Article

Static Polarizabilities at the Basis Set Limit: A Benchmark of 124 Species

Anders Brakestad, Stig Rune Jensen, Peter Wind, Marco D'Alessandro, Luigi Genovese, Kathrin Helen Hopmann, and Luca Frediani*

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆘 Supporting Information

ABSTRACT: Benchmarking molecular properties with Gaussian-type orbital (GTO) basis sets can be challenging, because one has to assume that the computed property is at the complete basis set (CBS) limit, without a robust measure of the error. Multiwavelet (MW) bases can be systematically improved with a controllable error, which eliminates the need for such assumptions. In this work, we have used MWs within Kohn−Sham density functional theory to compute static polarizabilities for a set of 92 closed-shell and 32 open-shell species. The results are compared to recent benchmark calculations employing the GTO-type aug-pc4 basis set. We observe discrepancies between GTO and MW results for several species, with open-shell systems showing the largest deviations. Based on linear response calculations, we show that these discrepancies originate from artifacts caused by the field strength and

BSIEs in GTO polarizabilities
MRChem
Powered by MultiWavelets

that several polarizabilies from a previous study were contaminated by higher order responses (hyperpolarizabilities). Based on our MW benchmark results, we can affirm that aug-pc4 is able to provide results close to the CBS limit, as long as finite difference effects can be controlled. However, we suggest that a better approach is to use MWs, which are able to yield precise finite difference polarizabilities even with small field strengths.
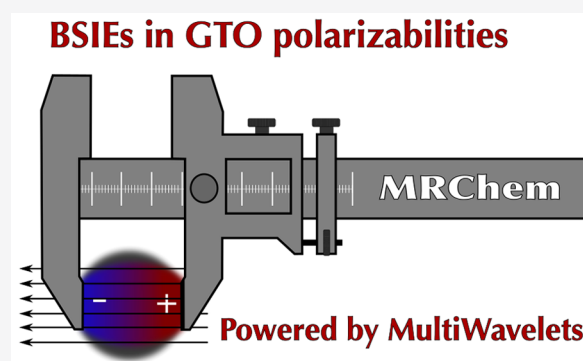
1. INTRODUCTION

Molecular electronic structure calculations are a widespread tool in chemistry, biology, and materials science. Such a diffusion across disciplines has been enabled by Kohn−Sham density functional theory (KS-DFT, hereafter just "DFT")[1] which brought about calculations with accuracy comparable to coupled cluster with singles and doubles (CCSD) at the computational cost of a single-determinant method like Hartree−Fock (HF). A large part of the current development of theoretical methods is concerned with obtaining accurate energies, which are essential to interpret and predict chemical reactivity.

Molecular properties constitute another important area of method development. Electric dipole polarizabilities are related to important processes in chemistry; for example, they hold a key role in our understanding of intra- and intermolecular interactions such as dispersion,[2,3] they are at the foundation of techniques such as Raman spectroscopy and Raman optical activity,[4] and they are employed in the development of accurate force fields for molecular simulations.[5,6] It is therefore highly relevant to assess the accuracy of polarizability predictions within the density functional theory (DFT) framework.

The quality of a given DFT calculation depends on two factors: the density functional approximation (DFA) and the basis set. In order to fairly assess the performance of functionals and basis sets, we must distinguish between these two sources of error. While an ideal (nonexact) functional should be *accurate* and yield a result as close as possible to the corresponding full configuration interaction (FCI) calculation, an ideal basis should be *precise* and minimize the error with respect to the complete basis set (CBS) limit. Most functionals and basis sets are developed so as to provide the best possible energies, which sometimes rely on fortuitous error cancellation. The assessment of accuracy (functional) and precision (basis set) for molecular properties, such as polarizabilites, is therefore challenging.

Hait and Head-Gordon[7] benchmarked the accuracy of electric dipole polarizability predictions for a large number of DFAs against coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) calculations, for a set of 132 species. They employed the aug-pc4 basis set[8−11] for all DFT calculations, assuming that the obtained quantities were close
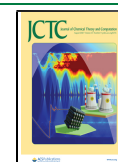
to the CBS limit, although they noted that this assumption may not hold for certain DFAs. We refer the reader to their excellent paper for details.[7]

Even the largest practical Gaussian-type orbital (GTO) basis sets are far from complete in the mathematical sense. In general, it must be *assumed* that GTO basis sets deliver results close to the CBS limit, even for the large aug-pc4 basis set: one cannot know how close to the limit a given basis set is without a reference value, and simply comparing with a larger GTO basis set does *not* in general guarantee that one converges to the CBS limit. For energies, the variational principle serves as a guide, but quantifying the basis set incompleteness error (BSIE) for other molecular properties is not a trivial task, and two issues lie at the heart of the challenge: (i) atomic orbital (AO) bases are generally developed by minimization of the total energy as the guiding principle and may therefore not be optimal for molecular properties, and (ii) hierarchical constructions of AO bases do not *guarantee* any rate of convergence of the molecular properties. While the Hylleraas−Undheim theorem[12] proves that the polarizability for non-dipolar molecules is a lower bound of the CBS limit, there is no guarantee that a systematic extension of an AO basis will in practice reach the CBS limit, unless the basis can formally be extended to completeness.

Multiwavelets (MWs)[13] have recently emerged as a powerful alternative to the traditional AO bases and are not subject to the same shortcomings as AO bases. MWs are a particular choice of wavelets[14] used to represent functions and operators on a real-space grid. To overcome the hurdles posed by real-space methods, such as large memory footprint and computational cost, MWs exploit adaptive grid refinement[15−17] and Cartesian separated representation of the required operators.[18] Such features are combined with a rigorous formalism with strict error control.[19−21] For molecular energies, it is possible to request a predefined precision with respect to the CBS limit, and for molecular properties such as polarizabilities, a steady progression toward the corresponding limit is observed.[22] MW calculations of polarizabilities performed at high precision can be employed as a true reference because they can be assumed complete to within the given precision. Such capabilities have been recently exploited in our group to perform two extensive benchmark studies on total and atomization energies[20] and on magnetizabilities and NMR shielding constants.[23]

The objective of the present paper has been to use MWs to assess whether aug-pc4 indeed is capable of delivering polarizabilities at the CBS limit, by comparing MW-based polarizabilities to the recent aug-pc4 benchmark.[7] We start by describing the mathematical framework for computing molecular properties with MWs. Next, we report the computational details and present and discuss our results. We also touch upon additional benefits of MWs related to the finite differences (FD) approach, and finish by summarizing our findings.

