to cc-pV nZ does not change the convergence behavior considerably for aug-cc-pV nZ .

The FHI-aims-2009 basis sets were established using the ground-state LDA total energy of symmetric dimers. The minimal basis of the NAO basis sets is by construction exact for a spherically symmetric, non-spin polarized free atom and a given DFT functional [57]. Thus, total energies for free atoms of any kind are already very close to converged when using the same DFT functional. Beyond the minimal basis, the ‘tiers’ in FHI-aims-2009 were by construction optimized to capture the DFT total energy contributions arising from the bonding environment. However, in RPA, the absolute energies of free atoms are not yet converged (unlike in the DFT case). As a result, NAO basis sets constructed based on semilocal or hybrid DFT do not necessarily provide the flexibility to converge atomic total energies in RPA, and this is exactly what is shown in figures 2 and 3.

The NAO-VCC- nZ basis sets provide a consistent convergence behavior for these elements. As shown in the lower right panel in both figures 2 and 3, the convergence curves almost sit on top of each other for the elements in the same periodic table groups, which have significantly different total and RPA correlation energies. For example, the total and the RPA correlation energies calculated with NAO-VCC-5Z are -2004.204 and -8.427 eV for the oxygen atom, but $-10\,824.019$ and -7.331 eV for the sulfur atom, respectively. Reviewing the performance

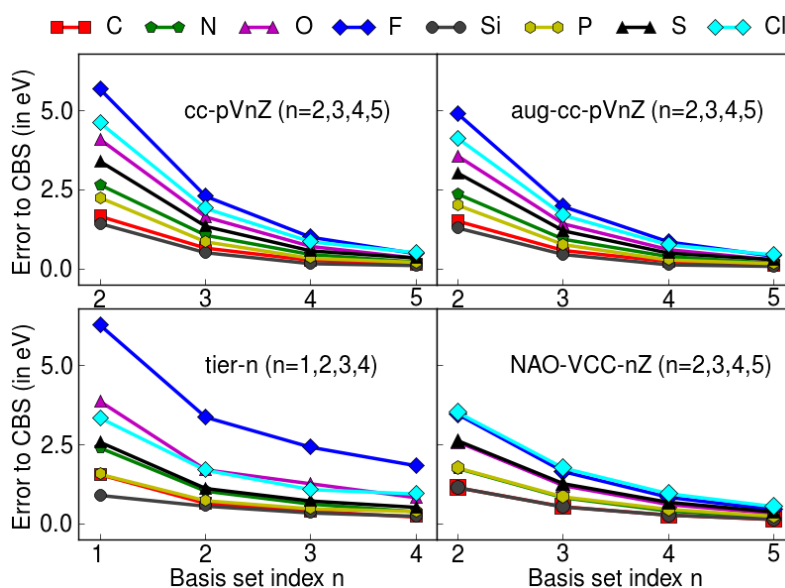


Figure 2. Basis set errors of RPA@PBE total energies for several first- and second-row atoms. As references, we take the complete-basis-set (CBS) values, extrapolated from quintuple- ζ (5Z) and sextuple- ζ (6Z) in the sequence of aug-cc-pVnZ, CBS-TE[56]/aug-cc-pVnZ. Tier- n denotes different tiers of the FHI-aims-2009 basis sets (in eV).

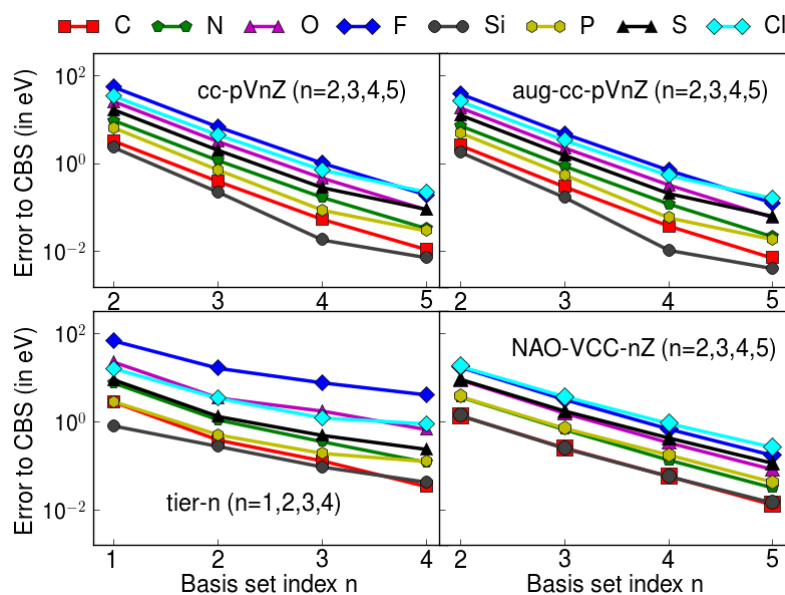


Figure 3. Basis set errors of figure 2, but on a logarithmic scale (in eV).

of (aug)-cc-pVnZ, the basis set convergence curves are quite close for C and Si. However the deviation increases for the pair of N and P. And the maximum discrepancy occurs for F and Cl. Figure 3 presents the behavior of the basis set errors of figure 2 on a logarithmic scale. Among the cases shown, NAO-VCC- n Z exhibits excellent VCC for RPA calculations.

Table 4. RPA@PBE total energies of cc-pV2Z and corresponding energy lowerings of aug-cc-pV2Z and NAO-VCC- n Z for elements from C to F and from Si to Cl. The energy lowering of a basis set is defined as difference between the total energy of the other three basis sets and cc-pV2Z, $E_{\text{total}}^{\text{other}} - E_{\text{total}}^{\text{cc-pV2Z}}$ (in eV).

Elements	cc-pV2Z	aug-cc-pV2Z	NAO-VCC-2Z
C	-1029.0505	-0.1532	-0.5169
N	-1483.8917	-0.2931	-0.9012
O	-2040.3463	-0.5293	-1.5003
F	-2710.4529	-0.7828	-2.2165
Si	-7862.8566	-0.1418	-0.2857
P	-9274.3840	-0.2382	-0.4774
S	-10 820.7206	-0.3737	-0.7674
Cl	-12 508.0326	-0.4940	-1.0936

The NAO-VCC- n Z basis sets extend the ‘valence-correlation consistency’ strategy upon including the enhanced minimal basis. Considering that the only difference of the elements in the same periodic table groups comes from the core shells, the equally good convergence behavior of NAO-VCC- n Z for these elements confirms the importance of saturating s and p functions for valence-correlation consistency.

The quality of NAO-VCC- n Z is further demonstrated in table 4, which lists the absolute RPA total energies of cc-pV2Z, and the corresponding energy lowerings of aug-cc-pV2Z, NAO-VCC- n Z for elements from C to F, and Si to Cl. Comparing to the RPA total energies of cc-pV2Z, consistent energy lowerings can be achieved by introducing a set of diffuse basis functions, (one s, p and d primitive GTO each for first- and second-row elements). The aug-cc-pV2Z basis sets aim to saturate the basis set in the outer space. In contrast, the enhanced minimal basis sets within NAO-VCC- n Z are designed to saturate the (sp) basis functions in the inner space. Table 4 shows that dramatically larger energy lowerings can be gained by NAO-VCC-2Z, indicating that a large contribution to the error of double- ζ basis sets can be removed by saturating the inner (sp) basis function space. More importantly, these errors are not simply a function of the number of valence electrons. For example, the energy lowering of NAO-VCC-2Z for F is -2.2165 eV, which is more than double that of Cl. The elimination of these errors is the key to the excellent performance of NAO-VCC- n Z in figures 2 and 3, where we only focus on the issue of valence correlation.

MP2 is the simplest finite-order perturbation theory method, counting only corrections at second-order. As a systematic extension of RPA, renormalized second-order perturbation theory (rPT2) [12, 13] adds second-order screened exchange [6, 10, 79] and renormalized single-excitation (rSE) corrections [14] to the standard RPA method. To demonstrate the transferability of NAO-VCC- n Z, we also show frozen-core MP2 and rPT2@PBE results. The corresponding basis set errors are plotted as a function of the index ‘ n ’ in figures 4 and 5. Since all these methods are explicit correlation methods which share the same difficulty in approaching the CBS limit, our basis sets also provide a performance similar to RPA for MP2 and rPT2@PBE. In particular, NAO-VCC- n Z performs similarly well for systems with the same number of valence electrons, highlighting once again the importance of the enhanced minimal basis to saturate the (sp) basis function space for valence correlation.

