Thermochemical evaluation of adaptive and fixed density functional theory quadrature schemes

Andreas Hesselmann,[1, a)] Hans-Joachim Werner,[1, b)] and Peter J. Knowles[2, c)]

[1)] *Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany*

[2)] *School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK*

(Dated: November 18, 2022)

A systematic study is made of the accuracy and efficiency of a number of existing quadrature schemes for molecular Kohn-Sham Density-Functional Theory (DFT), using 408 molecules and 254 chemical reactions. Included are the fixed SG-x (x=0-3) grids of Gill *et al.*, Dasgupta and Herbert, the 3-zone grids of Treutler and Ahlrichs, a fixed five-zone grid implemented in Molpro, and a new adaptive grid scheme. While all methods provide a systematic reduction of errors upon extension of the grid sizes, significant differences are observed in the accuracies for similar grid sizes with various approaches. For the tests in this work the SG-x fixed grids are less suitable to achieve high accuracies in the DFT integration, while our new adaptive grid performed best among the schemes studied in this work. The extra computational time to generate the adaptive grid scales linearly with molecular size and is negligible compared with the time needed for the self-consistent field (SCF) iterations for large molecules. A comparison of the grid accuracies using various density functionals shows that meta-GGA functionals need larger integration grids than GGA functionals to reach the same degree of accuracy, confirming previous investigations of the numerical stability of meta-GGA functionals. On the other hand, the grid integration errors are almost independent of the basis set, and the basis set errors are mostly much larger than the errors caused by the numerical integrations, even when using the smallest grids tested in this work.

[a)] Electronic mail: hesselmann@theochem.uni-stuttgart.de

[b)] Electronic mail: werner@theochem.uni-stuttgart.de

[c)] Electronic mail: KnowlesPJ@cardiff.ac.uk

I. INTRODUCTION

Kohn-Sham density functional methods (KS DFT)[1,2] have emerged as the most popular approach to describe electronic structure during the past decades.[3] The reason for this lies in the fact that KS methods are much simpler than most quantum chemical wave function methods. They can be easily implemented and exhibit a moderate scaling of $\mathcal{O}(\mathcal{N}^3)$ (which can be brought down to linear scaling[4–6]) with the system size $\mathcal{N}$, so that KS calculations can be performed for very large systems (in the range of 100 to 1000 atoms) with modern computer hardware. However, unlike ab initio wave function methods KS DFT methods lack the property of being systematically improvable[7,8] due to the fact that the underlying exchange-correlation functional is unknown and must be approximated. As a consequence, care needs to be exercised when applying DFT methods for calculating certain molecular properties, since large errors might occur, depending on the functional chosen.[7] The most crucial shortcomings of original formulations of DFT functionals, however, are nowadays understood and can be corrected with various approaches that have been developed over the years. These involve methods to correct the dispersion energy problem,[9] the overestimation of polarizabilities of long-conjugated acene molecules,[10] the inability to describe excitation energies of charge-transfer[11] or Rydberg character,[12] and the incorrect description of band structures of periodic systems.[13] A large number of density functionals is available in most software implementations, including in Molpro's internal library[14] and in the library libxc.[15]

The basic quantities that need to be calculated in a (ground-state) KS calculation are the exchange-correlation (xc) functional

$$E_{\mathrm{xc}} = \int \varepsilon_{\mathrm{xc}}[\rho, \nabla\rho, \nabla^2\rho, \dots](\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{1}$$

and the exchange-correlation potential, which is normally represented in terms of matrix elements of atom centered Gaussian type basis functions (GTOs) $\chi_\mu(\mathbf{r})$:[16,17]

$$[v_{\mathrm{xc}}]_{\mu\nu} = \int \chi_\mu(\mathbf{r}) v_{\mathrm{xc}}(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} = \int \chi_\mu(\mathbf{r}) \frac{\delta E_{\mathrm{xc}}}{\delta\rho(\mathbf{r})} \chi_\nu(\mathbf{r}) d\mathbf{r} \tag{2}$$

As displayed in Eq. (1) the xc functional is usually described in terms of an xc energy density function $\varepsilon_{\mathrm{xc}}$ that depends on the electron density $\rho(\mathbf{r})$ and its derivatives. Density functional methods can then be categorised with respect to the quantities that are involved, namely local density approximation (ρ), generalised gradient approximation (ρ, $\nabla\rho$), or (ρ, $\nabla\rho$, τ)

(LDA, GGA, and meta-GGA methods, respectively), the lattermost sometimes containing the Laplacian of the density instead of the local kinetic energy density (cf. Ref. 18) because both quantities are related to each other.[8] Two of the central steps in KS programs are therefore the evaluation of Eqs. (1) and (2). In general, these integrals cannot be evaluated analytically. Therefore most DFT programs based on local basis functions make use of numerical quadrature techniques.

Due to the fact that the integrand is dominated by cusps at the atomic positions, the most common approach is to first decompose the whole molecular volume into atomic cells and then to define integration grids and integration weights for each atom individually. This was first proposed in the ground-breaking work of Becke.[19] The simplest approach is to choose non-overlapping discrete cells such that the atomic regions are strictly separated from each other. Becke found, however, that a much more accurate description of the integrals is achieved if one allows the atomic cells to overlap, namely, by choosing fuzzy overlapping cells instead of discrete ones.[19]

Rodriguez *et al.*[20] derived an alternative numerical integration scheme that is based on transformed sparse grids and does not follow the standard recipe of Becke. Furthermore, in valence-only atomic-orbital codes employing pseudopotentials the integrals can be evaluated with uniform 3D grids.[21] Such approaches are also used in combination with all-electron approaches in the multi-resolution methods described in Refs. 22–24.

There also exist a number of grid-free DFT methods where the integration problem has been solved either by fitting various powers of the density to linear combinations of atom-centered functions,[25,26] or by exploiting the fact that any matrix element of a function of the density can be written in terms of the corresponding eigenvalues and eigenvectors of the matrix

$$M_{\mu\nu} = \int \chi_\mu(\mathbf{r})\rho(\mathbf{r})\chi_\nu(\mathbf{r})d\mathbf{r}. \tag{3}$$

With GTOs, this can be computed analytically from a contraction of four centre overlap integrals and the density matrix.[27–31] The latter approach by Zheng *et al.*[27,28] and Glaesemann *et al.*[29–32] involves, however, resolution of the identity (RI) approximations, and therefore the error of the method depends on the size of the orbital and RI basis sets. Glaesemann and Gordon have designed auxiliary basis sets which minimize the errors arising from the RI.[31] It was found that the errors can be nicely reduced for LDA type functionals with

such adapted auxiliary basis sets. However, noticable errors still remained for the Becke88 exchange functional,[33] even if extended auxiliary basis sets are used.[31]

In the current work we use numerical quadrature techniques using fuzzy cells as originally introduced by Becke[19] and further improved in later works.[4,34–36] For each cell, the integration grid is conventionally generated from combinations of radial[34,36–44] and angular[34,36,45–51] grids. The choice of these grids is essential for the accuracy and efficiency. This will be reviewed and discussed in section II.

Taking just the direct product of a radial and angular grid may result in an enormous number of grid points for larger systems. It is therefore important to "prune" the grids by omitting points that hardly affect the accuracy. This can be achieved either by using different angular grids for each radial point (or ranges of them) or vice versa. Various "fixed" and "adaptive" grid pruning schemes will be discussed in section III. In the fixed grid methods[35,36,39,52–55] the number of angular grid points is predetermined for ranges of radial grid points, while in the adaptive methods[56,57] the angular grid is adjusted on the fly for each radial grid point according to some accuracy criterion. In Section IV a new adaptive grid method which is particularly efficient and accurate is proposed.

In spite of the large number of developments to improve the numerical quadrature for DFT methods over the years it seems that, to the best of our knowledge, a comprehensive work that analyses the accuracy of the various quadrature grids on equal footing has not been published so far. Very recently, Lehtola and Marques have studied the accuracy of various grid schemes.[58] They found considerable variation in the grid sensitivity of density functionals. Such behavior has been known before to exist for certain classes of functionals.[8,59]

Extensive benchmark calculations using several grid schemes that are implemented in the Molpro program[60–62] will be presented in section V. We will compare grid errors of absolute and relative energies for various functionals and basis sets. The effect on relative energies will be investigated for 254 chemical reactions, which are taken from the GMTKN30 database by Grimme *et al.*.[63] Finally, section VI summarizes the results of this work.

II. NUMERICAL INTEGRATION METHODS IN DFT

The molecular integral over a multi-center quantity, such as the exchange-correlation energy density $F(\mathbf{r})$

$$I = \int F(\mathbf{r}) d\mathbf{r} \tag{4}$$

is commonly calculated by decomposing the molecular space into fuzzy cells,[19] each associated with one of the atoms A

$$\begin{aligned} F(\mathbf{r}) &= \sum_A w_A(\mathbf{r}) F(\mathbf{r}) = \sum_A F_A(\mathbf{r}) \\ \sum_A w_A(\mathbf{r}) &= 1, \qquad w_A(\mathbf{r}) \geq 0, \end{aligned} \tag{5}$$

where w_A is an atomic weight function that is equal to one when $\mathbf{r}$ is close to the nucleus of A, and decays to zero the further the point lies away from the atomic center. Using Eq. (5) the integral in Eq. (4) then decomposes into atomic contributions

$$I = \sum_A I_A, \tag{6}$$

which can be evaluated in a spherical coordinate system centered on A using

$$I_A = \int_0^\infty dr_A \int_0^\pi d\theta_A \int_0^{2\pi} d\phi_A F_A(r_A, \theta_A, \phi_A) r_A^2 \sin\theta_A. \tag{7}$$

Due to the complicated form of the integrand F_A in DFT, which can depend on powers of the density and its derivatives, it is convenient to evaluate Eq. (7) by the means of numerical quadrature. The radial and angular integrations can be done successively to give:

$$I_A = \sum_{i=1}^{N_{\text{rad}}^A} w_i^{\text{rad}} \sum_{j=1}^{N_{\text{ang}}^{A,i}} w_j^{\text{ang}} F_A(r_{A,i}, \theta_{A,j}, \phi_{A,j}), \tag{8}$$

where N_{rad}^A denotes the number of radial grid points, $N_{\text{ang}}^{A,i}$ the number of angular grid points at center A for a radial point i, and $w_i^{\text{rad}}, w_j^{\text{ang}}$ are the corresponding quadrature weights. In pruned grid methods, $N_{\text{ang}}^{A,i}$ is a function of the radial grid index i, as discussed in section III. Note that the angular grid may be different for each radial grid point i, as indicated in Eq. (8) by the upper bound $N_{\text{ang}}^{A,i}$ of the summation over j. Thus, the weights w_j^{ang} and the associated angles $\theta_{A,j}$ and $\phi_{A,j}$ depend implicitly on i. Since the shape of F_A can have characteristic features in certain spatial regions, such as the valence region in molecular

systems, a number of quadrature methods have been developed that are designed to improve the accuracy of the integration compared to standard schemes such as Gauss-Legendre or Gauss-Chebyshev methods.

A. Partitioning into atomic contributions

The partitioning of molecular integrals into atomic contributions sketched in Eq. (4) is a crucial step in DFT integration methods because it involves atomic weight functions that may be chosen arbitrarily apart from the aforementioned constraint that $w_A(\mathbf{r}) \underset{r \to r_A}{=} 1$ and $w_A(\mathbf{r}) \underset{|r-r_A| \gg 0}{=} 0$ ($w_A \geq 0$ otherwise). As argued by Becke[19] as well as Treutler and Ahlrichs,[36] a general further criterion for the derivation of the function w_A is that it should be smooth between the transition from the nucleus of A to the external region, because it is coupled directly to the smoothness of the function $w_A \cdot F$ to be integrated.

A popular scheme to decompose the molecular volume into atomic cells is due to Becke which consists of two steps.[19] In the first step the molecular space is partitioned into non-overlapping Voronoi polyhedra using the functions

$$P_A(\mathbf{r}) = \prod_{B \neq A} s(\mu_{AB}) \tag{9}$$

with

$$s(\mu_{AB}) = \begin{cases} 1 & -1 \leq \mu_{AB} \leq 0 \\ 0 & 0 < \mu_{AB} \leq 1 \end{cases} \tag{10}$$
$$\mu_{AB} = \frac{r_A - r_B}{R_{AB}}$$

where r_A, r_B denote the distances from the nuclei and R_{AB} is the distance between the nuclei of atoms A and B, respectively. The function P_A in Eq. (9) obeys the constraints that $P_A(\mathbf{r}) \approx 1$ if r_A is small and that it goes to zero if r_A is large. Becke introduced an approach to soften the boundaries of the resulting polyhedra to create fuzzy cells.[19] The step discontinuity at $\mu_{AB} = 0$ is softened, and rendered differentiable, by replacing s in Eq.

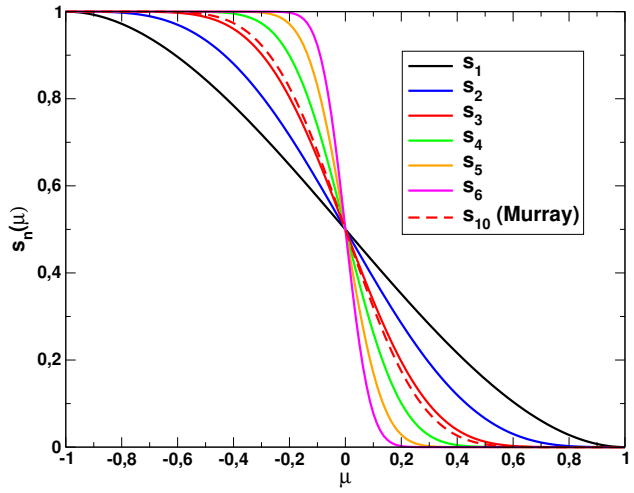


Figure 1. The step function $s_n(\mu)$ by Becke (Eq. (11)) for different recursion levels n. The corresponding step function by Murray (Eq. (15)) at the default iteration level $n = 10$ is displayed by the red dashed line.

(10) by

$$\begin{aligned} f_0(\nu) &= \nu, \quad -1 \le \nu \le 1 \\ f_{n+1}(\nu) &= \frac{3}{2} f_n(\nu) - \frac{1}{2} f_n(\nu)^3, \\ s_n(\mu_{AB}) &= \frac{1}{2}(1 - f_n(\mu_{AB})). \end{aligned} \tag{11}$$

With increasing n, $s_n(\mu_{AB})$ approaches the step function in Eq. (10), as shown in Fig. 1. A sufficiently large value of n is required to achieve effective partitioning; however, the sharp step that would arise from very large n leads to significant oscillations in the overall integral when grid points move from the cell of one atom to another as a consequence of a change in molecular geometry. Becke found that $n = 3$ is appropriate for general applications,[19] and this was adopted by other authors as well.[34,36,56,57]

A disadvantage of the partitioning method described above is that it does not take into account the atomic sizes for heteronuclear molecular systems. Becke has therefore modified the step function by an atomic size adjustment correction defined by

$$\mu_{AB} \to \mu_{AB} + a_{AB}(1 - \mu_{AB}^2) \tag{12}$$

with

$$a_{AB} = \frac{u_{AB}}{u_{AB}^2 - 1}$$
$$u_{AB} = \frac{\chi_{AB} - 1}{\chi_{AB} + 1}$$
$$\chi_{AB} = \frac{R_A}{R_B} \tag{13}$$

where R_A, R_B denote the Bragg-Slater radii of atoms A, B. The size adjustment scheme of Eq. (12) has been investigated by Treutler and Ahlrichs[36] and was found to pose a problem for certain chemical systems, such as magnesocene. They have therefore modified the scheme above by replacing χ_{AB} in Eq. (13) with the corresponding square root over the ratio of the Bragg radii of atoms A and B

$$\chi_{AB} = \sqrt{\frac{R_A}{R_B}}. \tag{14}$$

We will denote this approach as the Treutler scheme in in the following, to distinguish it from the original size-adjustment method by Becke given in Eqs. (12)-(13).