2. THEORETICAL FRAMEWORK

2.1. Molecular Properties as Energy Derivatives.
Molecular energies are affected by the presence of external fields. In particular, when a static electric field is applied, the total energy of the molecule can be expressed as a Taylor expansion with respect to the external field F[24]

$$E = E_0 + \sum_a \mu_a F_a + \frac{1}{2} \sum_{ab} \alpha_{ab} F_a F_b + \frac{1}{6} \sum_{abc} \beta_{abc} F_a F_b F_c$$
$$+ \frac{1}{24} \sum_{abcd} \gamma_{abcd} F_a F_b F_c F_d + ... \tag{1}$$

where a, b, c, ... are Cartesian directions. Such an expansion implicitly defines the components of the dipole moment (μ), the polarizability (α), and the hyperpolarizabilities (here limited to the first and second hyperpolarizabilities, β and γ, respectively).

The dipole moment components μ_a can be obtained as a simple expectation value of the corresponding dipole operators $\hat{\mu}_a$. Several approaches can be employed to compute (hyper)polarizabilities. Hait and Head-Gordon[7] have employed a second-order FD expression of the energy. For the diagonal components of the polarizability tensor, the expression reads

$$\alpha_{aa} = -\frac{E(F_a) + E(-F_a) - 2E_0}{F_a^2} + O(F^2) \tag{2}$$

Such an expression is formally equivalent to taking the derivative of eq 1 with respect to the external field and then applying a linear finite difference formula to the dipole moment:

$$\alpha_{ab} = \frac{\mu_a(F_b) - \mu_a(-F_b)}{2F_b} + O(F^2) \tag{3}$$

Both formulas have a leading error term that is quadratic in the applied field and proportional to the second hyperpolarizability. To minimize the error, it is therefore necessary to employ small fields, especially for molecules with large γ.

2.2. Multiwavelets.
Wavelet theory is a relatively recent branch of mathematics, dating back to the 1980s.[14,25] It constructs functions with the following properties: they are localized in both real and Fourier space, they achieve completeness as a limit in the L^2 sense, and they provide rigorous error control. *Multi*wavelets are a particular kind of wavelets that include several functions in one interval, as the "multi" prefix suggests. For the construction of MW bases and details about their properties, we refer to the literature on the subject.[13,15]

Finite Field Polarizability with Multiwavelets. In 2004, Harrison and co-workers[19] for the first time used MWs to solve the Kohn−Sham (KS) equations of DFT, demonstrating that arbitrary precision with respect to the CBS limit can be achieved also for general molecular systems;[19,26,27] previously, this was possible only for very small and highly symmetric molecules.[28]

Due to the large number of primitive MW basis functions necessary for the precise representation of the molecular orbitals, it is not practical to solve the KS equations in the traditional way by constructing the primitive Fock matrix and solving the corresponding Roothaan equations. Instead, the equations are rewritten in integral form

$$\varphi_i = -\hat{G}_i \hat{V} \varphi_i \tag{4}$$

using the bound-state Helmholtz integral operator, which is given as the inverse of the kinetic energy operator shifted by the orbital energy $\hat{G}_i = (\hat{T} - \epsilon_i)^{-1}$. There are several benefits with this reformulation: (i) it avoids the explicit construction and diagonalization of the primitive Fock matrix; (ii) it allows

for different and adaptive primitive basis sets for each orbital; (iii) it avoids the application of the kinetic operator as a second derivative, which is not numerically stable in the discontinuous MW basis; and (iv) the implicit construction of a huge virtual orbital space is not necessary, and one solves instead only for the occupied orbitals by iterating eq 4 to self-consistency.

In the presence of a uniform electric field $\vec{F}$, the KS potential operator in eq 4 reads

$$\hat{V} = \hat{V}_{\mathrm{nuc}} + \hat{J} + \hat{V}_{\mathrm{xc}} - \vec{F}\cdot\hat{\vec{\mu}} \qquad (5)$$

which features the nuclear ($\hat{V}_{\mathrm{nuc}}$), Hartree ($\hat{J}$), and exchange-correlation ($\hat{V}_{\mathrm{xc}}$) potentials. By solving the corresponding KS equations to obtain the ground state density $\rho = \sum_i |\varphi_i|^2$, we can compute the electric dipole moment as a function of field strength from the expectation value

$$\mu_a(\vec{F}) = \int \hat{\mu}_a \rho(\vec{F})\, \mathrm{d}r, \qquad a = x, y, z \qquad (6)$$

This procedure can be used to approximate the electric polarizability through the finite difference formula in eq 3.

Linear Response Polarizability with Multiwavelets. The starting point to obtain linear response properties with multiwavelets is standard perturbation theory. A small perturbation $\hat{h}^{(1)}$ is introduced, and all terms in eq 4 are expanded to first order into a set of Sternheimer equations,[22,29−31] which can be written in integral form

$$\varphi_i^{(1)} = -\hat{G}_i[\hat{V}^{(0)}\varphi_i^{(1)} + (1 - \hat{\rho}^{(0)})(\hat{h}^{(1)} + \hat{V}^{(1)})\varphi_i^{(0)}] \qquad (7)$$

where $\hat{G}_i$ is the same as that for the ground state problem and $\hat{\rho}^{(0)}$ is the ground state density projector, while $\hat{V}^{(0)} = \hat{V}_{\mathrm{nuc}}^{(0)} + \hat{J}^{(0)} + \hat{V}_{\mathrm{xc}}^{(0)}$ and $\hat{V}^{(1)} = \hat{J}^{(1)} + \hat{V}_{\mathrm{xc}}^{(1)}$ are the unperturbed and first-order perturbed potential operators, respectively. At this point, all unperturbed quantities are already known from solving the ground state problem, whereas first-order quantities are obtained by iterating eq 7 to self-consistency. The perturbed orbitals are then used to build the corresponding density perturbation

$$\rho^{(1)} = 2 \sum_i \varphi_i^{(1)} \varphi_i^{(0)} \qquad (8)$$

Here we have assumed real, time-independent perturbations: only one set of real, perturbed orbitals is obtained, which simplify the expression for the perturbed density.

The polarizability tensor is computed as the expectation value of the dipole operator $\hat{\vec{\mu}}$, on a density perturbed by the same operator

$$\alpha_{ab} = \int \hat{\mu}_a \rho_b^{(1)}\, \mathrm{d}r, \qquad a, b = x, y, z \qquad (9)$$

For details about the general derivation of time-dependent and imaginary (magnetic) perturbations in a MW framework, we refer the reader to the works by Sekino et al.[22] and Jensen et al.,[23] respectively.

