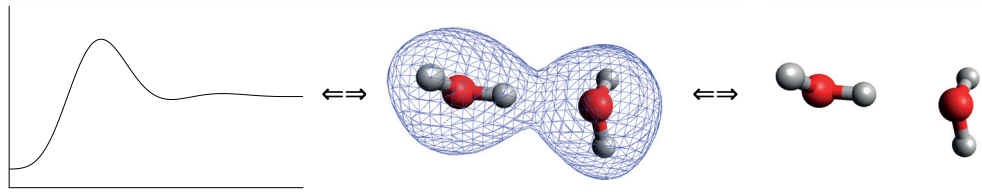
Report Series in Physics
HU-P-D202



COMPUTATIONAL MODELING OF THE ELECTRON MOMENTUM DENSITY

Susi Lehtola

UNIVERSITY OF HELSINKI REPORT SERIES IN PHYSICS

HU-P-D202

COMPUTATIONAL MODELING OF THE ELECTRON MOMENTUM DENSITY

Susi Lehtola

Laboratory of Materials Physics
Department of Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium E204 of the Physicum building (Gustaf Hällströmin katu 2a), on 22 February 2013, at 12 o'clock noon.

Helsinki 2013

Supervisors

Dr. Mikko Hakala
Department of Physics
University of Helsinki
Helsinki, Finland

Prof. Keijo Hämäläinen
Department of Physics
University of Helsinki
Helsinki, Finland

Pre-examiners

Prof. Kirk Peterson
Department of Chemistry
Washington State University
Washington, USA

Prof. Tapio Rantala
Department of Physics
Tampere University of Technology
Tampere, Finland

Opponent

Prof. Lars Pettersson
Department of Physics
Stockholm University
Stockholm, Sweden

Custos

Prof. Keijo Hämäläinen
Department of Physics
University of Helsinki
Helsinki, Finland

This thesis has been written on Fedora Linux with LyX, using the TeXLive LaTeX compiler. The figures have been prepared with xfig and GLE, with Octave for preprocessing the raw data.

Report Series in Physics HU-P-D202
ISSN 0356-0961
ISBN 978-952-10-8090-6 (printed version)
978-952-10-8091-3 (pdf version)

http://ethesis.helsinki.fi
Helsinki University Print
Helsinki 2013

Abstract

The properties and the functionality of materials are determined to a large extent by their electronic structure. The electronic structure can be examined through the electron momentum density, which is classically equivalent to the velocity distribution of the electrons. Changes in the structure of materials induce changes on their electronic structure, which in turn are reflected as changes in the electron momentum densities that can be routinely measured using, *e.g.*, x-ray Compton scattering.

The changes in the momentum density can be linked back to the structural changes the system has experienced through the extensive use of computational modeling. This procedure naturally requires using a model matching the accuracy of the experiment, which is constantly improving as the result of the ongoing development of synchrotron radiation sources and beam line instrumentation. However, the accuracies of the current computational methods have not been hitherto established.

This thesis focuses on developing the methods used to compute the electron momentum density in order to achieve an accuracy comparable to that of the experiment. The accuracies of current quantum chemical methods that can be used to model the electron momentum density are established. The completeness-optimization scheme is used to develop computationally efficient basis sets for modeling the electron momentum density at the complete basis set limit. A novel, freely available software program that can be used to perform all of the necessary electronic structure calculations is also introduced.

I wanted to study physics, because I was confused. Now I am still confused, but on a much higher level.

Adapted from Earl C. Kelley

Preface

I initially came to study physics in order to learn about the general theory of relativity. I did take the course a few years ago, but as it often happens, during my studies before that I found my interests to lie elsewhere. Of the many people who have had influence over my choice of specialization I want to acknowledge especially Keijo Kajantie, whose insight and childlike enthusiasm towards physics – even now as a pensioner – inspired me already as a freshman, Jouni Niskanen for deeply topical courses in quantum mechanics, and Antti Kuronen for courses in scientific programming.

I ended up in the field of materials physics somewhat by coincidence, having looked for a topic for my master's thesis after four years of studies in theoretical physics. I was especially interested in computational science and quantum mechanics. My supervisor Keijo Hämäläinen hired thus me to work on numerical quantum mechanics, *i.e.*, to compute inelastic x-ray scattering spectra. Although the beginning of my work as a theoretician in an experimental group was quite rough, after a few years I remembered why I initially came to the x-ray laboratory in the first place and started following this interest. The result of this decision is this thesis, completed in a time span of roughly three years.

I am in gratitude to Keijo Hämäläinen and my advisor Mikko Hakala, who indefatigably steered me into the subtleties of scientific writing. I am also very much grateful to the Jenny and Antti Wihuri foundation, whose financial support has been essential for the making of this thesis.

I thank my coauthors Pekka Manninen, Arto Sakko, and Juha Vaara, who have provided me incisive comments on my manuscripts and thus helped me to hone the texts towards perfection. As can be seen from the composition of this thesis, collaboration with Pekka has been especially fruitful. Thanks to the pre-examiners of this thesis, Kirk Peterson and Tapio Rantala, are also in order.

Last but not least I owe gratitude for my family for support. The PhD experiences of my parents and my brother have comforted me when facing seemingly gargantuan problems. Also the free time spent with my friends has been indispensable. Thanks to you all.

Susi Lehtola
Espoo, January 2013

Table of contents

List of figures

List of Tables

List of publications

Publications included in this thesis and the author's contribution

I **J. Lehtola**, M. Hakala, J. Vaara and K. Hämäläinen, Calculation of isotropic Compton profiles with Gaussian basis sets, *Phys. Chem. Chem. Phys.* 13, 5630 (2011), doi:10.1039/C0CP02269A, on page 47.

II **J. Lehtola**, P. Manninen, M. Hakala and K. Hämäläinen, Completeness-optimized basis sets: Application to ground-state electron momentum densities, *J. Chem. Phys.* 137, 104105 (2012), doi:10.1063/1.4749272, on page 61.

III **S. Lehtola**, P. Manninen, M. Hakala and K. Hämäläinen, Contraction of completeness-optimized basis sets: Application to ground-state electron momentum densities, *J. Chem. Phys.* 138, 044109 (2013), doi:10.1063/1.4788635, on page 71.

IV **J. Lehtola**, M. Hakala, A. Sakko and K. Hämäläinen, ERKALE – A Flexible Program Package for X-ray Properties of Atoms and Molecules, *J. Comp. Chem.* 33, 1572 (2012), doi:10.1002/jcc.22987, on page 81.

SL (formerly JL) wrote the computer programs, performed all of the calculations, carried out the analysis of the results and wrote the first versions of the manuscripts **I**-**IV**.

Other publications by the author related to this thesis

V **J. Lehtola**, M. Hakala and K. Hämäläinen, Structure of Liquid Linear Alcohols, *J. Phys. Chem. B* 114 (2010), 6426.

VI T. Pylkkänen, **J. Lehtola**, M. Hakala, A. Sakko, G. Monaco, S. Huotari and K. Hämäläinen, Universal Signature of Hydrogen Bonding in the Oxygen K-Edge Spectrum of Alcohols, *J. Phys. Chem. B* 114 (2010), 13076.

List of abbreviations

BO	Born–Oppenheimer approximation
BSSE	Basis set superposition error
CBS	Complete basis set
CC	Coupled cluster
CCSD	Coupled cluster theory with single and double excitations
CI	Configuration interaction
CISD	Configuration interaction with single and double excitations
CP	Compton profile
CS	Compton scattering
ED	Electron density
EES	$(e, 2e)$ spectroscopy
EMD	Electron momentum density
ESRF	European synchrotron radiation facility
FCI	Full configuration interaction
GTO	Gaussian type orbital
HF	Hartree–Fock
IXS	Inelastic x-ray scattering
KS-DFT	Kohn–Sham density functional theory
LCAO	Linear combination of atomic orbitals
MP	Møller–Plesset perturbation theory
MPn	Møller–Plesset perturbation theory, truncated at the n:th order
MP2	Møller–Plesset perturbation theory, truncated at the second order
MP3	Møller–Plesset perturbation theory, truncated at the third order
MP4	Møller–Plesset perturbation theory, truncated at the fourth order
PAS	Positron annihilation spectroscopy
SCF	Self-consistent field
STO	Slater type orbital

1. Introduction

Materials science is focused on the study of materials – how the structure at the atomic level is connected to various microscopic and macroscopic properties, such as the electronic band gap in a semiconductor or the heat capacity of a substance. Practically all of the properties[1] of a material one might be interested in are caused by (or majorly affected by) its electronic structure. To understand the function of a material it is thus necessary to understand its electronic structure.

The Hohenberg–Kohn theorems[1] inextricably link the electronic structure to the electron density (ED). Knowing one, it is in theory possible to computationally reproduce the other in an exact fashion. The ED thus contains all the physical and chemical information about a system there is to know. In consequence, changes in the ED retail of changes in the chemical structure; for instance, the formation of chemical bonds.

The ED can routinely be measured by, *e.g.*, x-ray diffraction techniques[2]. However, investigating the formation of chemical bonds with x-ray diffraction requires a very high accuracy, as the ED, and accordingly the x-ray diffraction pattern, is dominated by the dense nuclear regions that are relatively insensitive to chemical bonding.

This thesis focuses on a quantity very much analogous to the electron density – the electron momentum density (EMD)[3]. Whereas the ED $\rho(\mathbf{r})$ gives the amount of electrons at point $\boldsymbol{r}$ as $\rho(\boldsymbol{r})d^3r$, the EMD $n(\boldsymbol{p})$ yields the amount of electrons with momentum $\boldsymbol{p}$ as $n(\boldsymbol{p})d^3p$. The two quantities are reciprocal in the sense that the relevant length scales are inverted: in an atom centered at the origin, the core electrons are found at small r but large p, whereas the valence electrons have large r but small p. In contrast to the ED, the EMD is much flatter in form: while the ED decays exponentially[4] when $r \to \infty$, the EMD only decays polynomially[5] when $p \to \infty$. The EMD undergoes larger changes than the ED when chemical bonds are formed, making it an ideal tool for studying chemical structure.

The EMD can be experimentally measured with x-ray Compton scattering[6,7] (CS), $(e, 2e)$ spectroscopy[8] (EES) and positron annihilation spectroscopy[9] (PAS). EES can be used to extract the momentum densities of individual orbitals[8], while x-ray CS yields the EMD of the whole system[2]. However, even though EES thus yields more intricate information, due to the prevalence of multiple scattering the method is only applicable to small molecules in the gas phase[7].

Unlike the Coulombic interactions in EES and PAS, the interaction of the probe (x-ray photons) with the electrons in the sample is weak[3], meaning that multiple scattering events that plague EES experiments are rare in x-ray CS. Also, compared to PAS, the probe in x-ray CS is independent of the studied system[4]. This feature is the main

[1] Excluding, *e.g.*, the mass, which mostly comes from the atomic nuclei. However, the density depends greatly on the electronic properties.

[2] under the impulse approximation[14], see Section 2

[3] The coupling constant is small, compared to *e.g.* absorption.

[4] PAS depends on the positronic wave function, which must be modeled in addition to the electronic one.[10,11]

advantage of x-ray CS: it is a simple bulk probe of the sample. For the reasons given above only x-ray CS is considered in the rest of the current work.

As the EMD is a sensitive probe of the electronic wave function, measurements of the EMD can be used to benchmark quantum chemical theories. Access to the EMD through experiment also makes it possible to measure, *e.g.*, the purely electronic contribution to the heat capacity[12,13]. Furthermore, since the molecular structure links to the electronic structure, measurements of the EMD can also be used to extract information about structural changes occurring in the system through the use of computational modeling, as illustrated in the cover. The changes in the EMD can be routinely measured with x-ray CS at modern synchrotron radiation facilities. This approach has been recently used in many structural studies[13,15–26] as a complementary tool for other methods of structural determination.

The above-mentioned use of structural modeling clearly requires that the models used for the interpretation have an accuracy equal to or better than that of the experiment. The current statistical accuracy of the experiment is of the order of 0.02% for difference Compton profiles of systems composed of light elements[13,16,17,24,26]. The ongoing development of synchrotron radiation sources and beam line instrumentation will further improve the accuracy of experiments, requiring more and more precise calculations to be used.

However, there had been little work performed in this direction before this thesis. Accurate mean-field calculations[5] had been performed for the moments of the EMD of diatomic and linear molecules[27–29], but studies with more complicated geometries had used Gaussian basis sets[30–36], for which computational accuracy had not been truly established. Basis set requirements for calculations of the moments of the EMD had been estimated[30] and some studies had been made into the level of theory required for accurate modeling of the radial[6] EMD[33] and of Compton profiles[23,26,30,34,37], but no systematical benchmark showing the accuracies of different basis sets and levels of theory for the reproduction of EMD properties had been performed.

The main contributions of the current work are in the development of algorithms for numerically exact calculations of the radial EMD within the used basis set and the determination of the level of theory and the quality of the basis set necessary for accurate calculations of the moments of the EMD and of CS (paper **I**).

Significant contributions have also been made concerning basis set development. We have introduced novel algorithms that can perform automatic black-box optimization of contracted basis sets tuned for computing any property at any level of theory. Applying the algorithms, we have developed new computationally efficient basis sets with known accuracy properties for the modeling of the EMD (papers **II** and **III**).

As part of the work, we have developed a freely available software program[38] for modeling non-resonant inelastic x-ray scattering that is able, among other features, to calculate the EMD and the Compton profile (paper **IV**).

The structure of this thesis is the following. Section 2 briefly presents the experimental method this work was motivated by – x-ray Compton scattering. Section 3 discusses the theoretical groundwork for computational modeling of the EMD. The publications

[5] (Near) Hartree–Fock limit. However, in paper **II** these calculations were shown to be insufficient.
[6] Directionally averaged EMD, see eqn (2.6).

included in this thesis (list on page v) are summarized in Section 4. The thesis concludes in Section 5.

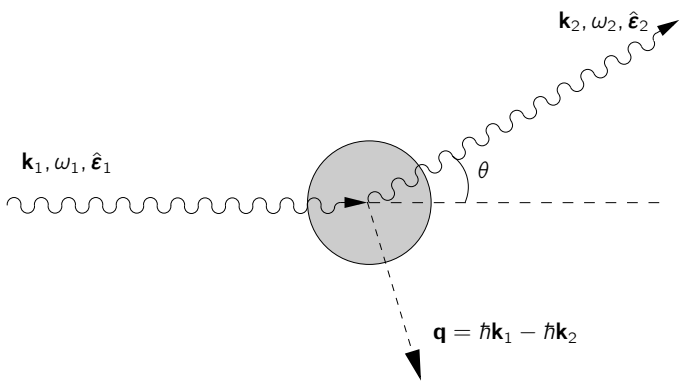


Figure 1: Schematic view of inelastic x-ray scattering. $\boldsymbol{k}_i$, ω_i and $\hat{\boldsymbol{\epsilon}}_i$ mark the momentum, energy and polarization of the photons. The scattering vector $\boldsymbol{q}$ is the recoil momentum received by the electron.

2. Inelastic x-ray scattering

In inelastic x-ray scattering (IXS), an incoming photon with momentum $\hbar\boldsymbol{k}_1$, energy $\hbar\omega_1$ and polarization $\hat{\boldsymbol{\epsilon}}_1$ scatters from an electron in the target, losing the energy $\hbar\omega = \hbar\omega_1 - \hbar\omega_2$ and exchanging an amount $\boldsymbol{q} = \hbar\boldsymbol{k}_1 - \hbar\boldsymbol{k}_2$ of momentum with the electron in the process. The outgoing photon has then momentum $\hbar\boldsymbol{k}_2$, energy $\hbar\omega_2$ and polarization $\hat{\boldsymbol{\epsilon}}_2$. This process is illustrated schematically in Figure 1.