Based on earlier work by Handy and Boys,[64] Murray *et al.*[34] proposed an alternative smoothing function that is derived from the derivative

$$\frac{ds}{d\mu} = A_{m_\mu}(1 - \mu^2)^{m_\mu} \tag{15}$$

which is zero at the end points $\mu = \pm 1$. The value of A_{m_μ} in Eq. (15) is chosen such that $s(-1) = 1$ and $s(1) = 0$. This leads to a power series expansion of the step function in $(1 - \mu^2)$ which is given by

$$s_{m_\mu}(\mu) = \frac{1}{2} - \frac{1}{2}\mu - \frac{1}{2}\mu \sum_{k=1}^{m_\mu} \left[(1 - \mu^2)^k \prod_{l=0}^{k-1} f(l) \right] \tag{16}$$

with

$$f(l) = \frac{2l + 1}{2l + 2}. \tag{17}$$

Measuring the errors in the integrated density and the Slater-Dirac functional of the N_2 molecule for different exponents m_μ in Eq. (15) Murray *et al.* observed that the lowest errors are achieved for exponent values in the range between $m_\mu = 4$ to $m_\mu = 12$. They concluded that the setting $m_\mu = 10$ is most suitable because it turned out that larger exponents of m_μ in Eq. (15) gave better accuracies for larger grid sizes. This scheme is used

in most calculations of this work. In figure 1 the $m_\mu = 10$ function of Murray is compared with the corresponding Becke step functions from Eq. (11). As can be seen, the Murray step function is very similar and only slightly sharper than the corresponding Becke step function at its default level of $n = 3$.

Once the step functions have been determined and inserted in Eq. (9) to obtain the fuzzy atomic cells, the atomic weight functions in Eq. (5) can be determined with

$$w_A(\mathbf{r}) = \frac{P_A(\mathbf{r})}{\sum_B P_B(\mathbf{r})} \tag{18}$$

where the sum in the denominator goes over all atoms.

While the Becke partitioning scheme is still most widely used in computer implementations of DFT quadratures, it should be noted that alternative schemes were developed by Stratmann *et al.*[4] and more recently by Laqua and Ochsenfeld[65] in order to overcome some shortcomings of the Becke scheme for certain systems as for example weakly bound complexes. Since the current work mainly focusses on the adaptive angular grid generation these methods are not considered here.

B. Radial quadrature

The radial integration part of the quadrature, which can be written generally as

$$\mathcal{R} = \int_0^\infty F(r) r^2 dr, \tag{19}$$

is often done such that the interval $[0, \infty]$ is mapped on a finite interval $[a, b]$ so that standard quadrature schemes like those by Gauss-Chebyshev[19,36] or Euler-Maclaurin[34] can be used. Many different radial integration methods have been developed with the aim of optimizing the accuracy of the radial grid for different integral types.[34,36,37,39–44] The specific choice of this mapping can have a great impact on the accuracy of the integration, because the radial points should be distributed such that the core and bonding regions are adequately described for various chemical situations, such as atoms in different oxidation states or environments.

Early radial quadrature schemes focussing on this aspect were developed by Becke,[19] Murray *et al.*,[34] Treutler and Ahlrichs,[36] Gill et al.,[35] and Mura and Knowles.[37] These quadrature

methods are still used as the standard radial integration schemes in many quantum chemistry programs including Turbomole,[66] Molpro,[60–62] and Q-Chem.[67]

More recently Gräfenstein and Cremer[41] have developed a radial grid scheme that can explicitly augment the grid in certain local regions. The motivation of their work was to be able to improve the accuracy of DFT integrations when meta-GGA functionals are employed, since then the accuracy often deteriorates due to an insufficient radial resolution in the valence region. Gräfenstein *et al.* showed that their method is able to yield the same degree of accuracy for the argon dimer than a more than two times larger non-augmented grid when the Van Voorhis-Scuseria functional[68] is used in the calculation. The locally augmented radial grids by Gräfenstein *et al.* are therefore suitable alternatives to the more commonly used radial grids if meta-GGA functionals are used. In this work we do not consider the method of Gräfenstein *et al.* since we mainly focus on the comparison of different angular grid pruning methods that are used in conjunction with various standard radial grids.

Other notable radial quadrature schemes were developed by Kakhiani *et al.*,[42] Lindh *et al.*,[38] and Gill and Chien.[39] The MultiExp grid by Gill and Chien[39] is used within the fixed SG-0 standard grid developed by the same authors.[52] The method was analysed by El-Sherbiny and Poirier[40] along with the radial quadratures from Becke,[19] Treutler and Ahlrichs,[36] and the radial grid that is used in the SG-1 method of Gill *et al.*.[35] It was observed that the MultiExp grids provide the best accuracies for the KS energies for a large range of different molecules.[40]

Table I. Mappings for radial quadrature schemes[a)]

method	mapping	roots	used in	ref.
Treutler-Ahlrichs (TA)	$r_i = \frac{\alpha}{\ln 2}(1+x_i)^m \ln\left(\frac{2}{1-x_i}\right)$	$x_i = \cos\left(\frac{i\pi}{N+1}\right)$	3-zone and 5-zone	36
Mura-Knowles (MK)	$r_i = -\alpha \ln(1-x_i)^m$	$x_i = \frac{i}{N}$	AMG	37
MultiExp	$r_i = -\alpha \ln x_i$	see Ref. 52	SG-0	39,52
Euler-Maclaurin (EM)	$r_i = \alpha \frac{x_i^m}{(1-x_i)^m}$	$x_i = \frac{i}{N+1}$	SG-1	34,35
DoubleExp	$r_i = \exp(\alpha x_i - e^{-x_i})$	$x_i = -1 + (i-1)\frac{2}{N-1}$	SG-2,3	43,44

a) The parameters α are atom-dependent scaling factors and are chosen different for each scheme, as discussed in the text

We now describe the radial quadrature methods which are used in this work in more detail.

They are summarised in Table I, which gives for each case the mapping formulas to distribute an evenly spaced grid onto the atomic radial grid.

The Treutler-Ahlrichs (TA) mapping is combined with a second-kind Chebyshev scheme, and depends on a global exponent m as well as on atom-dependent parameters α (adopted from Becke's radial grids[19]) that were optimized for isolated atoms and have values between 0.8 and 2.0 (see Table I in Ref. 36). Best performance for molecular calculations was obtained with $m = 0.6$.

Mura and Knowles[37] proposed a quadrature scheme, termed MK scheme in the following, based on a logarithmic mapping function and Euler-Maclaurin quadrature. The mapping is defined by a scale parameter α and a power m that serves to adjust the relative weight of inner and outer regions. In this work, we adopt the α, m values recommended in Table IV in Ref. 37.

Gill *et al.* have combined the logarithmic transformation $r_i = -\alpha \ln x_i$ with the so called log-square quadrature to obtain the multiexponential (MultiExp) grid.[39,52] This method integrates certain linear combinations of exponentials exactly. Optimised values of α for atoms H-Cl of this method can be found in Table 1 in Ref. 52. The MultiExp scheme has been used as the radial quadrature scheme of the fixed SG-0 standard grid developed by Chien and Gill[52] and is further discussed in section III A. A downside of the MultiExp scheme is that the quadrature nodes cannot be obtained analytically. They must be determined by solving a set of linear equations, which become increasingly ill-conditioned as the number of grid points is increased.

The Euler-Maclaurin quadrature scheme developed by Murray *et al.* is based on a transformation that enforces the derivatives of the integrand to become zero at the end points (utilising the same approach that Murray *et al.* also used for deriving their novel step function, see section II A), thus yielding faster convergence to the exact integral.[34,64]

The scaling factors of α in the EM quadrature are chosen to be the Bragg-Slater radii. The EM quadrature has been used as the radial quadrature scheme for the fixed SG-1 standard grid by Gill et al..[35]

The double exponential (DoubleExp) quadrature scheme was developed by Mitani *et al.*.[43,44,69] They studied the convergence of the electron number N obtained by integrat-

ing over the density $\rho(\mathbf{r})$ (in the following "electron number integral")

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \tag{20}$$

and observed that the accuracy with the DoubleExp grids is less sensitive to grid sizes and atomic species than with the TA and MK radial grids.[44] However, Köster et al.[57] have found that a reliable numerical integration of the electron density is not sufficient to ensure the reliable numerical integration of the exchange-correlation potential. Nevertheless, Dasgupta and Herbert have adopted the DoubleExp method in their fixed SG-2,3 standard grids, which will also be considered in the tests in this work; see Ref. 53 and section III A.

In practical calculations the size of the radial grid needs to be sufficiently large to achieve a certain accuracy as chosen by the user. In principle this could be achieved by varying the grid size within an adaptive method that also optimises the angular grid to reach the desired degree of accuracy. However, it was noted by Termath *et al.*[70] as well as by Krack and Köster[56] that this can cause instabilities in the adaptive procedure, since the change of N_{rad}^{A} shifts all radial grid points of atom A. As is argued by Krack and Köster, adapting the radial grid is not necessary if enough radial points are provided in conjunction with an angular grid adaptation subsequently performed to reach the desired accuracy.[56] Krack *et al.*[56] proposed the following equation to relate the number of grid points N_{rad} to a requested tolerance T_{rad}

$$N_{\mathrm{rad}}(T_{\mathrm{rad}}) = \max\left(20, -15\log_{10}\left(T_{\mathrm{rad}}\right) + 5 \cdot \kappa_{\mathrm{row}}[Z] - 40\right)) \tag{21}$$

where $\kappa_{\mathrm{row}}[Z]$ stands for the row in the periodic table the atom belongs to. A similar formula is used in the Molpro program to compute the number of radial grid points:

$$N_{\mathrm{rad}}(T_{\mathrm{rad}}) = -10\log_{10}(T_{\mathrm{rad}}) + 15 \cdot k_{\mathrm{row}}[Z] - 30 \tag{22}$$

(rounded to the nearest integer number if the mantissa of T_{rad} is unequal to one). The factor $k_{\mathrm{row}}[Z]$ is dependent on the nuclear charge Z and is defined as

$$k_{\mathrm{row}}[Z] = \begin{cases} 0 & \text{if } Z < 3 \\ 1 & \text{if } 3 \leq Z \leq 10 \\ 2 & \text{if } 11 \leq Z \leq 18 \\ 3 & \text{if } Z > 18 \end{cases} \tag{23}$$

For typical threshold values of $T_{\mathrm{rad}} = 10^{-6}$ and $T_{\mathrm{rad}} = 10^{-12}$ this formula yields 45 and 105 grid points, respectively, for a first-row element like carbon (for comparison, the equation (21) by Krack *et al.* yields 55 and 145 grid points, respectively, for these threshold values). The radial grid is then determined using the method of Mura and Knowles[37] with $m = 3$, cf. Table I.

C. Angular quadrature

The angular grids are nowadays mostly generated by the quadrature method of Lebedev[45–48] that constructs spherical grids with octahedral and inversion invariance. The reason for this originates from the favorable size-cost ratio of Lebedev grids that can integrate spherical harmonics of order L with only approximately $(L+1)^2/3$ grid points. Lebedev grids up to the order of $L = 131$ (5810 grid points) can be computed using a program developed by Laikov.[50] Such high orders, however, are not needed in the framework of DFT integration. Our program Molpro currently has Lebedev grids implemented up to $L = 83$ (2354 grid points) and limits the maximum possible order to $L_{\mathrm{max}} = 59$ in all calculations presented in this work.

While the employment of the Lebedev quadrature for the angular integration part is found in practically any quantum chemistry program, separate sub-integrations over θ and ϕ can also be performed using Gauss-Legendre or Lobatto integration techniques.[34,36] The supplementation of the Lebedev grids with open-ended quadrature schemes is useful for high-precision calculations and has been implemented in some quantum chemistry codes.[60,66,71]

Treutler and Ahlrichs have developed a method to reduce the number of grid points of the Lobatto grid at a given order L such that the number of points increases only approximately as $L(L-1)/2$.[36] Schemes to reduce the number of angular grid points using symmetry have been discussed by Murray *et al.*.[34]

Daul and Daul[49] as well as Ahrens and Beylkin[51] have defined alternative angular quadrature methods whose rules go beyond the octahedral point groups of Lebedev. The spherical quadrature method by Ahrens and Beylkin[51] stays invariant within the icosahedral group and needs slightly more grid points than the Lebedev scheme to integrate spherical harmonics to a given order. To our knowledge the methods by Daul and Daul and Ahrens and Beylkin

have, however, not been investigated for DFT quadratures so far, and it is not clear whether they would offer a better performance regarding the trade of cost against accuracy.

In the current work we use only Lebedev grids.

III. GRID TRUNCATION METHODS

The computational effort for DFT calculations based on numerical quadrature methods described in section II and that employ local basis functions (see Eq. (2)) can be shown to scale linearly with the molecular size if the sparsity of underlying quantities[72] is exploited. Moreover, the methods can be parallelised straightforwardly over the loops over grid points. However, depending on the extent of the atomic radial and angular grids the total number of grid points can become enormous. For example, an integration grid with $N_{\mathrm{rad}} = 50$ and $N_{\mathrm{ang}} = 974$ ($L = 53$) for a molecule that consists of 10 atoms has about $5 \cdot 10^5$ points. Therefore brute-force integrations over full angular atomic grids can become the dominant part of the computational cost compared to the calculation of the Coulomb and exchange integrals required to build the Fock matrix. A simple strategy to reduce the number of grid points adopted in many DFT programs is therefore to truncate those grid points with very small weights, e.g., grid points very close to the nucleus or grid points for which the scaling factor from the Voronoi partitioning becomes very small. The cutoff threshold used for this approach should, however, not be chosen too loose since the grid errors can become large if the core regions are not accurately integrated. Due to this fact, only a small number of grid points can be usually saved in practice with this method.

Fortuitously, high Lebedev orders for the angular integration are not needed at all radial points to achieve high accuracies for the integration. For example, in calculations of ground state energies one can assume that in the atomic core region the shape of the integrand is close to spherical, and therefore low Lebedev orders are sufficient. Since the radial grid is most dense in the short range region this can lead to large savings. It should be noted, however, that for other properties as for example hyperfine coupling constants[73] much more dense grids may be needed close to the nucleus to obtain accurate results. A number of possible grid pruning methods have been devised, which can be categorised into "fixed grid methods"[35,52,53] and "adaptive grid methods".[56,57] These will be described in sections III A

and III B, respectively.

A. Fixed grids

The first type of fixed standard grid denoted as SG-1 was developed by Gill *et al.*.[35] It is based on partitioning the atomic volume into five spherical shells separated by four spherical surfaces with radii $\gamma_{1-4}R$, where R denotes an atomic radius and γ are fixed scaling parameters that depend on the row in the periodic system of elements. Note that Gill *et al.* used the same atomic size parameter also within the EM radial grid scheme used for these grids (denoted as α in table I). Angular grids with different sizes can then be chosen for each region. Gill *et al.* used the Lebedev grids with 6, 38, 86, 194 and 86 points, respectively, taking into account that the points close to the nucleus (first region in the scheme) require only a low Lebedev order.[35] The scaling parameters of γ were chosen such that the deviation from the electronic energies obtained with a full (50,194) grid is minimised. Here the short-hand notation (50,194) means that the grid is built from 50 radial points (using the EM method) and 194 Lebedev grid points. As an example, for the carbon atom the SG-1 grid contains 3816 grid points, and a reduction of 60 percent is achieved relative to the (50,194) grid, which has 9700 grid points.