3. COMPUTATIONAL DETAILS

Cartesian coordinates of the species studied here were obtained from Hait and Head-Gordon,[7] and a list of the species and their spin multiplicities is given in Table 1. The set of 124 species includes 92 closed-shell and 32 open-shell systems. The set is slightly smaller than the original one provided in the mentioned benchmark,[7] due to convergence issues encountered for the remaining species (missing species:

Table 1. 124 Species and Their Spin Multiplicities Used in This Study, Sorted Alphabetically[a]

^{1}AlF	^{1}Ar	^{1}Be	^{2}BeH	^{1}BeH$_2$	^{1}BF	^{1}BH$_2$Cl
^{1}BH$_2$F	^{1}BH$_3$	^{1}BHF$_2$	^{3}BN	^{2}BO	^{2}BS	^{2}C$_2$H
^{1}C$_2$H$_2$	^{2}C$_2$H$_3$	^{1}C$_2$H$_4$	^{1}CH$_2$BH	^{2}CH$_2$F	^{1}CH$_2$NH	^{1}CH$_2$PH
^{1}CH$_3$BH$_2$	^{1}CH$_3$Cl	^{1}CH$_3$F	^{1}CH$_3$NH$_2$	^{1}CH$_3$OH	^{1}CH$_3$SH	^{1}CH$_4$
^{1}Cl$_2$	^{1}ClCN	^{1}ClF	^{2}CN	^{1}CO	^{1}CO$_2$	^{1}CS
^{1}CSO	^{1}F$_2$	^{1}FCN	^{2}FCO	^{2}FH−OH	^{1}FNO	^{2}H
^{1}H$_2$	^{2}H$_2$CN	^{1}H$_2$O	^{2}H$_2$O−Li	^{1}HBO	^{1}HBS	^{1}HCCCl
^{1}HCCF	^{1}HCHO	^{1}HCHS	^{1}HCl	^{1}HCN	^{2}HCO	^{1}HCONH$_2$
^{1}HCOOH	^{1}HCP	^{1}He	^{1}HF	^{1}HNC	^{1}HNO	^{1}HNS
^{1}HO$_2$	^{1}HOCl	^{1}HOF	^{1}HOOH	^{2}Li	^{1}Li$_2$	^{1}LiBH$_4$
^{1}LiCl	^{1}LiCN	^{1}LiH	^{1}Mg	^{1}Mg$_2$	^{4}N	^{1}N$_2$
^{1}N$_2$H$_2$	^{1}N$_2$H$_4$	^{2}Na	^{1}Na$_2$	^{1}NaCl	^{1}NaCN	^{1}NaH
^{1}NaLi	^{2}NCO	^{1}Ne	^{3}NH	^{2}NH$_2$	^{1}NH$_2$Cl	^{1}NH$_2$F
^{1}NH$_2$OH	^{1}NH$_3$	^{1}NH$_3$O	^{1}NOCl	^{1}NP	^{3}O$_2$	^{1}O$_3$
^{2}OCl	^{1}OCl$_2$	^{2}OF	^{1}OF$_2$	^{2}OH	^{4}P	^{1}P$_2$
^{1}P$_2$H$_4$	^{3}PH	^{2}PH$_2$	^{1}PH$_2$OH	^{1}PH$_3$	^{1}PH$_3$O	^{1}S$_2$H$_2$
^{2}SCl	^{1}SCl$_2$	^{2}SF	^{1}SF$_2$	^{1}SH$_2$	^{2}SiH$_3$	^{1}SiH$_3$Cl
^{1}SiH$_3$F	^{1}SiH$_4$	^{1}SiO	^{3}SO	^{1}SO$_2$		

[a]The numbers of closed-shell and open-shell species are 92 and 32, respectively. Closed-shell species are indicated in blue, while open-shell species are indicated in red.

CH$_3$O, PS, CH$_3$, NO, CH$_2$, BH$_2$, SH, S$_2$). All coordinates and spin multiplicities are available in the form of *XYZ* files in the Supporting Information, together with Python scripts in the form of Jupyter Notebooks for our data analyses and figure generation.

3.1. Multiwavelet Calculations. MW calculations were performed with a prereleased version (1.0.0-alpha) of the MRC$_{\mathrm{hem}}$ program package.[32−34] The relative numerical precision was set to $\epsilon_{\mathrm{rel}} = 1 \times 10^{-7}$, the MW polynomial order to 11, and the norms of the orbital residuals between consecutive iterations ($\|\phi_i^{n+1} - \phi_i^n\|$) were converged to within $\epsilon_{\mathrm{mo}} = 1 \times 10^{-6}$, both for unperturbed and perturbed orbitals.

In general, it is expected that the converged total energy should be correct at least within ϵ_{rel} with respect to the CBS limit (relative precision). The orbital convergence necessary to reach this precision in total energy is roughly $\epsilon_{\mathrm{mo}} = \sqrt{\epsilon_{\mathrm{rel}}}$ because of quadratic error propagation. However, since we are also interested in properties with linear error propagation (dipole moment and polarizability), we converge the orbitals well beyond this point in order to get the maximum number of digits out of the chosen numerical precision ϵ_{rel},[35] and we then expect around ϵ_{mo} absolute precision in dipole moment and polarizability.

Static polarizabilities were computed with DFT using the LDA[36] and PBE[37] functionals, provided by the XCFun library.[38] Closed-shell species were treated with the spin-restricted formalism, and open-shell species, with the spin-unrestricted formalism. We used the central two-point finite difference formula of eq 3 to compute the diagonal elements α_{aa} of the polarizability tensor. Field strengths of ± 0.001 au were used for all species. Calculations without an applied electric field were first performed to generate initial orbitals for the FD calculations. Initial orbitals for zero-field calculations were generated by the superposition of atomic densities (SAD) method.[39] All MW calculations benefited from the Krylov subspace accelerated inexact Newton (KAIN) convergence accelerator.[40]