CS corresponds to the case in which $qr \gg 1$ and $\hbar\omega \gg E_0$, where r and E_0 are the relevant length and energy scales in the system (*e.g.*, the radius of the orbital and the binding energy). As the energy transfer is $\hbar\omega$ is thus large compared to the energy scale of the system, the scattering is non-resonant. As another consequence, the impulse approximation[14] holds, and the double differential cross-section[7] for Compton scattering can be found out to be[6,7]

$$\frac{d^2\sigma}{d\Omega d\omega_2} = \left(\frac{d\sigma}{d\Omega}\right)_{\mathrm{Th}} \frac{\omega_2}{\omega_1}\frac{m}{q} \iint n(p_x, p_y, p_z)\, dp_x\, dp_y, \tag{2.1}$$

where $(d\sigma/d\Omega)_{\mathrm{Th}}$ is the classical Thomson scattering cross section, $n(\mathbf{p})$ is the electron momentum density and the direction of $\mathbf{q}$ defines the z axis. The p_z component is fixed by the conservation of energy in the scattering as[6,7]

$$\frac{p_z}{mc} = (\omega_1 - \omega_2) \pm \frac{\omega_1\omega_2}{mc}\frac{1-\cos\theta}{\sqrt{\omega_1^2 + \omega_2^2 + 2\omega_1\omega_2\cos\theta}}. \tag{2.2}$$

[7] Measures the amount of photons that scatter into the solid angle $d\Omega$ with an energy in the range $[\omega_2, \omega_2 + d\omega_2]$.

Eqn (2.1) can thus be written in the form

$$\frac{d^2\sigma}{d\Omega d\omega_2} = C(E_1, E_2, \theta) J(p_z), \tag{2.3}$$

where

$$J(p_z) = \iint n(p_x, p_y, p_z) dp_x dp_y \tag{2.4}$$

is the Compton profile (CP) and

$$C(E_1, E_2, \theta) = \left(\frac{e^2}{mc^2}\right)^2 |\hat{\boldsymbol{\varepsilon}}_1 \cdot \hat{\boldsymbol{\varepsilon}}_2|^2 \frac{\omega_2}{\omega_1} \frac{m}{q} \tag{2.5}$$

is a function that only depends on the setup of the experiment. Measuring the CP along all different directions of the sample it is in principle possible to reconstruct the three-dimensional electron momentum density, analogously to absorption tomography[8].

For the rest of the current work we switch to the atomic unit system. In isotropic or amorphous samples the molecules are found in random orientations, and thus only a spherical average is obtained. In this case the relevant quantities are the radial EMD

$$n(p) = \int n(\boldsymbol{p}) d\Omega_{\boldsymbol{p}} \tag{2.6}$$

and the isotropic Compton profile

$$J(q) = \frac{1}{2} \int_{|q|}^{\infty} p n(p) dp. \tag{2.7}$$

The Compton profile is tied to the moments of the EMD

$$\left\langle p^k \right\rangle = \int_0^{\infty} p^{2+k} n(p) dp, \ -2 \leq k \leq 4 \tag{2.8}$$

in addition to the trivial identity

$$J(0) = \frac{1}{2} \left\langle p^{-1} \right\rangle \tag{2.9}$$

via the sum rules[12,30,39]

$$\left\langle p^{-2} \right\rangle = 2 \int_0^{\infty} q^{-2} \left(J(q) - J(0)\right) dq, \tag{2.10}$$

$$\left\langle p^k \right\rangle = 2 (k+1) \int_0^{\infty} q^k J(q) dq, \ 0 \leq k \leq 4. \tag{2.11}$$

The other moments, *i.e.* $\left\langle p^k \right\rangle$ for $k < 2$ or $k > 4$, diverge for the exact momentum density[3,5].

[8] Absorption tomography yields the electron density, not the momentum density.

2.1. Experimental setup

Due to the small cross-section (*i.e.,* probability) of inelastic x-ray scattering, experiments with laboratory equipment are challenging due to the small luminosity of laboratory sources. This is the main reason why inelastic x-ray scattering experiments are almost exclusively performed at modern third generation synchrotron radiation facilities, which commonly offer a 10^{13} fold improvement in brilliance compared to x-ray tubes[40]. Most CS experiments are currently performed at the European Synchrotron Radiation Facility (ESRF, located in Grenoble, France) or at SPring-8 (Hyōgo, Japan).

According to the theory of electrodynamics, an accelerating charge emits electromagnetic radiation. The emitted power can be found out to be inversely proportional to the mass of the radiating particle[41]. In particle physics facilities radiation losses are a parasitic hindrance to the proper function of the accelerator. In contrast, in synchrotron radiation facilities these radiation losses are actually aimed at. For this reason, synchrotron radiation facilities are operated with electrons (or their antiparticles, positrons) as their small mass allows for large accelerations to be produced, resulting in the generation of powerful radiation. However, this leads to the electron beam continuously losing large amounts of energy and thus needing to be constantly re-accelerated to produce a stable light source.

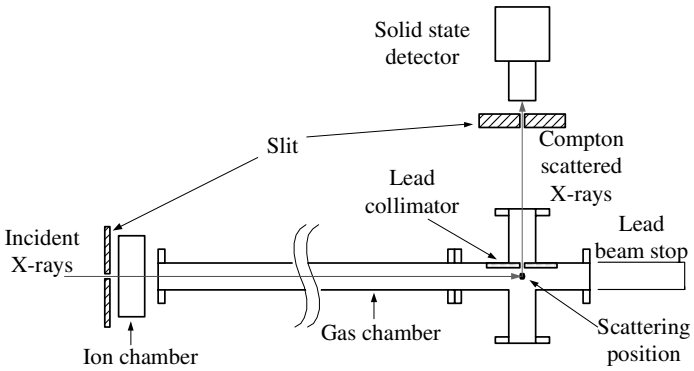


Figure 2: The experimental setup of the xenon measurement of reference 43.

A typical experimental setup from a gas phase experiment by Sakurai *et al.*[43] is shown in Figure 2. The x-rays generated by the synchrotron source enter from the left and are finely collimated with a slit. The x-rays scatter from the sample and enter the detector. Another slit in front of the detector allows only the radiation scattered by the sample to be measured. The resulting spectrum for a measurement on xenon is shown in Figure 3. At energies below the Compton peak there are five peaks corresponding to x-ray fluorescence, notably the Xe $K\alpha$ and $K\beta$ peaks from the sample, and the Pb $K\alpha$ and $K\beta$ peaks from the slit and radiation shielding. Also an elastic peak is seen in the spectrum.

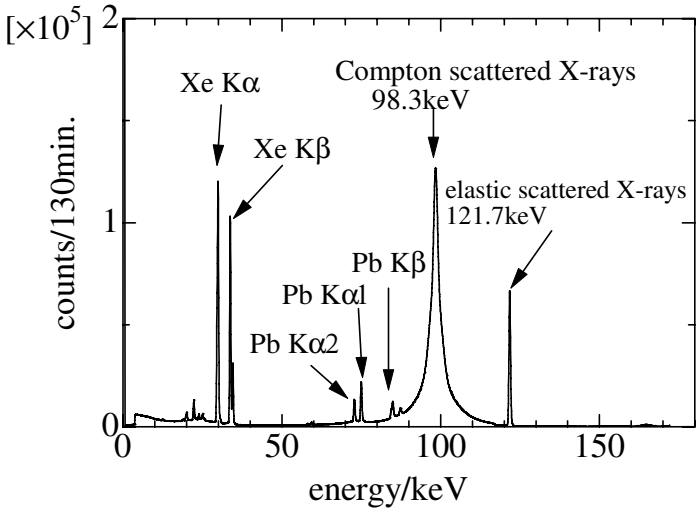


Figure 3: The measured raw spectrum in the CS experiment of reference 43.

In this particular experiment the requirements of the impulse approximation are not fulfilled for the innermost electrons of Xe. While the incoming photon energy (121.7 keV) is large compared to their binding energies, the energy *transfer* isn't ($\sim$ 23.4 keV vs a binding energy of 34.6 keV for K shell electrons and 5.5 keV for L shell electrons[44]).

3. Electronic structure theory

Eine exakte Lösung der Schrödingerschen Wellengleichung im $3N$-dimensionalen Konfigurationsraum (N = Anzahl der Elektronen) bietet aber unüberwindliche matematische Schwierigkeiten, und bei der Behandlung dieses Problems ist man auf Näherungsmethoden angewiesen.

An exact solution to the Schrödinger wave equation in $3N$ dimensional configuration space (N = number of electrons) poses, however, insurmountable mathematical difficulties, and one relies on approximate methods for the treatment of this problem.

– V. Fock, Z. Phys. 61, 126 (1930).

The current section discusses the framework on which the research undertaken in this thesis is founded on: how is the molecular structure interconnected with the electronic structure, what kind of approximations are used in solving the electronic structure in practice, and how the electron momentum density can be obtained from these calculations.

3.1. Molecular structure

In the systems studied in the current work relativistic effects on the electron momentum density are of small importance[43,45]. Thus, for simplicity, only non-relativistic methods are considered. Furthermore, as the electron momentum density is measured in the ground state, we will only consider stationary[9] states.

Although important in exceptional conditions such as temperatures near absolute zero or extremely large[10] magnetic fields[46,47], in normal situations magnetic interactions have no effect on the structure of matter; it is only determined by the interplay of Coulombic forces – the attraction of the electrons with the nuclei vs the repulsion of the electrons and the nuclei with themselves – as according to the so-called molecular Hamiltonian

$$\hat{H}_{\text{mol}} = -\sum \frac{\nabla_I^2}{2M_I} + \hat{H}_{\text{el}}, \tag{3.1.1}$$

$$\hat{H}_{\text{el}} = -\sum_i \frac{\nabla_i^2}{2} + \sum_{J<I} \frac{Z_I Z_J}{r_{IJ}} + \sum_{j<i} \frac{1}{r_{ij}} - \sum_{iI} \frac{Z_I}{r_{iI}}, \tag{3.1.2}$$

[9] time-independent

[10] $\sim 10^9$ T, which can be found on the surface of neutron stars.

where small and capital indices label electrons and nuclei, respectively. The first terms in eqns (3.1.1) and (3.1.2) describe accordingly the kinetic energy of the nuclei and the electrons. The last terms in eqn (3.1.2) represent the Coulombic repulsions of the nuclei with each other, of the electrons with each other, and the Coulombic attraction of the electrons with the nuclei.

As electrons and nuclei are distinct types of particles, a reasonable Ansatz for the total wave function of the system is[4]

$$\Psi(\boldsymbol{r}, \boldsymbol{R}) = \Phi(\boldsymbol{r}; \boldsymbol{R})\varphi(\boldsymbol{R}), \tag{3.1.3}$$

where $\mathbf{r}$ and $\mathbf{R}$ are shorthand for the set of electronic and nuclear coordinates $\{\boldsymbol{r}_i\}$ and $\{\boldsymbol{R}_I\}$, respectively. This is known as the Born–Oppenheimer (BO) approximation, in which Φ_k is an electronic eigenstate in the Coulomb field generated by the fixed nuclei clamped at positions $\{\boldsymbol{R}_I\}$

$$\hat{H}_{\text{el}}\Phi_k(\boldsymbol{r}; \boldsymbol{R}) = V_k(\boldsymbol{R})\Phi_k(\boldsymbol{r}; \boldsymbol{R}), \tag{3.1.4}$$

where k labels the electronic state. The corresponding nuclear wave function φ_k, which describes the spatial probability distribution of the nuclei, is in turn determined by the potential energy surface $V_k(\mathbf{R})$ generated by the electrons as

$$\left[-\sum \frac{\nabla_I^2}{2M_I} + V_k(\boldsymbol{R})\right] \varphi_k(\boldsymbol{R}) = E_k\varphi_k(\boldsymbol{R}). \tag{3.1.5}$$

The BO approximation relies on the fact that nuclei are much heavier than electrons, which often decouples their dynamics. The vibro-electronic coupling terms that arise from the nuclear derivative ∇_I operating on the electronic wave function are ignored; this approximation can be usually made around the equilibrium configuration[4].

The quantum nature of the nuclei in the BO picture needs to be taken into account for, *e.g.*, obtaining zero-point energy corrections arising from the vibrational ground state of a molecule, which may affect its stability, or properly describing "wobbly" molecules which do not have a clearly defined molecular structure – such as ammonia, the NH_3 molecule.

Assuming that the potential energy surfaces are well separated (as is often the case), the system stays on a single electronic state for which the nuclear Schrödinger equation (eqn (3.1.5)) then needs to be solved. This is usually done by constructing a local model of the potential energy surface around the minimum geometry, involving the solution of the electronic problem (eqn (3.1.4)) in slightly perturbed geometries around the equilibrium. Models of the potential energy surface can also be used for classical simulations of, *e.g.*, the evolution of structures in molecular liquids, as done in paper **V**.

For a more exact treatment, the non-BO wave function can always be solved in terms of BO wave functions. The coupling of different vibro-electronic states may need to be taken into account, *e.g.*, when studying[50,51] x-ray absorption or x-ray Raman scattering[11], , as the excitation of an electron may result in an excitation of the nuclear

[11] When the energy transfer $\hbar\omega$ in the inelastic scattering is similar to a core electron binding energy, the core electron is excited to an unoccupied state, yielding information similar to x-ray absorption spectroscopy.[48,49]

wave function as well[12].

However, as the electron momentum density is studied in the ground state, in the current work we invoke the widely used molecular structure approximation, which assumes that the potential wells (in which the nuclei are in) are deep, as well as steep. In this case the nuclear probability distribution approaches a Dirac delta function, meaning that the nuclei have well defined positions in the system and thus justifying the classical concept of a molecular structure. As a consequence, the electronic problem only needs to be solved for the minimum nuclear geometry[13].

3.2. Hartree–Fock theory

As we have simplified the full quantum mechanical problem into solving the nuclear wave functions from the electronic ones, we now proceed to the purely electronic problem. Because electrons are fermions, the electronic wave function must be antisymmetric in all of the electronic coordinates:

$$\Phi(\dots, \boldsymbol{x}_m, \dots, \boldsymbol{x}_n, \dots) = -\Phi(\dots, \boldsymbol{x}_n, \dots, \boldsymbol{x}_m, \dots), \tag{3.2.1}$$

where the generalized coordinate $\boldsymbol{x}_i = (\boldsymbol{r}_i, \sigma_i)$ contains both position $\boldsymbol{r}_i$ and spin σ_i. Expanding the N-electron wave function in an arbitrary, complete set of single-electron states (a.k.a. orbitals) $\{\phi_i\}$ one obtains

$$\Phi(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) = \sum_{n_1,\dots,n_N=1}^{\infty} c_{n_1\dots n_N}\phi_{n_1}(\boldsymbol{x}_1)\dots\phi_{n_N}(\boldsymbol{x}_N). \tag{3.2.2}$$

Here $c_{n_1\dots n_N}$ is a completely antisymmetric tensor, which makes the wave function fulfil the Pauli exclusion principle. However, the orbitals and the elements of c are unknown.