More recently, Chien and Gill[52] have developed another standard grid termed SG-0, which even more strongly reduces the number of grid points than SG-1, and which is designed for calculations on large molecules. The SG-0 grid uses the MultiExp[39] quadrature scheme and reduces the grid sizes roughly to one half of those in SG-1, i.e., an element of the 2nd and 3rd row is described by about 1500 grid points. As an example, the SG-0 grid for the carbon atom is defined by the notation

$$6^6 18^2 26^1 38^2 50^2 86^1 110^1 146^1 170^2 146^2 86^1 38^1 18^1$$

where the notation x^y indicates that the x-point Lebedev grid is used at y successive radial points. This means that a 6-point Lebedev grid is assigned to the first six radial points, an 18-point Lebedev grid to the next two radial points, etc.. The grid optimisation was performed by starting from the full parent (23,170) Euler-Maclaurin grid. Then the angular grid was pruned at each radial point as long as the energy difference stayed within an

acceptable range; see Ref. 52 for details. The SG-0 and SG-1 grids have been shown to yield acceptable results for total energies, atomisation energies or molecular geometries.[35,52]

Very recently Dasgupta and Herbert[53] have employed a similar pruning technique as for SG-0, i.e. using fixed sized radial grids and optimising the order of the angular grid assigned to each radial grid point. They employed the DoubleExp radial quadrature method by Mitani,[43,44] and generated two grids denoted SG-1 and SG-2. The grid sizes for carbon are 7790 and 17674 grid points, respectively. This means they are roughly twice (SG-2) and five times (SG-3) as large as the SG-1 grids. For SG-2 the radial grid size is $N_{\mathrm{rad}} = 75$ and for SG-3 $N_{\mathrm{rad}} = 99$, and so about 50 percent and 100 percent larger than the corresponding SG-1 grids, see Ref. 53.

These grids were designed for applications that require higher accuracy. For example, potential energy curves of rare gas dimers exhibit significant oscillations with some modern meta-GGA functionals if the integration grid is too coarse,[53,59] and for such functionals it is not recommended to use small standard grids. In a systematic study of atomisation energies, isomerization energies, geometrical parameters and harmonic vibrational frequencies it was shown that SG-2 and SG-3 yield very small errors relative to the unpruned (75,302) and (99,590) grids, respectively. Moreover, it could be shown that the artificial imaginary frequencies for certain vibrational modes of the ethynyl radical found with SG-0 and SG-1 for the M06-HF functional are removed by the extended SG-2 and SG-3 grids; see Table 6 in Ref. 53.

The 3-zone grid of Treutler and Ahlrichs[36] can also be regarded as a fixed grid with a different assignment of the Lebedev quadrature order L to the radial grid points. Here, the values of L are determined by the total number of radial quadrature points N_{rad} in three zones using

$$L = \begin{cases} L_{\mathrm{short}} & \text{for } r_i \text{ with } i \leq \frac{N_{\mathrm{rad}}}{3} \\ L_{\mathrm{middle}} & \text{for } r_i \text{ with } \frac{N_{\mathrm{rad}}}{3} < i \leq \frac{N_{\mathrm{rad}}}{2} \\ L_{\mathrm{long}} & \text{for } r_i \text{ with } i > \frac{N_{\mathrm{rad}}}{2} \end{cases} \tag{24}$$

This method is implemented in the Turbomole program[66] (using a corresponding second-kind Chebyshev radial quadrature[36]) and defines seven grid levels using the values of L_{short}, L_{middle}, L_{long} as shown in Table II.

The 3-zone grid by Treutler thus uses few angular functions in the region close to the nucleus,

Table II. Angular L values of the 3-zone grids

grid level	1	2	3	4	5	6	7
L_{short}	7	7	7	7	7	7	7
L_{middle}	11	17	17	17	17	17	17
L_{long}	17	23	29	35	41	53	59

a medium sized Lebedev order of L_{middle} for the valence region, and a large Lebedev order (mostly adjustable by the global grid level that can be chosen) for the far range. The default level of 3 of the 3-zone grid sets the number of radial grid points to $N_{\text{rad}} = 35$ for second row elements.[36]

Another kind of fixed DFT quadrature scheme denoted "5-zone grid" is implemented in the Molpro quantum chemistry program. It is similar to the one used in the Orca4 quantum chemistry package[54,55] but differs in the choice of the radial grid (see below). As in the case of the SG-1 grids, the atomic radial grids are divided into five domains that are defined through the numbers γ_1-γ_4. These are scaled by the Clementi radius R_{max}[74] of the outermost valence orbital to define the boundaries between the five partitioning segments. Each segment then is assigned the Lebedev order of

$$L_{\text{max}} - 18,\ L_{\text{max}} - 12,\ L_{\text{max}} - 6,\ L_{\text{max}},\ L_{\text{max}} - 6$$

from inner to outer. The maximum Lebedev order is computed from a grid level index k that can be varied in the range $k = 3 - 13$ using the formula

$$L_{\text{max}}(k) = 5 + 6 \cdot k \tag{25}$$

As an example, for the grids with indices $k = 4$ and $k = 9$, the Lebedev orders for the five radial regions are, respectively, $(11, 17, 23, 29, 23)$ and $(41, 47, 53, 59, 53)$. The size for the radial grid itself is computed by the formula of Krack and Köster shown in Eq. (21) and each grid level is assigned a particular T_{rad} threshold value that defines the size for each row of the periodic table according to eq. (21). This value is computed by the formula

$$T_{\text{rad}}(k) = 4.34 + 0.33 \cdot \min(0, k - 3) \tag{26}$$

which results to $N_{\text{rad}} = 35$ radial grid points for second row elements for the default level of $k = 4$. By default, the Gauss-Chebyshev grid as described in the work of Treutler and

Ahlrichs[36] is used. Note that in contrast to the Molpro implementation the Orca4 program uses a modified TA grid in which the atomic scaling factors are omitted,[55] as proposed by Krack and Köster.[56]

B. Adaptive angular grid generation

Unlike the fixed grid methods described in the previous section, adaptive grid methods try to take the chemical environment of an atom into account when pruning the grid. Therefore it can be expected that they deliver higher accuracies than the fixed grids at comparable grid sizes.

A central quantity in adaptive grid methods is the functional that is used for pruning the grid relative to a complete grid with the maximum number of angular functions per radial grid point. Krack and Köster (KK) assumed that the numerical integration of the overlap matrix may be a suitable measurement for the pruning error due to the fact that the overlap matrix also enters the calculation of the electron number integral[56] (cf. Eq. (20))

$$N = \int \rho(\mathbf{r}) d\mathbf{r} = \sum_{\mu\nu} P_{\mu\nu} \int \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} = \sum_{\mu\nu} P_{\mu\nu} S_{\mu\nu} \tag{27}$$

where $S_{\mu\nu}$ and $P_{\mu\nu}$ are the overlap and density matrices, respectively, in the atomic orbital basis χ_μ. Due to the observation that the convergence of the numerically integrated overlap matrix elements is directly related to their absolute values, and that large values often converge more slowly than small ones, Krack and Köster consider only the diagonal elements $S_{\mu\mu}$ of the overlap matrix to measure the accuracy for the pruned grid. The pruning method by KK then starts at small Lebedev orders L and increases L until the deviation of the angular part of the overlap integral for successive values of L is lower than some predefined threshold for all diagonal matrix elements; see Ref. 56 for more details. Later, Köster et al.[57] stated that a reliable numerical integration of the electron density is not sufficient to ensure reliable numerical integration of the exchange correlation potential. They therefore replaced $S_{\mu\mu}$ by the diagonal elements of the exchange correlation matrix $V_{\mu\mu}$. Since this depends on the density, two successive KS calculations are carried out. First an approximate grid is used to determine an approximate density, which is then used to compute $V_{\mu\mu}$ in the pruning step in the second calculation. It was concluded that the resulting adaptive scheme performs

very well, particularly for transition metal systems for which sub-microhartree accuracies are obtained with grids that are smaller by a factor of about 20 to 25 in size compared to full reference grids; see tables VI and VII of Ref. 57.

An alternative adaptive grid method, which is used in the Molpro program[60–62] and avoids the repeated KS calculations, will be outlined in section IV and benchmarked in section V.

Adaptations of the DFT quadrature grid to the chemical environment can also be made for the radial grid instead of the angular grid; see for example the method by Kakhiani *et al.*.[42] As noted in section II B, however, a simultaneous adaptation of radial and angular grids can lead to numerical instabilities,[56,70] and so one of the two grids should be kept fixed in an adaptive grid pruning scheme.

IV. THE ADAPTIVE MOLPRO GRID (AMG)

Our pruning approach is based on a model function $\tilde{\mathcal{I}}_{A_i}(L_i)$, which is evaluated for a given radial grid point r_i^A for increasing values of L_i until the error relative to a reference value is smaller than a value that is designed to lead to an error in the complete (radial-summed) integral that is smaller than a pre-defined threshold. This determines a specific order L_i for each radial grid point. In principle, the model should be the result of angular quadrature of the target density-functional kernel, together with full fuzzy-Voronoi weight factors, but this leads to computational costs for determining the grid that can easily exceed those from all remaining parts of the DFT calculation. We therefore approximate both the functional kernel and the Voronoi weight factors, with the aim that the model function should accurately estimate the real function, but in a way that it can be rapidly evaluated. Before discussing these approximations, we consider how to measure the sensitivity of quadrature errors to changes in the angular grid. The overall approach is to assert that the lowest error in the integral for a given grid size is likely to arise when the errors introduced by truncated angular quadrature at each of the radial points are approximately equal, rather than being concentrated in one part of the radial range. We then estimate the overall error in terms of these radial-point errors using statistical arguments, and invert the analysis to obtain a prescription for angular quadrature at each radial point targetting a given desired overall

error. We first define

$$I_{A_i}(L_i) = \sqrt{N^A_{\text{rad}}}\, w^{\text{rad}}_i \sum_j^{N^i_{\text{ang}}} \widetilde{w}_{ij}(L_i)\, F(\rho_{ij}), \quad i \in A \tag{28}$$

Here and in the following, indices i and j refer to radial and angular grid points, respectively. $i \in A$ means that index i only refers to grid points on atom A. Thus, all quantities carrying index i are specific to atom A, and for the sake of simplicity this will be implied in all subsequent equations. N^i_{ang} is the number of angular grid points for Lebedev order L_i. Here $F(\rho)$ is assumed to be the same density-functional kernel as used in the Kohn-Sham calculation, but for efficiency reasons it will be replaced by an approximate kernel $\tilde{F}(\rho)$ later on (cf. Sec. IV A).

$\rho_{ij} \equiv \rho(\mathbf{r}_{ij})$ are density values evaluated at the grid points $\mathbf{r}_{ij}$, which are defined by the position of atom A, the radius r^A_i (distance to center of A), and the corresponding Lebedev grid with order L_i. The adoption of the factor $\sqrt{N^A_{\text{rad}}}$ is informed by the statistical analysis below. The weights $\widetilde{w}_{ij}$ include the full Voronoi weight factors [cf. Eq. (18)] at each grid point,

$$\widetilde{w}_{ij}(L) = w^{\text{ang}}_j(L)\, \frac{P_A(\mathbf{r}_{ij})}{\sum_B P_B(\mathbf{r}_{ij})}, \tag{29}$$

and $w^{\text{ang}}_j(L)$ is the bare angular weight factor for grid point j in the Lebedev grid of order L. The total approximate integral is then

$$I(\mathbf{L}) = \sum_A I_A(\mathbf{L}^A) = \sum_A (N^A_{\text{rad}})^{-1/2} \sum_{i \in A} I_{A_i}(L_i) \tag{30}$$

The individual radial errors, $I_{A_i}(L_i) - I_{A_i}(\infty)$ are randomly distributed samples of a population with a mean of zero and standard deviation η. Then, by the central limit theorem,[75] their mean,

$$(N^A_{\text{rad}})^{-1} \sum_{i \in A} (I_{A_i}(L_i) - I_{A_i}(\infty)) = (N^A_{\text{rad}})^{-1/2}\, I_A(\mathbf{L}^A), \tag{31}$$

can be viewed as being drawn randomly from a population of mean zero and standard deviation $(N^A_{\text{rad}})^{-1/2}\eta$. Thus the standard error of I_A is η, and the corresponding extensive error of I is $N_{\text{atom}}\eta$. We then arrange that a fixed factor times the standard deviation of the individual radial point errors is less than a target intensive error measure ϵ by choosing each L_i such that $|I_{A_i}(L_i) - I_{A_i}(\infty)| \leq \epsilon$. If we assume the worst case that, once L_i is chosen, the local errors are distributed uniformly in the range $\pm\epsilon$, this condition actually leads to a

standard deviation of $\eta = \epsilon/\sqrt{3}$, meaning that I_A is expected to lie within $\pm\epsilon$ of the exact value with a probability of about 92%.

The functional in eq. (28) can in principle be used to determine an optimum angular grid order L_i for each grid point. However, in practice some approximations are introduced in order to minimize the computational effort and to recover linear scaling. First, rather than using in eq. (28) the same functional F as in the Kohn-Sham calculation, a simplified functional $\tilde{F}$ can be chosen. This will be discussed in section IV A. Second, in IV B we propose a simplified model functional, which leads to linear scaling. Finally, the detailed angular pruning procedure will be described in section IV C.

A. Choice of the trial integrand function

A simple choice for the trial integrand function $\tilde{F}$ is the Slater-Dirac (SD) exchange functional; because of its simplicity, it can be evaluated rapidly. However, it has been observed by Knizia[76] that with this choice often the number of angular grid points in the long range region is too small to obtain a good accuracy in the evaluation of typical modern GGA functionals, where the exchange energy density decays as reciprocal distance rather than exponentially. Therefore, instead we use the full PW92 exchange-correlation local density functional.[77] Furthermore, Knizia has supplemented the PW92 functional kernel with an additional term in order to sample integration of the density itself:

$$\tilde{F}(\rho) = -\frac{1}{2}\varepsilon_{\mathrm{xc}}^{\mathrm{PW92}}(\rho)\rho + \frac{3}{20}\exp\left(\frac{1}{72}\Big(\ln\rho\Big)^2\right)\rho^{7/6}, \tag{32}$$

where the exponent $\frac{7}{6}$ in the second term was chosen halfway between 1 (bare density) and $\frac{4}{3}$ (exchange functional). The second term in Eq. (32) becomes dominant for small density values and therefore further augments the importance of the angular grid for low density regions. Note that the negative sign for $E_{\mathrm{xc}}^{\mathrm{PW92}}$ ensures that both terms are positive and not mutually cancelling. We denote the angular integral of this kernel $\tilde{I}_{A_i}$, and work with the assumption that a quadrature scheme that obtains $\tilde{I}_{A_i}$ to a given accuracy will deliver the same accuracy in the target integral I_{A_i}.

B. Linear-scaling adaptive integration

The functional in eq. (28) with the integration kernel in eq. (32) works well for small molecules. Results for this approach are presented in the supplementary material. However, for large systems the computational effort for the pruning procedure becomes significant because it scales cubically with system size. We can obtain $\tilde{\mathcal{I}}_{A_i}(L_i) \approx \tilde{I}_{A_i}(L_i)$ as a more efficiently-calculated function for the adaptive determination of L_i in the following way. Denoting a general grid point as $\mathbf{r}_k$, the total density $\rho(\mathbf{r}_k)$ can be well approximated as

$$\rho(\mathbf{r}_k) = \sum_A \rho_A(\mathbf{r}_k), \tag{33}$$

where $\rho_A(\mathbf{r}_k)$ are the Slater atomic densities.[78,79] In Ref. 79 these are defined for elements up to $n = 6$ with effective principal numbers of $n^\star = 3.7, 4.0, 4.2$ for $n = 4, 5, 6$. The implementation in Molpro uses $n^\star = 4.4$ for seventh row elements based on a simple linear extrapolation of the $n^\star$ values given by Leach.[79] The advantage of using Slater's densities over tabulated numerical densities is that they are given analytically and can be easily computed at any given grid point.