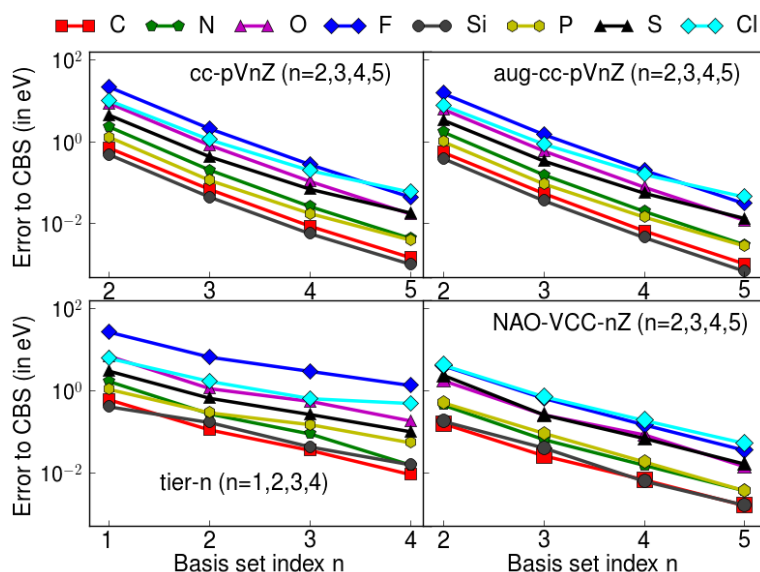


Figure 4. Basis set errors of MP2 total energies on a logarithmic scale for several first- and second-row elements. As references, we take the CBS values, extrapolated from quintuple- ζ (5Z) and sextuple- ζ (6Z) by the sequence of aug-cc-pVnZ. Tier- n denotes different tiers of the FHI-aims-2009 basis set (in eV).

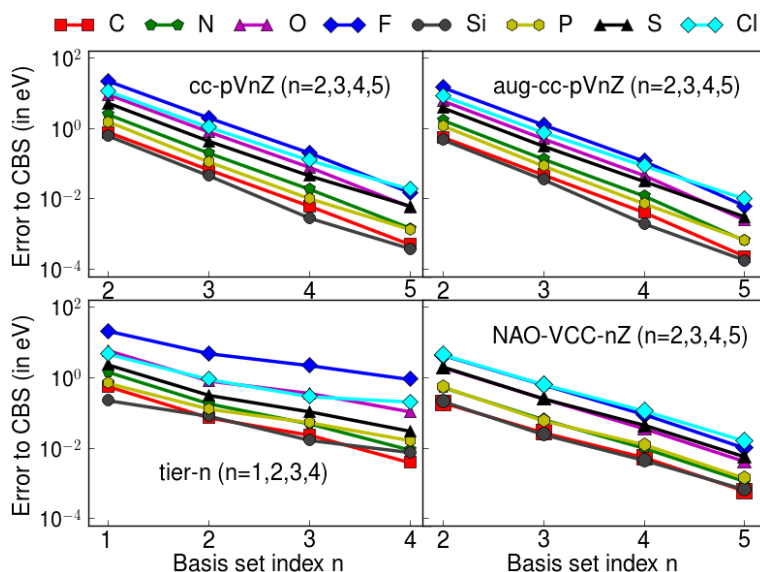


Figure 5. Basis set errors of rPT2@PBE total energies on a logarithmic scale for several first- and second-row elements. As references, we take the CBS values, extrapolated from quintuple- ζ (5Z) and sextuple- ζ (6Z) by the sequence of aug-cc-pVnZ. Tier- n denotes different tiers of the FHI-aims-2009 basis sets (in eV).

3.3. Basis set convergence for energy differences

We next focus on the homonuclear diatomic molecules N_2 and F_2 to illustrate the performance of NAO-VCC- nZ for covalent binding energies. In figure 6, the RPA binding energies of N_2

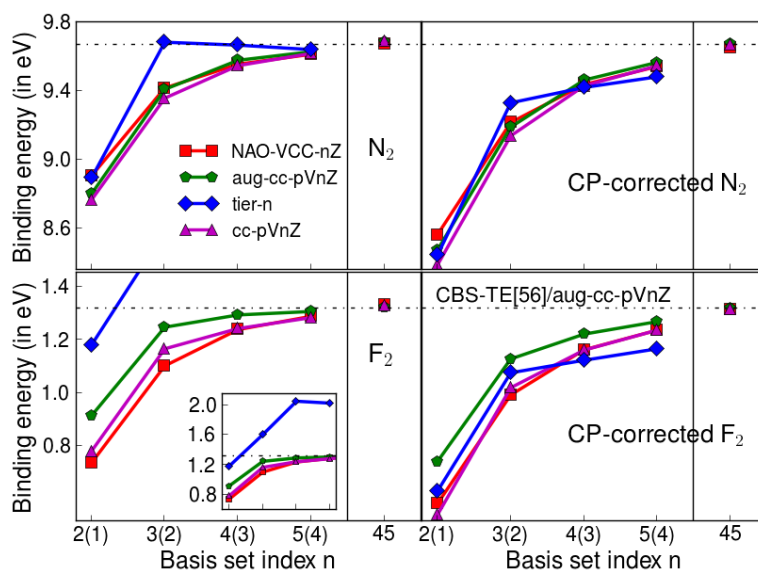


Figure 6. Convergence of the RPA@PBE binding energies of N_2 and F_2 as a function of the index of various basis sets. $n = 2, 3, 4, 5$ for NAO-VCC- nZ and (aug)-cc-pV nZ , and $n = 1, 2, 3, 4$ within the parentheses for tier- n in FHI-aims-2009. The results labeled ‘45’ on the x -axis are extrapolated using the CBS-TE scheme, CBS-TE[45]. The dotted line marks the CP-corrected CBS-TE[56]/aug-cc-pV nZ values for N_2 and F_2 , respectively (in eV).

and F_2 with (right) and without (left) CP correction are shown. The dotted line marks the CBS limit of CP-corrected RPA@PBE binding energies extrapolated by CBS-TE[56]/aug-cc-pV nZ (9.670 eV for N_2 , and 1.318 eV for F_2).

Upon increasing the basis size, the computed RPA binding energies of N_2 and F_2 converge quickly to the CBS limits for the sequences of (aug)-cc-pV nZ and NAO-VCC- nZ . The CP correction decelerates the basis set convergence. However, most importantly, the extrapolated values (CBS-TE[45]) with or without CP correction converge to almost the same CBS limit, demonstrating that the notorious basis set superposition error (BSSE) is eliminated naturally without having to resort to CP corrections. This is the strategy of VCC. In fact, the maximum deviation is only about 20 meV or 0.2% (1.5%) of the N_2 (F_2) binding energies, among six CP corrected and uncorrected CBS limit values extrapolated by (aug)-cc-pV nZ and NAO-VCC- nZ for both N_2 and F_2 . For comparison, the FHI-aims-2009 basis sets (‘tier- n ’), which were not constructed to converge unoccupied-state sums, require a CP correction to obtain qualitatively correct energy differences. Without such a correction, systematic error cancellation between the unoccupied-state sums entering the energy difference would not be guaranteed.

To further inspect the basis set convergence, figure 7 plots the basis set errors of RPA binding energies on a logarithmic scale. We take CBS-TE[56]/aug-cc-pV nZ values as the reference. The basis set error is defined as the absolute deviation between the calculated binding energy (BE) in a finite basis set and the reference $|BE - BE^{\text{Ref}}|$. The diffuse (augmentation) functions in aug-cc-pV nZ are very helpful for the description of the chemical bond. The sequence of aug-cc-pV nZ yields fast basis set convergence, especially for the F_2 case. As shown in figure 7, NAO-VCC- nZ also provides a very good convergence behavior. The logarithmic scale reveals that the NAO-VCC- nZ basis set errors decrease nearly exponentially

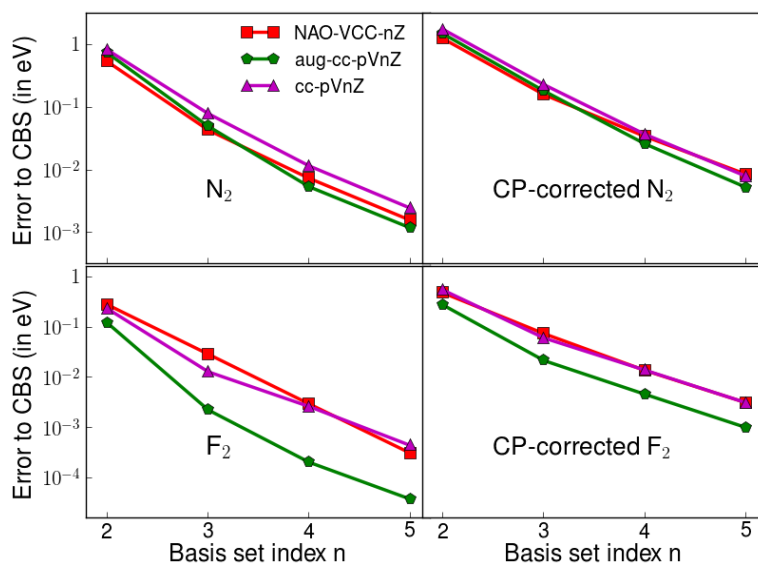


Figure 7. Basis set errors of RPA binding energies on a logarithmic scale for N_2 and F_2 dimers. As references, we take the CBS values, extrapolated from quintuple- ζ (5Z) and sextuple- ζ (6Z) by the sequence of aug-cc-pVnZ (in eV).

with the cardinal number, n . We argue that the (sp) enhanced minimal basis makes an essential contribution to this excellent convergence behavior of NAO-VCC- nZ , which highlights again the importance of saturating the (sp) basis function space for valence correlation. Figure 7 also illustrates that the CP corrections can significantly improve the convergence quality of the sequences of (aug)-cc-pVnZ, even though they decelerate the basis set convergence.