To validate our results, we also used MWs to compute static polarizabilities with linear response (LR) for a subset of the species. PBE response calculations were performed for 17 of the 124 species (closed-shell only), while LDA response
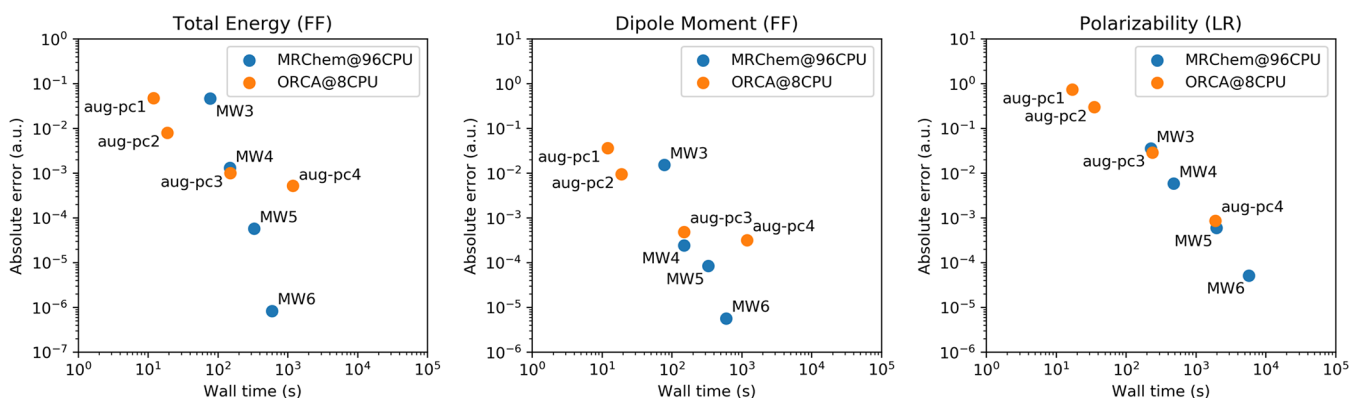
Figure 1. Scaling of computation time with precision for sequences of calculations on SiH$_3$Cl using MRC$_{hem}$ and ORCA. MWn correspond to ϵ_{rel} = 10^{-n} and $\epsilon_{mo} = 10\epsilon_{rel}$, and all errors are measured against the corresponding MW7 calculation, which is the parameter chosen for the full benchmark study below. The left and center plots show timings for a *single* finite field (FF) calculation with field strength 0.001 au in the z direction (along the Si−Cl bond) vs errors in total energy and the z component of the dipole moment, respectively. The right-hand plot shows the timings for the full polarizability tensor from linear response (LR) vs the error in its isotropic average. Note: ORCA calculations are on 8 CPU cores, while MRC$_{hem}$ are on 96 cores, so the computational cost is not directly comparable.

calculations were performed for 114 species. Numerical instabilities for GGA functionals at low-density values affected the convergence of the PBE LR calculations, explaining the low success rate. However, the cases that *did* converge are as precise as the corresponding LDA calculations: they converged to within 1×10^{-6}, indicating that they are not affected by these instabilities.

3.2. GTO Calculations. All FD results are taken from the work of Hait and Head-Gordon.[7] They used the energy expression in eq 2 to estimate the polarizability using a field strength of 0.01 au. However, they identified a few cases that were contaminated by hyperpolarizabilities, for which they reduced the field strength to 0.001 au, but this diagnosis was performed only at Hartree−Fock level and simply transferred to the DFT calculations.

In order to assess if further contamination could be present in the DFT results of Hait and Head-Gordon,[7] we performed *analytical* polarizability calculations using the ORCA program package, version 4.1.2,[41] with the PBE functional and the aug-pc4 basis set.[8−11] All species were treated with the spin-unrestricted formalism, and the integrals were computed over an angular Lebedev grid consisting of 770 points and a radial grid consisting of 50, 55, and 60 points for first, second, and third row elements, respectively ("grid7"). Self-consistent field convergence was accelerated by the direct inversion of the iterative subspace (DIIS) method.[42,43] The total energy change was converged to within 1×10^{-9} E_h, and the one-electron energy change to within 1×10^{-6} E_h (as defined by the "VeryTightSCF" ORCA keyword). The (default) resolution of identity (RI) approximation was turned off for all calculations in order to guarantee benchmark quality of the results (some initial test runs indicated a large dependence on the choice of auxiliary basis set).

3.3. Data Analysis. For all error analyses, we used the average polarizability, $\bar{\alpha}$, defined as

$$\bar{\alpha} = \frac{1}{3} \sum_{a=x,y,z} \alpha_{aa} \tag{10}$$

Polarizabilities from different calculations were compared using the relative error (RE) metric, which for species n was given by

$$RE_n = \frac{\bar{\alpha}_n - \bar{\alpha}_n^{REF}}{\bar{\alpha}_n^{REF}} \tag{11}$$

where the reference value may change depending on the comparison. The mean relative error (MRE) over N molecules was defined as

$$MRE = \frac{1}{N} \sum_{n=1}^{N} |RE_n| \tag{12}$$

3.4. Linear and Degenerate Open-Shell Systems. We have given special treatment to seven species in our data analysis (*vide infra*). In order to motivate this decision, it will be useful with a reminder of the electronic structure of linear and open-shell systems with a degenerate ground state. Such systems are particularly challenging to model for mean-field methods such as DFT. Let us consider NO as a prototypical molecule. It has an unpaired electron in a π orbital. Ideally, π_x and π_y are degenerate, but mean-field approaches break the symmetry as soon as one of the two orbitals is populated (the density and hence the KS potential become non-totally symmetric). For such systems, Hait and Head-Gordon[7] reported identical values for α_{xx} and α_{yy}, which is not what we observed: our MW-FD polarizabilities show that one component (the larger one) is virtually identical to the GTO-FD value, whereas the other is slightly smaller. According to Hait and Head-Gordon,[44] the smaller component should in this case be discarded as being unphysical, in connection to the symmetry breaking occurring for mean-field approaches.[45] Since the main objective of the present paper is to quantify the BSIE for the GTO basis set aug-pc4, we decided that the fairest analysis could be made by performing the same procedure as that by Hait and Head-Gordon.[7] We therefore explicitly set α_{xx} = α_{yy} in our MW data set by selecting the component closest to the xx/yy component reported by Hait and Head-Gordon.[7]

The seven species that received the above treatment in our data analysis are SCl, OCl, OH, OF, SF, BN, and NCO. To qualify for the special treatment, they had to fulfill the following three criteria (to within a tolerance of 1×10^{-4}):