Due to the summations over all atoms in eqs. (29) and (33), the evaluation of the functional I_{A_i} for all atoms A scales cubically with molecular size (Note that the evaluation of the P_A functions in Eq. (29) themselves involve products over all atoms (Eq. (9)) that needs to be taken into account for the scaling factor). It is possible to reduce the summations over the full number of atoms in Eq. (33) to those which are within a distance

$$r^A_{\text{cut}} = \sqrt{f_{\text{neigh}}}\, r^A_{\text{min}} \tag{34}$$

from atom A. Here r^A_{min} is the distance to the closest neighbor atom of A in the molecule and f_{neigh} is an empirical scaling factor, typically between 2 and 8. This means that Eq. (33) is approximated as

$$\rho(\mathbf{r}_{ij}) \approx \rho_A(\mathbf{r}_{ij}) + \sum_{B \neq A}^{N^A_{\text{neigh}}} \rho_B(\mathbf{r}_{ij}) \tag{35}$$

with N^A_{neigh} being the number of neighbor atoms defined by the cutoff r^A_{cut}. This should be a good approximation because the atomic densities decay exponentially with the distance from the center.

Limiting also the summation in Eq. (29) to the closest neighbor atoms is a less accurate approximation since the Voronoi weights decay more slowly, as discussed in section II A. The scaling can, however, be reduced to linear if the exact the functional I_{A_i} of Eq. (28) is replaced by an approximate model functional $\tilde{\mathcal{I}}_{A_i}$ which only includes one-center and two-center contributions ($i \in A$ is implied):

$$\tilde{\mathcal{I}}_{A_i}[L_i] = \sqrt{N^A_{\mathrm{rad}}}\, w^{\mathrm{rad}}_i \left[\sum_j^{N^i_{\mathrm{ang}}} w^{\mathrm{ang}}_j(L_i) \tilde{F}(\rho^A_{ij}) + \sum_{B\neq A}^{N^A_{\mathrm{neigh}}} \sum_j^{N^i_{\mathrm{ang}}} \widetilde{w}^{AB}_{ij} \tilde{F}(\rho^{AB}_{ij}) \right]. \tag{36}$$

We will denote the grid pruning method based on evaluating Eq. (36) for different values of L as AMG (Adaptive Molpro Grid) method in the following. The sum over B in the equation is restricted to all neighbor atoms of A as defined by r^A_{cut}, and $\widetilde{w}^{AB}_{ij}(L)$ are the angular weight factors for a given Lebedev order L, defined in analogy to Eq. (29) as

$$\widetilde{w}^{AB}_{ij}(L) = w^{\mathrm{ang}}_j(L) \frac{P_A(\mathbf{r}_{ij})}{P_A(\mathbf{r}_{ij}) + P_B(\mathbf{r}_{ij})}. \tag{37}$$

The integrands $\tilde{F}(\rho^A_{ij})$ and $\tilde{F}(\rho^{AB}_{ij})$ are computed as in Eq. (28) for the density points on a sphere, except that either the atomic density $\rho^A_{ij} = \rho_A(\mathbf{r}_{ij})$, or the superposition of two atomic densities $\rho^{AB}_{ij} = \rho_A(\mathbf{r}_{ij}) + \rho_B(\mathbf{r}_{ij})$, is used.

C. The pruning procedure

With the definition of the model functional in eqs. (28) or (36), and using the integration kernel in Eq. (32), the generation of an atomic quadrature grid is carried out as follows. First, a radial grid is determined from a given number of radial points as described at the end of section II B. Then for each point in the radial grid, the model functional is evaluated for a maximum Lebedev order $L_{\max}$ (we choose $L_{\max} = 59$) to obtain the target values $\tilde{\mathcal{I}}_{A_i}[L_{\max}]$. Then a search is performed over all Lebedev orders L, starting from the lowest possible up to $L_{\max}$, and the deviations from the reference values are computed:

$$\Delta\tilde{\mathcal{I}}_{A_i}(L) = \left| \tilde{\mathcal{I}}_{A_i}[L] - \tilde{\mathcal{I}}_{A_i}[L_{\max}] \right| \tag{38}$$

The most straightforward approach would be to take the lowest value of L for which the condition

$$\Delta\tilde{\mathcal{I}}_{A_i}(L) \leq T_{\mathrm{ang}} \tag{39}$$

is satisfied, where the threshold T_{ang} can be chosen by the user. However, $\tilde{\mathcal{I}}_{A_i}[L]$ is not a smooth function, since random fluctuations are introduced by the specific positions of the angular points, and it can happen that for larger values of L the condition is not always satisfied. We therefore adopt the value of L for which it, and the three higher L values, satisfy the criterion, and assuming that for all $L \geq L_{\mathrm{max}}$ the criterion is fulfilled.

The two grid threshold T_{rad} and T_{ang} which determine the sizes of the radial and the angular grids [Eqs. (22) and (39)], respectively, are by default determined in Molpro from a single threshold value of T_{grid} as

$$T_{\mathrm{grid}} \rightarrow \begin{cases} T_{\mathrm{rad}} = f_{\mathrm{rad}} T_{\mathrm{grid}} \\ T_{\mathrm{ang}} = f_{\mathrm{ang}} T_{\mathrm{grid}} \end{cases} \tag{40}$$

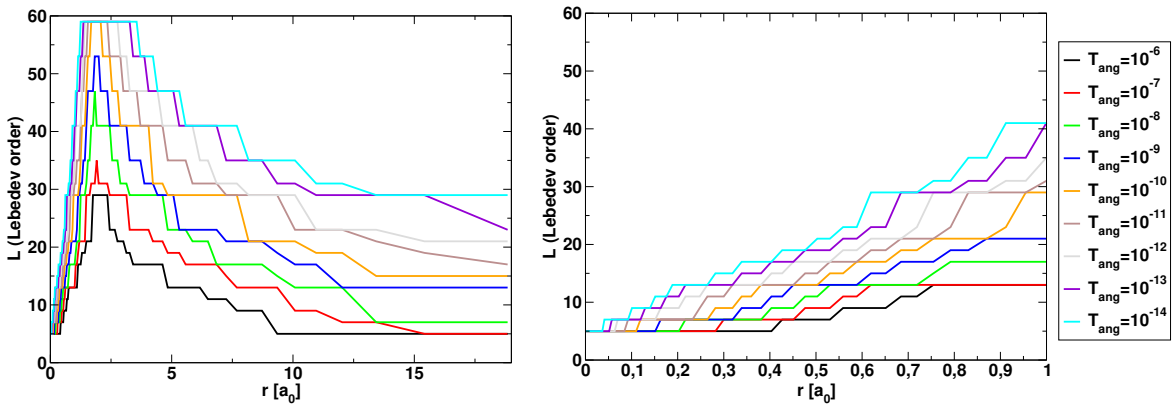
with f_{rad} and f_{ang} being two appropriately chosen pre-factors. These factors are set to $f_{\mathrm{rad}} = \frac{1}{2}$ and $f_{\mathrm{ang}} = \frac{1}{10}$, respectively. A grid threshold of $T_{\mathrm{grid}} = 10^{-6}$ is used by default in the Molpro program. Alternative presets `MEDIUM`, `HIGH`, or `VERYHIGH` correspond to T_{grid} values of 10^{-7}, 10^{-8}, and 10^{-10}, respectively.

A critical issue in the above pruning method is the numerical implementation of the test in Eq. (39) when T_{ang} is very small (10^{-13} or less). If the summation in Eq. (36) is carried out in a straightforward way, strong oscillations in the close nuclei range are observed for the $L(r_i)$ curves; see Fig. 1 in the supplementary material. The reason for this are rounding errors in the numerical integration. One of the most popular approaches to reduce rounding errors in sums of floating point numbers has been derived by Kahan,[80] whose algorithm keeps track of a separate running compensation (that accumulates small errors) for the sum. When this correction is applied, the corresponding oscillations disappear and the curves monotonically increase in the short range region for all chosen values T_{ang}.

Even with the employment of the Kahan method to correct for rounding errors it can happen that sudden steps occur for the L values obtained with the pruning method for neighbored radial grid points. This may happen due to numerical issues with tight or even very tight threshold values. Therefore a simple smoothing procedure is carried out across all radial grid points r_i with $i \geq 3$ if $L^{i-1} < \min(L^i, L^{i-2})$, i.e., when the intermediate value of L^{i-1} is a minimum (the exponents here indicating the radial grid index value). Then this value is shifted to $L^{i-1} \rightarrow \min(L^i, L^{i-2})$ to make the shape convex. After this first smoothing step has been applied for the L profile, the resulting set of L values that are ordered according to

the radial grid points (from small values of r_i to large ones) are searched for local one-point or two-point maxima and minima. In case of a one-point minimum/maximum at L_i the point is shifted to $\max(L^{i-1}, L^{i+1})$ and in case of a two-point minimum/maximum for the pair L^i, L^{i+1} the two points are shifted to $\max(L^{i-1}, L^{i+2})$ to make the L profile smooth.

The resulting pruned $L(r_i)$ curves based on the model function in eq. (36) are shown in Fig. 2 for the oxygen atom in H_2O. Here we used a fixed radial grid with 128 points ($T_{\text{rad}} = 5 \times 10^{-15}$). For comparison, we have also used eq. (28) for pruning the angular grid. The corresponding $L(r_i)$ curves are shown in Fig. 2 of the supplementary material. They are qualitatively similar to those in Fig. 2, but the $L(r_i)$ are somewhat larger at greater distances.



(a) total range (b) short range

Figure 2. Optimised L values obtained with the pruning method based on $\tilde{\mathcal{I}}_{A_i}$ for different angular threshold values T_{ang} for the oxygen atom (H_2O) using a radial grid with 128 points ($T_{\text{rad}} = 5 \times 10^{-15}$). 2(a): total range of radial grid; 2(b): close-up view of the short range part. $L_{\max} = 59$ was chosen. The Kahan summation algorithm[80] was used for the numerical integration of the model functional of Eq. (36) to reduce rounding errors.

D. Timings for the adaptive grid pruning

The adaptive grid generation requires some extra computational work as compared to the fixed grid methods. Fig. 3 shows for linear alkane chains the CPU times for the AMG grid generation (GRID) and for 10 iterations (ITER), using the def2-TZVP basis set and the BLYP functional. This demonstrates that the grid generation time scales linearly with

molecular size. The iteration time formally scales cubically due to the Coulomb contribution; in practice the scaling is lower due to screening procedures. In the current calculations the scaling exponent is between 1.8 ($T_{\text{grid}} = 10^{-10}$) and 2.2 ($T_{\text{grid}} = 10^{-6}$). For the smallest

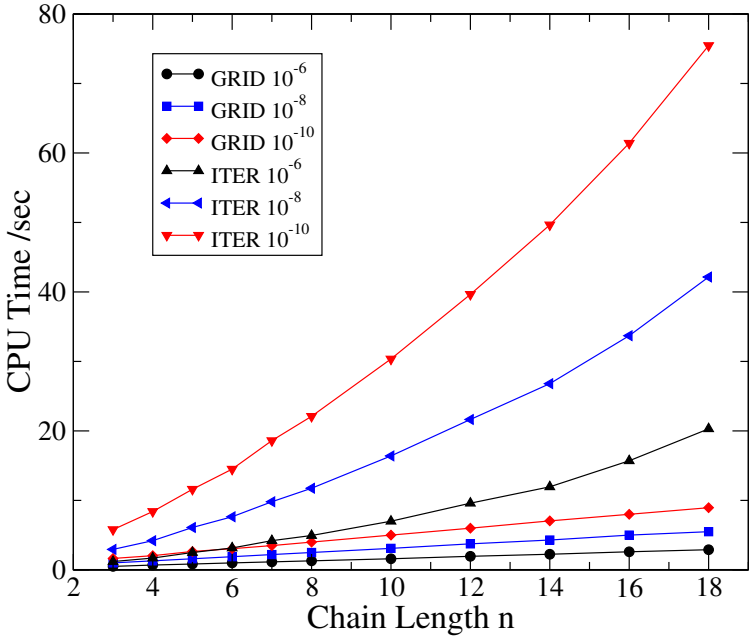


Figure 3. Computation times for linear alkanes C_nH_{n+2} (basis def2-TZVP, BLYP functional). Shown are the times for the adaptive grid generation (GRID) and the iteration time (ITER, 10 iterations) for 3 different values of T_{grid}. The calculations were run using a single core on a MacBook Pro laptop with M1 Max processor and 64 GB of memory.

systems the grid generation time is somewhat larger than the time for one iteration. However, due to the steeper scaling the iterations quickly become strongly dominant for longer chains. Of course, the linear chains are model systems and only used to reach the asymptotic scaling with relatively small molecular sizes. Timings for a transition metal complex with 175 atoms ($FeC_{72}N_2H_{100}$, structure taken from Ref. 81) are presented in Table III. In this case the AMG generation takes only about 2% of the time for one KS iteration with BLYP. Since the grid generation is independent of the basis set and functional used, this fraction becomes even smaller for calculations with larger basis sets and more complicated functionals, in particular if exact exchange is included (see the timings for PBE0).

Table III. Timings[a] for $FeC_{72}N_2H_{100}$ (^5A) using the def2-TZVP basis

T_{grid}	10^{-6}	10^{-8}	10^{-10}
Number of grid points	333968	947524	1909694
Time for grid generation	0.8	1.5	2.3
Time for one iteration (BLYP)	40	68	111
Time for one iteration (PBE0)	446	474	515

Timings on a MacBook (cf. Fig. 3) using 8 cores

(7 MPI processes and one helper process)

V. BENCHMARKS

A. Computational details

The following grid pruning methods described in the previous sections are used for the test calculations in this work:

- **SG-0,1,2,3** fixed grids by Gill *et al.*[35,52] and Dasgupta and Herbert[53]
- **3-zone** fixed grid by Treutler and Ahlrichs[36]
- **5-zone** fixed grid implemented in Molpro[60–62] based on the SG-1 grid developed by Gill *et al.*[35] and modifications to SG-1 by Neese[54,55]
- **AMG** adaptive grid implemented in Molpro

The latter three grid methods allow to choose various grid levels, and these have been varied in our benchmark calculation. The recommended default settings are summarized in Table IV.

The calculations with the SG grids were done by orienting the molecular coordinates using the approach described in Ref. 35. Such a reorientation of the molecular geometry was not performed for the calculations with the other grid methods, where the atomic coordinates were taken directly from Refs. 63 and 82. It should also be noted that the fixed SG grids were developed in conjunction with the Becke smoothing function of Eq. (11) which was

Table IV. Default grid level settings for different grid pruning methods.

pruning method	possible levels	default level	reference
3-zone	1-7	3	36
5-zone[a)]	3-9	4 (for 1st+2nd row elements)	this work
AMG	$T_{\text{grid}} = 10^{-x}$, $x \in \mathbb{N}$	$T_{\text{grid}} = 10^{-6}$	this work

a) A similar grid is used in the Orca4 program[54,55] but with a different radial grid, cf. Section III A

therefore used in the calculations in this work for these grids. In all other grid methods the smoothing function of Murray [Eq. (15)] has been applied.

The 74-point Lebedev grid for order $L = 13$ from Ref. 50 contains negative weights and therefore has been replaced in our implementation by the 78-point Lebedev grid from Heo and Xu[83] that contains only positive weights, with further refinement to improve accuracy; see the supplementary material for details.

Unless otherwise noted, calculations were done with the def2-TZVP basis set.[84] Density fitting approximations were employed to compute the Coulomb and exact exchange contributions to the Fock matrix. For the def2-TZVP orbital basis the universal QZVPP/JK fitting basis sets of Weigend[85] were used. To test the impact of the basis set on the accuracies of the numerical integrations, some calculations were also done using the Dunning's aug-cc-pVxZ basis sets ($x = \text{D}, \text{T}, \text{Q}$),[86–89] using the corresponding fitting basis sets.[90]

In most calculations the BLYP (Becke, Lee, Yang, Parr) energy functional[91,92] was employed. This functional was chosen because the B88 exchange energy density decays asymptotically correctly as[91] $e_x \approx -\frac{1}{2}\rho r^{-1}$. It should therefore properly sample the contributions far from the nuclei. Some calculations were also done using the PBE0 (Perdew, Burke, Ernzerhof)[93] hybrid-GGA, as well as with the TPSS (Tao, Perdew, Staroverov,Scuseria),[94] r2SCAN[95], and M06[96] meta-GGA functionals to demonstrate the influence of the functional type on the grid accuracies.