3.4. Role of the confining potential

As discussed in the context of equation (8), the default confining potential $v_{\text{cut}}(r)$ is employed with the onset radius $r_{\text{onset}} = 4 \text{ \AA}$ in the optimization of NAO-VCC- nZ basis sets for most of the elements. For Li, Na, Mg and Al, a slightly larger value $r_{\text{onset}} = 5 \text{ \AA}$ is used. We next quantify the influence of the onset radii on the results computed using NAO-VCC- nZ . Taking the calculated total energies for atoms and binding energies for dimers (no CP correction applied) together with larger onset radii ($r_{\text{onset}} = 6 \text{ \AA}$ for all elements) as reference, table 5 lists the absolute errors computed by the default onset radii.

The NAO-VCC- nZ basis sets are optimized for the corresponding RPA@PBE total energies of the 18 elements. Table 5 reveals that the default onset radii are a reasonable choice for the basis set optimization. The mean absolute errors (MAE) are within 2 meV. The largest deviations compared to $r_{\text{onset}} = 6 \text{ \AA}$ are $\sim 6 \text{ meV}$ (Li) and $\sim 10 \text{ meV}$ (Na). This deviation does not change with increasing the basis size.

In table 5, we also show the influence of the onset radii on the calculation of binding energies of 18 homonuclear diatomic molecules from H_2 to Ar_2 . In all cases, the change in binding energy with onset radius is of the order of a few meV only. The MAEs are also within 2 meV. The errors in the total energy calculations of Li and Na are mostly canceled out for the binding energies. Only the case of Si_2 is an exception. Here, the more confined double- ζ basis set (NAO-VCC-2Z, $r_{\text{onset}} = 4 \text{ \AA}$) apparently describes the dimer total energy better (lower total energy by $\approx 6 \text{ meV}$) than the same set of radial functions with the larger confinement radius

Table 5. Absolute errors of the default onset radii ($r_{\text{onset}} = 5 \text{ \AA}$ for metal elements and $r_{\text{onset}} = 4 \text{ \AA}$ for the remaining 14 elements), where the reference values are the total energies and binding energies calculated with $r_{\text{onset}} = 6 \text{ \AA}$ for all elements (in meV).

NCC ^a	Total energies				Binding energies				
	2Z	3Z	4Z	5Z	2Z	3Z	4Z	5Z	
H	1	0	0	0	H ₂	0	1	0	0
He	0	0	0	0	He ₂	0	0	0	0
Li	6	5	6	6	Li ₂	2	2	3	4
Be	4	3	3	4	Be ₂	2	3	1	0
B	1	2	2	1	B ₂	1	1	0	1
C	5	3	3	1	C ₂	3	2	3	2
N	2	1	1	0	N ₂	1	0	0	0
O	1	0	1	0	O ₂	0	0	0	0
F	0	0	0	0	F ₂	0	1	0	0
Ne	0	0	0	0	Ne ₂	0	0	0	0
Na	9	9	10	10	Na ₂	3	2	2	3
Mg	2	1	1	1	Mg ₂	0	0	0	0
Al	1	0	1	0	Al ₂	5	2	1	1
Si	6	5	3	4	Si ₂	17	11	10	5
P	1	1	1	1	P ₂	5	5	3	3
S	1	0	1	1	S ₂	2	3	4	2
Cl	0	0	1	1	Cl ₂	1	1	2	2
Ar	0	0	0	3	Ar ₂	0	0	0	0
MAE ^b	2	2	2	2	MAE ^b	2	2	2	1

^a NCC: the sequence of NAO-VCC- n Z basis sets.

^b MAE: mean absolute errors for all cases.

($r_{\text{onset}} = 6 \text{ \AA}$). For an incomplete basis set, residual effects of this magnitude are possible (there is no variational principle for a fixed basis set size and change of basis function shape with increasing confinement radius). Importantly, the effect is small and decreases with increasing basis set size, as it should. Thus, for the quintuple- ζ basis set (NAO-VCC-5Z), the maximum error (MAX) amongst 18 binding energies is only about 5 meV. This observation is consistent with the previous work by Ren *et al* [63] that the onset radii only affect the binding energies for small basis sets. Since the onset radius of the cutoff is always available as an explicit convergence parameter in a practical calculation, this issue is not particularly difficult to address. Nevertheless, even in small basis sets, such numerical uncertainty is negligible compared to the basis set incompleteness error for explicitly correlated methods, e.g. RPA and MP2.

4. Results and discussion

4.1. Homonuclear diatomic molecules from H₂ to Ar₂, HDM18

We choose homonuclear diatomic molecules from H₂ to Ar₂ (HDM18) as the first test set to examine the performance of the new basis set NAO-VCC- n Z. The experimental equilibrium

Table 6. Mean absolute errors (MAE) due to basis set incompleteness for HDM18 in RPA@PBE for sequences of (aug)-cc-pVnZ, NAO-VCC-nZ and tier-n in the FHI-aims-2009 basis sets. Values in parentheses are the absolute maximum errors (MAX) for each basis set and extrapolation scheme (in meV).

	aug-cc-pVnZ		cc-pVnZ		NAO-VCC-nZ		Tier-n	
	CP ^b		CP ^b		CP ^b		CP ^b	
$n = 2(1)^a$	348 (867)	476 (1193)	373 (911)	531 (1304)	327 (761)	463 (1104)	191 (774)	364 (1223)
$n = 3(2)^a$	113 (338)	196 (535)	139 (392)	229 (603)	135 (365)	213 (549)	62 (287)	144 (339)
$n = 4(3)^a$	51 (165)	96 (276)	64 (194)	112 (311)	60 (173)	113 (302)	80 (737)	107 (249)
$n = 5(4)^a$	23 (63)	57 (126)	26 (78)	58 (151)	27 (86)	61 (168)	81 (710)	87 (204)
CBS-TE[23]	27 (121)	82 (263)	47 (198)	105 (308)	53 (203)	108 (326)		
CBS-TE[34]	16 (45)	23 (86)	15 (50)	27 (97)	13 (43)	40 (122)		
CBS-TE[45]	12 (61)	6 (31)	16 (59)	4 (16)	10 (26)	7 (28)		
CBS-OPT[23]	46 (150)	35 (105)	42 (117)	36 (124)	37 (135)	35 (148)		
CBS-OPT[34]	16 (57)	17 (80)	15 (53)	13 (41)	14 (51)	18 (83)		
CBS-OPT[45]	14 (66)	7 (38)	19 (69)	6 (30)	14 (29)	4 (13)		

^a $n = 2, 3, 4, 5$ denote the indices of the sequences of (aug)-cc-pVnZ and NAO-VCC-nZ, while $n = 1, 2, 3, 4$ in parentheses denote the ‘tier’ number of FHI-aims-2009 (tier- n).

^b Values in these columns are CP corrected.

geometries of these diatomic molecules are collected from the ‘Computational Chemistry Comparison and Benchmark DataBase’ (CCCBDB) web site⁶. All RPA calculations are based on self-consistent PBE ground state orbitals. The CP-corrected CBS-TE extrapolated results from aug-cc-pV5Z and 6Z (CBS-TE[56]/aug-cc-pVnZ) are taken as the CBS reference for most of the dimers. For Be₂ and Mg₂, we take CBS-TE[34]/aug-cc-pVnZ values, and for Li₂ and Na₂ CBS-TE[45]/cc-pVnZ, because corresponding basis sets with higher quality are not available for these elements.

The MAEs and unsigned MAX of the basis set incompleteness error in the RPA@PBE binding energies for HDM18 are listed in table 6. Confirming the above observation, both MAEs and MAXs present a systematic convergence with increasing basis size for the (aug)-cc-pVnZ and NAO-VCC-nZ sequences. For the FHI-aims-2009 tier- n sequence, the MAX values show that energy differences without a CP corrections should not be used. However, with CP correction, a gradual decrease of the basis set incompleteness error is found also for tier- n .

To converge their RPA calculations, the authors of [22] and [21] extrapolated correlation consistent basis sets. As shown in table 6, when extrapolating from 4Z and 5Z (45), the CBS-TE scheme yields the CBS limit in good agreement with other sequences of basis sets. The statistical uncertainty is within 16 meV.

As observed in section 3.3, the CP correction can improve the regularity of the basis set convergence for RPA energy differences. One would therefore expect that better extrapolated values from 2Z and 3Z should be obtained using the CP correction. Unfortunately, the performance of the CBS-TE scheme from 2Z and 3Z becomes dramatically worse when doing CP corrections. A possible remedy is the more sophisticated extrapolation scheme,

⁶ 2011 Computational Chemistry Comparison and Benchmark Database (cccbdb), <http://cccbdb.nist.gov/default.htm>.