The simplest starting point for solving the electronic Schrödinger equation (eqn (3.1.4)) is to replace the exact expansion of eqn (3.2.2) with a Slater determinant, which automatically fulfills the antisymmetry requirement: $\Phi \to \Psi_0$, where

$$\Psi_0 = \frac{1}{\sqrt{N!}} \sum_{n_1,\dots,n_N=1}^{N} \epsilon_{n_1\dots n_N}\psi_{n_1}(\boldsymbol{x}_1)\dots\psi_{n_N}(\boldsymbol{x}_N), \tag{3.2.3}$$

in which ϵ is the N-dimensional Levi–Civita symbol. Now the optimal set of orbitals $\{\psi_i\}$ can be determined by minimizing the expectation value of the electronic energy

$$\langle E_{\text{el}} \rangle = \left\langle \Psi_0 \left| \hat{H}_{\text{el}} \right| \Psi_0 \right\rangle. \tag{3.2.4}$$

This approximation is known as Hartree–Fock (HF), or self-consistent field (SCF) theory. The theory is variational in the sense that the better the description of the orbitals is,

[12] The nuclear wave functions corresponding to different electronic states are non-orthogonal. When the electronic state is excited $k \to k'$, the ground-state nuclear wave function corresponding to the potential energy surface V_k has multiple components in terms of the eigenfunctions of $V_{k'}$.

[13] Naturally, finding the minimum geometry may require calculations in multiple points on the potential energy surface.

the lower the resulting energy will be. The theory leads to a set of coupled one-particle Schrödinger equations of the form[52]

$$\left[-\frac{\nabla^2}{2} + V_C(\boldsymbol{r}) + V_x(\boldsymbol{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\boldsymbol{r}), \tag{3.2.5}$$

where V_C is the classical Coulomb potential generated by the orbital charge distributions

$$V_C(\boldsymbol{r}) = \sum_j \int \frac{|\psi_j(\boldsymbol{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} d^3r' - \sum_I \frac{Z_I}{|\mathbf{r}-\mathbf{R}_I|}, \tag{3.2.6}$$

V_x is the quantum mechanical "exchange" potential caused by the fermionic nature of electrons

$$V_x(\boldsymbol{r})\psi_i(\boldsymbol{r}) = -\sum_j \int \frac{\psi_j^*(\boldsymbol{r}')\psi_i(\boldsymbol{r}')}{|\mathbf{r}-\mathbf{r}'|}\psi_j(\boldsymbol{r}) d^3r_j, \tag{3.2.7}$$

and ϵ_i is the orbital energy. It turns out[53] that the minimum energy solution of the HF equation is always the one in which the orbitals are occupied in order of increasing orbital energy ϵ_i; this is known as the Aufbau principle.

As can be seen from the equations above, HF is a mean-field theory, since the electrons in the system feel only the average field generated by the nuclei and the other electrons[14] in the system. We will return to this in Subsection 3.4.

3.3. Basis sets

To actually solve for the orbitals, the problem must first be discretized[15] so that it can be implemented on a computer. Basis sets provide a handy tool of transforming the integro-differential Hartree–Fock equations into algebraic form. As is known from elementary quantum mechanics, when a complete, orthogonal basis set

$$\sum_{i=1}^{\infty} |\chi_i\rangle \langle\chi_i| = \mathbf{1} \tag{3.3.1}$$

is used, wave functions can be expanded exactly as

$$|\psi_i\rangle = \sum_{i=1}^{\infty} c_{ij} |\chi_i\rangle , \tag{3.3.2}$$

$$c_{ij} = \langle\chi_i |\psi_i\rangle . \tag{3.3.3}$$

For obvious reasons, in practice the basis sets that are used in calculations contain only a finite amount N of functions, and are thus incomplete:

$$\sum_{i=1}^{N} |\chi_i\rangle \langle\chi_i| \approx \mathbf{1}. \tag{3.3.4}$$

[14] An electron does not interact with itself as the contributions from eqns (3.2.6) and (3.2.7) cancel out.

[15] Written as a problem with a finite number of degrees of freedom.

The linear combination of atomic orbitals (LCAO) is the prevalent basis set by far in molecular calculations, as qualitatively correct results can be obtained with relatively few basis functions per atom. The interaction between atoms can be regarded as a perturbation on the atomic states; typical bond energies being of the order of eVs, while the ground state energy of a single atom is typically orders of magnitude larger[16]. Also, the use of post-HF methods, such as Møller–Plesset perturbation or coupled cluster theory that are discussed below, is computationally feasible when an LCAO basis set[17] is used.

As the electronic structure is only given the degrees of freedom present within the basis set, a calculation is only as good as the basis set it uses. A calculation at a very high level of theory with a bad basis set often gives much worse results than a low level calculation with a good quality basis set. Papers **II** and **III** focus explicitly on optimizing the basis set $\{|\chi_i\rangle\}_{i=1}^{N}$ so that all of the essential physics and chemistry (considering the studied property) is captured in eqn (3.3.4).

The drawback in the LCAO method is that approaching the complete basis set (CBS) limit is painstaking due to, *e.g.*, basis set superposition errors[18] (BSSE) and over-completeness effects which cause numerical problems when the size of the basis set is increased. A basis set free, numerical solution of the HF equations is possible for simple systems (atoms and dimers)[55–57]. Recently, progress has also been made into the use of adaptive, multiresolution real-space grid methods that provide uniform convergence and guaranteed precision[58–60]. However, there are still many performance problems that prevent a widespread application of these methods.

3.3.1. STO basis

The analytic solution of the one-electron hydrogenic atom shows that the wave functions are of the form

$$\psi_{nlm}(\boldsymbol{r}) = R_{nl}(r) Y_l^m(\theta, \phi), \tag{3.3.5}$$

$$R_{nl}(r) = N_{nl} \left(\frac{2Zr}{n}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right) e^{-Zr/n}, \tag{3.3.6}$$

$$N_{nl} = \sqrt{\left(\frac{2Z}{n}\right)^3 \frac{(n-l-1)!}{2n\left[(n+l)!\right]^3}}, \tag{3.3.7}$$

where Z is the nuclear charge, L_n^l are associated Laguerre polynomials and Y_l^m are spherical harmonics in the complex form. Also the many-electron HF equation (eqn (3.2.5)) retains the asymptotic behavior of exponential decay[4] seen in eqns (3.3.5) and (3.3.6).

[16] ~10–100 eV for second row atoms, ~200–500 eV for third row atoms

[17] Namely, the GTO basis.

[18] The electrons in an atom may "borrow" the basis functions on another atom, complicating the calculation of, *e.g.*, bonding energies[54]. Furthermore, because the basis functions are localized on the atomic nuclei, changing the geometry also changes the basis set, which needs to be taken into account when forces are computed.

As such it would thus be natural to use a basis set of Slater-type orbitals (STO)[61]

$$\chi_{nlm}(\boldsymbol{r}) = R_{nl}^{\mathrm{STO}}(r) Y_{lm}(\theta, \phi), \tag{3.3.8}$$

$$R_{nl}^{\mathrm{STO}}(r) = \frac{(2\zeta)^{3/2}}{\sqrt{\Gamma(2n+1)}} (2\zeta r)^{n-1} e^{-\zeta r}, \tag{3.3.9}$$

where ζ is a positive real number and Y_{lm} are spherical harmonics in the real form, as these constitute the analytic solution to the one-electron Schrödinger equation for the hydrogenic atom.

Unfortunately, the evaluation of integrals necessary for HF theory is rather difficult when an STO basis set is used: many integrals need to be computed numerically[62,63], which hinders the use of post-HF methods that require a large set of accurate integrals. Staying at the HF level of theory, calculations on molecular systems with STO basis sets can be routinely performed with, *e.g.*, the ADF code[63,64].

3.3.2. GTO basis

Boys suggested the use of Gaussian functions instead of Slater functions in quantum chemical calculations[65]. The method gained wider acceptance and soon became the standard tool of quantum chemistry, as it was shown that STOs can be analytically expanded in terms of Gaussian type orbitals (GTOs), since[66]

$$\begin{aligned} \exp(-\zeta r) &= \frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-t^2) \exp\left(- \left[\frac{\zeta r}{2t} \right]^2 \right) dt \\ &\equiv \int_0^\infty w(t) \exp\left[-\alpha(t) t^2 \right] dt, \end{aligned} \tag{3.3.10}$$

where $w(t) := 2\exp(-t^2)/\sqrt{\pi}$ has been identified as the weighting factor of the GTO with exponent $\alpha(t) := \zeta^2/4t^2$. More importantly, the full integral transform of eqn (3.3.10) is not necessary, as a good accuracy for the molecular integrals can be achieved with relatively few functions[67–69]:

$$\exp(-\zeta r) \approx \sum_{n=1}^{N} w_i \exp(-\alpha_i r^2). \tag{3.3.11}$$

This is illustrated by the expansion[69] of the radial part of the hydrogen ground state wave function in Figure 4. As can be seen, the finite expansion is problematic both near the nucleus (wrong cusp condition) and far away from it (decay is too fast). However, the calculation of the necessary integrals is much simpler[65] and can be performed efficiently using recursion relations[70–72] when a GTO basis set is used, because the basis functions and the integrals factorize.

The integrals in GTO basis sets are computed in terms of cartesian GTOs[61]

$$\chi_{klm}^{\mathrm{GTO}}(\mathbf{r}) = \chi_k^{\mathrm{GTO}}(x)\ \chi_l^{\mathrm{GTO}}(y)\ \chi_m^{\mathrm{GTO}}(z), \tag{3.3.12}$$

$$\chi_k^{\mathrm{GTO}}(x) = \left(\frac{2\alpha}{\pi} \right)^{1/4} \sqrt{\frac{(4\alpha)^k}{(2k-1)!!}} x^k \exp(-\alpha x^2), \tag{3.3.13}$$

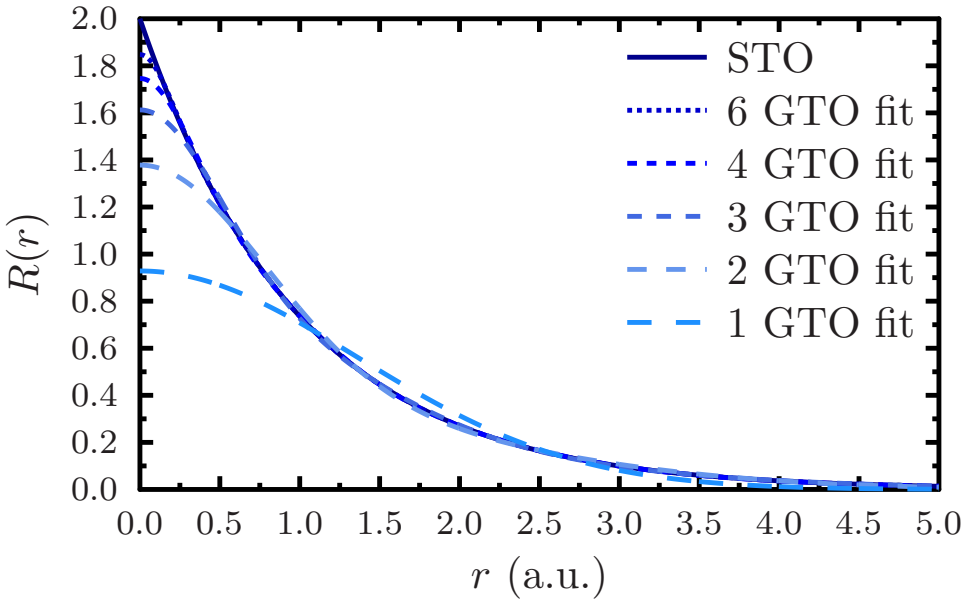


Figure 4: Expansion of the hydrogen ground state radial wave function (STO) in terms of GTOs. As can be seen, the expansion converges rather quickly.

the forms of $\chi_l^{\mathrm{GTO}}(y)$ and $\chi_m^{\mathrm{GTO}}(z)$ being analogous to eqn (3.3.13). As a consequence, quantum chemistry programs typically operate with shells of functions with the same cartesian angular momentum[19] $\lambda = k + l + m$. The functions with cartesian angular momentum λ span the spherical harmonics with angular momentum $l = \lambda$, but possibly contain also contaminants with $l = \lambda - 2$ (and lower ones as well)[20].

In order to avoid the problems caused by these contaminants and to reduce the amount of functions, most commonly spherical harmonics GTOs are used (also in papers **I**–**IV**). They can be expanded exactly in terms of cartesian GTOs[61,73], thus retaining the capability of fast integral evaluation via recursion relations. The general form of spherical harmonic GTOs is[61]

$$\chi_{nlm}^{\alpha}(\boldsymbol{r}) = R_{nl;\alpha}^{\mathrm{GTO}}(r) Y_{lm}(\hat{\boldsymbol{r}}), \tag{3.3.14}$$

where the radial function is

$$R_{nl;\alpha}^{\mathrm{GTO}}(r) = \frac{2\,(2\alpha)^{n-l/2-1/4}}{\pi^{1/4}} \sqrt{\frac{2^{2n-l-2}}{(4n-2l-3)!!}}\, r^{2n-l-2} e^{-\alpha r^2}. \tag{3.3.15}$$

[19] For example the $\lambda = 2$ shell with exponent α_i contains the functions $(k,l,m) \in \{(0,0,2),(0,1,1),(0,2,0),(1,1,0),(2,0,0),(1,0,1)\}$, each with the same exponent.

[20] For instance, a cartesian D shell has 6 functions, corresponding to the 5 pure D functions, plus one S-type contaminant with a higher primary quantum number n.

As varying exponents are used to span the necessary degrees of freedom, the factor n is $n = l + 1$ for spherical harmonic basis sets. In contrast, when cartesian GTOs are used as the basis set, the common exponent of r in eqn (3.3.15) can be identified as $2n - l - 2 = \lambda$. Solving for $n = (l + \lambda)/2 + 1$ gives upon substitution

$$\chi^{\alpha}_{\lambda l;m}(\boldsymbol{r}) = R^{\mathrm{GTO}}_{\lambda;\alpha}(r) Y_{lm}(\hat{\boldsymbol{r}}), \tag{3.3.16}$$

$$R^{\mathrm{GTO}}_{\lambda;\alpha}(r) = \frac{2\,(2\alpha)^{\lambda/2+3/4}}{\pi^{1/4}} \sqrt{\frac{2^{\lambda}}{(2\lambda+1)!!}}\, r^{\lambda} e^{-\alpha r^2}. \tag{3.3.17}$$

This result, necessary for the lower angular momentum contaminants, will be used in Subsection 3.8.

3.3.3. Formation of basis sets

The parametrization of basis sets traditionally begins by optimizing an *sp* set to produce the smallest atomic HF energy with a fixed amount of basis functions. Additional functions of higher angular momentum, depicting the polarization of bonds in molecules or effects of electron correlation, are then added using additional molecular or post-HF calculations of the energy. Finally, diffuse functions are added by hand[21] to properly describe other properties, such as the dipole moment – or the electron momentum density. This approach has been used, *e.g.*, in the parametrization of the Dunning-style correlation-consistent cc-pVXZ series[74–78] and Jensen's polarization-consistent pc-N series[22] [79–83].

It was noted early on that elements with similar electronic shell structures can be described with the same "universal" basis set[84,85]. The reason for this is that basis sets are just a mathematical tool for the representation of the electronic structure. The connection between the basis set and the energy criterion can thus be severed altogether, *e.g.,* by the process of completeness-optimization[86], a procedure based on the concept of completeness profiles[87].