All calculations were done using the Molpro quantum chemistry program.[60–62] Tight convergence thresholds of 10^{-9} a.u. for energies and 10^{-10} a.u. for density matrix changes have been used.

B. Test systems and error analysis

The accuracies of the six methods have been tested for electron number integrals (Eqs. (20) and (27)) and thermochemical properties, using an extended set of molecules and chemical reactions. These were taken from 14 subsets of the GMTKN30 database of Grimme *et al.*[63,82] comprising: BHPERI (barrier heights of pericyclic reactions), ISO34 and ISOL22 (isomerization energies of small, medium- sized and large organic molecules), DC9 (a database containing nine difficult cases for DFT), DARC (Diels-Alder reaction energies), BSR36 (bond separation reactions of saturated hydrocarbons), IDISP (intramolecular dispersion interactions), WATER27 (binding energies of water, $H^+(H_2O)_n$ and $OH^-(H_2O)_n$ clusters), S22 (noncovalent interaction energies), ADIM6 (interaction energies of n-alkane dimers), and the PCONF, ACONF, SCONF, and CYCONF databases (containing systems to study the performance for predicting relative energies of tri-peptide, alkane, sugar, and cysteine conformers). These test databases contain molecules with up to 68 atoms (folded/unfolded alkanes in IDISP) and mostly atoms from the first (long) rows of the PES. A few molecules with the second row elements Si, P, S, and Cl are contained in BHPERI (P, S), CYCONF (S), DC9 (S) and ISOL22 (Si, Cl). The geometries were taken from Refs. 63 and 82.

While the conventional quantities that are used to measure the statistical errors for a large set of data values give some idea about the total magnitudes of the grid errors for the test systems they might emphasize too much the databases with comparably large energy differences (like chemical reactions) and less those with small ones (intermolecular interactions and intramolecular conformer reactions). Therefore for assessing the quadrature errors across all 14 databases (comprising 408 molecules and 254 reactions), along with the $\bar{\Delta}$ (mean absolute errors), Δ^{rms} (root mean square errors) and $\Delta^{\max}$ (averaged maximum absolute errors per database) statistical quantities, we also measured the grid accuracies by calculating the quantity

$$\bar{\Delta}^{\mathrm{w}} = \frac{1}{N} \sum_{i=1}^{\mathrm{database}} \frac{\bar{\Delta}_i}{\bar{E}_i} \tag{41}$$

termed as weighted mean absolute error. Here, $N(=14)$ is the number of databases for the test, $\bar{\Delta}_i$ denotes the mean absolute error of the energy-difference results for database with label i, and $\bar{E}_i$ are the mean absolute reference values of database i; see also Refs. 82 and 97.

In section V G we demonstrate the convergence of absolute energies for the molecules from the FH51 data base of Friedrich and Hänchen[98,99] and the C_{60} molecule.

C. Reference energies

Unless otherwise noted, the reference grid is defined by the following specifications:

- Mura-Knowles radial quadrature scheme (MK) using $T_{\mathrm{rad}} = 5 \times 10^{-15}$ (Eq. (22)) (128 radial grid points for first row elements (Li-Ne))
- $L = L_{\mathrm{max}} = 59$ Lebedev quadrature at each radial grid point (1202 functions per radial grid point)
- Smoothing function of Murray (Eq. (15))
- Treutler size-adjustment scheme in the Voronoi partitioning (Eq. (14))

The reference energies used in the thermochemical energy tests are available in the supplementary material.

In order to estimate the accuracy of the reference grid used in our calculations we made calculations with much larger unpruned radial and angular grids up to a grid size of $(500, 2354)$ corresponding to a grid with 500 radial grid points and 2354 angular grid points ($L = 83$) for the FH51 set of molecules from Friedrich and Hänchen.[98,99] This test has been done both for the BLYP functional (GGA type functional)[91,92] and r2SCAN (meta-GGA functional).[95] The results are displayed in table V

showing both the *rms* errors and the *max* errors (in parenthesis) to the largest $(500, 2354)$ grids for the total energies (normalised over the number of atoms for each of the 87 molecules) and the 51 reaction energies. Note that the total energy errors are given in units of μE_h (microhartree) and the reaction energies in units of 10^{-3} kcal mol^{-1}.

It can be seen that the BLYP energies are practically converged for our reference grid, the energy difference to the $(500, 2354)$ grid results is below 0.01 μE_h and therefore of the order of the convergence threshold used in the calculations [$T_{\mathrm{SCF}} = 10^{-9}$ Hartree (E_h)]. Even the maximum error for the reference grid for BLYP is very small and amounts to 0.01 μE_h.

Table V. FH51 systems:[98,99] root mean square and maximum deviations of total energies and reaction energies for the reference grid used in our calculations (denoted as 'reference' in the table) and other large unpruned grids with radial grid sizes of $N_{\mathrm{rad}} = 200, 300, 400, 500$ and an angular grid of the size $N_{\mathrm{ang}} = 1202$ to a very large $(500, 2354)$ unpruned grid. Total energies are normalised over the number of atoms of each molecule and are in units of μE_h. The deviations for the reaction energies are given in units of 10^{-3} kcal mol^{-1}. The *max* errors are given in parenthesis in the table. The def2-TZVP basis set was used in the calculations.

quantity	functional	reference[a)]	(200, 1202)	(300, 1202)	(400, 1202)	(500, 1202)
total energy	BLYP	0.00	0.00	0.00	0.00	0.00
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	r2SCAN	0.12	0.01	0.01	0.01	0.01
		(0.23)	(0.06)	(0.05)	(0.05)	(0.05)
reaction energy	BLYP	0.08	0.08	0.08	0.08	0.08
		(0.28)	(0.28)	(0.28)	(0.28)	(0.28)
	r2SCAN	0.23	0.18	0.18	0.18	0.18
		(0.61)	(0.63)	(0.64)	(0.64)	(0.64)

a) AMG reference grid, $N_{\mathrm{rad}} = 113 + k_{\mathrm{row}}[Z] \cdot 15$, $N_{\mathrm{ang}} = 1202$.

In contrast to this, for r2SCAN a larger *rms* error of 0.12 μE_h is found for our reference grid which reduces by one order of magnitude upon an increase of the radial grid size to $N_{\mathrm{rad}} = 200$ (fourth column in table V). This indicates that the main source of error for meta GGA energies for our reference grid is due to the radial grid size. We think, however, that this accuracy is still good enough to be able to test grid errors for pruned grid methods. See for example table VI showing the grid errors for the AMG method which are orders of magnitudes larger than the 0.12 μE_h error estimate for our reference grid.

For the reaction energies the grid errors for the reference grid amount to $8 \cdot 10^{-5}$ and $2.3 \cdot 10^{-4}$ kcal mol^{-1} for BLYP and r2SCAN, respectively, see Table V. This is clearly below chemical accuracy and much lower than energy differences normally considered for comparing errors for pruned grid methods (see for example the work by Mardirossian and Head-Gordon[8]). As can be seen in the table, a further increase of the radial grid up to a size of $N_{rad} = 500$ almost has no effect for BLYP and only a small effect for r2SCAN, indicating that the main

source of error here appears to be the not fully converged angular grid (in this very small error range, though).

Another test for the convergence of the angular grid size of the reference grid has been made for the C_{60} molecule, see Fig. 13 in section V G. In this case an increase of the maximum L-value from $L_{\max} = 59$ to $L_{\max} = 83$ with the reference radial grid changed the absolute energy by less than 10^{-5} E_h.

In summary it can be concluded that the reference grid defined at the beginning of this subsection delivers very accurate total and relative KS energies and should therefore serve well to assess the accuracies of the pruned grid methods in the next sections.

D. Electron number test

Since the integral of the electron density is known exactly, the quantity

$$\Delta N = \left| N - \int \rho(\mathbf{r})\, d\mathbf{r} \right| . \tag{42}$$

(with N being the number of electrons of the molecule and ρ the electron density) is well suited to test the accuracy of the numerical quadrature. The ΔN shown below are averaged for the 408 molecules of the benchmark sets. In order to rule out secondary effects of the quadrature grids in KS calculations on the densities self consistent Hartree-Fock densities were used throughout.

First the effect of the specifications for the Voronoi partitioning (see section II A) on the accuracy of the integration grids is investigated. Fig. 4 shows the ΔN values obtained with the adaptive grid method of section IV B, using the five different global grid threshold values of 10^{-6}, 10^{-7}, 10^{-8}, 10^{-10}, and 10^{-14} (see section IV B) and either the Becke or Treutler size adjustment schemes [Eq. (13) and (14), respectively]. On the abscissa the base-10 logarithm of the average grid size per atom is plotted. With both schemes the values of ΔN approach zero with larger grid sizes (tighter threshold values), as expected. However, the Treutler scheme (Eq. (14)) performs clearly better than the Becke scheme (Eq. (13)) over the whole range of grid sizes. Particularly at large grids the ΔN values are smaller by one order of magnitude. However, neither method is able to reduce ΔN below about 10^{-6} – 10^{-7}. The reason for this can be attributed to the limit $L_{\max} = 59$, which was used as the

maximum Lebedev order in the calculations. Also secondary effects like screening, rounding errors in the sums for the numerical integration, or the numerical accuracies of the density matrices themselves might play a role. We did not investigate the actual source of this error further, since the main focus here lies on the comparison between the different integration grids rather than to reproduce the electron number by numerical integration as accurately as possible.

In summary, the results in Fig. 4 indicate that the Treutler scheme (Eq. (14)) delivers higher accuracies than the Becke scheme (Eq. (13)) for the integration over ρ. The Treutler scheme was therefore chosen in all further tests of the quadrature methods, except for the fixed SG grids which have been explicitly developed for the Becke scheme of Eq. (13).

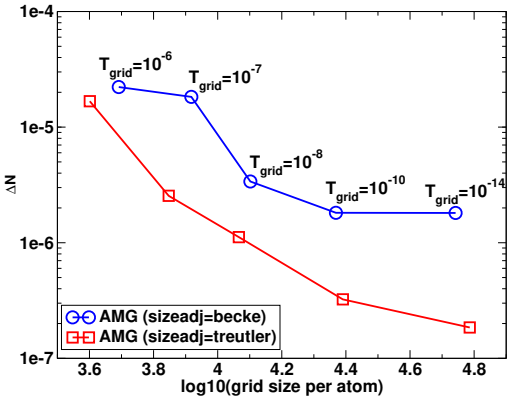


Figure 4. Benchmark test for electron number difference ΔN (cf. eq. 42) using the Becke (Eq. (13)) and Treutler (Eq. (14)) size adjustment schemes in the Voronoi partitioning. The global grid threshold values T_{grid} determining the grid size have been added to the respective points of the 'sizeadj=becke' curve; see section IV B for definition.

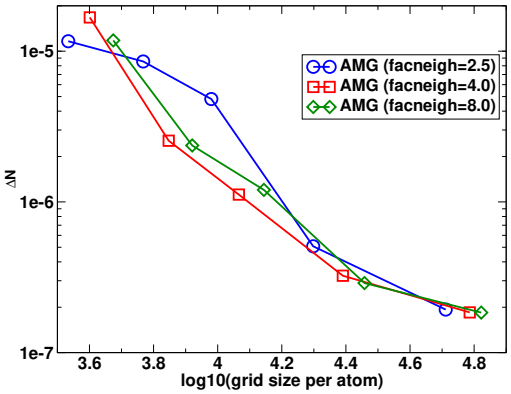


Figure 5. Benchmark test for electron number difference ΔN (cf. eq. 42) using different neighbor factor cutoff values (Eq. (34)) in the angular tester functional of Eq. (36). The Treutler (Eq. (14)) size adjustment scheme in the Voronoi partitioning was used. The five points of the AMG curves correspond to the grid thresholds as given in Fig. 4.

Another important parameter in the AMG grid pruning method is the neighbor factor f_{neigh} [cf. Eq. (34)], which is used to limit the sum over atoms in the second term of the model functional in Eq. (36) to the close neighbor atoms. Fig. 5 shows the ΔN values obtained with the adaptive grid method and neighbor factors of 2.5 (previously the default setting in Molpro), 4 and 8. Overall, the choice of $f_{\text{neigh}} = 4$ gives higher accuracies than $f_{\text{neigh}} = 2.5$. This might be due to better description of anisotropy of the density caused by the neighbor atoms. However, there is no further improvement in the ΔN values upon increase of the number of neighbor atoms in Eq. (36). Therefore, the value of $f_{\text{neigh}} = 4$ has been used as the default setting for the adaptive grid method in the following.

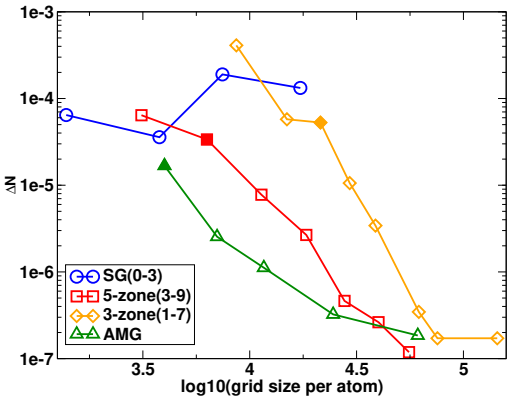


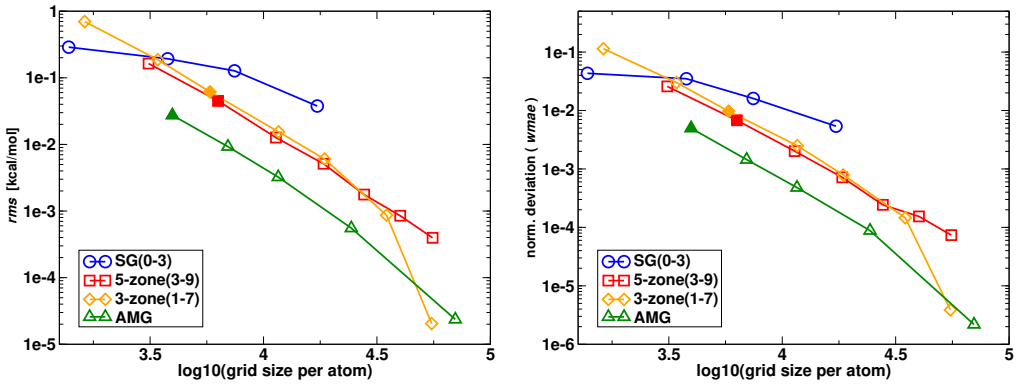
Figure 6. Benchmark test for electron number difference $|N - \int \rho(\mathbf{r})d\mathbf{r}|$ for different grid types. The grids that correspond to the default settings for each method are marked by filled symbols. The five points of the AMG curves correspond to the grid thresholds as given in Fig. 4.

Fig. 6 compares the averaged ΔN values obtained with the fixed and adaptive grid methods described in sections III A and IV, respectively. The SG 0-3 grids yield errors of the order of 10^{-4}, but there is no reduction of the error upon an increase of the grid size from SG-0 to SG-3. The 5-zone and 3-zone grids with default grid levels yield similar errors as SG-1, but for both grid methods a systematic improvement towards $\Delta N = 10^{-7}$ is observed if the grid levels are increased. The 5-zone grids yield consistently more accurate ΔN values than the 3-zone grids for comparable grid sizes. Most accurate values are obtained with the adaptive AMG method, which yields ΔN values that are for small and medium sizes one to two orders of magnitudes closer to zero than those of all other grid methods.

E. Thermochemical benchmarks

The thermochemical properties for the 14 subsets from the GMTKN30 database (see section V B for a description) have been computed with the BLYP, PBE0, and the TPSS functionals using the def2-TZVP basis set. To measure the accuracies of the different integration grids both the *rms* and weighted errors, Δ^{rms} and $\bar{\Delta}^{\mathrm{w}}$ [Eq. (41)], have been calculated with respect to the reference grid defined in section V C. Fig. 7 shows the Δ^{rms} and $\bar{\Delta}^{\mathrm{w}}$ values of relative energies for the different grid methods obtained using the BLYP functional. While the $\bar{\Delta}^{\mathrm{w}}$ measure reduces a possible overrating of the database subsets with larger mean energies,

one can see that there is no qualitative change in the behavior and relative orderings of the curves in the two figures. We therefore only consider the Δ^{rms} data in the following.