Table 7. The global parameters α , d in the CBS–OPT extrapolation scheme, $A/(n+d)^\alpha$, for (aug)-cc-pVnZ and NAO-VCC- n Z. These parameters are used for each level of extrapolation.

	NCC ^a	ACC ^a	CC ^a
α	4.49	4.47	4.51
d	2.85	2.39	2.65

^a CC, NCC and ACC are the abbreviations of cc-pVnZ, aug-cc-pVnZ and NAO-VCC- n Z, respectively.

CBS-OPT, that has two global parameters α and d (see equation (7)) [68, 69]. We optimized the parameters by minimizing the CP-corrected MAEs of CBS-OPT[23] in table 6. The same parameters are then applied to CBS-OPT[34] and CBS-OPT[45] without any reoptimization. The optimized α , d for (aug)-cc-pVnZ and NAO-VCC- n Z are listed in table 7, and the corresponding performance in table 6.

We see that the CBS-OPT[23] scheme performs well for the three correlation consistent basis set prescriptions if CP corrections are applied. The MAEs of CBS-OPT[23] are around 35 meV, comparable to those of the bare 5Z basis sets. While the parameters α , d are only determined in CBS-OPT[23], they are also suitable for high levels of the extrapolation. A substantial improvement can be found for CBS-OPT[34], and the performance of CBS-OPT[45] is comparable to the corresponding CBS-TE[45]. Inspecting table 6 also reveals that this CBS-OPT scheme optimized with CP corrections works for the calculations without CP corrections. For NAO-VCC- n Z, the non-CP corrected MAE obtained with the CBS-OPT[23] scheme is about 37 meV. In particular, a small MAE combined with a small MAX shows that both CBS-TE[45]/NAO-VCC- n Z and CBS-OPT[45]/NAO-VCC- n Z are accurate and robust methods to approach the CBS limit of RPA binding energies.

4.2. Reaction energies, G2RC

The so-called G2RC test set consists of 25 reactions of small organic molecules, one of which includes the element Li. Because the basis sets of (aug)-cc-pV6Z are unavailable for Li, we examine the basis sets using the remaining 24 reactions in this work. Figure 8 presents the mean error (ME), MAE and MAX in RPA@PBE reaction energies of the G2RC set. In this case, the reference reaction energies and geometries are the experimental(!) values taken from [80]. Thus, the errors do not go to zero: we are simultaneously exploring the basis set incompleteness error and the error of the chosen approximation to the calculated energy, RPA@PBE. A key message from figure 8 is that the basis set incompleteness error can be eliminated systematically using the NAO valence-correlation consistent NAO-VCC- n Z sequence. The three overall measures of the reaction energy error examined here exhibit a systematic convergence behavior toward the values that represent the intrinsic errors of the RPA@PBE method itself.

The MAE is an important criterion to judge the performance of a specific method. As shown in the upper right panel of figure 8, the MAE value given by cc-pV2Z—the smallest one in the cc-pVnZ sequence—suffers from a large basis set incompleteness error. This error dies away quickly with increasing basis size, and the corresponding MAEs approach the converged value. With a similar basis size, NAO-VCC- n Z basis sets provide a

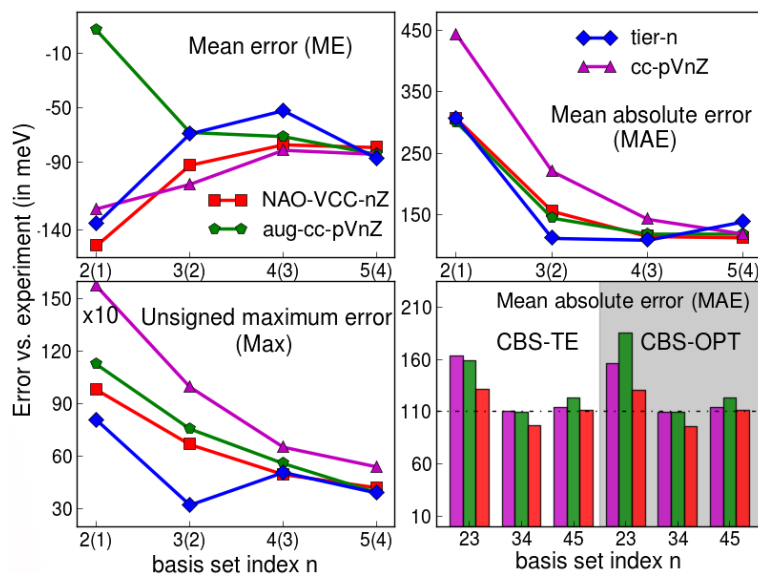


Figure 8. Basis set convergence of ME (upper left), mean absolute errors (MAE, upper right), and unsigned MAX (lower left) in RPA@PBE reaction energies of the G2RC test set compared to experimental(!) reference data. On the x -axis, $n = 2, 3, 4, 5$ we list the indices of (aug)-cc-pVnZ and NAO-VCC- n Z, while the number $n = 1, 2, 3, 4$ given in parentheses are those of FHI-aims-2009 tier- n . The lower right panel presents the MAEs according to different extrapolation schemes for the sequences of (aug)-cc-pVnZ and NAO-VCC- n Z. The dotted line in this panel marks the MAE of the CBS-TE[56]/aug-cc-pVnZ scheme as 110 meV. The ME is the averaged deviation of $RE^{\text{RPA}} - RE^{\text{Ref}}$, where RE^{Ref} is the experimental reference reaction energy (RE) taken from [80] (in meV).

better convergence behavior for this test set, and achieve convergence comparable to that of aug-cc-pVnZ basis sets.

The lower right panel of figure 8 shows the MAE values for two kinds of extrapolation schemes for the (aug)-cc-pVnZ and NAO-VCC- n Z sequences. As the benchmark reference, the dotted line marks the MAE of the CBS-TE[56]/aug-cc-pVnZ, which is 110 meV. The six CBS-limits are in good agreement with this reference. In particular, the NAO-VCC- n Z basis sets provide almost the same MAE as the reference (111 meV) for both CBS-TE[45] and CBS-OPT[45]. Therefore, we report an intrinsic MAE of 120 meV in RPA@PBE reaction energies of the whole G2RC set (*including 25 reactions*). This MAE is obtained by CBS-TE[45]/NAO-VCC- n Z.

Our results confirm the conclusion of Eshuis and Furche [21] that the reaction energies are converged well at the basis set level of aug-cc-pV4Z (MAE = 119 meV). While the MAE of the cc-pV4Z basis set is 143 meV, NAO-VCC-4Z gives a better performance (MAE = 115 meV) albeit at the size of cc-pV4Z. The basis set incompleteness contribution to the MAE is about 40 meV for NAO-VCC-3Z and aug-cc-pV3Z. Our results suggest that this triple- ζ error can be reduced efficiently to less than 20 meV using an extrapolation from 2Z and 3Z (both CBS-TE[23] and CBS-OPT[23]) without a substantial increase of computational cost, especially for the NAO-VCC- n Z sequence shown in the lower right panel of figure 8.

Table 8. Basis set convergence of the RPA@PBE ME, mean absolute errors (MAE) and unsigned MAX for the ISO34 set using four sequences of basis sets. ME are computed as the averaged deviation of $E^{\text{RPA}} - E^{\text{Ref}}$, where E^{Ref} is the experimental reference taken from [81](in meV).

$n^{\text{a}}=$	2(1)	3(2)	4(3)	5(4)	CBS-TE			CBS-OPT		
					23	34	45	23	34	45
ME										
CC ^b	-44	-31	-21	-20	-26	-13	-18	-22	-12	-18
ACC ^b	-22	-26	-20	-20	-27	-15	-19	-28	-15	-19
NCC ^b	-21	-20	-18	-19	-20	-16	-20	-20	-15	-20
Tier- n	-33	-26	-26	-24						
MAE										
CC ^b	87	50	42	42	45	40	44	47	40	44
ACC ^b	76	51	43	43	47	41	43	47	41	43
NCC ^b	61	45	44	44	43	44	44	45	45	44
Tier- n	78	56	47	44						
MAX										
CC ^b	399	215	150	154	146	133	158	162	129	159
ACC ^b	311	213	167	164	189	137	160	199	139	160
NCC ^b	220	151	153	162	136	154	171	136	154	173
Tier- n	293	197	178	152						

^a $n = 2, 3, 4, 5$ are the indices of the sequences of (aug)-cc-pV nZ and NAO-VCC- nZ , while $n = 1, 2, 3, 4$ within the parentheses denote the ‘tier’ number of FHI-aims-2009 (tier- n).

^b CC, ACC and NCC are abbreviations of cc-pV nZ , aug-cc-pV nZ , and NAO-VCC- nZ , respectively.