3.3.3.1. Completeness profiles

The completeness profile is a graphical tool for visualizing eqn (3.3.4) as[87]

$$Y(\alpha) = \sum_{\mu\nu} \langle \alpha | \mu \rangle\, S^{-1}_{\mu\nu}\, \langle \nu | \alpha \rangle\,, \tag{3.3.18}$$

where $|\alpha\rangle$ is a scanning function used to probe the completeness of the basis set, $|\mu\rangle$ and $|\nu\rangle$ are functions in the basis set and all of the functions are centered on the same atom. $S^{-1}_{\mu\nu}$ denotes the (μ, ν) element of the inverse overlap matrix, where the elements of the overlap matrix are $S_{\mu\nu} = \langle \mu | \nu \rangle$.

[21] Diffuse functions are not procured by energy optimization because of their small importance for energetics, unless very large basis sets are used.

[22] Jensen's basis sets have been parametrized and contracted using density-functional theory instead of Hartree–Fock discussed in the text, but the two theories share similar basis set requirements.

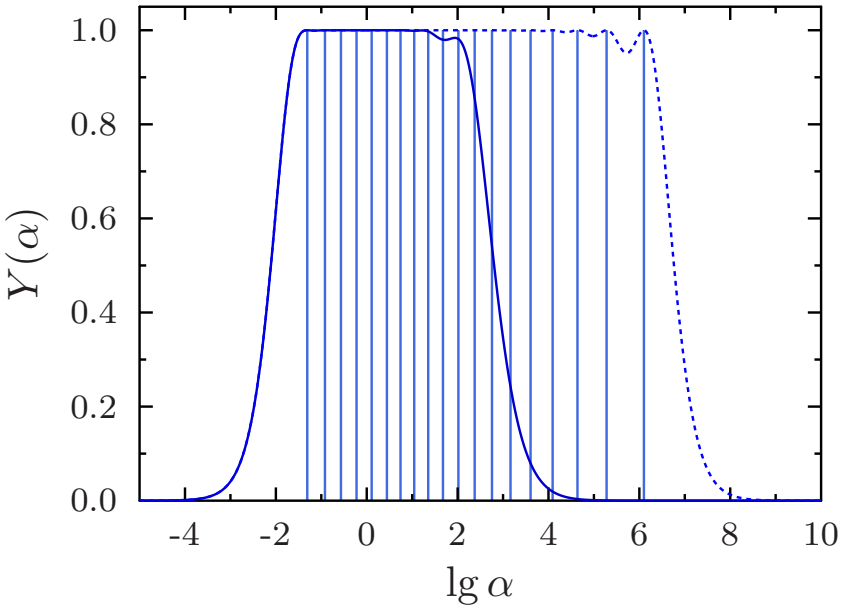


Figure 5: Completeness profile for the S shell of the aug-pc-4 basis set for oxygen. The solid line depicts the profile for the basis set in the normal, contracted form, whereas the dashed line shows the profile of the decontracted basis set. The placing of the exponents is shown as bars.

If the basis set is flexible enough to accurately represent the scanning function $|\alpha\rangle$, then $Y(\alpha) \approx 1$. On the other hand, if the basis set is unable to represent the scanning function, then $Y(\alpha) \approx 0$. When Gaussian basis sets are used, the scanning function conventionally used is

$$\langle \boldsymbol{r} \,|\alpha\rangle = \chi^{\alpha}_{\lambda m}(\boldsymbol{r}), \tag{3.3.19}$$

where the function $\chi^{\alpha}_{\lambda m}$ was defined in eqn (3.3.16). A separate profile is obtained for all the values of the angular momentum present in the used basis set, while all values of the z-component m yield an identical profile. The completeness profile is commonly visualized in a $(\lg\alpha, Y(\alpha))$ plot, as illustrated in Figure 5.

3.3.3.2. Completeness-optimization of exponents

As can be seen from Figure 5, a plateau where $Y(\alpha) \approx 1$ is commonly seen in the completeness profile of the primitive basis set[23], with dips occurring where the exponents

[23] The basis set in decontracted form.

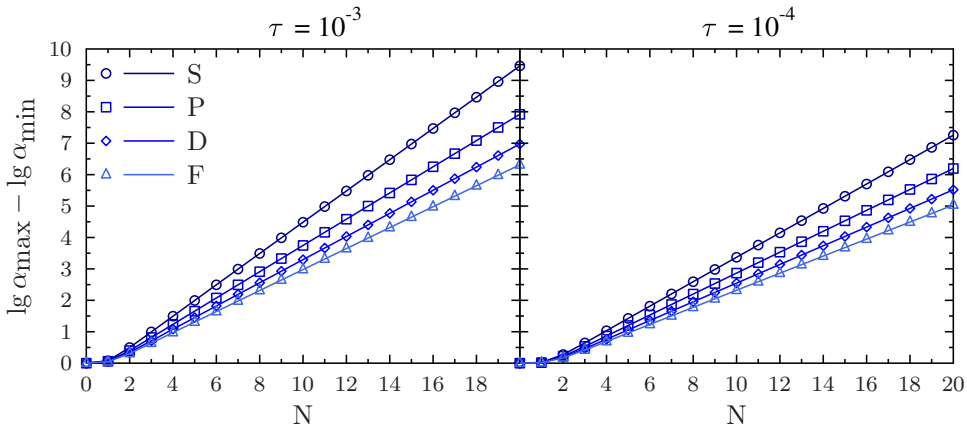


Figure 6: Width of the completeness plateau $\lg \alpha_{\max} - \lg \alpha_{\min}$ plotted as a function of the amount of functions N for two values of $\tau_1 = \tau$.

are placed more sparsely. The idea of completeness-optimization [86] is to optimize the primitives, *i.e.*, the exponents in the basis set so that the resulting completeness plateau is as flat as possible on the wanted exponent interval. The measure we use in the optimization is (paper **IV**)

$$\tau_n = \left(\frac{1}{\lg \alpha_{\max} - \lg \alpha_{\min}} \int_{\lg \alpha_{\min}}^{\lg \alpha_{\max}} [1 - Y(\alpha)]^n \, d \lg \alpha \right)^{1/n}, \tag{3.3.20}$$

in which $\alpha_{\min}$ and $\alpha_{\max}$ are the lower and upper limits of the used interval (for the current value of the angular momentum) and $n > 0$. The exponents which minimize the value of τ_n can be straightforwardly obtained with standard minimization techniques; in this work (papers **II**, **III** and **IV**) we have used the Nelder–Mead "*amoeba*" method [88].

The special case $n = 1$ in eqn (3.3.20) corresponds to the maximization of the area, while $n = 2$ corresponds to the minimization of the root-mean square deviation from completeness. However, as in the limit $\tau_n \to 0$ all values of n lead to a complete basis set[24], the value $n = 1$ is commonly used in practice[25] (references 86,89–92 and papers **II** and **III**). Accordingly, for the rest of the thesis we denote τ_1 with a plain τ. The proper functionality of the minimization algorithm is demonstrated by the smooth increase of the width of the completeness plateau with fixed τ, presented in Figure 6.

3.3.3.3. Completeness-optimization of basis sets

The completeness-optimization of the exponents provides a systematical way to approach the complete basis set in the limit $\alpha_{\min} \to 0$, $\alpha_{\max} \to \infty$ and $\tau \to 0$. This

[24] $Y(\alpha) \to 1$ when $\tau_n \to 0$ for $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, because x^n and $x^{1/n}$ are monotonically increasing functions for $n > 0$.

[25] $n = 2$ places more weight on the edges, leading to closer spacing of the exponents near the limits at $\alpha_{\min}$ and $\alpha_{\max}$, which can cause problems due to linear dependencies arising in molecular calculations.

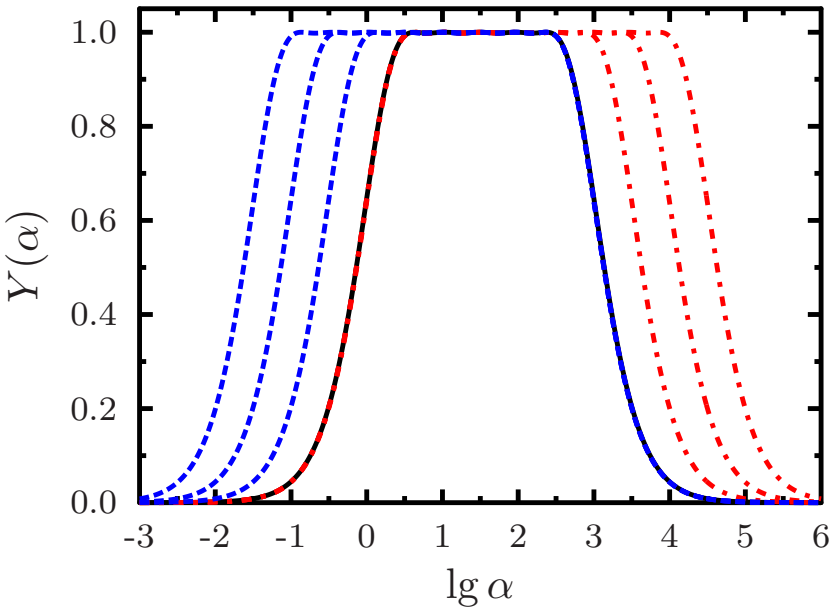


Figure 7: Illustration of the completeness-optimization scheme. The completeness profile of the starting point basis set is in black. The red dash-dotted curves on the right represent expansion of the completeness profile to the steep end, whereas the blue dashed curves on the left that to the diffuse end.

makes it straightforward to find out the necessary completeness of the exponent space for computing the wanted property at the wanted accuracy by the method of trial and error. This process is known as completeness-optimization [86], which is a method for forming computationally efficient basis sets for any property at any level of theory.

Starting with a minimal basis set, one can calculate the change in the computed property caused by an expansion of the basis set to the steep and diffuse areas (larger α_{max} or smaller α_{min}, illustrated in Figure 7) while keeping τ constant. The procedure is repeated for all values of angular momentum l. When the change in the property that is evaluated is small enough upon extension of the existing shells or the addition of new polarization shells, the CBS limit has been reached. Completeness-optimization is used extensively in papers **II** and **III**.

3.4. Configuration interaction

Once a basis set has been adopted, one can solve the HF equations and thus obtain a mean-field description of the physics in the system. Although the HF picture is often a good starting point, the method only accounts for the average interaction of the

electrons with each other – the correlation of the electrons[26] is not included in the theory. To describe correlation effects, it is necessary to go beyond HF theory to post-HF methods. As the HF orbitals span a complete space, the exact electronic wave function can be expanded in terms of the HF orbitals similarly to eqn (3.2.2) as[27]

$$|\Phi\rangle \propto |\Psi_0\rangle + \sum_{ia} c_a^i \left|\Psi_i^a\right\rangle + \sum_{ij,ab} c_{ab}^{ij} \left|\Psi_{ij}^{ab}\right\rangle + \ldots, \tag{3.4.1}$$

in which $|\Psi_0\rangle$ is the Hartree–Fock determinant, $\left|\Psi_i^a\right\rangle$ is the determinant with an electron excited from the occupied orbital i to the virtual orbital a and c_a^i is the relevant amplitude, and analogously for higher excitations. This is known as configuration interaction (CI) theory. If all possible determinants are included in the expansion, one obtains the full configuration interaction (FCI) theory.

Alas, the amount of excited determinants grows factorially in the amount of orbitals[61], which limits the application of FCI to very small systems. A recently introduced stochastic Monte-Carlo approach has made calculations on somewhat larger systems possible[93–102], the largest calculations performed so far containing 10^{108} Slater determinants[102].

To overcome the factorial scaling, the infinite series in eqn (3.4.1) is conventionally truncated at some excitation level, yielding for example the configuration interaction single and double excitations (CISD) theory, which has been used in the parametrization of the Dunning-style basis sets[74,76,78]. Although all versions of configuration interaction theory are variational, the truncation breaks size extensivity[28] and alternative methods (see below) are used instead.

3.5. Møller–Plesset perturbation theory

Møller–Plesset (MP) theory[103] handles the correlation problem by treating the difference of the exact two-electron interaction and the mean-field interaction as a perturbation to the system. The excitation amplitudes in eqn (3.4.1) are then obtained from a perturbation expansion. As is true for perturbation theories in general, the perturbation expansion may not always converge[29]. MP works well for single-reference systems[30], such as most molecules around their equilibrium geometry.

[26] Hartree–Fock theory predicts the probability of finding an electron at r and another of opposite spin at r′ to be $P(\mathrm{r},\mathrm{r}') \propto \rho_\alpha(\mathrm{r})\rho_\beta(\mathrm{r}') + \rho_\beta(\mathrm{r})\,\rho_\alpha(\mathrm{r}')$. However, when $|\mathrm{r}-\mathrm{r}'| \to 0$ the mean-field description of HF theory is no more valid. Because the Coulombic repulsion of the two electrons increases without limit when $|\mathrm{r}-\mathrm{r}'| \to 0$, in reality the probability of finding two electrons close to one another is smaller than HF theory predicts. The aim in post-HF methods is to properly describe this *correlation hole* around electrons.

[27] The proportionality sign is used in eqn (3.4.1) because of the use of intermediate normalization on the right hand side, the norm of which is $1 + \sum_{ia} |c_a^i|^2 + \sum_{ij,ab} |c_{ab}^{ij}|^2 + \ldots$.

[28] The performance of the model decays with increasing system size.

[29] Theories truncated at different orders give results that differ by an arbitrarily large amount from each other.

[30] A system which is described well by a single HF determinant. For systems with significant multireference character the use of multireference perturbation theory (MRPT) (or multireference coupled cluster theory) yields better convergence.

As a perturbation theory, truncated MP is non-variational. However, it is size extensive and is often used thanks to its favorable scaling. Whereas Hartree–Fock theory scales conventionally[31] as N^4, N representing the system size, Møller–Plesset theory truncated at the nth order (MPn) scales as[104] N^{n+3}. The use of MP methods truncated above the second order (MP2, used in papers **I** and **III**) such as MP3 or MP4 is generally considered fruitless in quantum chemistry, as coupled cluster theories, which yield better results, have the same or even better scaling properties.

3.6. Coupled cluster theory

To overcome the size extensivity problem of truncated CI, an alternative formulation is warranted. The CI equation ((3.4.1)) can be recast in another form as[105]

$$|\Phi\rangle = e^{\hat{T}} |\Psi_0\rangle \,, \tag{3.6.1}$$

in which the excitation operator is

$$\hat{T} = t_0 + t^1_{i,a}\hat{\tau}^{i,a}_1 + t^2_{ij,ab}\hat{\tau}^{ij,ab}_2 + t^3_{ijk,abc}\hat{\tau}^{ijk,abc}_3 + \ldots \,, \tag{3.6.2}$$

where $\hat{\tau}^{i,a}_1$ excites the electron from the occupied i orbital to the virtual a orbital, $t^1_{i,a}$ being the related cluster amplitude (and analogously for higher excitation orders), and summation over repeated indices is implied. This reformulation is known as coupled cluster (CC) theory. Thanks to the exponential parametrization, CC theory clearly is size extensive, and also leads to faster convergence in the truncation order than CI theory[61] as can be seen from the apparition of higher order excitations in a series expansion of eqn (3.6.1).

Unfortunately, solving the CC amplitudes is non-trivial. In the commonly used projection approach the amplitudes are obtained by projecting the CC wave function onto the HF ground state and the singly, doubly, *etc.*, excited determinants. Although this procedure yields workable equations, the variationality of the theory is lost[61]. The coupled cluster single and double excitations (CCSD, used in paper **I**) method scales conventionally as N^6, whereas if triple excitations are also included (CCSDT) the scaling is N^8. If triples are treated only perturbatively, yielding CCSD(T), the golden standard of quantum chemistry[61], the scaling is N^7.