1. $\alpha_{xx} = \alpha_{yy}$ in Hait and Head-Gordon's data set
2. $\alpha_{xx} \neq \alpha_{zz}$ in Hait and Head-Gordon's data set
3. $\alpha_{xx} \neq \alpha_{yy}$ in our data set
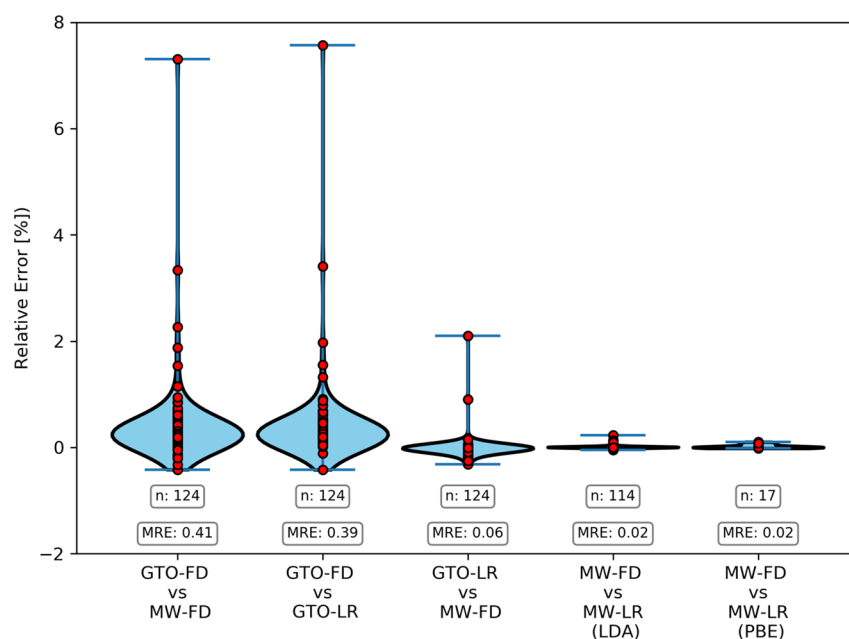
Figure 2. Violin plot summarizing the RE distributions discussed. Red dots indicate data points. The internal validation of our MW results demonstrates that the MW-FD polarizabilities are virtually free from field strength-related effects. GTO-FD polarizabilities display quite large errors, considering the size of the aug-pc4 basis set that was used, while GTO-LR display much smaller errors.

3.5. MW vs GTO: Practical Considerations. *Availability.*
MRC_{hem} is one of two programs currently available that offer an all-electron MW basis (the other is MADNESS[46]). Detailed instructions on how to obtain and compile the MRC_{hem} code, as well as a user manual, are available on the documentation web page.[34]

Ease of Use. From a user standpoint, selecting appropriate GTO basis sets for a particular application is not a simple task: The high parametrization of GTO basis sets means the user has to carefully evaluate several factors, for example, which family of GTOs to use (Pople, Jensen, Karlsruhe, Ahlrichs, Dunning, etc.), how many polarization functions to use, how many diffuse functions to use, whether to treat different atoms differently, and so on. Although the result can be converged to *a limit* within the given basis, no knowledge about the *CBS limit* can be inferred from it. Selecting the best basis is not trivial, and suboptimal choices based on "habit" and "popularity" are common (analogous to the "zoo" of DFAs[47]).

For MW calculations, all the user must do is to specify an overall numerical precision, in terms of convergence thresholds for energy and orbitals. This precision parameter is relative to the exact CBS limit, which is a key distinction from GTO. MWs can therefore provide the user with excellent precision and a quantifiable error without expert knowledge about basis sets.

Cost and Performance. At present, a calculation at moderate precision is cheaper to perform with GTOs because of a smaller prefactor. At very high precision, MW calculations become more competitive due to a linear scaling with respect to the precision. The most severe limitation of MWs is the memory requirements, which is rather demanding. For the molecules used in the present work, the total memory needed for the MW-FD calculations was typically between 50 and 150GB (Figure 4 in the Supporting Information), although this is rather efficiently distributed across several compute nodes on a cluster. The number of CPU hours needed for the MW-FD data set is presented in Figure 5 in the Supporting Information.

Figure 1 presents plots of computation time against increasingly larger GTO and MW basis sets for the calculation of total energy, dipole moment, and polarizability for the SiH_3Cl molecule: it shows that each additional digit of precision for MWs requires a predictable doubling of CPU time, while moving along the aug-pc*n* (n = 1, 2, 3, ...) series increases the computational cost by a larger factor, without a guarantee of gaining an additional digit in precision.

4. RESULTS AND DISCUSSION
A challenge in computational benchmark studies is the precision of the basis set: one has to assume that the computed reference property is at the CBS limit. The large GTO basis set aug-pc4[7,48] has been assumed to be close to the CBS limit for electrical properties. Here, we attempt to evaluate if aug-pc4 indeed is at the CBS limit for static polarizability predictions, by quantifying the BSIE associated with this basis set. We do this by comparing our reference MW polarizabilities to a recent aug-pc4 benchmark on DFT static polarizabilities.[7] All data presented herein are available via the Supporting Information accompanying this Article, or as a separate document at the Dataverse open-data repository.[49]

In order to isolate BSIEs, we need a detailed understanding of other potential errors, and in particular the error associated with using a finite field approach. In order to assign errors to the right source, we have considered the following types of calculations: (1) GTO-FD calculations; (2) MW-FD calculations; (3) GTO-LR calculations; (4) MW-LR calculatations.