4.3. Isomerization energies, ISO34

Isomerization is a well-defined reaction process in organic chemistry, and isomerization energies have been used to validate the performance of theoretical methods [81–83]. Table 8 presents the basis set convergence of RPA@PBE results for 34 isomerization energies of small organic molecules of the ISO34 set [81]. Recently, Eshuis and Furche [21] reported ME, MAE and MAX, of RPA@PBE isomerization energies for Dunning’s (aug)-cc-pV nZ ($n = 3, 4, 5$) using a development version of the TURBOMOLE program package⁷. In this work, we reproduce these values with a discrepancy of less than 1 meV using our own numerical framework (FHI-aims) [12, 57].

We again extrapolate the isomerization energies using CBS-TE[45] and CBS-OPT[45] with the (aug)-cc-pV nZ and NAO-VCC- nZ sequences. The converged MAE of RPA@PBE has an uncertainty of 1 meV and is reported in table 8. The MAX error is more sensitive to the basis set incompleteness error than ME and MAE. Even from the ‘34’ to the ‘45’ extrapolation levels, the investigated correlation consistent basis set types still show a change of 20–30 meV. In short, the convergence of the MAX error is comparatively poor, leaving an uncertainty of the order of 10 meV or more for all basis sets and extrapolation schemes shown in table 8.

⁷ 2011 Turbomole v6.3, Turbomole GmbH, Karlsruhe (www.turbomole.com).

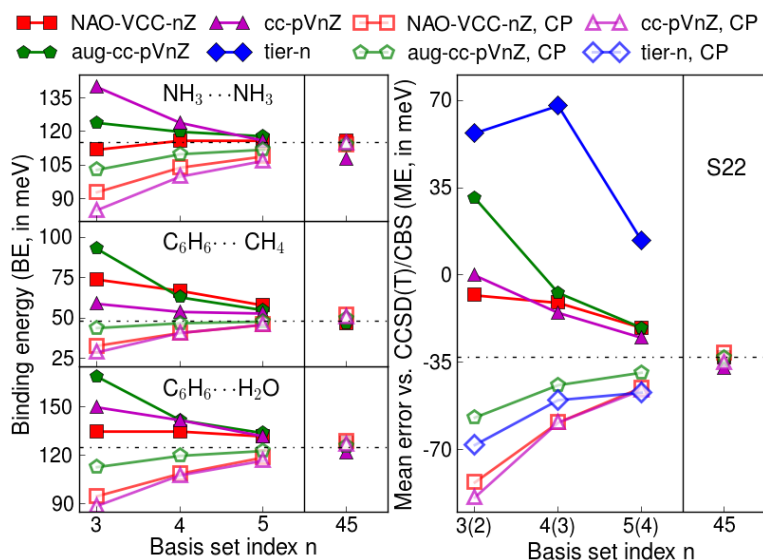


Figure 9. Left panels: basis set convergence of the RPA@PBE binding energies for $\text{NH}_3 \cdots \text{NH}_3$, $\text{C}_6\text{H}_6 \cdots \text{CH}_4$, and $\text{C}_6\text{H}_6 \cdots \text{H}_2\text{O}$. Right panel: basis set convergence of the ME in RPA@PBE interaction energies of the S22 set with and without CP corrections. On the x -axis, $n = 2, 3, 4, 5$ list the indices of (aug)-cc-pVnZ and NAO-VCC- n Z, while the number $n = 1, 2, 3, 4$ given in parentheses are those of FHI-aims-2009 tier- n . The CBS-TE[45] extrapolated values are labeled ‘45’. Geometries are taken from [84]. The ME is the averaged deviation of $\text{BE}^{\text{RPA}} - \text{BE}^{\text{Ref}}$, where BE^{Ref} is the CCSD(T) reference binding energy (BE) taken from [85]. Negative MEs imply systematic underbinding. The dotted line marks the CP-corrected CBS-TE[45]/aug-cc-pVnZ extrapolated data (in meV).

Nevertheless, our results demonstrate that for these isomerization energies, the quadruple- ζ basis sets will generally suffice, not only for Dunning’s (aug)-cc-pVnZ sequences, but also for the numeric NAO-VCC- n Z basis sets developed in this work. The bare tier- n basis sets of FHI-aims-2009 perform similarly well for the isomerization case, indicating that some of the basis set incompleteness error cancels here.

4.4. Non-covalent interactions, S22

The S22 set is the most widely used benchmark database for non-covalent interactions [84]. It includes 22 binding energies of seven hydrogen bonds, eight dispersion bonds and seven weak bonds with mixed bonding character. Molecular geometries in the S22 set are taken from [84], and the reference binding energies are CCSD(T) values in the CBS limit taken from [85]. The right panel of figure 9 shows the ME in RPA@PBE binding energies of the S22 test set using different sequences of basis sets. BSSE has the effect to overestimate the non-covalent interactions systematically. The CP correction removes the BSSE. The remaining basis set convergence error (BSCE) now underestimates the non-covalent interactions [86, 87]. For three kinds of valence-correlation consistent basis sets, the non-CP and CP corrected MEs converge to the CBS limit from opposite directions. In the CBS-TE[45] extrapolation scheme, the extrapolated MEs with (without) CP correction are -31 meV (-33 meV) for NAO-VCC- n Z, -32 meV (-36 meV) for cc-pVnZ, and -34 meV (-33 meV) for aug-cc-pVnZ respectively.

A negative value implies an underestimation of non-covalent bonds. Because the RPA@PBE method consistently underestimates all interactions in the S22 set, the extrapolated MAEs are simply the absolute values of the MEs.

For non-covalently bonded systems, even quintuple- ζ basis sets are not converged to better than 10 meV, regardless of whether CP corrections are applied or not. The non-CP corrected MEs are -21 meV (NAO-VCC-5Z), -25 meV (cc-pV5Z) and -21 meV (aug-cc-pV5Z), and the CP-corrected MEs are -45 meV, -46 meV and -39 meV, respectively. Furthermore, it is also insufficient to extrapolate from the smaller 2Z and 3Z basis sets (extrapolation not shown), even though a considerable improvement is found. The extrapolation from 3Z and 4Z is enough to produce converged MEs and MAEs (CBS-TE[34] values) for the aug-cc-pVnZ sequence, but not for cc-pVnZ and NAO-VCC-nZ. Therefore, we report the intrinsic RPA@PBE errors for the S22 set as CBS-TE[45]/aug-cc-pVnZ extrapolated ME and MAE of -33 meV and 33 meV with an uncertainty of ± 2 meV.

Our results confirm the observation of Blum *et al* [57] and Ren *et al* [11] that CP corrections are always necessary for the FHI-aims-2009 basis sets in RPA or MP2 calculations. The CP-corrected ME of tier-4 is -47 meV, similar to that of cc-pV5Z and NAO-VCC-5Z. Upon adding a set of diffuse functions obtained from aug-cc-pV5Z, the CP-corrected ME of ‘tier-4+a5Z-d’ is -39 meV reported by Ren *et al* [11], which is equal to that of aug-cc-pV5Z.

In the left panels of figure 9, we chose the interactions of $\text{NH}_3 \cdots \text{NH}_3$, $\text{C}_6\text{H}_6 \cdots \text{CH}_4$, and $\text{C}_6\text{H}_6 \cdots \text{H}_2\text{O}$ as illustrations for hydrogen bonding, dispersion bonding and mixed bonding. Dunning’s aug-cc-pVnZ sequence performs well for the three types of non-covalent interactions. Especially, the CP-corrected values of aug-cc-pVnZ present the fastest convergence for the three valence-correlation consistent basis sets. However, without the CP correction, the best performance is given by the NAO-VCC-nZ basis sets. For the $\text{NH}_3 \cdots \text{NH}_3$ case, the converged values are approached quickly by NAO-VCC-4Z. We also find a similar behavior for the $\text{C}_6\text{H}_6 \cdots \text{H}_2\text{O}$ case. Taking the CP-corrected CBS-TE[45]/aug-cc-pVnZ values as reference, the extrapolated NAO-VCC-nZ values have a tendency (within 4 meV or 3% of the binding energy) to overestimate mixed bonds. For $\text{NH}_3 \cdots \text{NH}_3$ and $\text{C}_6\text{H}_6 \cdots \text{CH}_4$, the deviations between the extrapolated values of NAO-VCC-nZ and aug-cc-pVnZ are less than 1 meV. In comparison, we can see that the CBS-TE[45]/cc-pVnZ scheme without CP corrections underestimates hydrogen bonds (8 meV for $\text{NH}_3 \cdots \text{NH}_3$) and mixed bonds (3 meV for $\text{C}_6\text{H}_6 \cdots \text{H}_2\text{O}$).

5. Conclusions and outlook

In this work, we introduce a sequence of numeric atom-centered basis sets with valence-correlation consistency, called NAO-VCC-nZ ($n = 2, 3, 4, 5$), for the elements from H to Ar. We demonstrate that NAO-VCC-nZ basis sets are suitable for explicit correlation methods, e.g. RPA or MP2. The basis set incompleteness error, including the notorious BSSE, is reduced gradually by increasing the index ‘ n ’, and can be removed directly using two-point extrapolation schemes for the total energy. Generally, the NAO-VCC-4Z basis sets are sufficient to produce satisfactory results for covalent bonds and isomerization energies. However, for non-covalent bonds, it is always necessary to extrapolate the NAO-VCC-nZ sequence from 4Z and 5Z.