3.7. Density-functional theory

In addition to the *ab initio*, wave function based methods discussed above, there is also another, completely different way of approaching the electronic structure problem. As was mentioned in Section 1, the Hohenberg–Kohn theorems state that the ground state density uniquely determines all of the properties of a system of N electrons[1]. This is the foundation of Kohn–Sham density-functional theory[106,107] (KS-DFT), in which the real electronic system is replaced with a fictitious system of *non-interacting* electrons,

[31] The theory can be made to scale (quasi)linearly in large non-metallic systems by using integral screening and multipole methods. Similar speedups can also be made for MP.[61]

the density of which is postulated to coincide with that of the real, interacting system. The theory leads to a group of single-particle equations as in HF, yielding notable speed benefits compared to the post-HF methods described above, while still capturing all of the features of the real electronic system.

KS-DFT works relatively well for many properties such as core electron spectroscopies (for instance in paper **VI**), but its use in modeling the absolute EMD is problematic. In addition to the general problem of KS-DFT – the exact exchange-correlation functional[32] is not known[33] – there is also another problem in the application of KS-DFT to modeling of the EMD. Even if the exact ED could be calculated, it is not clear how the EMD could be obtained from it[3]. The straightforward process of taking the square norm of the Fourier-transformed orbitals, which works in HF theory, yields unsatisfactory results in KS-DFT[108,109]. This feature was explained by Lam and Platzman[110], who argued that while the real-space density produced by KS-DFT

$$\rho(\boldsymbol{r}) = \sum_i n_i \left|\psi_i(\boldsymbol{r})\right|^2, \tag{3.7.1}$$

where ψ_i and n_i are the KS orbitals and their occupation numbers, is that of independent particles, in the momentum space the particles are *correlated*:

$$n(\boldsymbol{p}) \neq \sum_i n_i \left|\tilde{\psi}_i(\boldsymbol{p})\right|^2, \tag{3.7.2}$$

where the tilde denotes the Fourier transform. Thus, a correction term must be added to the right hand side of eqn (3.7.2) – the form of which is unknown[34]. However, even though KS-DFT fails to reproduce absolute EMD properties, it often reproduces the correct shape of difference Compton profiles [reference 37 and paper **I**].

3.8. Electron momentum density

The EMD can be obtained in atom-centered basis sets as[111]

$$n(\boldsymbol{p}) = \sum_{\mu\nu} P_{\mu\nu} \overline{\tilde{\chi}_\mu(\boldsymbol{p})} \tilde{\chi}_\nu(\boldsymbol{p}), \tag{3.8.1}$$

where $P_{\mu\nu}$ is the one-electron density matrix, $\chi_\mu(\mathbf{r})$ is the μ:th basis function and the overline denotes complex conjugation. In the following we will present an algorithm for computing the radial EMD in any atom-centered basis set, further developing the algorithm presented in paper **I**[35].
Expanding the basis functions around their centers $\mathbf{R}_\mu$ as

$$\chi_\mu(\boldsymbol{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_\mu^{lm} R_{\mu l}(|\boldsymbol{r} - \boldsymbol{R}_\mu|)\, Y_l^m\left(\frac{\boldsymbol{r} - \boldsymbol{R}_\mu}{|\boldsymbol{r} - \boldsymbol{R}_\mu|}\right), \tag{3.8.2}$$

[32] The functional maps the behavior of the non-interacting system to that of the interacting one.
[33] There are a variety of semi-empirical approximations to the exchange-correlation functional. Contrary to post-HF methods, there is no way to systematically approach the correct result.
[34] Lam and Platzman suggested an approximate correction in reference 110.
[35] In paper **I** only GTO basis sets were considered.

where $R_{\mu l}$ are normalized radial functions, Y_l^m are spherical harmonics in the complex form and c_μ^{lm} are complex expansion coefficients, the Fourier transforms can be calculated using the expansion

$$e^{i\boldsymbol{p}\cdot\boldsymbol{r}} = 4\pi \sum_{L=0}^{\infty} i^L j_L(pr) \sum_{M=-L}^{L} \overline{Y_L^M(\hat{\boldsymbol{r}})} Y_L^M(\hat{\boldsymbol{p}}), \tag{3.8.3}$$

where j_L is the spherical Bessel function of order L and $\hat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r}$. Using the orthonormality of spherical harmonics, we get

$$\tilde{\chi}_\mu(\boldsymbol{p}) = \frac{1}{(2\pi)^{3/2}} \int \chi_\mu(\boldsymbol{r}) e^{-i\boldsymbol{p}\cdot\boldsymbol{r}} d^3r \tag{3.8.4}$$

$$= e^{-i\boldsymbol{p}\cdot\boldsymbol{R}_\mu} \sum_{lm} c_\mu^{lm} \tilde{R}_{\mu l}(p) Y_l^m(\hat{\boldsymbol{p}}), \tag{3.8.5}$$

where the phase factor $\exp(-i\boldsymbol{p}\cdot\boldsymbol{R}_\mu)$ in eqn (3.8.5) comes from the change of origin and the transformed radial function is

$$\tilde{R}_{\mu l}(p) = \sqrt{\frac{2}{\pi}} (-i)^l \int_0^\infty r^2 R_{\mu l}(r) j_l(pr) dr. \tag{3.8.6}$$

The forms of the radial functions in STO and GTO basis sets are presented below. However, the same procedure can in principle also be applied to numerical atomic orbitals as used by, *e.g.*, the SIESTA and GPAW codes[112], for which the integral in eqn (3.8.6) needs to be calculated numerically. Inserting eqn (3.8.5) into eqn (3.8.1) we obtain the EMD as

$$n(\boldsymbol{p}) = \sum_{\mu\nu} P_{\mu\nu} e^{-i\boldsymbol{p}\cdot(\boldsymbol{R}_\nu - \boldsymbol{R}_\mu)} \sum_{lm} \overline{c_\mu^{lm} \tilde{R}_{n_\mu l}^{\zeta_\mu}(p) Y_l^m(\hat{\boldsymbol{p}})} \sum_{l'm'} c_\nu^{l'm'} \tilde{R}_{n_\nu l'}^{\zeta_\nu}(p) Y_{l'}^{m'}(\hat{\boldsymbol{p}}). \tag{3.8.7}$$

The radial EMD (eqn (2.6)) then becomes

$$n(p) = \sum_{\mu\nu} P_{\mu\nu} \sum_{ll'} \sum_{mm'} \overline{c_\mu^{lm} \tilde{R}_{n_\mu l}^{\zeta_\mu}(p)} c_\nu^{l'm'} \tilde{R}_{n_\nu l'}^{\zeta_\nu}(p) \int e^{-i\boldsymbol{p}\cdot(\boldsymbol{R}_\nu - \boldsymbol{R}_\mu)} \overline{Y_l^m(\hat{\boldsymbol{p}})} Y_{l'}^{m'}(\hat{\boldsymbol{p}}) d\Omega_{\boldsymbol{p}}. \tag{3.8.8}$$

We perform the angular integral in eqn (3.8.8) as

$$\begin{aligned}
&\sum_{mm'} \overline{c_\mu^{lm}} c_\nu^{l'm'} \int e^{i\boldsymbol{p}\cdot\boldsymbol{R}_{\mu\nu}} \overline{Y_l^m(\hat{\boldsymbol{p}})} Y_{l'}^{m'}(\hat{\boldsymbol{p}}) d\Omega_p \\
&= \sum_{mm'} \overline{c_\mu^{lm}} c_\nu^{l'm'} \left[4\pi \sum_{L=0}^{\infty} i^L j_L(pR_{\mu\nu}) \sum_{M=-L}^{L} \overline{Y_L^M(\hat{\boldsymbol{R}}_{\mu\nu})} \right] \int Y_L^M(\hat{\boldsymbol{p}}) \overline{Y_l^m(\hat{\boldsymbol{p}})} Y_{l'}^{m'}(\hat{\boldsymbol{p}}) d\Omega_{\boldsymbol{p}} \\
&= \sum_{mm'} 4\pi \overline{c_\mu^{lm}} c_\nu^{l'm'} \sum_{L=L_0}^{l+l'} i^L j_L(pR_{\mu\nu}) G_{lL l'}^{m(m-m')m'} \overline{Y_L^{m-m'}(\hat{\boldsymbol{R}}_{\mu\nu})},
\end{aligned} \tag{3.8.9}$$

where we have used eqn (3.8.3) and notated $\boldsymbol{R}_{\mu\nu} = \boldsymbol{R}_\mu - \boldsymbol{R}_\nu$. The lower limit of the sum is $L_0 = \max\{|l - l'|, |m - m'|\}$ and $G_{Ll l'}^{Mmm'}$ is the Gaunt coefficient

$$G_{Ll l'}^{Mmm'} = \int \overline{Y_L^M(\hat{\boldsymbol{r}})} Y_l^m(\hat{\boldsymbol{r}}) Y_{l'}^{m'}(\hat{\boldsymbol{r}}) d\Omega_{\boldsymbol{r}} \tag{3.8.10}$$

which can be obtained as

$$G_{Ll l'}^{Mmm'} = (-1)^M \sqrt{\frac{(2L+1)(2l+1)(2l'+1)}{4\pi}} \begin{pmatrix} L & l & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} L & l & l' \\ -M & m & m' \end{pmatrix}, \tag{3.8.11}$$

where $\begin{pmatrix} L & l & l' \\ M & m & m' \end{pmatrix}$ are Wigner 3j-symbols. Furthermore, by performing the sums with respect to m and m' as

$$C_{LM;ll'}^{\mu\nu} = 4\pi i^L \sum_{mm'} \overline{c_\mu^{lm}} c_\nu^{l'm'} G_{lLl'}^{mMm'} \tag{3.8.12}$$

the radial EMD is obtained in the form

$$n(p) = \sum_{\mu\nu} P_{\mu\nu} \sum_{ll'} \overline{\tilde{R}_{n_\mu l}^{\zeta_\mu}(p)} \tilde{R}_{n_\nu l'}^{\zeta_\nu}(p) \sum_{L=|l-l'|}^{l+l'} j_L(pR_{\mu\nu}) \sum_M C_{LM;ll'}^{\mu\nu} \overline{Y_L^M(\hat{\boldsymbol{R}}_{\mu\nu})}. \tag{3.8.13}$$

Now the Compton profile and the moments of the EMD can be integrated numerically from eqns (2.7) and (2.8).

It must be noted here that the radial expansion of eqn (3.8.4) only needs to be performed once in the calculation, and only for the non-equivalent basis functions (differing only in the phase factor $\exp(i\boldsymbol{p}\cdot\boldsymbol{R}_\mu)$). As a consequence, the necessary amount of radial functions $R_{n_\mu l}^{\zeta_\mu}$ and coupling coefficients $C_{LM;ll'}^{\mu\nu}$ is reduced tremendously. The computation can then be grouped into the final expression

$$\begin{aligned} n(p) = \sum_{\text{group } \tau} \sum_{\text{group } \sigma} \sum_{L=\min|l_\tau - l_\sigma|}^{\max l_\tau + \max l_\sigma} \sum_M \left(\sum_{l_\tau l_\sigma} \overline{\tilde{R}_{n_\tau l_\tau}^{\zeta_\tau}(p)} \tilde{R}_{n_\sigma l_\sigma}^{\zeta_\sigma}(p) C_{LM;l_\tau l_\sigma}^{\sigma\tau} \right) \\ \times \left(\sum_{\mu \text{ in } \tau} \sum_{\nu \text{ in } \sigma} P_{\mu\nu} j_L(p|\boldsymbol{R}_\mu - \boldsymbol{R}_\nu|) \overline{Y_L^M(\hat{\boldsymbol{R}}_{\mu\nu})} \right). \end{aligned} \tag{3.8.14}$$

The current implementation in ERKALE[38] is based on this formalism.

3.8.1. STO basis

The Fourier transform of the STO radial function (eqn (3.3.9)) is given by[113]

$$\begin{aligned} \tilde{R}_{nl}^{\zeta}(p) = (2\pi)^{3/2} \frac{2^{n-1}(n-l)!}{\pi^2} (-ip)^l \frac{\zeta^{n-l}(2\zeta)^{n+1/2}}{\sqrt{(2n)!}} \times \\ \sum_{k=0}^{\lfloor (n-l)/2 \rfloor} \left(-\frac{1}{4\zeta^2} \right)^k \frac{(n-k)!}{k!(n-l-2k)!} \frac{1}{(p^2+\zeta^2)^{n+1-k}}, \end{aligned} \tag{3.8.15}$$

Currently, the most commonly used STO program, ADF [64], does not use pure spherical functions; cartesian STOs are used, instead[36]. However, the cartesian STOs can be expanded in spherical harmonics (as was done in paper **I**), making it possible to use eqn (3.8.15) for the Fourier transforms of the cartesian basis functions.

3.8.2. GTO basis

The Fourier transform of the Gaussian function in eqn (3.3.16) is found using eqn (3.8.6) as

$$\tilde{R}^{\alpha}_{\lambda l}(p) = (-ip)^l 2^{(\lambda-l)/2} \left(\frac{2}{\pi}\right)^{1/4} \alpha^{-l/2-3/4} \times \frac{(l+\lambda+1)!!}{(2l+1)!!\sqrt{(2\lambda+1)!!}} {}_1F_1\left(\frac{l+\lambda}{2}+\frac{3}{2}; l+\frac{3}{2}; -\frac{p^2}{4\alpha}\right), \tag{3.8.16}$$

in which $n!!$ is the double factorial and ${}_1F_1$ is the confluent hypergeometric function given by

$${}_1F_1(a;b;z) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+k)} \frac{z^k}{k!}. \tag{3.8.17}$$

Here $\Gamma(z)$ is the Euler Gamma function. As shown in paper **I**, eqn (3.8.16) can generally be written as a product of a polynomial and an exponential. When $n = l+1$ the radial function simplifies to

$$\tilde{R}^{\alpha}_{l}(p) = (-ip)^l 2^{1/4} \alpha^{-l/2-3/4} \pi^{-1/4} \frac{1}{\sqrt{(2l+1)!!}} e^{-p^2/4\alpha}. \tag{3.8.18}$$

[36] The lower angular momentum contaminants are projected out at a later stage in ADF.

3.9. Illustrative calculation

The rather extreme difference between the change in the ED and that in the EMD upon the formation of chemical bonds is demonstrated by a HF calculation on the fluorine dimer, using the decontracted aug-pc-4 basis set [79–81]. The atoms are placed on the z axis, symmetrically about the origin, with the bond distance being 2.668 a.u.

The changes in the ED from the superposition of the two monomers to the bonded dimer (shown in Figure 8) are minute, only $\sim 0.02\%$ on the x and y axes (perpendicular to bond axis) and $\sim 0.2\%$ on the z axis (along the bond), as compared to the maximum electron density (at the nuclei). In comparison, the EMD undergoes a dramatic change (Figure 9): it changes by $\sim 7\%$ on the p_x and p_y axes and by $\sim 26\%$ on the p_z axis, as compared to its maximal value.