The comparison of GTO-FD vs MW-FD is the central point of this contribution. The comparison of GTO-FD vs GTO-LR will shed light on potential errors due to finite field effects with GTOs; the comparison of MW-FD vs MW-LR will show how much MW results are affected by the FD approach, and the comparison of MW-FD with GTO-LR will be used to double-check that the discrepancies observed have been attributed correctly. The RE distributions listed here are summarized in Figure 2.
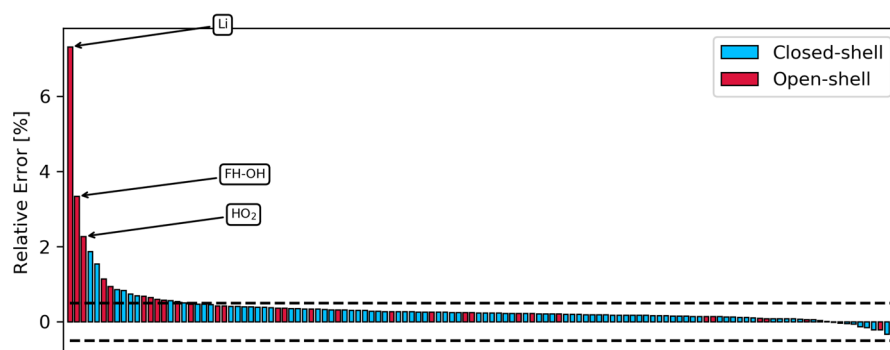
Figure 3. Distribution of REs of PBE polarizabilities for the 124 species, comparing GTO-FD with MW-FD, using the latter as a reference. The dashed lines are located at $\pm 0.5\%$ RE.
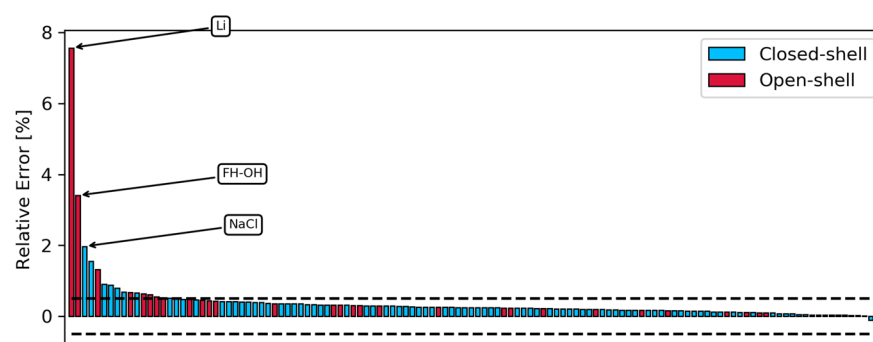
Figure 4. Distribution of REs of PBE polarizabilities for the 124 species, comparing GTO-FD and GTO-LR (both aug-pc-4), using the latter as a reference. The dashed lines are located at $\pm 0.5\%$ RE.
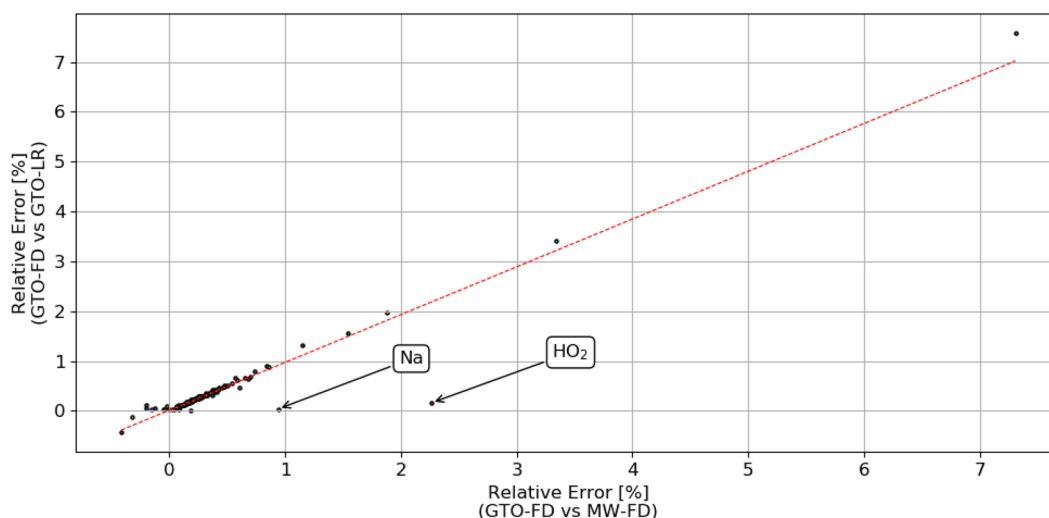
Figure 5. Correlation between the RE distributions presented in Figure 4 (x-axis) and Figure 3 (y-axis). The red dashed line indicates a least-squares linear fit with $r^2 = 0.97$.

4.1. Consistency of the MW Calculations: MW-FD vs MW-LR. To make sure that our MW-FD polarizabilities were not contaminated by field-related effects, we compared the MW-FD polarizabilities to MW-LR results, using both LDA and PBE, as shown in the right-most plots in Figure 2. The LDA validation included 114 of the 124 species, while the PBE validation included 17 closed-shell species. All results about these validations, including the list of species with converged LR values, are reported in the Supporting Information.

The maximum absolute RE and MRE of the LDA validation were 0.23 and 0.011%, respectively. The PBE validation yielded similar statistics. The PBE set was limited due to convergence problems. Nevertheless, we see no indication that LDA and PBE behave differently.

Based on the very high numerical precision that has been used throughout, we expect that the remaining discrepancy between MW-FD and MW-LR is due to the field strength of 0.001 au, although this has not been verified numerically. We conclude that our MW-FD polarizabilities have field-related errors *at least* below 1% but usually much lower than this.

4.2. How Good Are FD Results with the aug-pc4 Basis? To judge the quality of the FD aug-pc4 results, we
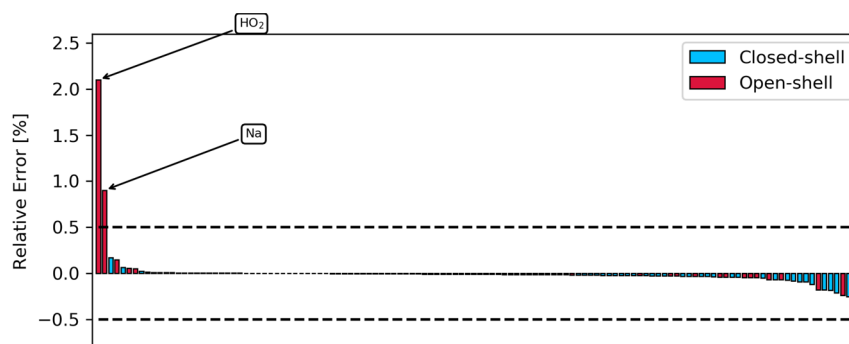
Figure 6. Distribution of REs of PBE polarizabilities for the 124 species, comparing GTO-LR with MW-FD, using the latter as a reference. The dashed lines are located at ±0.5% RE.

compared our MW-FD polarizabilities to the published GTO-FD polarizabilities.[7] The distribution of REs for all 124 species, as defined in eq 11 and with MWs as a reference, are presented in the left-most plot in Figure 2 and in more detail in Figure 3. Several features are revealed:

1. The error distribution suggests that FD aug-pc4 on average performs quite well, yielding a RE smaller than ±0.5% for most species.

2. GTOs seem to overestimate static polarizabilities, which may be counterintuitive as *analytical* polarizabilities are variationally approached from below.[12]

3. The most challenging species have open-shell electronic structures.

4. Six species have an RE larger than 1%, which, when considering the size of the basis set employed, should be considered significant errors: Li (7.3%); FH−OH (3.3%); HO_2 (2.3%); NaCl (1.9%); NaCN (1.5%); BeH (1.2%).

4.3. Is It a Basis Set Issue or an FD Issue? In order to evaluate whether the errors in Figure 3 arose from the FD approach, we compared the GTO-FD polarizabilities to computed GTO-LR values. The RE distribution for this comparison, using the analytical polarizabilities as a reference, is presented in the second plot in Figure 2 and in more detail in Figure 4. At first sight, the distribution is very similar to the one presented in Figure 3, indicating that GTO-FD polarizabilities have been contaminated by external field-related effects (the aug-pc4 benchmark study[7] used a field strength of 0.01 au for most species). To rule out the possibility that the two distributions incidentally show similar shapes, we plotted the two distributions against each other, species for species in Figure 5. Linear regression yielded an r^2 value of 0.97. Here it is clear that the error for one species is more or less the same across the two distributions, further indicating that the GTO-FD polarizabilities have been contaminated. Thus, our results show that the observed deviations between MW and GTO (aug-pc4) polarizabilities in Figure 3 are predominantly field-strength-related errors in the GTO-FD values.