(a) Δ^{rms} values (b) $\bar{\Delta}^{\mathrm{w}}$ values (Eq. (41))

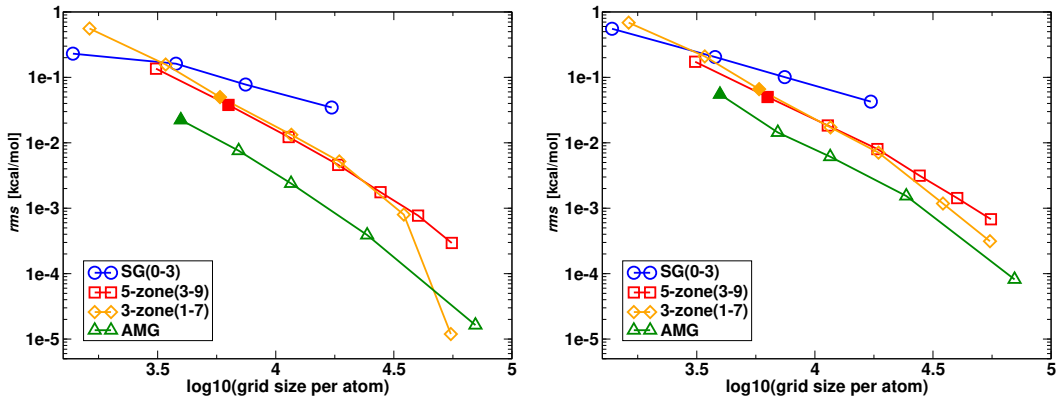
Figure 7. Benchmark test for thermochemical energies for different grid types. The BLYP functional was used in the calculations. The grids that correspond to the default settings for each method are marked by filled symbols. The five points of the AMG curves correspond to the grid thresholds as given in Fig. 4.

The comparison of the different quadrature methods in figures 7 shows a clear hierarchy in performance. While all methods yield higher accuracies upon increasing the grid sizes (according to the individual grid levels that can be chosen in each method), the SG grids exhibit only a moderate improvement at larger grid levels. The 5-zone and 3-zone grids yield very similar errors at the same grid sizes. For the default grid levels the weighted mean error is close to 0.01 kcal mol^{-1} for both methods. An increase of the grid levels leads to a systematic improvement of the accuracies.

Since the 3-zone grids use a small number of angular grid points close to the nucleus, the good performance shows that this region is well represented by low L values, and that the main impact for the overall accuracy stems from the number of angular grid points farther away from the nucleus. This can be seen from the improvement of the 3-zone grid results from grid level 6 to 7 (last two points), which differ only in the highest order of the Lebedev grids (53 and 59, respectively) in the outermost radial zone.

The best overall performance is found for the adaptive grid method. Already with small grid sizes of $\log_{10}(N_{\mathrm{grid}}/N) \approx 3.5$ the average error is smaller than 0.1 kcal mol^{-1}, which means

that influences of the grid truncation errors can be neglected in most KS calculations. Only for the largest grid size considered the 3-zone grid slightly outperforms the AMG one. This is mainly due to the approximations described in section IV B; if the functional in Eq. (28) is used, the errors of the AMG scheme are smallest for all grid sizes (see the supplementary material). In any case, the integration errors for the largest grid sizes [$\log_{10}(N_{\mathrm{grid}}/N) > 4.5$] are completely negligible. Errors caused by the choice of the functional or the basis set are much larger, as discussed further in section V F.



(a) PBE0 (b) TPSS

Figure 8. Benchmark test for thermochemical energies for different grid types for a hybrid-GGA (PBE0) and a meta-GGA (TPSS) functional. The grids that correspond to the default settings for each method are marked by filled symbols. The five points of the AMG curves correspond to the grid thresholds as given in Fig. 4.

Fig. 8 shows the performances of the different quadrature methods using two different functionals, PBE0, and TPSS. PBE0 is a hybrid GGA functional and includes 25% exact exchange, in contrast to the BLYP functional whose results were presented above. The contribution of the exact exchange energy is computed analytically, and can therefore only have a secondary influence on the accuracies of the DFT integrations due to modifications in the shape of the density; see also Ref. 100 (Fig. 2). In contrast, the PBE GGA part of the PBE0 functional is evaluated by numerical integration and therefore depends directly on the size and distribution of grid points for the different quadrature methods. A comparison of the BLYP and PBE0 curves in Figs. 7 and 8, respectively, shows that these are qualitatively and to some degree even quantitatively similar. This indicates that the relative accuracies for the numerical integration found with the BLYP method are transferable to different

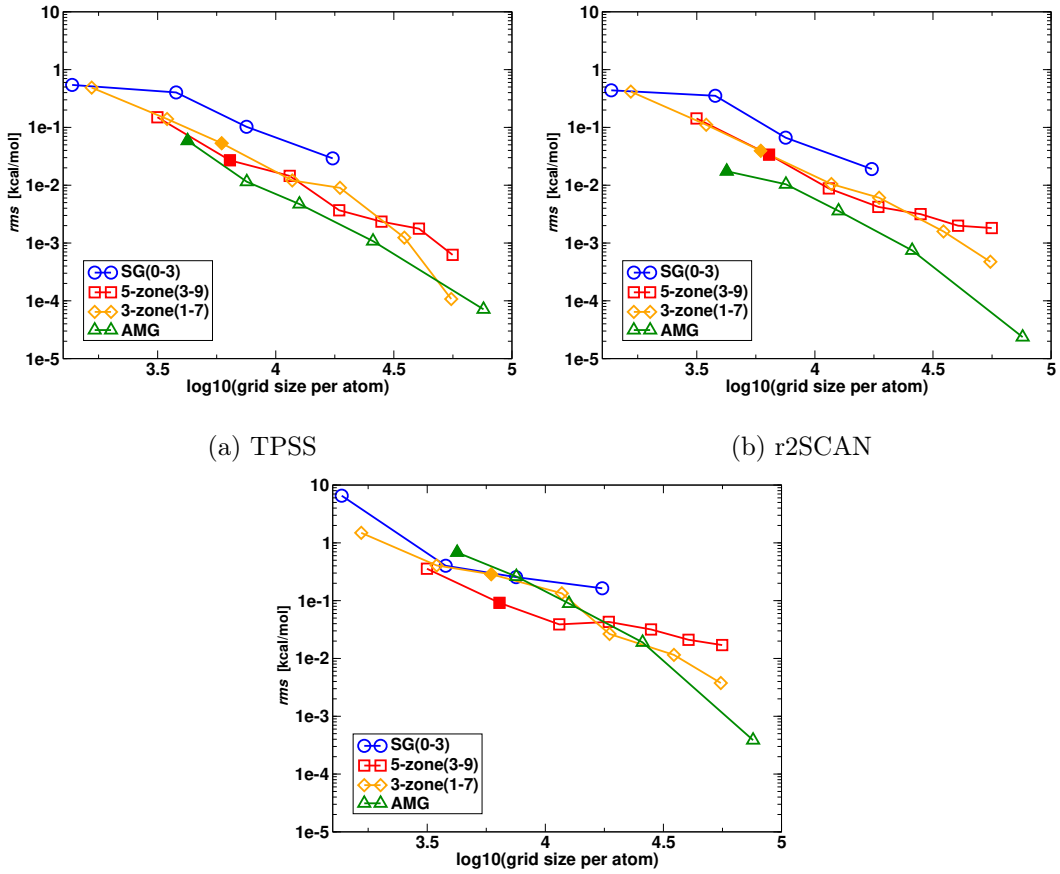
(hybrid) GGA functionals.

This, however, is no longer true when comparing the results for the TPSS functional in subfigure 8(b). TPSS is a meta-GGA functional and depends on the local kinetic energy density $\tau(\mathbf{r}) = \frac{1}{2}\sum_i^{\mathrm{occ}} |\nabla\phi_i(\mathbf{r})|^2$ (tau-dependent functional).[18] As shown in Fig. 8, overall the accuracies achieved with all tested quadrature methods with the TPSS functional are worse than with BLYP and PBE0. The calculations with the largest grids for the respective methods (except SG-3) only approach an accuracy of 10^{-3} to 10^{-4} kcal mol^{-1} at $\log_{10}(N_{\mathrm{grid}}/N) \approx 5$, which is one to two orders of magnitude larger than the Δ^{rms} values obtained with BLYP and PBE0. The relative ordering of the average errors for the different methods is the same as found for the corresponding BLYP and PBE0 results.

The difficulty to achieve accurate numerical integrations with meta-GGA functionals has been demonstrated earlier by other authors. For example, in a review Mardirossian and Head-Gordon[8] a mean absolute percent integration grid error of 75.9% was found for the NC15 (noncovalent interaction benchmark) data set using the SG-1 grid and M06 functional compared to results obtained with a (500,974) reference grid. A recent paper by Lehtola and Marques[58] stated that many recent meta-GGA functionals are ill-behaved with respect to accurate numerical integration. They considered spherical atoms and found that the convergence with radial grid size can be extremely slow, making such functionals unsuitable for practical applications as well as benchmark studies. Another finding is that the energies may converge very slowly to the basis set limit with increasing basis set sizes.[8]

While a detailed study of these problems is beyond the scope of the present work, we have carried out a number preliminary calculations for the PCONF and FH51 test sets, using the TPSS, r2SCAN, and M06 functionals. The results for PCONF are shown in Fig. 9 and the results for the FH51 test is presented in section V G (tables VI and VII).

As can be seen in Fig. 9, in all cases there is a similar slow reduction of the *rms* errors with respect to the grid size. The behavior of the curves for TPSS similar to the corresponding one for the whole thermochemical test in Fig. 8(a), indicating that the selection of the PCONF test is representative for the full thermochemical test. A further check of the particular performance of the AMG grid method for meta GGA functionals has been made for the FH51 reaction energy benchmark systems[98,99] and is presented in section V G.



(a) TPSS

(b) r2SCAN

(c) M06

Figure 9. PCONF conformer energy test for the different grid methods and using the three different meta GGA functionals TPSS, r2SCAN and M06. The five points of the AMG curves correspond to the grid thresholds as given in Fig. 4.

For the M06 functional the relative orderings of the curves are different than for TPSS and r2SCAN, see Fig. 9. In particular, the AMG grid for $T_{\text{grid}} = 10^{-6}$ delivers accuracies that are only slightly smaller than 1 kcal mol^{-1} for the M06 functional, which is one and two orders of magnitude larger than what is found for TPSS and r2SCAN. Clearly, such large errors are unacceptable, and with the grid scheme of this work it can only be reduced to a reasonable degree by reducing T_{grid} to 10^{-8} or even lower, which leads to a significant increase of the grid size. In order to reach an accuracy of about 0.03 kcal mol^{-1} the grid sizes of the AMG, 3-zone and 5-zone grids are comparable. However, the 5-zone grid performs somewhat better for smaller grid sizes and reaches the target accuracy of 0.1 kcal mol^{-1} already with the default grid level. This indicates that there is some potential to achieve better accuracies for smaller grids also with an adaptive grid method. Future work will be

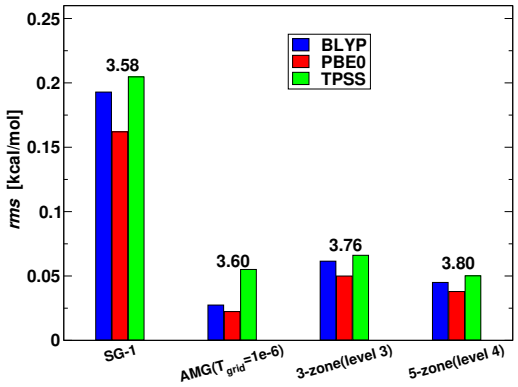


Figure 10. Root mean squared errors from the reference grid results using the default settings for each grid method. In case of the SG grids the SG-1 was chosen for this comparison. The base-10 logarithm of the average sizes of the grids per atom are displayed above the respective groups.

needed to investigate this problem in more detail. It could for example arise from the simple form of the test functional in Eq. (36) or the approximate densities used in it.

Finally, in Fig. 10 the *rms* errors obtained with the default settings for each grid pruning method (cf. Table IV) are summarized. In case of the fixed SG grids we chose the intermediate sized SG-1 grid for this comparison. The methods are ordered with respect to the average grid sizes per atom, which are displayed above the various groups in the diagram (in $\log_{10}$ format). All grids except for SG-1 yield *rms* errors below 0.1 kcal mol^{-1}. In case of the 3-zone and 5-zone grid the grid errors are rather similar for the three different functional types, while for the AMG method it depends more strongly on the functional and is twice as large for the meta-GGA TPSS functional as for the GGA-type functionals BLYP and PBE0. The *rms* errors for SG-1 are of the order of 0.2 kcal mol^{-1}, and even larger errors are expected with SG0. These grids are therefore less suitable for high accuracy calculations.

F. Dependence of the accuracies of the adaptive grids on the basis set

The results presented in the previous section V E were obtained by using the def2-TZVP basis sets to represent the density and orbitals in the self-consistent KS calculations. While basis sets of triple-zeta quality may be sufficient to obtain results that are resonably converged with respect to the basis set size, it will be useful to take a look at the influence of the basis set on the accuracies for the DFT integration. For this, calculations with the adaptive

grid method of section IV B were done using the aug-cc-pVXZ (X=D,T,Q) basis sets of Dunning.[88,89] While these basis sets were not particularly designed for DFT methods and the larger ones certainly contain more high angular momentum basis functions than needed in DFT, they can still be used to study the convergence towards the basis set limit. In fact, as shown below, the aug-cc-pVTZ basis yields for our test set much more accurate results than the def2-TZVP basis. This is mainly attributed to the diffuse basis functions, which are important in molecules with polar groups. We assume that aug-cc-pVQZ basis gives DFT results quite close to the basis set limit.

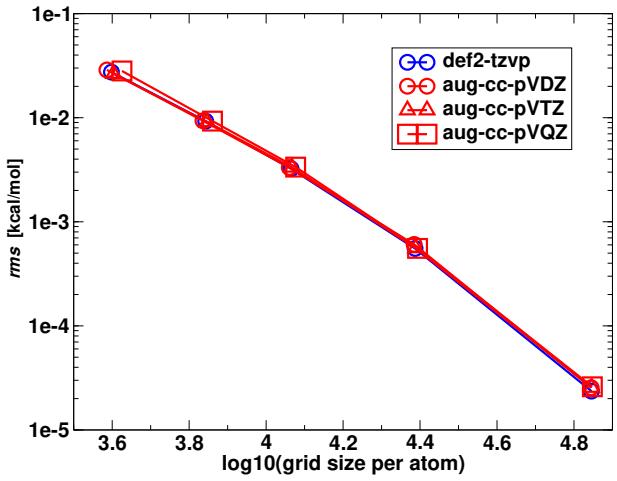


Figure 11. Grid errors of thermochemical energies using different basis sets using the adaptive grid obtained with the model functional of Eq. (36).

Fig. 11 shows the average errors of the adaptive grid method for all four basis sets, using otherwise the same calculation details as chosen in the previous section. The reference values obtained with large grids were determined independently for each basis set. On the scale of the diagram there is hardly any difference between the four curves. This shows that the grid errors are almost independent of the basis set and orders of magnitude smaller than the basis set errors. For example, the results obtained with the aug-cc-pVDZ, def2-TZVP, and aug-cc-pVTZ basis sets differ on the average from the aug-cc-pVQZ ones by about 2.0, 1.5, and 0.14 kcal mol^{-1}, respectively. Likely these are lower bounds for the corresponding true basis set errors.