The isotropic Compton profile of the fluorine dimer is shown in Figure 10. However, as experimental determination of absolute Compton profiles is difficult, difference profiles are often used, instead. Difference profiles between the same sample in different environmental conditions, or between two different samples can be routinely measured with Compton scattering experiments at modern synchrotron radiation facilities. The difference profile of the F_2 molecule above, which we define as

$$\Delta J(q) = \frac{J_{\text{dimer}}(q) - J_{\text{monomer 1}}(q) - J_{\text{monomer 2}}(q)}{J_{\text{dimer}}(0)} \cdot 100\%, \qquad (3.9.1)$$

is shown in Figure 11. As the covalent bond between the fluorine atoms is quite strong, the amplitude of the difference profile is unusually large. The differences measured in experiments are usually an order of magnitude smaller, as there are usually no changes in the strong chemical bonds.

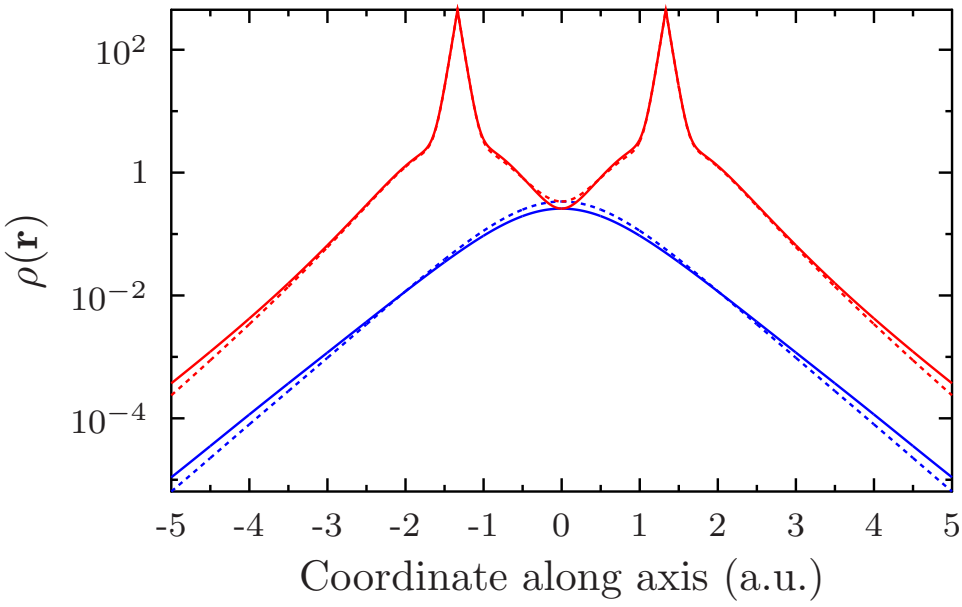


Figure 8: Electron density of the fluorine dimer. The solid line represents the dimer quantity, whereas the dashed line represents the superposition of the monomer quantities. Quantities along the z axis are plotted in red, whereas quantities orthogonal to the bond axis, *i.e.*, along the x and y axes are in blue. Note the logarithmic scale.

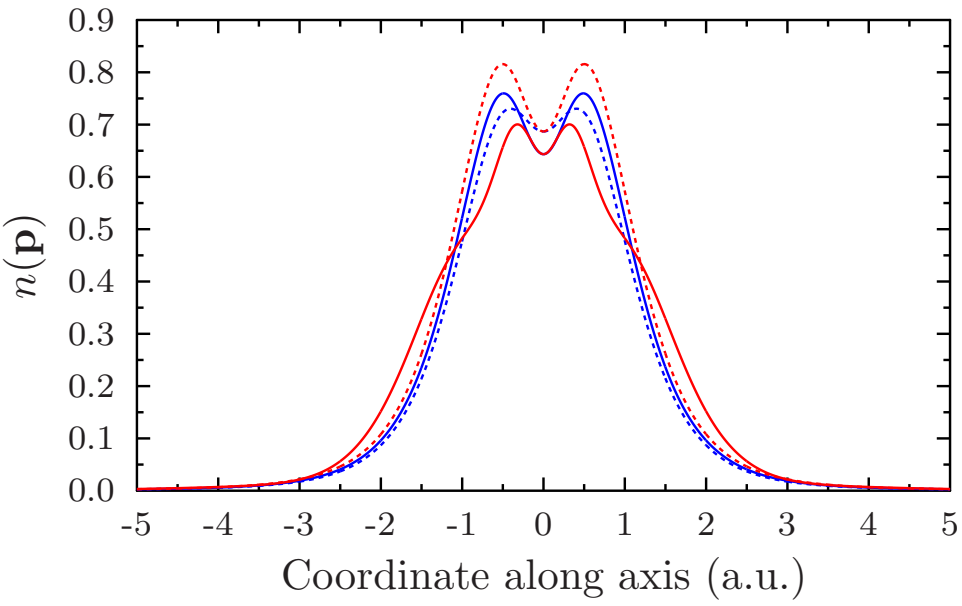


Figure 9: Electron momentum density of the fluorine dimer. The notation is analogous to the one in Figure 8.

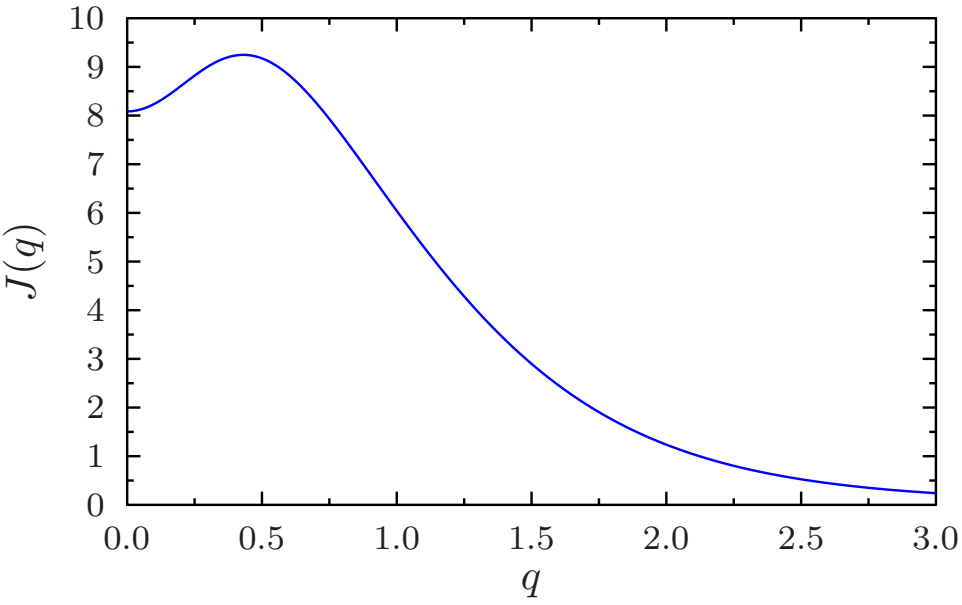


Figure 10: Isotropic Compton profile of the fluorine dimer.

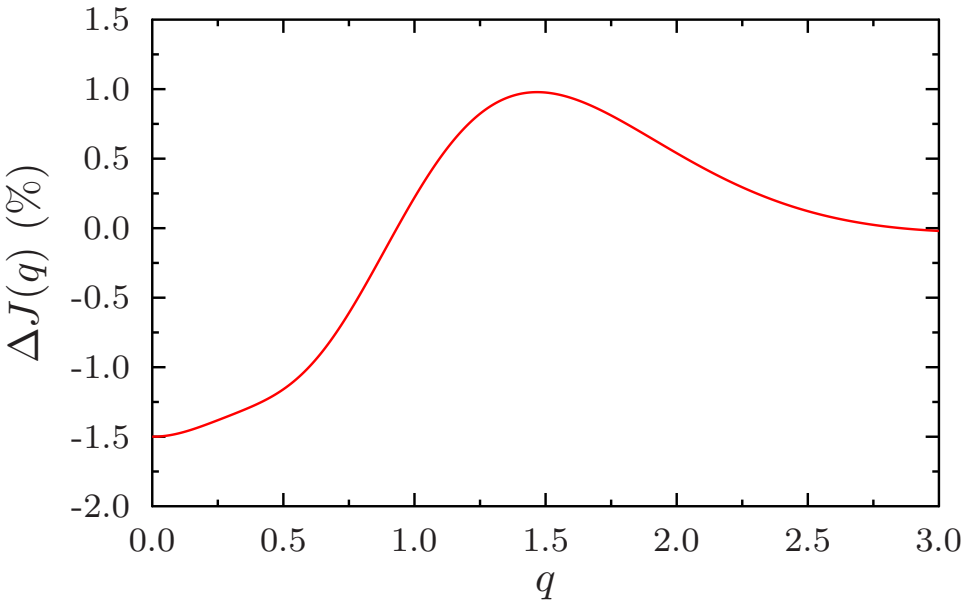


Figure 11: Difference Compton profile of the fluorine dimer.

4. Summary of papers

This thesis includes four publications on the computational modeling of the electron momentum density. Paper **I** focuses on developing a novel method for calculating the radial electron momentum density, and studies the electron momentum density produced by different basis sets at different levels of theory. Papers **II** and **III** focus on developing novel basis sets especially adapted for electron momentum density studies through the use of completeness-optimization. Paper **IV** presents the computer program that can be used, among other features, for computing the electron momentum density.

4.1. Paper I: Calculation of isotropic Compton profiles with Gaussian basis sets

We devised an algorithm for the calculation of the Compton profile on an adaptive grid, and introduced an algorithm for computing the radial electron momentum density analytically in Gaussian basis sets of arbitrary angular momentum. Using these, we were able to achieve machine accuracy for the calculation of isotropic Compton profiles within the used basis sets.

We chose two model systems for studying the basis set convergence of the Compton profile: the water dimer and the helium dimer. Water was chosen due to the recent interest in its experimental study using Compton scattering [13,15–17,20,23–25,37], whereas helium was chosen to represent weakly bound systems in which dispersion effects are dominant. Calculations were performed using HF and KS-DFT, but also with CC and MP theory. Only the valence electrons were correlated in the post-HF calculations.

We studied the basis set convergence, and noted the usual behavior: HF and KS-DFT behave similarly to each other, converging the fastest with regard to the basis set. Differently to the former methods, post-HF methods are more stringent about the used basis set, but share similar convergence properties within themselves.

We showed that an estimated accuracy of 0.01% of $J(0)$ in the Compton profile could be achieved with the aug-cc-pVTZ basis set at all levels of theory. Whereas KS-DFT fails for absolute profiles and the difference profiles of helium, in the case of water KS-DFT was found to produce similar difference profiles to those of wave-function methods – some hybrid KS-DFT functionals reproducing the difference profile even better than MP2, when compared to the CCSD reference.

As extremely accurate calculations are possible for two-electron systems such as the helium atom, we compared our highest level calculations on the helium atom to a large basis set FCI calculation performed in the literature. The difference in the calculations was around 0.04% of $J(0)$, hinting towards a substantial remaining basis set error even with quintuple zeta basis sets.

1 H							2 He
3 Li	4 Be	5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg	13 Al	14 Si	15 S	16 P	17 Cl	18 Ar

Table 1: Assembly of the elements of the first three rows of the periodic table into five groups with similar electronic shell structures. The coloring illustrates the grouping.

4.2. Paper II: Completeness-optimized basis sets: Application to ground-state electron momentum densities

We decided to tackle the basis set convergence problem using completeness-optimization, which was presented Subsection 3.3.3. As conventional basis sets are energy-optimized, they are often sub-optimal for computing other (such as magnetic) properties. Completeness-optimized basis sets offer a systematical and cost-effective approach to the complete basis set limit. The purpose of our study was twofold: to establish the CBS limit and to form computationally efficient basis sets with known levels of accuracy.

We introduced a novel algorithm for the automatic formation of basis sets adapted for any property at any level of theory, realizing the original procedure of Manninen and Vaara [86], and applied it to the electron momentum densities of the atoms and homoatomic dimers of the first three rows of the periodic table (elements H–Ar). For simplicity, the HF level of theory was used and only primitive basis sets were considered.

Test calculations showed that the CBS limit of $\langle p^n \rangle$ is attainable in practice with four digit accuracy only for $n \in \{-1, 1, 2, 3\}$ due to numerical problems. Concentrating on these values we proceeded with the completeness-optimization. Inspired by the idea of universal basis sets, the elements were assembled into five groups with similar electronic shell structures, the same basis set being used for all the elements in the group. The grouping is illustrated in Table 1.

Motivated by the current experimental accuracy of difference Compton profiles of 0.02% of $J_0 = \frac{1}{2}\langle p^{-1} \rangle$, the CBS limit *coemd-ref* basis set was parametrized so that the addition of a single exponent at any value of angular momentum would result in a relative change in the values of the moments of any element in the group by less than 10^{-4}. Additionally, the CBS set was reduced to form the *coemd-4*, *coemd-3*, *coemd-2* and *coemd-1* capable of reproducing the CBS limit values within the relative accuracies of 10^{-4}, 10^{-3}, 10^{-2} and 10^{-1}, respectively. Comparison of the results obtained with the completeness-optimized *coemd* basis sets with the Dunning-style correlation-consistent cc-pVXZ [74–77] and Jensen's polarization-consistent pc-N [79–82] basis sets proved that the completeness-optimized sets are notably more computationally efficient than the standard, energy-optimized basis sets. The contracted versions of the pc-N basis sets were found to converge to incorrect values, which we interpreted as an artifact of using contraction coefficients obtained from KS-DFT calculations.

4.3. Paper III: Contraction of completeness-optimized basis sets: Application to ground-state electron momentum densities

We continued the work started in paper **II** to forming contracted basis sets. Minor improvements were introduced in the automatic completeness-optimization algorithms presented in paper **II**, but as a larger difference to the preceding work, the SCF solutions were diagnosed for saddle point convergence [114] during the optimization. Due to this, reparametrization of the polarization-consistent basis sets of uncontracted primitives suggested in paper **II** resulted in minor changes to the composition of the completeness-optimized basis sets. However, test calculations did not show major differences between the results reproduced by the new *un-pcemd* and the old *coemd* parametrizations of the basis sets.

An automatic algorithm, applicable for any property, was suggested for forming the contractions. Contracting the primitive *un-pcemd* sets to the *pcemd* sets resulted in major reductions in the number of functions, while not significantly compromising the accuracy of the basis set.

The work was then continued to the post-HF level of theory. The MP2 level of theory was chosen due to its good accuracy in reproducing the moments of the EMD [30] and its lighter computational requirements compared to higher level theories. While the previously proposed algorithms could be straightforwardly applied to the basis set optimization, it was found out that the 10^{-4} relative accuracy is computationally unfeasible, as the resulting basis sets would be too large for any use in practice.

Accordingly, the target accuracy for the correlation-consistent *un-ccemd-ref* CBS limit basis set was increased to $\epsilon_t = 5 \times 10^{-4}$. The primitive basis set was then reduced to form the *un-ccemd-3* and *un-ccemd-2* basis sets, reproducing the estimated CBS limit with the relative accuracies of 10^{-3} and 10^{-2}, respectively. Contrary to the case of the polarization-consistent *pcemd* basis sets, the contraction to the correlation-consistent *ccemd* basis sets was not found to produce a notable reduction in the amount of functions in the basis set due to the higher basis set requirements of the level of theory.

Benchmark calculations were performed on a set of 45 diatomic molecules. The results unequivocally showed that the completeness-optimized basis sets outperform conventional basis sets both at the SCF and MP2 levels of theory.