4.4. What Is the True BSIE? In order to return to our original objective, the estimation of BSIEs in static polarizability predictions with the aug-pc4 basis set, we ultimately chose to compare GTO-LR and MW-FD values. Based on the above discussion, this comparison should yield the fairest estimation of the BSIE of aug-pc4. The RE distribution is presented in the center plot in Figure 2, and in more detail in Figure 6, and it is clear that the REs have been dramatically reduced for almost all species. Two species stand out with REs

larger than 0.5%: HO_2 (2.1%) and Na (0.9%). While both have open-shell (doublet) electronic structures, it is not clear what the origin of their relatively large REs may be. Despite the two outliers, the comparison in Figure 6 shows that the BSIE for aug-pc4 is very small.

4.5. Multiwavelets Can Handle Smaller Field Strengths than GTOs. Using FD calculations to estimate molecular response properties is a very simple approach, but it requires a careful consideration of the applied field strength. A weak field is required in order to stay within the linear regime, but this at the same time leads to the amplification of numerical errors due to cancellation of significant digits in the nominator of eq 3; a large field reduces numerical noise but simultaneously increases nonlinear effects from higher-order responses, leading to deviations from the correct result. The optimal compromise is therefore the weakest possible field that induces a sufficiently large first-order response in the dipole to obtain a sufficient number of digits in the polarizability. Examples of nonlinear behavior for a few species are presented and briefly discussed in the Supporting Information.

The MW framework guarantees that the computed dipoles are at the CBS limit with a controlled and systematically improvable precision:[50] the number of correct digits in the calculated polarizabilities can be improved systematically by tightening the precision thresholds. Therefore, MWs can make use of very small fields (10^{-3} or less) to eliminate higher-order responses, while still controlling the numerical noise by making use of tighter thresholds. As shown in Figure 1, even aug-pc4 has an error of roughly 10^{-3} in the energy as well as in the dipole moment, whereas the best MW calculation (MW7) yields three additional digits (10^{-6}). Making use of such a small field for GTOs will therefore heavily rely on error cancellation.

5. CONCLUSIONS

We have shown that GTO-FD polarizabilities presented by Hait and Head-Gordon[7] display quite large errors, considering the size of the aug-pc4 basis set used. However, we conclude that these errors mainly originate from field strength-related effects and not from BSIEs. Indeed, GTO-LR polarizabilities computed with aug-pc4 are very close to the CBS limit, which we have confirmed by comparing to very precise MW reference calculations. Specifically, we show that the observed errors exceeding 1% in GTO-FD polarizabilities are attributed to a field strength of 0.01 au, while the MRE of 0.06% in GTO-LR polarizabilities is attributed to the BSIE of aug-pc4.

The internal validation of our MW results demonstrates that MW-FD polarizabilities can be made virtually free from field strength-related effects, because numerical issues arising from using very small fields can be countered by tightening the MW thresholds. Our MW-FD polarizabilities using a field strength of 0.001 au show a MRE of 0.02% relative to a MW-LR reference.

For future benchmarks of any property, we recommend to validate that the reference data indeed is at the CBS limit by comparing to MW results.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00128.

> Cartesian coordinates of all species, data sets for all presented results, Python code for all data analyses and figure generation, and a brief discussion of nonlinearity effects and MW-FD validations (ZIP)

AUTHOR INFORMATION

Corresponding Author

Luca Frediani − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0003-0807-682X; Email: luca.frediani@uit.no

Authors

Anders Brakestad − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0001-9361-2759

Stig Rune Jensen − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0002-2175-5723

Peter Wind − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, 9037 Tromsø, Norway*

Marco D'Alessandro − *Istituto di Struttura della Materia, Consiglio Nazionale delle Ricerche, 00133 Roma, Italia*

Luigi Genovese − *Laboratoire de Simulation Atomistique, Université Grenoble Alpes, CEA, INAC-MEM, 38000 Grenoble, France*

Kathrin Helen Hopmann − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT, The Arctic University of Norway, 9037 Tromsø, Norway;* orcid.org/0000-0003-2798-716X

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00128

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(2) Hermann, J.; DiStasio, R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117*, 4714−4758.

(3) Zhan, Y.-Y.; Jiang, Q.-C.; Ishii, K.; Koide, T.; Kobayashi, O.; Kojima, T.; Takahashi, S.; Tachikawa, M.; Uchiyama, S.; Hiraoka, S. Polarizability and isotope effects on dispersion interactions in water. *Communications Chemistry* **2019**, *2*, 2399−3669.

(4) Parchaňský, V.; Kapitán, J.; Bouř, P. Inspecting chiral molecules by Raman optical activity spectroscopy. *RSC Adv.* **2014**, *4*, 57125−57136.

(5) Bedrov, D.; Piquemal, J.-P.; Borodin, O.; MacKerell, A. D.; Roux, B.; Schröder, C. Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields. *Chem. Rev.* **2019**, *119*, 7940−7995.

(6) Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Computational Molecular Science* **2015**, *5*, 241−254.

(7) Hait, D.; Head-Gordon, M. How accurate are static polarizability predictions from density functional theory? An assessment over 132 species at equilibrium geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800−19810.

(8) Jensen, F. Polarization consistent basis sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113−9125.

(9) Jensen, F. Polarization consistent basis sets: II. Estimating the Kohn-Sham basis set limit. *J. Chem. Phys.* **2002**, *116*, 7372−7379.

(10) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234−9240.

(11) Jensen, F.; Helgaker, T. Polarization consistent basis sets. V. The elements Si-Cl. *J. Chem. Phys.* **2004**, *121*, 3463−3470.

(12) Hylleraas, E. A.; Undheim, B. Numerische Berechnung der 2 S-Terme von Ortho- und Par-Helium. *Eur. Phys. J. A* **1930**, *65*, 759−772.

(13) Alpert, B. K. A Class of Bases in L^2 for the Sparse Representation of Integral Operators. *SIAM Journal on Mathematical Analysis* **1993**, *24*, 246−262.

(14) Keinert, F. *Wavelets and Multiwavelets*, 1st ed.; Chapman and Hall: 2003.