The relative insensitivity to the grid size on thermochemical energies also offers the application of approaches to speed up KS calculations by changing the integration grid during the

SCF cycles. This is well known and used by default in various programs, e.g., Turbomole[66,101] and Orca.[54,55] An early approach where different adaptive grids were used in the SCF iterations and in the final energy calculation was investigated by Tozer *et al.*.[102] In Molpro such a method can be activated with the `coarsegrid` option.[60] When this is done the SCF iterations are performed with a coarse grid whose threshold is by default set to $f_{\mathrm{coarse}} \cdot T_{\mathrm{grid}}$, with $f_{\mathrm{coarse}} = 10^3$ and T_{grid} being the threshold value for the fine grid. After convergence the energy is recomputed with the fine grid. Fig. 12 shows the Δ^{rms} errors for the AMG grid method for the thermochemical benchmarks both with and without this `coarsegrid` option (using the def2-TZVP basis). The `coarsegrid` option works very well for larger grids with $\log_{10}(N_{\mathrm{grid}}/N_{\mathrm{atoms}}) \geq 4$ ($T_{\mathrm{grid}} \leq 10^{-7}$) for which the errors are almost indistinguishable from the full grid calculations. However, for smaller grids slight deviations between the curve for the full grid calculations and the coarse grid calculations exist, but the errors are still negligibly small. It should be noted, however, that a significant speedup can be achieved only for functionals without exact exchange, since the calculations of the exchange matrix takes usually much longer than the numerical integrations.

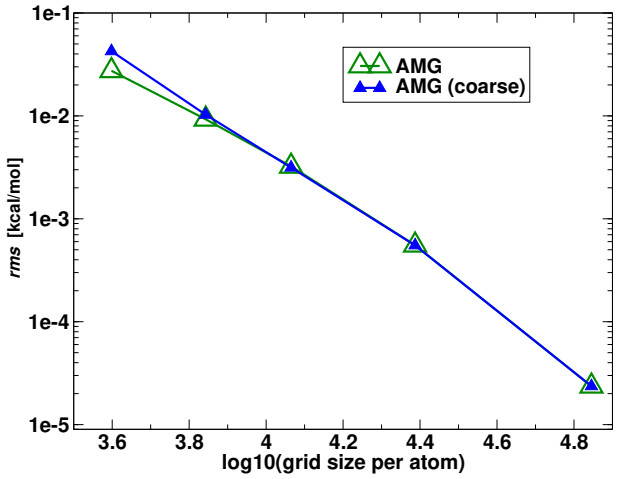


Figure 12. Grid errors of thermochemical energies using the full grids or coarse grids in each SCF cycle using the def2-TZVP basis set (see text)

G. Convergence of absolute energies

In general, only relative energies are relevant in chemistry, and therefore we focussed on thermochemical data in the previous sections. However, when calculating energy differences

the number and types of atoms do not change, and it can be expected that errors in the absolute energies largely cancel out in the relative energies. It is therefore of some interest to look at the convergence of the absolute energies with the grid threshold (or grid level) as well.

In Table VI the root mean square and maximum absolute errors of the 87 molecules in the FH51 reaction energy test set[98,99] are shown for various functionals and a range of AMG grid thresholds. The reference energies are computed as described in section V C. Since the errors of absolute energies are extensive quantities, i.e. they depend on the molecular size, the errors per atom are presented. The individual energies of all molecules can be found in the supplementary material.

Using LDA or GGA functionals (e.g. BLYP) the absolute grid errors are small and converge smoothly to the reference values. For the Molpro default threshold $T_{\mathrm{grid}} = 10^{-6}$ the *rms* error per atom amounts to about 1 μE_h, which corresponds to 0.0015 kcal mol^{-1}. However, as already seen for the relative energies, the errors are 1-2 orders of magnitude larger for meta-GGA functionals like M06 or r2SCAN, and they also converge more slowly towards zero with decreasing grid threshold. The largest errors are found for the M06 functional. In this case the *rms* error still amounts to about 42 μE_h (0.026 kcal mol^{-1}) for a grid threshold of 10^{-8}, and the maximum error for this threshold is nearly an order of magnitude larger.

Table VII shows the error statistics for the reaction energies of the FH51 test set. In most cases the errors are tiny and negligible for all tested thresholds. Only with M06 the errors are unacceptably large (up to 2 kcal mol^{-1}) with default thresholds. To reduce them below 0.1 kcal mol^{-1}, a threshold $T_{\mathrm{grid}} = 10^{-10}$ is needed, which makes the calculations significantly more expensive than with other functionals.

Additional tables with grid errors per atom for molecules containing second-row atoms, transition metals, and some molecular anions are presented in the supplementary material. A number of these molecules were also used in the work of Köster et al.[57] and considered to be difficult therein. Overall, the absolute errors per atom in this set are about 1.5 times larger than for the FH51 set, but otherwise the conclusions are very similar.

In order to compare the absolute energy grid errors for different grid methods we have considered the C_{60} molecule using the BLYP and TPSS functionals, cf. Fig. 13. The total

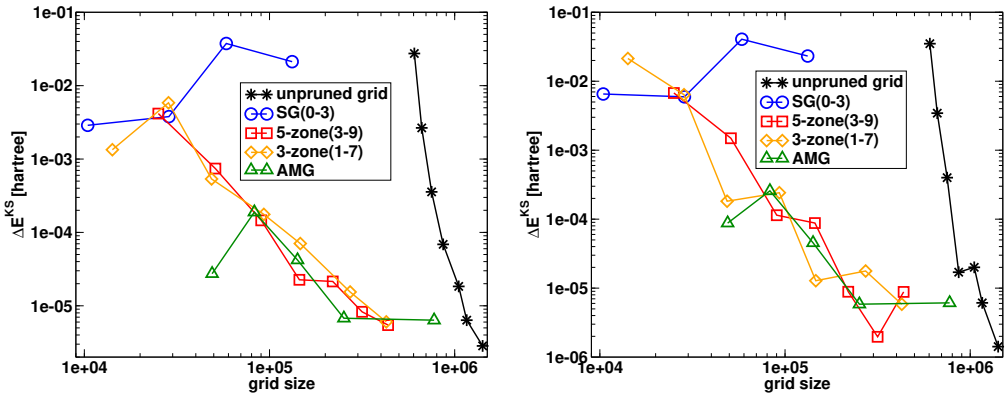
Table VI. Root mean square (*rms*) and maximum (*max*) energy differences for the total energies of BLYP, TPSS, M06 and r2SCAN for the FH51 systems[98,99] for AMG with different thresholds to the reference grid values. Energies are normalised over the number of atoms for each system and are shown in μE_h. The total energies for the 87 molecules range from -1 E_h to -1050 E_h. Basis set: def2-TZVP.

functional	error	$T_{\text{grid}} = 10^{-6}$	$T_{\text{grid}}10^{-7}$	$T_{\text{grid}} = 10^{-8}$	$T_{\text{grid}} = 10^{-10}$	$T_{\text{grid}} = 10^{-14}$
BLYP	*rms*	0.99	0.25	0.06	0.01	0.0
	max	3.32	0.68	0.21	0.05	0.01
TPSS	*rms*	1.83	0.57	0.30	0.06	0.0
	max	5.42	1.83	0.83	0.24	0.0
M06	*rms*	52.68	42.92	41.58	1.66	0.02
	max	329.77	379.38	380.31	5.97	0.11
r2SCAN	*rms*	20.73	6.30	3.90	0.90	0.0
	max	39.44	18.33	9.73	3.65	0.01

Table VII. Root mean square (*rms*) and maximum (*max*) energy differences for the reaction energies of BLYP, TPSS, M06 and r2SCAN for the FH51 systems[98,99] for AMG with different thresholds to the reference grid values. Energies are shown in $\text{kcal}\,\text{mol}^{-1}$. Reaction energies for the 51 reactions range from 1 to 150 $\text{kcal}\,\text{mol}^{-1}$ in magnitude (about 30 $\text{kcal}\,\text{mol}^{-1}$ on average). Basis set: def2-TZVP.

functional	error	$T_{\text{grid}} = 10^{-6}$	$T_{\text{grid}} = 10^{-7}$	$T_{\text{grid}} = 10^{-8}$	$T_{\text{grid}} = 10^{-10}$	$T_{\text{grid}} = 10^{-14}$
BLYP	*rms*	0.016	0.004	0.001	0.0	0.0
	max	0.046	0.012	0.004	0.001	0.0
TPSS	*rms*	0.028	0.008	0.004	0.0	0.0
	max	0.110	0.024	0.010	0.002	0.0
M06	*rms*	0.541	0.361	0.333	0.021	0.0
	max	2.000	1.521	1.578	0.052	0.0
r2SCAN	*rms*	0.019	0.011	0.004	0.0	0.0
	max	0.046	0.040	0.014	0.002	0.0

energy errors obtained with the SG-x grids and the BLYP functional are above 1 kcal mol^{-1} and there is no convergence from SG-0 to SG-3. Even larger errors are found for the TPSS functional. For the other grid methods the errors nicely converge to 10^{-5} E_h or below with increasing grid size. The black curves show the convergence as a function of the Lebedev order L used at each grid point (unpruned). The radial grid is the same as in the reference AMG calculations. It is found that the convergence with L is rather fast, and an absolute energy error below 0.1 kcal mol^{-1} (0.16 mH) is reached with $L = 41$.



(a) BLYP (b) TPSS

Figure 13. Convergence of the KS energies with respect to the grid size for the C_{60} molecule. The black curve in the figure shows the deviations for an unpruned grid for the Lebedev orders $L = 11, 21, 31, 41, 53, 59, 71$ and $T_{\mathrm{rad}} = 5 \cdot 10^{-15}$ compared to the largest unpruned grid with $L = 83$. (units: Hartree).

H. Dipole moments

The performances of the different integration grids were analysed for the electron number integrals and thermochemical energy values in the previous sections V D – V F. While for the electron number integrals Hartree-Fock densities were used to rule out any secondary effect from the exchange-correlation potential contributions that determine the shape of the density, this effect is to some extent contained in the KS energies since these are computed with self-consistent densities. A more direct analysis of the grid errors on the self-consistent density can be made, however, by considering first-order electric molecular properties such as the dipole moments. Fig. 14 shows the mean absolute errors for the benchmark tests

for the dipole moments using the different quadrature methods (the BLYP functional was chosen for this test). It can be seen that overall the errors are quite small. Even with the smallest grids the average dipole moments do not differ by more than 10^{-3} a.u. from the reference grid results. The relative hierarchy in accuracy found for the electron number and energy benchmark tests in the previous sections is the same also for the dipole moment. The fixed SG grids are less well suited to achieve high accuracies for dipole moments than the other grid methods considered. Just as for the electron number and energy tests presented in the previous sections, the AMG method yields the highest accuracies also for the dipole moments for given grid sizes.

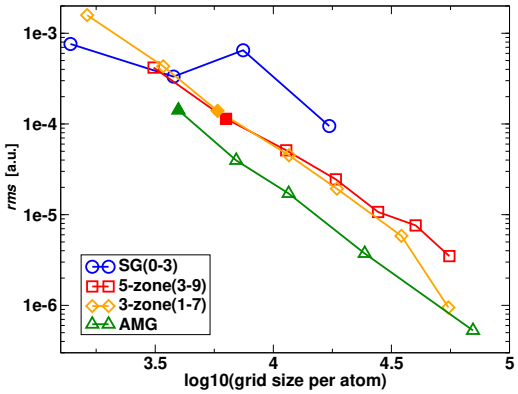


Figure 14. Average absolute grid errors for the dipole moments of the molecules from the benchmark tests. Calculations were done using the BLYP functional and def2-TZVP basis set. The grids that correspond to the default settings for each method are marked by filled symbols.

VI. CONCLUSION

In this work a number of DFT quadrature methods have been reviewed and tested for a number of molecular properties, namely electron numbers obtained by numerical integration, thermochemical relative energies from a large set of 254 different chemical reactions from the GMTKN30 databases, as well as absolute total energies and molecular dipole moments. The range of quadrature methods considered includes the fixed SG grids by Gill *et al.*,[35,52] the 3-zone grids by Treutler *et al.*,[36] a 5-zone grid that is similar to the grid used in Orca4,[54,55] as well as the adaptive grid scheme AMG implemented in the Molpro program.[60–62] The fixed grid methods have the advantage that they are quite simple and available in several electronic

structure codes. While the 3-zone and 5-zone grids can be enlarged to systematically reduce the errors to negligible values, this is not the case for the SG-x (x=0-3) grids. In contrast to the fixed grids methods, the AMG method takes into account the molecular environment for pruning the angular grids at each radial grid point. This leads to significantly reduced errors for given grid sizes. For the default AMG grids, which are mostly smaller than all other default fixed grids (except for SG-0), the errors are typically 1-2 orders of magnitude smaller than those of the default fixed grids. The additional computation time needed for the pruning scales linearly with molecular size and is negligible compared to the time needed for one SCF cycle (even without exact exchange).

With the recommended default grids sizes, the SG-1 method yielded the largest root mean squared errors in the thermochemical benchmarks, on the average about 0.2 kcal mol^{-1}. For the BLYP and PBE0 functionals the maximum errors for the different grid methods were found to be of the order of 0.45 kcal mol^{-1} for the SG-1 grid, about 0.15 kcal mol^{-1} for the 3-zone(level=3) and 5-zone(level=4) grids, and about 0.06 kcal mol^{-1} for the AMG($T_{\mathrm{grid}} = 10^{-6}$) default grids. The grid errors of (hybrid-)GGA type functionals were found to be insensitive to the functional used. However, for meta-GGAs like TPSS the errors are approximately one order of magnitude larger than for GGAs and hybrid GGAs and converge more slowly with increasing grid sizes. Most critical are M06, r2SCAN and similar functionals. If these are used it should be carefully checked that the grid errors are sufficiently small.

The grid errors have also been compared for various basis sets. It was found that they are largely independent of the underlying basis set, and that the basis set errors themselves are orders of magnitude larger than the grid errors. For example, the average deviations of the augmented correlation-consistent double-zeta and triple zeta basis sets results from the corresponding quadruple zeta ones amount to about 2.0 and 0.14 kcal mol^{-1}, respectively, for the 254 chemical reactions considered. Compared to this, the average grid errors of less than 0.01 kcal mol^{-1} with the default adaptive grid ($T_{\mathrm{grid}} = 10^{-6}$) are negligible.

A possible drawback of adaptive grid schemes is that the grid depends on the molecular geometry. This can lead to potential energy surfaces that are not microscopically smooth, and it can also affect the convergence of geometry optimizations. Extensive tests for geometry optimizations of 41 organic molecules from the GDB-13 database with rather crude starting

geometries showed that with medium grid size ($T_{\mathrm{grid}} = 10^{-7}$) convergence of the geometry optimizations is similar as with the fixed SG-x grids, and further reduction of the grid threshold does not reduce the iteration count. It is also possible to freeze the grid definition at each atom (denoted "meta-grid"), so that during geometry optimizations it behaves like a fixed grid. Such freezing may slightly change the final energies, but the effects are very small (typically in the μE_h range), and the optimized structures are hardly affected. These issues will be investigated in more detail in future work.

SUPPLEMENTARY MATERIAL

The supplementary material contains parameters for the 78-point Lebedev grid used in this work and diagrams showing the perfomances of the grid pruning methods for other statistical quantities (*mae* and *max* deviations) as well as total energies for all molecules obtained with the reference grid. Furthermore, performances of the pruning methods for the C_{60} molecule and the CN^- anion are shown and the grid errors for the meta GGA functionals TPSS, M06 and r2SCAN are shown for each molecule from the FH51 data base.

ACKNOWLEDGMENTS

We would like to thank John Herbert and Saswata Dasgupta for their help regarding the implementation of the SG-2 and SG-3 grids and Filipp Furche and Yannick Franze for their help regarding the implementation of the 3-zone grids. Particularly, we thank Yannick Franze for his effort in scanning pages of Oliver Treutler's PhD thesis from 1996 (Karlsruhe) which is not available in electronic form. We are also grateful to Peter Bygrave, Gerald Knizia, and Fred Manby for important contributions to earlier versions of the DFT program in Molpro.

REFERENCES

[1]P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," Phys. Rev. **136**, B864–B871 (1964).