The NAO-VCC-nZ basis sets developed here are only designed for valence correlation. We expect them to fail for core correlation and extended cases, like anions. Core and core–valence correlation effects generally become important when relatively small errors in energetics or

spectroscopic constants are required [88–90]. Also, more diffuse functions might be required for the calculation of electronic affinities and other properties associated with anions [50, 51]. These aspects were not the target of our present work, but it is clear that suitable core or diffuse functions could be added to NAO-VCC- nZ if needed. Furthermore, at present the NAO-VCC- nZ basis sets only provide access to the elements from H to Ar. In fact, so far there is no valence-correlation consistent basis set that covers all elements of the periodic table [90]. Now that we have established a solid basis with NAO-VCC- nZ , work to address these remaining shortcomings is ongoing in our group.

Appendix. Accuracy of numerical integrations

In FHI-aims, there is a hierarchy of numerical integration settings ‘light’, ‘tight’ and ‘really tight’, catering to various numerical precision requirements. These settings simultaneously increase the accuracy of the following aspects of the code: (i) number of radial functions; (ii) confinement radius; (iii) Hartree potential expansion; and (iv) the grids used for three-dimensional integrations. The technical aspects behind these choices are described in [57]. What is important for the present development, however, is that the appropriate choice of integration grid can of course depend on the choice of basis set. Here, we encounter a substantial difference between standard quantum-chemical GTO basis sets, and NAO basis sets. Specifically, our experience suggests that Dunning’s (aug)-cc-pV nZ basis sets demand denser radial grids near the nucleus and denser angular grids on faraway shells than would be required for NAOs.

In FHI-aims, spherical shells of integration grid points (‘radial grid shells’) are positioned around each atom. In practice, these shells are determined as follows:

1. A basic radial grid is determined as proposed by Baker and co-workers [91]

$$r(s) = r_{\text{outer}} \frac{\log\{1 - [s/(N_r + 1)]^2\}}{\log\{1 - [N_r/(N_r + 1)]^2\}} \quad (\text{A.1})$$

and $s = 1, \dots, N_r$. Note that the hypothetical shell indices $s = 0$ and $N_r + 1$ correspond to $r = 0$ and ∞ , respectively. Typical values for the parameters are $r_{\text{outer}} = 7 \text{ \AA}$ (FHI-aims ‘tight’ and ‘really_tight’ species defaults) and $N_r = 1.2 \times 14 \times (\text{atomic number} + 2)^{1/3}$, again according to Baker *et al* [91]. In practice, this leads to between 24 (hydrogen) and approx. 80 grid shells, but in principle, any other combination (r_{outer}, N_r) is possible.

2. A simple way to increase the radial grid density uniformly is to place additional grid shells according to equation (A.1) at fractional values s , e.g. $s = 1/2, 3/2, \dots, (N_r + 1)/2$. The grid is thus extended both toward $r = 0$ and ∞ . The fractional spacing is called *radial_multiplier*; in the example, *radial_multiplier*=2, which is employed in standard FHI-aims ‘tight’ settings. Examples of the location of grid shells for the C atom and *radial_multiplier*=2, 4 and 6 are shown in the topmost panel of figure A.1.

For each radial grid shell, the angular integration points are distributed based on so-called Lebedev grids [91–97]. These grids are tabulated to integrate successively higher angular momentum orders around a center exactly on the unit sphere. In FHI-aims, the number of angular grid points increases from the innermost to the outermost grid shells. The number of angular grid points on faraway shells is 434 in the ‘tight’ settings and 590 in the ‘really_tight’ settings. We denote this maximum number of angular grid points by ‘*outer_grid*’ in the following.

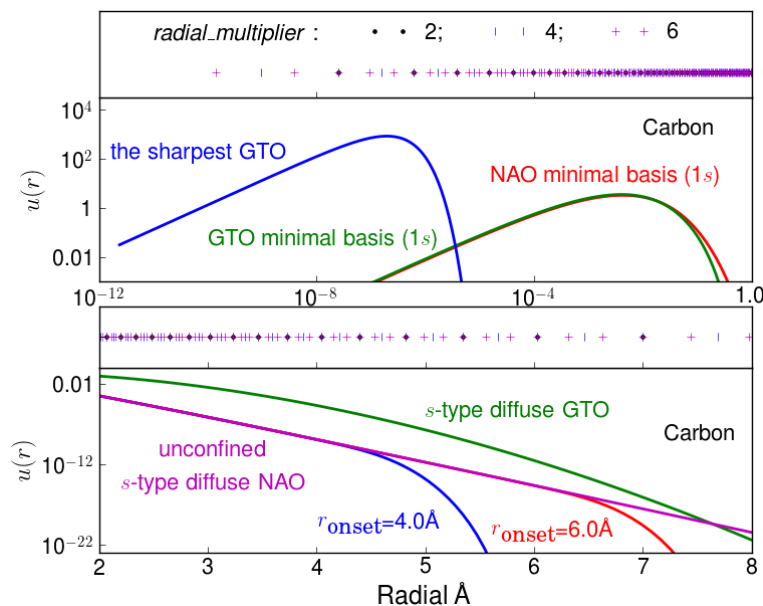


Figure A.1. The radial shapes $u(r)$ of some basis functions for carbon. Upper panel: the 1s contracted GTO in the minimal basis of cc-pV5Z and 1s orbital in the minimal basis of NAO-VCC- nZ . The sharpest primitive GTO in the 1s contracted GTO is also plotted. Grid shell locations for three radial grids with increasing *radial_multiplier* are shown in the topmost part of this panel. Lower panel: the radial tails of the most diffuse s-type GTO in cc-pV5Z and s-type hydrogen-like NAO in NAO-VCC-5Z. Three radial grids with increasing accuracy are shown above this panel.

The minimal basis of Dunning’s ‘correlation consistent’ basis sets is composed of combinations of primitive GTOs. These combinations are called the contracted GTOs. Using the 1s orbital of carbon as an illustration, the contracted GTO in Dunning’s cc-pV5Z consists of ten primitive GTOs. The exponential factors (EFs) from maximum to minimum are 96 770.0000, 14 500.0000, 3300.0000, 935.8000, 306.2000, 111.3000, 43.9000, 18.4000, 8.0540 and 3.6370. Figure A.1 shows the radial shapes of the 1s contracted GTO of cc-pV5Z, and the sharpest primitive GTO ($EF = 96\,770$) in this contracted GTO on a logarithmic scale. The contracted GTOs in cc-pV nZ are constructed to approximate the exact solutions of the free atoms. In contrast, accurate numerical versions of the free-atom solutions are used as the minimal basis in NAO-VCC- nZ . In figure A.1, the 1s orbital from NAO minimal basis for the PBE functional is also plotted for comparison.

Table A.1 lists the numerical error in the DFT-PBE and RPA@PBE total energies of the carbon dimer. Using the default ‘tight’ setting with *radial_multiplier*=2, the numerical error of cc-pV5Z is more than 300 meV for both DFT-PBE and RPA@PBE. The magnitude of this error is surprising, all the more since already *radial_multiplier*=4 (i.e. doubling the grid density) reduces this error to below 1 meV.

In figure A.2, we trace the origin of the discrepancy to the 1s function of the minimal basis of Dunning’s cc-pV5Z GTO basis set. In fact, practically the entire error can be traced to the 1s Kohn–Sham eigenvalue already in DFT-PBE. At a first glance, the GTO 1s function and the

Table A.1. Numerical error in DFT–PBE and RPA@PBE total energies of the carbon dimer for three levels of radial integration grids. The total energies calculated with very dense radial grids, ‘*radial_multiplier=8*’, are taken as the reference E_{total}^8 . The columns show the energy deviations $E_{\text{total}}^n - E_{\text{total}}^8$ for *radial_multiplier* values $n = 2, 4, 6$ (in meV).

<i>radial_multiplier</i> ^a	2	4	6
cc-pV5Z			
PBE	324.485	−0.833	−0.068
RPA@PBE	320.473	−0.826	−0.067
cc-pV5Z but using the NAO minimal basis			
PBE	0.090	−0.001	0.002
RPA@PBE	0.036	−0.004	0.001

^a ‘Tight’ grids and the default numerical thresholds are used, except for *radial_multiplier*. The default *radial_multiplier* is 2 in ‘tight’.