(15) Alpert, B.; Beylkin, G.; Gines, D.; Vozovoi, L. Adaptive solution of partial differential equations in multiwavelet bases. *J. Comput. Phys.* **2002**, *182*, 149−190.

(16) Beylkin, G.; Cheruvu, V.; Perez, F. Fast adaptive algorithms in the non-standard form for multidimensional problems. *Applied and Computational Harmonic Analysis* **2008**, *24*, 354−377.

(17) Frediani, L.; Fossgaard, E.; Flå, T.; Ruud, K. Fully adaptive algorithms for multivariate integral equations using the non-standard form and multiwavelets with applications to the Poisson and bound-state Helmholtz kernels in three dimensions. *Mol. Phys.* **2013**, *111*, 1143−1160.

(18) Beylkin, G.; Cramer, R.; Fann, G.; Harrison, R. J. Multi-resolution separated representations of singular and weakly singular operators. *Applied and Computational Harmonic Analysis* **2007**, *23*, 235−253.

(19) Harrison, R. J.; Fann, G. I.; Yanai, T.; Gan, Z.; Beylkin, G. Multiresolution quantum chemistry: Basic theory and initial applications. *J. Chem. Phys.* **2004**, *121*, 11587.

(20) Jensen, S. R.; Saha, S.; Flores-Livas, J. A.; Huhn, W.; Blum, V.; Goedecker, S.; Frediani, L. The Elephant in the Room of Density Functional Theory Calculations. *J. Phys. Chem. Lett.* **2017**, *8*, 1449−1457.

(21) Kato, T.; Yokoi, Y.; Sekino, H. Basis set limit computation of dynamic polarizability at near-resonance region. *Int. J. Quantum Chem.* **2013**, *113*, 286−289.

(22) Sekino, H.; Maeda, Y.; Yanai, T.; Harrison, R. J. Basis set limit Hartree−Fock and density functional theory response property

evaluation by multiresolution multiwavelet basis. *J. Chem. Phys.* **2008**, *129*, 034111.

(23) Jensen, S. R.; Flå, T.; Jonsson, D.; Monstad, R. S.; Ruud, K.; Frediani, L. Magnetic properties with multiwavelets and DFT: the complete basis set limit achieved. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21145−21161.

(24) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley: 2007.

(25) Daubechies, I.; Grossmann, A.; Meyer, Y. Painless non-orthogonal expansions. *J. Math. Phys.* **1986**, *27*, 1271−1283.

(26) Yanai, T.; Fann, G. I.; Gan, Z.; Harrison, R. J.; Beylkin, G. Multiresolution quantum chemistry in multiwavelet bases: Hartree−Fock exchange. *J. Chem. Phys.* **2004**, *121*, 6680.

(27) Yanai, T.; Fann, G. I.; Gan, Z.; Harrison, R. J.; Beylkin, G. Multiresolution quantum chemistry in multiwavelet bases: Analytic derivatives for Hartree−Fock and density functional theory. *J. Chem. Phys.* **2004**, *121*, 2866.

(28) Kobus, J.; Laaksonen, L.; Sundholm, D. A numerical Hartree-Fock program for diatomic molecules. *Comput. Phys. Commun.* **1996**, *98*, 346−358.

(29) Mahan, G. Modified Sternheimer equation for polarizability. *Phys. Rev. A: At., Mol., Opt. Phys.* **1980**, *22*, 1780−1785.

(30) Andrade, X.; Botti, S.; Marques, M. A. L.; Rubio, A. Time-dependent density functional theory scheme for efficient calculations of dynamic (hyper)polarizabilities. *J. Chem. Phys.* **2007**, *126*, 184106.

(31) D'Alessandro, M.; Genovese, L. Locality and computational reliability of linear response calculations for molecular systems. *Physical Review Materials* **2019**, *3*, 023805.

(32) MRCPP library 2020. DOI: 10.5281/zenodo.3749748.

(33) MRC$_{hem}$ program package 2020. DOI: 10.5281/zenodo.3786996.

(34) MRC$_{hem}$ documentation 2020. https://mrchem.readthedocs.io/en/latest/.

(35) It is not feasible to converge beyond ϵ_{rel} due to numerical noise, so we leave a factor of 10 between ϵ_{mo} and ϵ_{rel} to be on the safe side.

(36) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(38) Ekström, U.; Visscher, L.; Bast, R.; Thorvaldsen, A. J.; Ruud, K. Arbitrary-Order Density Functional Response Theory from Automatic Differentiation. *J. Chem. Theory Comput.* **2010**, *6*, 1971−1980.

(39) Van Lenthe, J. H.; Zwaans, R.; Van Dam, H. J. J.; Guest, M. F. Starting SCF calculations by superposition of atomic densities. *J. Comput. Chem.* **2006**, *27*, 926−932.

(40) Harrison, R. J. Krylov subspace accelerated inexact Newton method for linear and nonlinear equations. *J. Comput. Chem.* **2004**, *25*, 328−334.

(41) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.

(42) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

(43) Pulay, P. Improved SCF convergence acceleration. *J. Comput. Chem.* **1982**, *3*, 556−560.

(44) Hait, D.; Head-Gordon, M. Private communication, 2019.

(45) However, we believe that the reason for the two different components is not a consequence of the mean-field treatment but that it naturally derives from elementary considerations of degenerate perturbation theory. We are currently investigating this issue and will report on it elsewhere.

(46) Harrison, R. J.; Beylkin, G.; Bischoff, F. A.; Calvin, J. A.; Fann, G. I.; Fosso-Tande, J.; Galindo, D.; Hammond, J. R.; Hartman-Baker, R.; Hill, J. C.; Jia, J.; Kottmann, J. S.; Yvonne Ou, M.-J.; Pei, J.; Ratcliff, L. E.; Reuter, M. G.; Richie-Halford, A. C.; Romero, N. A.; Sekino, H.; Shelton, W. A.; Sundahl, B. E.; Thornton, W. S.; Valeev, E. F.; Vázquez-Mayagoitia, A.; Vence, N.; Yanai, T.; Yokoi, Y. MADNESS: A Multiresolution, Adaptive Numerical Environment for Scientific Simulation. *SIAM Journal on Scientific Computing* **2016**, *38*, S123−S142.

(47) Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User. *Aust. J. Chem.* **2019**, *72*, 563−573.

(48) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969−1981.

(49) Brakestad, A.; Jensen, S. R.; Wind, P.; D'Alessandro, M.; Genovese, L.; Hopmann, K. H.; Frediani, L. Replication Data for: Static polarizabilities at the basis set limit: A benchmark of 124 species. 2020. DOI: 10.18710/KLQVOK.

(50) The inherent limits are dictated by the double precision in the numerical representation.