[2]W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," Phys. Rev. **140**, A1133–A1138 (1965).

[3]F. M. Bickelhaupt and E. J. Baerends, "Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry," in Reviews in Computational Chemistry (John Wiley & Sons, Ltd, 2000) pp. 1–86.

[4]R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, "Achieving linear scaling in exchange-correlation density functional quadratures," Chem. Phys. Lett. **257**, 213 (1996).

[5]C. Ochsenfeld, J. Kussmann, and D. Lambrecht, "Reviews in computational chemistry," (Wiley-VCH, John Wiley & Sons, Inc., 2007) Chap. Linear-Scaling Methods in Quantum Chemistry.

[6]D. R. Bowler and T. Miyazaki, "O(n) methods in electronic structure calculations," Rep. Prog. Phys. **75**, 036503 (2012).

[7]A. D. Becke, "Perspective: Fifty years of density-functional theory in chemical physics," J. Chem. Phys. **140**, 18A301 (2014).

[8]N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry: an overviewand extensive assessment of 200 density functionals," Mol. Phys. **115**, 2315 (2017).

[9]S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, "Dispersion-corrected mean-field electronic structure methods," Chem. Rev. **116**, 5105 (2016).

[10]A. Heßelmann, "Polarisabilities of long conjugated chain molecules with density functional response methods: The role of coupled and uncoupled response," J. Chem. Phys. **142**, 164102 (2015).

[11]O. Gritsenko and E. J. Baerends, "Asymptotic correction of the exchange-correlation kernel of time-dependent density functional theory for long-range charge-transfer excitations," J. Chem. Phys. **121**, 655 (2004).

[12]D. J. Tozer and N. C. Handy, "Improving virtual Kohn-Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities," J. Chem. Phys. **1998**, 10180 (1998).

[13]M. Städele, J. A. Majewski, P. Vogl, and A. Görling, "Exact exchange Kohn-Sham formalism applied to semiconductors," Phys. Rev. Lett. **79**, 2089 (1997).

[14]R. Strange, F. R. Manby, and P. J. Knowles, "Automatic code generation in density functional theory," Comp. Phys. Commun. **136**, 310 (2001).

[15]S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, "Recent developments in libxc - a comprehensive library of functionals for density functional theory," SoftwareX **7**, 1–5 (2018).

[16]J. A. Pople, P. M. Gill, and B. G. Johnson, "Kohn–Sham density–functional theory within a finite basis set," Chem. Phys. Lett. **199**, 557 (1992).

[17]S. Lehtola, F. Blockhuys, and C. Van Alsenoy, "An Overview of Self-Consistent Field Calculations Within Finite Basis Sets," Molecules **25** (2020), 10.3390/molecules25051218.

[18]A. V. Arbuznikov and M. Kaupp, "The self-consistent implementation of exchange-correlation functionals depending on the local kinetic energy density," Chem. Phys. Lett. **381**, 495 (2003).

[19]A. D. Becke, "A multicenter numerical integration scheme for polyatomic molecules," J. Chem. Phys. **88**, 2547 (1988).

[20]J. I. Rodriguez, D. C. Thompson, P. W. Ayers, and A. M. Köster, "Numerical integration of exchange- correlation energies and potentials using transformed sparse grids," J. Chem. Phys. **128**, 224103 (2008).

[21]T. L. Beck, "Real–space mesh techniques in density-functional theory," Rev. Mod. Phys. **72**, 1041 (2000).

[22]J. Kong, S. T. Brown, and L. Fusti-Molnar, "Efficient computation of the exchange-correlation contribution in the density functional theory through multiresolution," The Journal of Chemical Physics **124**, 094109 (2006).

[23]N. J. Russ, C. Chang, and J. Kong, "Fast computation of DFT nuclear gradient with multiresolution," Canadian Journal of Chemistry **89**, 657–662 (2011).

[24]C.-M. Chang, N. J. Russ, and J. Kong, "Efficient and accurate numerical integration of exchange-correlation density functionals," Phys. Rev. A **84**, 022504 (2011).

[25]K. S. Werpetinski and M. Cook, "Grid–free density–functional technique with analytical energy gradients," Phys. Rev. A **52**, R3397 (1995).

[26]K. S. Werpetinski and M. Cook, "A new grid-free density-functional technique: Application to the torsional energy surfaces of ethane, hydrazine, and hydrogen peroxide," J. Chem. Phys. **106**, 7124 (1997).

[27]Y. C. Zheng and J. Almlöf, "Density functionals without meshes and grids," Chem. Phys. Lett. , 397 (1993).

[28]Y. C. Zhenh and J. E. Almlöf, "A grid–free DFT implementation of non–local functionals and analytical energy derivatives," J. Molecular Struct. (Theochem) **388**, 277 (1996).

[29]K. R. Glaesemann and M. S. Gordon, "Investigation of a grid–free density functional theory (dft) approach," J. Chem. Phys. **108**, 9959 (1998).

[30]K. R. Glaesemann and M. S. Gordon, "Evaluation of gradient corrections in grid–free density functional theory," J. Chem. Phys. **110**, 6580 (1999).

[31]K. R. Glaesemann and M. S. Gordon, "Auxiliary basis sets for grid–free density functuonal theory," J. Chem. Phys. **112**, 10738 (2000).

[32]G. Berghold, J. Hutter, and M. Parrinello, "Grid–free DFT implementation of local and gradient–corrected XC functionals," Theor. Chem. Acc. **99**, 344 (1998).

[33]A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098–3100 (1988).

[34]C. W. Murray, N. C. Handy, and G. J. Laming, "Quadrature schemes for integrals of density functional theory," Mol. Phys. **78**, 997 (1993).

[35]P. M. W. Gill, B. G. Johnson, and J. A. Pople, "A standard grid for density functional calculations," Chem. Phys. Lett. **209**, 506 (1993).

[36]O. Treutler and R. Ahlrichs, "Efficient molecular numerical integration schemes," J. Chem. Phys. **102**, 346 (1995).

[37]M. E. Mura and P. J. Knowles, "Improved radial grids for quadrature in molecular density-functional calculations," J. Chem. Phys. **104**, 9848–9858 (1996).

[38]R. Lindh, P.-Å. Malmqvist, and L. Gagliardi, "Molecular integrals by numerical quadrature. I. Radial integration," Theor. Chem. Acc. **106**, 178 (2001).

[39]P. M. W. Gill and S.-H. Chien, "Radial quadrature for multiexponential integrands," J. Comput. Chem. **24**, 732 (2003).

[40]A. El-Sherbiny and R. A. Poirier, "An evaluation of the radial part of numerical integration commonly used in dft," J. Comput. Chem. **25**, 1378 (2004).

[41]J. Gräfenstein and D. Cremer, "Efficient density-functional theory integrations by locally augmented radial grids," J. Chem. Phys. **127**, 164113 (2007).

[42]K. Kakhiani, K. Tsereteli, and P. Tsereteli, "A program to generate a basis set adaptive radial quadrature grid for density functional theory," Comput. Phys. Comm. **180**, 256 (2009).

[43]M. Mitani, "An application of double exponential formula to radial quadrature grid in density functional calculation," Theor. Chem. Acc. **130**, 645 (2011).

[44]M. Mitani and Y. Yoshioka, "Numerical integration of atomic electron density with double exponential formula for density functional calculation," Theor. Chem. Acc. **131**, 1169 (2012).

[45]V. I. Lebedev, Zh. Vychisl. Mat. Mat. Fiz. **15**, 48 (1975).

[46]V. I. Lebdev, Zh. Vychisl. Mat. Mat. Fiz. **16**, 293 (1976).

[47]V. I. Lebdev, Sibirsk. Mat. Zh. **18**, 132 (1977).

[48]V. I. Lebedev and A. L. Skorokhodov, Russ. Acad. Sci. Docl. Math. **45**, 587 (1992).

[49]C. Daul and S. Daul, "Symmetrical nonproduct quadrature rules for a fast calculation of multicenter integrals," Int. J. Quant. Chem. **61**, 219 (1997).

[50]V. I. Lebdedev and D. N. Laikov, "A quadrature formula for the sphere of the 131st algebraic order of accuracy," Doklady Mathematics **59**, 477 (1999).

[51]C. Ahrens and G. Beylkin, "Rotationally invariant quadratures for the sphere," Proc. R. Soc. A **465**, 3103 (2009).

[52]S.-H. Chien and P. M. W. Gill, "SG-0: A small standard grid for DFT quadrature on large systems," J. Comput. Chem. **27**, 730 (2006).

[53]S. Dasgupta and J. M. Herbert, "Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3," J. Comput. Chem. **38**, 869 (2017).

[54]F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The orca quantum chemistry program package," J. Chem. Phys. **152**, 224108 (2020).

[55]F. Neese, ORCA – An ab initio, DFT and semiempirical SCF-MO package. Version 4.2.1, Max-Planck Institut für Kohlenforschung, Kaiser-Wilhelm Platz, 45470 Mühlheim, Germany.

[56]M. Krack and A. M. Köster, "An adaptive numerical integrator for molecular integrals," J. Chem. Phys. **108**, 3226 (1998).

[57]A. M. Köster, R. Flores-Moreno, and J. U. Reveles, "Efficient and reliable numerical integration of exchange-correlation energies and potentials," J. Chem. Phys. **121**, 681 (2004).

[58]S. Lehtola and M. A. L. Marques, "Many recent density functionals are numerically ill–behaved," J. Chem. Phys. **157**, 174114 (2022).

[59]E. R. Johnson, A. D. Becke, C. D. Sherrill, and G. A. DiLabio, "Oscillations in meta-generalized-gradient approximation potential energy surfaces for dispersion-bound complexes," J. Chem. Phys. **131**, 03411 (2009).

[60]H.-J. Werner, P. J. Knowles, A. Hesselmann, G. Knizia, D. A. Kreplin, Q. Ma, et al., "MOLPRO, version 2022.1, a package of ab initio programs," (2022), see http://www.molpro.net.

[61]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package," WIREs Comput Mol Sci **2**, 242 (2012).

[62]H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, T. F. Miller, III, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut, and M. Sibaev, "The molpro quantum chemistry package," J. Chem. Phys. **152**, 144107 (2020).

[63]L. Goerigk and S. Grimme, "Efficient and accurate double-hybrid-meta-GGA density functionalssevaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions," J. Chem. Theory Comput. **7**, 291 (2011).

[64]N. C. Handy and S. F. Boys, "Integration points for the reduction of boundary conditions," Theor. Chim. Acta **31**, 195 (1973).

[65]H. Laqua, J. Kussmann, and C. Ochsenfeld, "An improved molecular partitioning scheme for numerical quadratures in density functional theory," J. Chem. Phys. **149**, 204111 (2018).

[66]R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, "Electronic structure calculations on workstation computers: the program system TURBOMOLE," Chem. Phys. Lett. **162**, 165 (1989).

[67]M. Head–Gordon et al., "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," Mol. Phys. **113**, 184 (2015).

[68]T. V. Voorhis and G. E. Scuseria, "A novel form for the exchange–correlation energy functional," J. Chem. Phys. **109**, 400 (1998).

[69]H. Takahashi and M. Mori, "Double exponential formulas for numerical integration," Publ. RIMS, Kyoto Univ. **9**, 721 (1974).

[70]V. Termath and J. Sauer, "Optimized molecular integration schemes for density functional theory ab initio molecular dynamics simulations," Chem. Phys. Lett. **255**, 187 (1996).

[71]J. Lehtola, M. Hakala, A. Sakko, and K. Hämäläinen, "Erkale-a flexible program package for x-ray properties of atoms and molecules," Journal of Computational Chemistry **33**, 1572–1585 (2012).

[72]E. Rudberg, E. H. Rubensson, and P. Salek, "Kohn-sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage," J. Chem. Theory Comput. **7**, 340 (2011).

[73]M. Kaupp, A. V. Arbuznikov, A. Heßelmann, and A. Görling, "Hyperfine coupling constants of the nitrogen and phosphorus atoms: A challenge for exact–exchange density–functional and post-HartreeâĂŞFock methods," J. Chem. Phys. **132**, 184107 (2010).

[74]E. Clementi and D. Raimondi, IBM Res. Note , NJ–27 (1963).

[75]R. J. Barlow, Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Wiley, 1993).

[76]G. Knizia, (2014), *private communication*.

[77]J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," Phys. Rev. B **45**, 13244 (1992).

[78]J. C. Slater, "Atomic shielding constants," Phys. Rev. **36**, 57 (1930).

[79]A. R. Leach, Molecular Modelling: Principles and Applications (Addison Wesley Publishing Company, 1997).

[80]W. Kahan, "Further remarks on reducing truncation errors," Communications of the ACM **8**, 40 (1965).

[81]Y. Guo, K. Sivalingam, E. F. Valeev, and F. Neese, "SparseMaps - A systematic infrastructure for reduced-scaling electronic structure methods. III. Linear-scaling multireference domain-based pair natural orbital N-electron valence perturbation theory," The Journal of Chemical Physics **144**, 094111 (2016).

[82]L. Goerigk and S. Grimme, "A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions," Phys. Chem. Chem. Phys. **13**, 6670 (2011).

[83]S. Heo and Y. Xu, "Constructing fully symmetric cubature formulae for the sphere," Math. Comp. **70**, 269 (2001).

[84]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[85]F. Weigend, "Hartree-fock exchange fitting basis sets for h to rn," Journal of Computational Chemistry **29**, 167–175 (2008).

[86]T. H. Dunning, Jr., "Gaussian basis functions for use in molecular calculations. I. contraction of (9s5p) atomic basis sets for the first–row atoms," J. Chem. Phys. **53**, 2823 (1970).

[87]T. H. Dunning, Jr., "Gaussian basis functions for use in molecular calculations. III. contraction of (10s6p) atomic basis sets for the first–row atoms," J. Chem. Phys. **55**, 716 (1971).

[88]T. H. Dunning Jr., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007 (1989).

[89]R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first–row atoms revisited. Systematic basis sets and wave functions," J. Chem. Phys. **96**, 6796 (1992).

[90]F. Weigend, "A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," Phys. Chem. Chem. Phys. **4**, 4285–4291 (2002).

[91]A. D. Becke, "Density–functional exchange–energy approximation with correct asymptotic behavior," Phys. Rev. A **38**, 3098 (1988).

[92]C. Lee, W. Yang, and R. G. Parr, "Development of the Colle–Salvetti correlation–energy formula into a functional of the electron density," Phys. Rev. B **37**, 785 (1988).

[93]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).

[94]J. P. Perdew, V. N. Staroverov, J. Tao, and G. E. Scuseria, "Density functional with full exact exchange, balanced nonlocality of correlation, and constraint satisfaction," Phys. Rev. A **78**, 052513 (2008).

[95]J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, "Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation," J. Phys. Chem. Lett. **11**, 8208 (2020).

[96]Y. Zhao and D. G. Truhlar, "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals," Theor. Chem. Acc. **120**, 215 (2006).

[97]A. Heßelmann, "Assessment of a nonlocal correction scheme to semilocal density functional theory methods," J. Chem. Theory Comput. **9**, 273 (2013).

[98]J. Friedrich and J. Hänchen, "Incremental CCSD(T)(F12*)|MP2: A Black Box Method To Obtain Highly Accurate Reaction Energies," J. Chem. Theory Comput. **9**, 5381 (2013).

[99]J. Friedrich, "Efficient calculation of accurate reaction energies – assessment of different models in electronic structure theory," J. Chem. Theory Comput. **11**, 3596 (2015).

[100]G. Jansen, "Symmetry–adpated perturbation theory based on density functional theory for noncovalent interactions," WIREs Comput. Mol. Sci. **4**, 127 (2013).

[101]K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, "Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials," Theor. Chem. Acc. **97**, 119 (1997).

[102]D. J. Tozer, M. E. Mura, R. D. Amos, and N. C. Handy, "Implementation of analytic derivative density functional theory codes on scalar and parallel architectures," AIP Conference Proceedings **330**, 3 (1995).