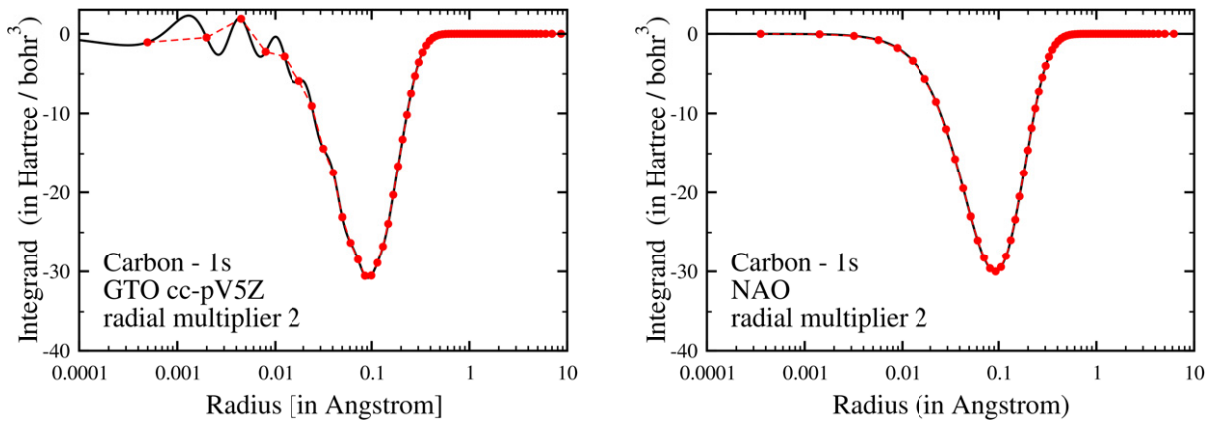


Figure A.2. Integrand $f(r)$ as defined in equation (A.3) for a GTO 1s radial function of the cc-pV5Z basis set (left panel), a NAO 1s radial function of the spherically symmetric, non-spinpolarized C atom (right panel), in either case evaluated for the potential of the spherically symmetric, non-spinpolarized C atom. The black lines are the continuous integrands $f(r)$ for GTO and NAO 1s orbitals. The red dash lines are the corresponding tabulated grids with the standard ‘tight’ setting for *radial_multiplier=2* as described in the text.

DFT-PBE NAO 1s function look very similar in figure A.1. The difference can be understood by analyzing the actual radial integrand $f(r)$ in the expression (assuming a spherical potential)

$$\int d^3r \varphi_{1s}(\mathbf{r}) \hat{H} \varphi_{1s}(\mathbf{r}) = \int dr [f(r)]. \quad (\text{A.2})$$

In the non-relativistic case, the integrand then reads

$$f(r) = u(r) \left[-\frac{1}{2}u''(r) + \frac{1}{2} \frac{l(l+1)}{r^2} u(r) + v(r)u(r) \right]. \quad (\text{A.3})$$

Table A.2. Influence of successively denser angular grids on the non-covalent interaction energy of the phenol dimer (C_6H_5OH)₂, calculated by the PBE, HF@PBE and RPA@PBE methods. (in meV).

Outer_grid ^a	434	590	770	974	1202	
	PBE	176	176	176	176	176
5Z ^b	HF ^c	43	43	44	44	44
	RPA ^c	393	280	276	276	276
	PBE	170	170	171	171	171
A5Z ^b	HF ^c	41	41	41	41	41
	RPA ^c	4574	332	327	281	280
	PBE	171	171	172	172	172
N5Z ^b	HF ^c	42	42	43	43	43
	RPA ^c	318	281	280	280	280

^a ‘Tight’ grids and the default numerical thresholds are used, except for *radial_multiplier* and *outer_grid*. For (aug)-cc-pV5Z, *radial_multiplier*=6, but for NAO-VCC-5Z, *radial_multiplier*=2. The number of angular grid points on faraway radial shells, *outer_grid*, changes from 434 to 1202. The default *outer_grid* is 434 in ‘tight’, and 590 in ‘really_tight’.

^b 5Z stands for cc-pV5Z, A5Z for aug-cc-pV5Z, and N5Z for NAO-VCC-5Z.

^c HF@PBE and RPA@PBE are abbreviated as HF and RPA in the table. They denote Hartree–Fock-like total energies and RPA total energies, each evaluated for the same self-consistent KS orbitals of DFT–PBE.

In figure A.2, the integrand $f(r)$ is plotted as a function of r for the Kohn–Sham Hamiltonian \hat{H} of a spherically symmetric, non-spinpolarized free C atom. The top panel shows the result for the minimal-basis 1s function from cc-pV5Z. The lower panel shows the same integrand, but for the accurate NAO solution of the problem. In each case, the black line is $f(r)$, while the red dots denote the a standard radial integration grid for *radial_multiplier*= 2.

In short, the near-nuclear integration is simple for the NAO, whereas $f(r)$ for the contracted GTO presents a significant problem. Obviously, for the specific shape of GTOs, the integration difficulty itself could easily be circumvented by textbook analytical expressions for the kinetic energy part and the nuclear Coulomb singularity in the Hamiltonian. However, the integrand itself is also not ‘physical’ in the sense that the correct behavior of an actual Kohn–Sham eigenfunction near the nucleus is not built in. For an all-electron framework, the fact that the NAO-VCC- nZ allow one to work with physically well behaved near-nuclear radial functions is a significant advantage.

Figure A.1 also plots the most diffuse s-type GTO in cc-pV5Z and s-type hydrogen-like NAO in NAO-VCC-5Z for carbon on a logarithmic scale. Formally speaking, the GTOs decay more quickly than the hydrogen-like functions, but the crossover point occurs around 7.6 Å in this case. More importantly, the flexibility of NAO allows us to control the shape of the radial tails explicitly using the confining potential $v_{\text{cut}}(r)$ (see equation (8)). We also plot the corresponding confined s-type hydrogen-like NAOs with the onset radius $r_{\text{onset}} = 4$ and 6 Å. The former is the default setting for carbon in NAO-VCC- nZ . Figure A.1 shows that both of them tend to zero rigorously before the crossover point.

It is easy to understand that the unconfined GTO tails require tighter numerical grids for both radial and angular integrations on faraway shells. Taking the binding energy of the phenol

dimer as example, table A.2 presents the sensitivity of DFT-PBE, HF@PBE, and RPA@PBE to *outer_grid*. Three quintuple- ζ (5Z) basis sets are examined here. Since increasing the angular grids in the near-nuclear shells has no effect on the accuracy, calculations are performed with the ‘tight’ grids and the default numerical thresholds in FHI-aims.

For the first- and second-row elements, *outer_grid* is set to 434 in ‘tight’. We can see that RPA@PBE gives a nonsensical binding energy of 4.574 eV for the default tight setting and the most diffuse aug-cc-pV5Z set investigated in this work. For cc-pV5Z and NAO-VCC-5Z, the default ‘tight’ setting offers plausible results, but still deviate by 117 and 38 meV, respectively. Such errors in cc-pV5Z can be reduced to about 4 meV with tighter grid setting of *outer_grid*=590 with which a numerically converged result (error within ± 1 meV) is achieved for NAO-VCC-5Z. *outer_grid*=770 is a safe setting for cc-pV5Z, and *outer_grid*=974 for aug-cc-pV5Z. The tier-*n* basis sets behave similar to NAO-VCC-*n*Z.

Consistent with previous work [11, 57], these results demonstrate that the default ‘*outer_grid*=434’ is good enough for conventional DFAs and HF. The PBE and HF@PBE values do not change (uncertainty within ± 1 meV) by increasing the angular grids on faraway shells.

References

- [1] Furche F 2008 *J. Chem. Phys.* **129** 114105
- [2] Furche F and Van Voorhis T 2005 *J. Chem. Phys.* **122** 164106
- [3] Harl J and Kresse G 2009 *Phys. Rev. Lett.* **103** 056401
- [4] Ren X, Rinke P and Scheffler M 2009 *Phys. Rev. B* **80** 045402
- [5] Scuseria G E, Henderson T M and Sorensen D C 2008 *J. Chem. Phys.* **129** 231101
- [6] Grüneis A, Marsman M, Harl J, Schimka L and Kresse G 2009 *J. Chem. Phys.* **131** 154115
- [7] Harl J, Schimka L and Kresse G 2010 *Phys. Rev. B* **81** 115126
- [8] Heßelmann A 2011 *J. Chem. Phys.* **134** 204107
- [9] Hesselmann A and Göerling A 2010 *Mol. Phys.* **108** 359–72
- [10] Paier J, Janesko B G, Henderson T M, Scuseria G E, Grüneis A and Kresse G 2010 *J. Chem. Phys.* **132** 094103
Paier J, Janesko B G, Henderson T M, Scuseria G E, Grüneis A and Kresse G 2010 *J. Chem. Phys.* **133** 179902
(erratum)
- [11] Paier J, Ren X, Rinke P, Scuseria G E, Grüneis A, Kresse G and Scheffler M 2012 *New J. Phys.* **14** 043002
- [12] Ren X, Rinke P, Joas C and Scheffler M 2012 *J. Mater. Sci.* **47** 7447
- [13] Ren X, Rinke P, Scuseria G E and Scheffler M 2013 *Phys. Rev. B* **88** 035120
- [14] Ren X, Tkatchenko A, Rinke P and Scheffler M 2011 *Phys. Rev. Lett.* **106** 153003
- [15] Toulouse J, Gerber I C, Jansen G, Savin A and Ángyán J G 2009 *Phys. Rev. Lett.* **102** 096404
- [16] Harl J and Kresse G 2008 *Phys. Rev. B* **77** 045136
- [17] Zhu W, Toulouse J, Savin A and Ángyán J G 2010 *J. Chem. Phys.* **132** 244108
- [18] Eshuis H, Bates J and Furche F 2012 *Theor. Chem. Account.* **131** 1084
- [19] Schimka L, Harl J, Stroppa A, Grüneis A, Marsman M, Mittendorfer F and Kresse G 2010 *Nature Mater.* **9** 741
- [20] Casadei M, Ren X, Rinke P, Rubio A and Scheffler M 2012 *Phys. Rev. Lett.* **109** 146402
- [21] Eshuis H and Furche F 2012 *J. Chem. Phys.* **136** 084105
- [22] Furche F 2001 *Phys. Rev. B* **64** 195120
- [23] Janesko B G, Henderson T M and Scuseria G E 2009 *J. Chem. Phys.* **130** 081105
- [24] Noga J, Kutzelnigg W and Klopper W 1992 *Chem. Phys. Lett.* **199** 497–4
- [25] Tew D P, Klopper W and Helgaker T 2007 *J. Comput. Chem.* **28** 1307
- [26] Hättig C, Klopper W, Khn A and Tew D P 2012 *Chem. Rev.* **112** 4–74
- [27] Kong L, Bischoff F A and Valeev E F 2012 *Chem. Rev.* **112** 75–107