4.4. Paper IV: ERKALE – A Flexible Program Package for X-ray Properties of Atoms and Molecules

Fundamentally a standard HF and KS-DFT code operating with Gaussian basis sets, ERKALE includes special features not available in other program packages, such as the calculation of Compton profiles, using the algorithms introduced in paper **I**. Furthermore, as accurate modeling of the EMD requires the use of post-HF methods currently not available in ERKALE, the program includes interfaces to, *e.g.*, Gaussian [115], which was applied in papers **I** and **III**.

The program also includes tools for completeness-optimization of basis sets, which were used in papers **II** and **III**. The minimization algorithm used in the completeness-optimization tools of ERKALE (Subsection 3.3.3.2) is more efficient than the ones in

the Kruununhaka program suite,[116] which has been used in previous work applying completeness-optimization[86,89–92].

5. Discussion and conclusions

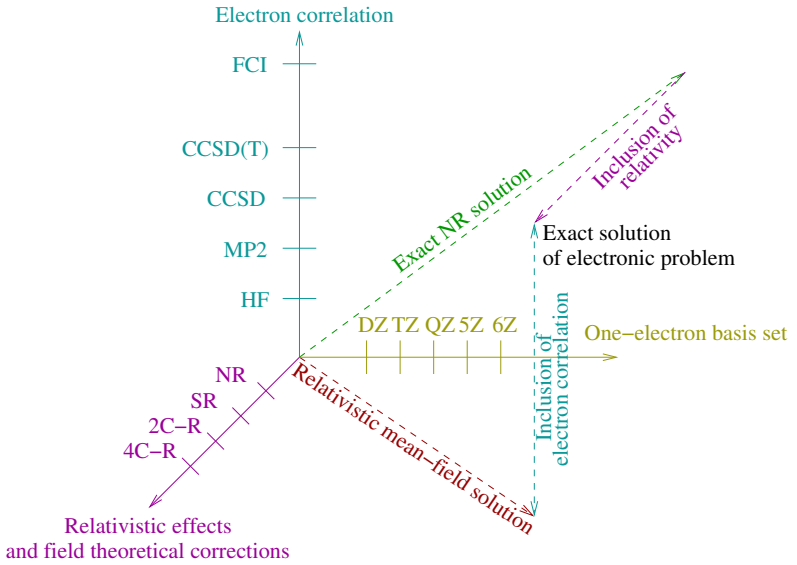


Figure 12: The strive towards the exact solution to the electronic problem.

The methodology of quantum chemistry is often illustrated as the cube shown in Figure 12. The figure has three axes. The basis set, which describes the quality of the one-electron picture from the double-ζ (2Z) to the sextuple-ζ (6Z) basis sets, continuing to the CBS limit at infinity, is on the horizontal axis. The vertical axis describes the quality of the N-electron picture, *i.e.*, the description of electron correlation. On the perpendicular axis is the Hamiltonian used to portray the physics of the system, ranging from the non-relativistic (NR) level of theory to scalar (spin-free) relativistic (SR), two-component relativistic (2C-R) and full four-component relativistic (4C-R) levels of theory, and going further on to corrections from quantum field theory, which result in, *e.g.*, the Lamb shift.

In this work we have only studied systems consisting of light elements, thus sticking to the back-lying plane of the non-relativistic level of theory, and strived towards the exact NR solution. We have investigated the computational requirements for accurate modeling of the electron momentum density by determining what kinds of basis sets and levels of theory are necessary for achieving sufficient convergence in the Compton profile for comparison with experimental results. The next logical step in the research would be to continue to heavier elements, where also relativistic effects need to be taken into account.

A notable part of the work has dealt with the development of novel basis sets through the completeness-optimization procedure. We have introduced automatic algorithms for

black-box formation of computationally efficient contracted basis sets that are straightforwardly applicable to any property at any level of theory. Applying the algorithms, we have developed novel, computationally efficient basis sets for calculating the EMD near the CBS limit at the SCF and MP2 levels of theory.

Furthermore, as a result of this thesis, we have introduced a novel, freely available software program, ERKALE[38], which can be used to model non-resonant inelastic x-ray scattering experiments.

The results obtained in this thesis should be of great practical importance in the modeling of Compton scattering experiments, as the basis sets developed in this work yield significant speed improvements to calculations, and the program is freely available. Due to the general nature of the program and the basis set formation algorithms we have suggested, we believe that the work performed as part of this thesis will also have an impact outside the field of non-resonant inelastic x-ray scattering spectroscopy.

Bibliography

[1] P. Hohenberg and W. Kohn. *Inhomogeneous Electron Gas*. **Phys. Rev. Lett. 136**, B864 (1964). doi:10.1103/PhysRev.136.B864.

[2] P. Coppens. *X-ray Charge Densities and Chemical Bonding*. Oxford University Press, 1997.

[3] A. J. Thakkar. *The momentum density perspective of the electronic structure of atoms and molecules*. **Adv. Chem. Phys. 128**, 303 (2004). doi:10.1002/0471484237.ch5.

[4] I. Mayer. *Simple Theorems, Proofs and Derivations in Quantum Chemistry*. Kluwer Academic / Plenum Publishers, New York, 2003. ISBN 0-306-47409-3.

[5] A. J. Thakkar. *Asymptotic behavior of atomic momentals*. **J. Chem. Phys. 86**, 5060 (1987). doi:10.1063/1.452648.

[6] M. J. Cooper, P. E. Mijnarends, N. Shiotani, N. Sakai, and A. Bansil. *X-ray Compton Scattering*, volume 5 of Oxford Series on Synchrotron Radiation. Oxford University Press, 2004.

[7] M. J. Cooper. *Compton scattering and electron momentum determination*. **Rep. Prog. Phys. 48**, 415 (1985). doi:10.1088/0034-4885/48/4/001.

[8] M. Coplan, J. Moore, and J. Doering. *(e,2e) spectroscopy*. **Rev. Mod. Phys. 66**, 985 (1994). doi:10.1103/RevModPhys.66.985.

[9] R. W. Siegel. *Positron Annihilation Spectroscopy*. **Ann. Rev. Mater. Sc.** 10, 393 (1980). doi:10.1146/annurev.ms.10.080180.002141.

[10] M. Alatalo, B. Barbiellini, M. Hakala, H. Kauppinen, T. Korhonen, M. J. Puska, K. Saarinen, P. Hautojärvi, and R. M. Nieminen. *Theoretical and experimental study of positron annihilation with core electrons in solids*. **Phys. Rev. B 54**, 2397 (1996). doi:10.1103/PhysRevB.54.2397.

[11] B. Barbiellini, M. Hakala, M. J. Puska, R. M. Nieminen, and A. A. Manuel, *Correlation effects for electron-positron momentum density in solids*, **Phys. Rev. B 56**, 7136 (1997). doi:10.1103/PhysRevB.56.7136.

[12] I. R. Epstein. *Calculation of Atomic and Molecular Momentum Expectation Values and Total Energies from Compton-Scattering Data*. **Phys. Rev. A 8**, 160 (1973). doi:10.1103/PhysRevA.8.160.

[13] K. Nygård, M. Hakala, S. Manninen, M. Itou, Y. Sakurai, and K. Hämäläinen. *Configurational Energetics in Ice Ih Probed by Compton Scattering*. **Phys. Rev. Lett. 99**, 197401 (2007). doi:10.1103/PhysRevLett.99.197401.

[14] P. Eisenberger and P. M. Platzman. *Compton Scattering of X Rays from Bound Electrons*. **Phys. Rev. A 2**, 415 (1970). doi:10.1103/PhysRevA.2.415.

[15] K. Nygård, M. Hakala, S. Manninen, K. Hämäläinen, M. Itou, A. Andrejczuk, and Y. Sakurai. *Ion hydration studied by x-ray Compton scattering*. **Phys. Rev. B 73**, 24208 (2006). doi:10.1103/PhysRevB.73.024208.

[16] K. Nygård, M. Hakala, T. Pylkkänen, S. Manninen, T. Buslaps, M. Itou, A. Andrejczuk, Y. Sakurai, M. Odelius, and K. Hämäläinen. *Isotope quantum effects in the electron momentum density of water.* **J. Chem. Phys. 126**, 154508 (2007). doi:10.1063/1.2723093.

[17] K. Nygård, M. Hakala, S. Manninen, A. Andrejczuk, M. Itou, Y. Sakurai, L. Pettersson, and K. Hämäläinen. *Compton scattering study of water versus ice Ih: Intra- and intermolecular structure*. **Phys. Rev. E 74**, 031503 (2006). doi:10.1103/PhysRevE.74.031503.

[18] I. Juurinen, K. Nakahara, N. Ando, T. Nishiumi, H. Seta, N. Yoshida, T. Morinaga, M. Itou, T. Ninomiya, Y. Sakurai, E. Salonen, K. Nordlund, K. Hämäläinen, and M. Hakala. *Measurement of Two Solvation Regimes in Water-Ethanol Mixtures Using X-Ray Compton Scattering*. **Phys. Rev. Lett. 107**, 197401 (2011). doi:10.1103/PhysRevLett.107.197401.

[19] B. Barbiellini and A. Shukla. *Ab initio calculations of the hydrogen bond*. **Phys. Rev. B 66**, 235101 (2002). doi:10.1103/PhysRevB.66.235101.

[20] P. Sit, Ch. Bellin, B. Barbiellini, D. Testemale, J.-L. Hazemann, T. Buslaps, N. Marzari, and A. Shukla. *Hydrogen bonding and coordination in normal and supercritical water from x-ray inelastic scattering*. **Phys. Rev. B 76**, 245413 (2007). doi:10.1103/PhysRevB.76.245413.

[21] B. Barbiellini, Ch. Bellin, G. Loupias, T. Buslaps, and A. Shukla. *How the hydrogen bond in NH4F is revealed with Compton scattering*. **Phys. Rev. B 79**, 155115 (2009). doi:10.1103/PhysRevB.79.155115.

[22] B. Barbiellini, A. Koizumi, P. Mijnarends, W. Al-Sawai, Hsin Lin, T. Nagao, K. Hirota, M. Itou, Y. Sakurai, and A. Bansil. *Role of Oxygen Electrons in the Metal-Insulator Transition in the Magnetoresistive Oxide La2-2xSr1+2xMn2O7 Probed by Compton Scattering*. **Phys. Rev. Lett. 102**, 206402 (2009). doi:10.1103/PhysRevLett.102.206402.

[23] M. Hakala, S. Huotari, K. Hämäläinen, S. Manninen, Ph. Wernet, A. Nilsson, and L. Pettersson. *Compton profiles for water and mixed water-neon clusters: A measure of coordination*. **Phys. Rev. B 70**, 125413 (2004). doi:10.1103/PhysRevB.70.125413.

[24] M. Hakala, K. Nygård, S. Manninen, L. Pettersson, and K. Hämäläinen. *Intra- and intermolecular effects in the Compton profile of water*. **Phys. Rev. B 73**, 035432 (2006). doi:10.1103/PhysRevB.73.035432.

[25] M. Hakala, K. Nygård, S. Manninen, S. Huotari, T. Buslaps, A. Nilsson, L. G. M. Pettersson, and K. Hämäläinen. *Correlation of hydrogen bond lengths and angles in liquid water based on Compton scattering.* **J. Chem. Phys. 125**, 084504 (2006). doi:10.1063/1.2273627.

[26] M. Hakala, K. Nygård, J. Vaara, M. Itou, Y. Sakurai, and K. Hämäläinen. *Charge localization in alcohol isomers studied by Compton scattering.* **J. Chem. Phys. 130**, 034506 (2009). doi:10.1063/1.3059421.

[27] R. K. Pathak, B. S. Sharma, and A. J. Thakkar. *Approximate relationships between density power integrals, moments of the momentum density, and interelectronic repulsions in diatomic molecules.* **J. Chem. Phys. 85**, 958 (1986). doi:10.1063/1.451252.

[28] A. J. Thakkar and W. A. Pedersen. *Local density functional approximations and conjectured bounds for momentum moments.* **Int. J. Quant. Chem. Symp. 24**, 327 (1990). doi:10.1002/qua.560382433.

[29] A. K. Roy and A. J. Thakkar. *MacLaurin expansions of electron momentum densities for 78 diatomic molecules: a numerical Hartree–Fock study.* **Chem. Phys. Lett. 362**, 428 (2002). doi:10.1016/S0009-2614(02)01101-6.

[30] J. R. Hart and A. J. Thakkar. *Moments of the electron momentum density: Requirements for ab initio and density functional theory calculations.* **Int. J. Quant. Chem. 102**, 673 (2005). doi:10.1002/qua.20444.

[31] C. Pisani, R. Dovesi, and R. Orlando. *Near-Hartree-Fock wave functions for solids: The case of crystalline silicon.* **Int. J. Quant. Chem. 42**, 5 (1992). doi:10.1002/qua.560420104.

[32] J. Wang and V. H. Smith. *Spherically averaged molecular electron densities and radial moments in position and momentum spaces.* **J. Phys. B: At. Mol. Opt. Phys. 27**, 5159 (1994). doi:10.1088/0953-4075/27/21/010.

[33] B. Miguel and J. M. García de la Vega. *Influence of electronic correlation in monoelectronic density in p-space.* **Theor. Chem. Acc. 118**, 723 (2007). doi:10.1007/s00214-007-0354-y.

[34] A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz, and D. Usvyat. *MP2 versus density-functional theory study of the Compton profiles of crystalline urea.* **Phys. Rev. B 81**, 165108 (2010). doi:10.1103/PhysRevB.81.165108.

[35] A. Erba, M. Itou, Y. Sakurai, R. Yamaki, M. Ito, S. Casassa, L. Maschio, A. Terentjevs, and C. Pisani. *Beyond a single-determinantal description of the density matrix of periodic systems: Experimental versus theoretical Compton profiles of crystalline silicon.* **Phys. Rev. B 83**, 125208 (2011). doi:10.1103/PhysRevB.83.125208.

[36] C. Pisani, A. Erba, S. Casassa, M. Itou, and Y. Sakurai. *Anisotropy of the electron momentum distribution in α-quartz investigated by Compton scattering and ab initio simulations*. **Phys. Rev. B 84**, 245102 (2011). doi:10.1103/PhysRevB.84.245102.

[37] T. K. Ghanty, V. N. Staroverov, P. R. Koren, and E. R. Davidson. *Is the Hydrogen Bond in Water Dimer and Ice Covalent?* **J. Am. Chem. Soc. 122**, 1210 (2000). doi:10.1021/ja9937019.

[38] S. Lehtola. *ERKALE - HF/DFT from Hel*, 2012. URL `http://erkale.googlecode.com`.

[39] A. J. Thakkar, A. M. Simas, and V. H. Smith. *Extraction of momentum expectation values from Compton profiles*. **Mol. Phys. 41**, 1153 (1980). doi:10.1080/00268978000103851.

[40] `http://www.esrf.eu/about/synchrotron-science/synchrotron-light`, cited Jan 10 2013.

[41] J. D. Jackson. *Classical Electrodynamics*. Wiley, New York, third edition, 1998.

[42] `http://www.diamond.ac.uk/Home/About/FAQs/Science.html`, cited Jan 19 2013.

[43] H. Sakurai, H. Ota, N. Tsuji, M. Itou, and Y. Sakurai. *Accurate Compton scattering measurements of noble gases: the importance of electron correlations in heavy atoms*. **J. Phys. B: At. Mol. Opt. Phys. 44**, 065001 (2011). doi:10.1088/0953-4075/44/6/065001.