- [28] Kutzelnigg W and Klopper W 1991 *J. Chem. Phys.* **94** 1985–2001
- [29] Handy N C 1971 *Mol. Phys.* **817** 817
- [30] Umezawa N and Tsuneyuki S 2004 *Phys. Rev. B* **69** 165102
- [31] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566–9
- [32] Lester W A Jr, Mitas L and Hammond B 2009 *Chem. Phys. Lett.* **478** 1–10
- [33] Luechow A and Anderson J B 2000 *Adv. Quantum Chem.* **51** 501
- [34] Nguyen H-V and de Gironcoli S 2009 *Phys. Rev. B* **79** 205114
- [35] Nguyen H-V and Galli G 2010 *J. Chem. Phys.* **132** 044109
- [36] Nakatsuji H, Nakashima H, Kurokawa Y and Ishikawa A 2007 *Phys. Rev. Lett.* **99** 240402
- [37] Fabiano E and Della Sala F 2012 *Theor. Chem. Account.* **131** 1278–8
- [38] Feller D, Peterson K A and Grant Hill J 2011 *J. Chem. Phys.* **135** 044102
- [39] Grüneis A, Marsman M and Kresse G 2010 *J. Chem. Phys.* **133** 074107
- [40] Shepherd J J, Grüneis A, Booth G H, Kresse G and Alavi A 2012 *Phys. Rev. B* **86** 035111
- [41] Varandas A J C 2007 *J. Chem. Phys.* **126** 244105
- [42] Björkman T, Gulans A, Krasheninnikov A V and Nieminen R M 2012 *Phys. Rev. Lett.* **108** 235502
- [43] Aulbur W G, Jönsson L and Wilkins J W 2000 *Solid State Phys.* **54** 1
- [44] Hedin L 1965 *Phys. Rev.* **139** A796
- [45] Onida G, Reining L and Rubio A 2002 *Rev. Mod. Phys.* **74** 601
- [46] Rinke P, Qteish A, Neugebauer J and Scheffler M 2008 *Phys. Status Solidi b* **245** 929
- [47] Berger J A, Reining L and Sottile F 2010 *Phys. Rev. B* **82** 041103
- [48] Bruneval F and Gonze X 2008 *Phys. Rev. B* **78** 085125
- [49] Dunning T H 1989 *J. Chem. Phys.* **90** 1007–23
- [50] Kendall R A, Dunning T H and Harrison R J 1992 *J. Chem. Phys.* **96** 6796–806
- [51] Woon D E and Dunning T H 1993 *J. Chem. Phys.* **99** 1914–29
- [52] Peterson K A and Dunning T H 2002 *J. Chem. Phys.* **117** 10548
- [53] Woon D E and Dunning T H 1995 *J. Chem. Phys.* **103** 4572–85
- [54] Van Lenthe E and Baerends E J 2003 *J. Comput. Chem.* **24** 1142–56
- [55] Velde G t, Bickelhaupt F M, Baerends E J, Guerra C F, Gisbergen S J A v, Snijders J G and Ziegler T 2001 *J. Comput. Chem.* **22** 931–67
- [56] Averill F W and Ellis D E 1973 *J. Chem. Phys.* **59** 6412–8
- [57] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 *Comput. Phys. Commun.* **180** 2175–96
- [58] Delley B 1990 *J. Chem. Phys.* **92** 508–17
- [59] Koepf K and Eschrig H 1999 *Phys. Rev. B* **59** 1743–57
- [60] Soler J M, Artacho E, Gale J D, García A, Junquera J, Ordejón P and Sánchez-Portal D 2002 *J. Phys. Condens. Matter* **14** 2745–79
- [61] Talman J D 2007 *Int. J. Quantum Chem.* **107** 1578–84
- [62] Zunger A and Freeman A J 1977 *Phys. Rev. B* **15** 4716–37
- [63] Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, Reuter K and Scheffler M 2012 *New J. Phys.* **14** 053020–60
- [64] Feller D 1992 *J. Chem. Phys.* **96** 6104–14
- [65] Feller D and Sordo J J A 2000 *Chem. Phys.* **113** 485–93
- [66] Xantheas S S and Dunning T H 1993 *J. Phys. Chem.* **97** 18–9
- [67] Hill R N 1985 *J. Chem. Phys.* **83** 1173
- [68] Halkier A, Helgaker T, Jorgensen P, Klopper W, Koch H, Olsen J and Wilson A K 1998 *Chem. Phys. Lett.* **286** 243–52
- [69] Helgaker T, Klopper W, Koch H and Noga J 1997 *J. Chem. Phys.* **106** 9639–46
- [70] Klopper W, Bak K L, Jorgensen P, Koch H, Olsen J and Helgaker T 1999 *J. Phys. B: At. Mol. Opt. Phys.* **32** R103
- [71] Weigend F, Furche F and Ahlrichs R 2003 *J. Chem. Phys.* **119** 12753

- [72] Ahlrichs R, Bär M, Häser M, Horn H and Kölmel C 1898 *Chem. Phys. Lett.* **162** 165
- [73] Jensen F 2005 *Theor. Chem. Account.* **113** 267
- [74] Eschrig H and Bergert I 1978 *Phys. Status Solidi b* **90** 621
- [75] Ozaki T and Kino H 2004 *Phys. Rev. B* **69** 195113
- [76] Raffenetti R C 1973 *J. Chem. Phys.* **59** 5936–49
- [77] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865–8
- [78] Nocedal J 2006 *Numerical Optimization* 2nd edn (Berlin: Springer)
- [79] Freeman D L 1977 *Phys. Rev. B* **15** 5512
- [80] Goerigk L and Grimme S 2011 *Phys. Chem. Chem. Phys.* **13** 6670
- [81] Grimme S, Steinmetz M and Korth M 2007 *J. Org. Chem.* **72** 2118–26
- [82] Repasky M P, Chandrasekhar J and Jorgensen W L 2002 *J. Comput. Chem.* **23** 1601–22
- [83] Sattelmeyer K W, Tirado-Rives J and Jorgensen W L 2006 *J. Chem. Phys. A* **110** 13551–9
- [84] Jurečka P, šponer J, Černý J and Hobza P 2006 *Phys. Chem. Chem. Phys.* **8** 1985–93
- [85] Takatani T, Hohenstein E G, Malagoli M, Marshall M S and Sherrill C D 2010 *J. Chem. Phys.* **132** 144104
- [86] Schwenkes D and Truhlar D 1985 *J. Chem. Phys.* **82** 2418–26
- [87] Sheng X W, Mentel L, Gritsenko O V and Baerends E J 2011 *J. Comput. Chem.* **32** 2896–901
- [88] Bauschlicher C W, Langhoff S R and Taylor P R 1988 *J. Chem. Phys.* **88** 2540–6
- [89] Fast P L and Truhlar D G 1999 *J. Phys. Chem. A* **103** 3802–3
- [90] Hill J G 2013 *Int. J. Quantum Chem.* **113** 21–34
- [91] Baker J, Andzelm J, Scheiner A and Delley B 1994 *J. Chem. Phys.* **101** 8894
- [92] Becke A 1988 *J. Chem. Phys.* **88** 2547
- [93] Delley B 1995 *J. Comput. Chem.* **17** 1152
- [94] Lebedev V and Laikov D 1999 *Dokl. Math.* **59** 477
- [95] Lebedev V and Vychisl Z 1975 *Mat. Mat. Fiz.* **15** 48
- [96] Lebedev V and Vychisl Z 1976 *Mat. Mat. Fiz.* **16** 293
- [97] Treutler O and Ahlrichs R 1995 *J. Chem. Phys.* **102** 102