[44] A. Thompson, D. Attwood, E. Gullikson, M. Howells, K.-J. Kim, J. Kirz, J. Kortright, I. Lindau, Y. Liu, P. Pianetta, A. Robinson, J. Scofield, J. Underwood, G. Williams, and H. Winick. **X-ray Data Booklet**. Lawrence Berkeley National Laboratory, University of California, 3rd edition, 2009. `http://xdb.lbl.gov/`.

[45] P. Jaiswal and A. Shukla. *Kinetically balanced Gaussian basis-set approach to relativistic Compton profiles of atoms*. **Phys. Rev. A 75**, 022504 (2007). doi:10.1103/PhysRevA.75.022504.

[46] D. Lai. *Matter in strong magnetic fields*. **Rev. Mod. Phys. 73**, 629 (2001). doi:10.1103/RevModPhys.73.629.

[47] K. K. Lange, E. I. Tellgren, M. R. Hoffmann, and T. Helgaker. *A Paramagnetic Bonding Mechanism for Diatomics in Strong Magnetic Fields*. **Science 337**, 327 (2012). doi:10.1126/science.1219703.

[48] Y. Mizuno and Y. Ohmura. *Theory of X-Ray Raman Scattering*. **J. Phys. Soc. Japan 22**, 445 (1967). doi:10.1143/JPSJ.22.445.

[49] W. Schülke. *Electron Dynamics by Inelastic X-Ray Scattering*, volume 7 of Oxford Series on Synchrotron Radiation. Oxford University Press Inc., New York, 2007. ISBN 978-0-19-851017-8.

[50] M. Dierksen and S. Grimme. *Density functional calculations of the vibronic structure of electronic absorption spectra.* **J. Chem. Phys. 120**, 3544 (2004). doi:10.1063/1.1642595.

[51] A. Sakko, S. Galambosi, J. Inkinen, T. Pylkkänen, M. Hakala, S. Huotari, and K. Hämäläinen. *Inelastic X-ray scattering and vibrational effects at the K-edges of gaseous N2, N2O, and CO2.* **Phys. Chem. Chem. Phys. 13**, 11678 (2011). doi:10.1039/c1cp20295b.

[52] E. Merzbacher. *Quantum Mechanics*. John Wiley & Sons, Inc., third edition, 1998. ISBN 0-471-88702-1.

[53] K. J. H. Giesbertz and E. J. Baerends. *Aufbau derived from a unified treatment of occupation numbers in Hartree-Fock, Kohn-Sham, and natural orbital theories with the Karush-Kuhn-Tucker conditions for the inequality constraints* $n_i \leq 1$ *and* $n_i \geq 0$. **J. Chem. Phys. 132**, 194108 (2010). doi:10.1063/1.3426319.

[54] S. F. Boys and F. Bernardi, *The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors*. **Mol. Phys. 19**, 553 (1970).

[55] L. Laaksonen, P. Pyykkö, and D. Sundholm. *Two-dimensional fully numerical solutions of molecular Schrödinger equations. I. One-electron molecules*. **Int. J. Quant. Chem. 23**, 309 (1983). doi:10.1002/qua.560230126.

[56] L. Laaksonen, P. Pyykkö, and D. Sundholm. *Two-Dimensional fully numerical solutions of molecular Schrödinger equations. II. Solution of the Poisson equation and results for singlet states of* H_2 *and* HeH^+. **Int. J. Quant. Chem. 23**, 319 (1983). doi:10.1002/qua.560230127.

[57] L. Laaksonen, P. Pyykkö, and D. Sundholm. *Two-dimensional fully numerical solutions of molecular Hartree-Fock equations: LiH and BH*. **Chem. Phys. Lett. 96**, 1 (1983). doi:10.1016/0009-2614(83)80104-3.

[58] T. Yanai, G. I. Fann, Z. Gan, R. J. Harrison, and G. Beylkin. *Multiresolution quantum chemistry in multiwavelet bases: Hartree-Fock exchange.* **J. Chem. Phys. 121**, 6680 (2004). doi:10.1063/1.1790931.

[59] R. J. Harrison, G. I. Fann, T. Yanai, Z. Gan, and G. Beylkin. *Multiresolution quantum chemistry: basic theory and initial applications.* **J. Chem. Phys. 121**, 11587 (2004). doi:10.1063/1.1791051.

[60] F. A. Bischoff and E. F. Valeev. *Low-order tensor approximations for electronic wave functions: Hartree-Fock method with guaranteed precision.* **J. Chem. Phys. 134**, 104104 (2011). doi:10.1063/1.3560091.

[61] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, Ltd., 2000. ISBN 0 471 96755 6.

[62] O. Vahtras, J. Almlöf, and M. W. Feyereisen. *Integral approximations for LCAO-SCF calculations*. **Chem. Phys. Lett. 213**, 514 (1993). doi:10.1016/0009-2614(93)89151-7.

[63] M. A. Watson, N. C. Handy, and A. J. Cohen. *Density functional calculations, using Slater basis sets, with exact exchange*. **J. Chem. Phys. 119**, 6475 (2003). doi:10.1063/1.1604371.

[64] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler. *Chemistry with ADF*. **J. Comp. Chem. 22**, 931 (2001). doi:10.1002/jcc.1056.

[65] S. F. Boys. *Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System*. **Proc. R. Soc. A 200**, 542 (1950). doi:10.1098/rspa.1950.0036.

[66] R. Kikuchi. *Gaussian Functions in Molecular Integrals*. **J. Chem. Phys. 22**, 148 (1954). doi:10.1063/1.1739831.

[67] S. Huzinaga. *Gaussian-Type Functions for Polyatomic Systems. I*. **J. Chem. Phys. 42**, 1293 (1965). doi:10.1063/1.1696113.

[68] H. Taketa, S. Huzinaga, and K. O-ohata. *Gaussian-Expansion Methods for Molecular Integrals*. **J. Phys. Soc. Japan 21**, 2313 (1966). doi:10.1143/JPSJ.21.2313.

[69] K. O-ohata, H. Taketa, and S. Huzinaga. *Gaussian Expansions of Atomic Orbitals*. **J. Phys. Soc. Japan 21**, 2306 (1966). doi:10.1143/JPSJ.21.2306.

[70] L. E. McMurchie and E. R. Davidson. *One- and two-electron integrals over cartesian gaussian functions*. **J. Comp. Phys. 26**, 218 (1978). doi:10.1016/0021-9991(78)90092-X.

[71] S. Obara and A. Saika. *Efficient recursive computation of molecular integrals over Cartesian Gaussian functions*. **J. Chem. Phys. 84**, 3963 (1986). doi:10.1063/1.450106.

[72] M. Head-Gordon and J. A. Pople. *A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations*. **J. Chem. Phys. 89**, 5777 (1988). doi:10.1063/1.455553.

[73] H. B. Schlegel and M. J. Frisch. *Transformation between Cartesian and pure spherical harmonic Gaussians*. **Int. J. Quant. Chem. 54**, 83 (1995). doi:10.1002/qua.560540202.

[74] T. H. Dunning. *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*. **J. Chem. Phys. 90**, 1007 (1989). doi:10.1063/1.456153.

[75] 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). doi:10.1063/1.462569.

[76] A. K. Wilson, T. van Mourik, and T. H. Dunning. *Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets for boron through neon*. **J. Mol. Struct. (Theochem) 388**, 339 (1996). doi:DOI: 10.1016/S0166-1280(96)80048-0.

[77] K. A. Peterson, D. E. Woon, and T. H. Dunning. *Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the H+H_2 → H_2+H reaction*. **J. Chem. Phys. 100**, 7410 (1994). doi:10.1063/1.466884.

[78] D. E. Woon and T. H. Dunning. *Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon*. **J. Chem. Phys. 98**, 1358 (1993). doi:10.1063/1.464303.

[79] F. Jensen. *Polarization consistent basis sets: Principles*. **J. Chem. Phys. 115**, 9113 (2001). doi:10.1063/1.1413524.

[80] F. Jensen. *Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit*. **J. Chem. Phys. 116**, 7372 (2002). doi:10.1063/1.1465405.

[81] F. Jensen. *Polarization consistent basis sets. III. The importance of diffuse functions*. **J. Chem. Phys. 117**, 9234 (2002). doi:10.1063/1.1515484.

[82] F. Jensen. *Polarization consistent basis sets. 4: the elements He, Li, Be, B, Ne, Na, Mg, Al, and Ar.* **J. Phys. Chem. A 111**, 11198 (2007). doi:10.1021/jp068677h.

[83] F. Jensen and T. Helgaker. *Polarization consistent basis sets. V. The elements Si-Cl*. **J. Chem. Phys. 121**, 3463 (2004). doi:10.1063/1.1756866.

[84] D. M. Silver and W. C. Nieuwpoort. *Universal atomic basis sets*. **Chem. Phys. Lett. 57**, 421 (1978). doi:10.1016/0009-2614(78)85539-0.

[85] D. M. Silver, S. Wilson, and W. C. Nieuwpoort. *Universal basis sets and transferability of integrals*. **Int. J. Quant. Chem. 14**, 635 (1978). doi:10.1002/qua.560140510.

[86] P. Manninen and J. Vaara. *Systematic Gaussian basis-set limit using completeness-optimized primitive sets. A case for magnetic properties*. **J. Comp. Chem. 27**, 434 (2006). doi:10.1002/jcc.20358.

[87] D. P. Chong. *Completeness profiles of one-electron basis sets*. **Can. J. Chem. 73**, 79 (1995). doi:10.1139/v95-011.

[88] J. A. Nelder and R. Mead. *A simplex method for function minimization*. **Comp. J. 7**, 308 (1965). doi:10.1093/comjnl/7.4.308.

[89] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara. *Laser-induced nuclear magnetic resonance splitting in hydrocarbons.* **J. Chem. Phys. 129**, 124102 (2008). doi:10.1063/1.2977741.

[90] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara. *NMR tensors in planar hydrocarbons of increasing size.* **Phys. Chem. Chem. Phys. 11**, 11404 (2009). doi:10.1039/b919860a.

[91] S. Ikäläinen, M. Romalis, P. Lantto, and J. Vaara. *Chemical Distinction by Nuclear Spin Optical Rotation.* **Phys. Rev. Lett. 105**, 153001 (2010). doi:10.1103/PhysRevLett.105.153001.

[92] S. Ikäläinen, P. Lantto, and J. Vaara. *Fully Relativistic Calculations of Faraday and Nuclear Spin-Induced Optical Rotation in Xenon.* **J. Chem. Theor. Comp. 8**, 91 (2012). doi:10.1021/ct200636m.

[93] G. H. Booth, A. J. W. Thom, and A. Alavi. *Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in Slater determinant space.* **J. Chem. Phys. 131**, 054106 (2009). doi:10.1063/1.3193710.

[94] G. H. Booth and A. Alavi. *Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: a study of ionization potentials.* **J. Chem. Phys. 132**, 174104 (2010). doi:10.1063/1.3407895.

[95] D. Cleland, G. H. Booth, and A. Alavi. *Communications: Survival of the fittest: accelerating convergence in full configuration-interaction quantum Monte Carlo.* **J. Chem. Phys. 132**, 041103 (2010). doi:10.1063/1.3302277.

[96] G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi. *Breaking the carbon dimer: the challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods.* **J. Chem. Phys. 135**, 084104 (2011). doi:10.1063/1.3624383.

[97] D. M. Cleland, G. H. Booth, and A. Alavi. *A study of electron affinities using the initiator approach to full configuration interaction quantum Monte Carlo.* **J. Chem. Phys. 134**, 024112 (2011). doi:10.1063/1.3525712.

[98] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse. *Natural Orbitals for Wave Function Based Correlated Calculations Using a Plane Wave Basis Set.* **J. Chem. Theor. Comp. 7**, 2780 (2011). doi:10.1021/ct200263g.

[99] D. Cleland, G. H. Booth, C. Overy, and A. Alavi. *Taming the First-Row Diatomics: A Full Configuration Interaction Quantum Monte Carlo Study.* **J. Chem. Theor. Comp. 8**, 4138 (2012). doi:10.1021/ct300504f.

[100] J. J. Shepherd, G. H. Booth, and A. Alavi. *Investigation of the full configuration interaction quantum Monte Carlo method using homogeneous electron gas models.* **J. Chem. Phys. 136**, 244101 (2012). doi:10.1063/1.4720076.

[101] C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi. *Full Configuration Interaction Excitations of Ethene and Butadiene: Resolution of an Ancient Question.* **J. Chem. Theor. Comp. 8**, 4441 (2012). doi:10.1021/ct300486d.

[102] J. Shepherd, G. Booth, A. Grüneis, and A. Alavi. *Full configuration interaction perspective on the homogeneous electron gas*. **Phys. Rev. B 85**, 081103 (2012). doi:10.1103/PhysRevB.85.081103.

[103] Chr. Møller and M. S. Plesset. *Note on an Approximation Treatment for Many-Electron Systems*. **Phys. Rev. 46**, 618 (1934). doi:10.1103/PhysRev.46.618.

[104] D. Cremer. *Møller-Plesset perturbation theory: from small molecule methods to methods for thousands of atoms*. **Wiley Interdisciplinary Reviews: Computational Molecular Science 1**, 509 (2011). doi:10.1002/wcms.58.

[105] Jiří Čížek. *On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods*. **J. Chem. Phys. 45**, 4256 (1966). doi:10.1063/1.1727484.

[106] W. Kohn and L. J. Sham. *Self-Consistent Equations Including Exchange and Correlation Effects*. **Phys. Rev. 140**, A1133 (1965). doi:10.1103/PhysRev.140.A1133.

[107] K. Burke. *Perspective on density functional theory*. **J. Chem. Phys. 136**, 150901 (2012). doi:10.1063/1.4704546.

[108] A. A. Jarzecki and E. R. Davidson. *Kinetic and potential energy of isoelectronic atomic ions from density functional theory compared with exact values*. **Mol. Phys. 98**, 1089 (2000). doi:10.1080/00268970050080456.

[109] S. Ragot. *Exact Kohn-Sham versus Hartree-Fock in momentum space: Examples of two-fermion systems*. **J. Chem. Phys. 125**, 014106 (2006). doi:10.1063/1.2212935.

[110] L. Lam and P. M. Platzman. *Momentum density and Compton profile of the inhomogeneous interacting electronic system. I. Formalism*. **Phys. Rev. B 9**, 5122 (1974). doi:10.1103/PhysRevB.9.5122.

[111] P. Kaijser and V. H. Smith. *Evaluation of Momentum Distributions and Compton Profiles for Atomic and Molecular Systems*. **Adv. Quantum Chem. 10**, 37 (1977). doi:10.1016/S0065-3276(08)60578-X

[112] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal. *The SIESTA method for ab initio order-N materials simulation*. **J. Phys: Cond. Mat. 14**, 2745 (2002). doi:10.1088/0953-8984/14/11/302.

[113] D. Belkić and H. S. Taylor. *A Unified Formula for the Fourier Transform of Slater-Type Orbitals*. **Phys. Scr. 39**, 226 (1989). doi:10.1088/0031-8949/39/2/004.

[114] R. Seeger and J. A. Pople. *Self-consistent molecular orbital methods. XVIII. Constraints and stability in Hartree–Fock theory*. **J. Chem. Phys. 66**, 3045 (1977). doi:10.1063/1.434318.

[115] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, Jr. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. **Gaussian 09 Revision C.01**, 2009.

[116] *Kruununhaka basis set tool kit*, written by P. Manninen and S. Lehtola, 2011. URL `http://www.chem.helsinki.fi/manninen/kruununhaka/`

Paper I

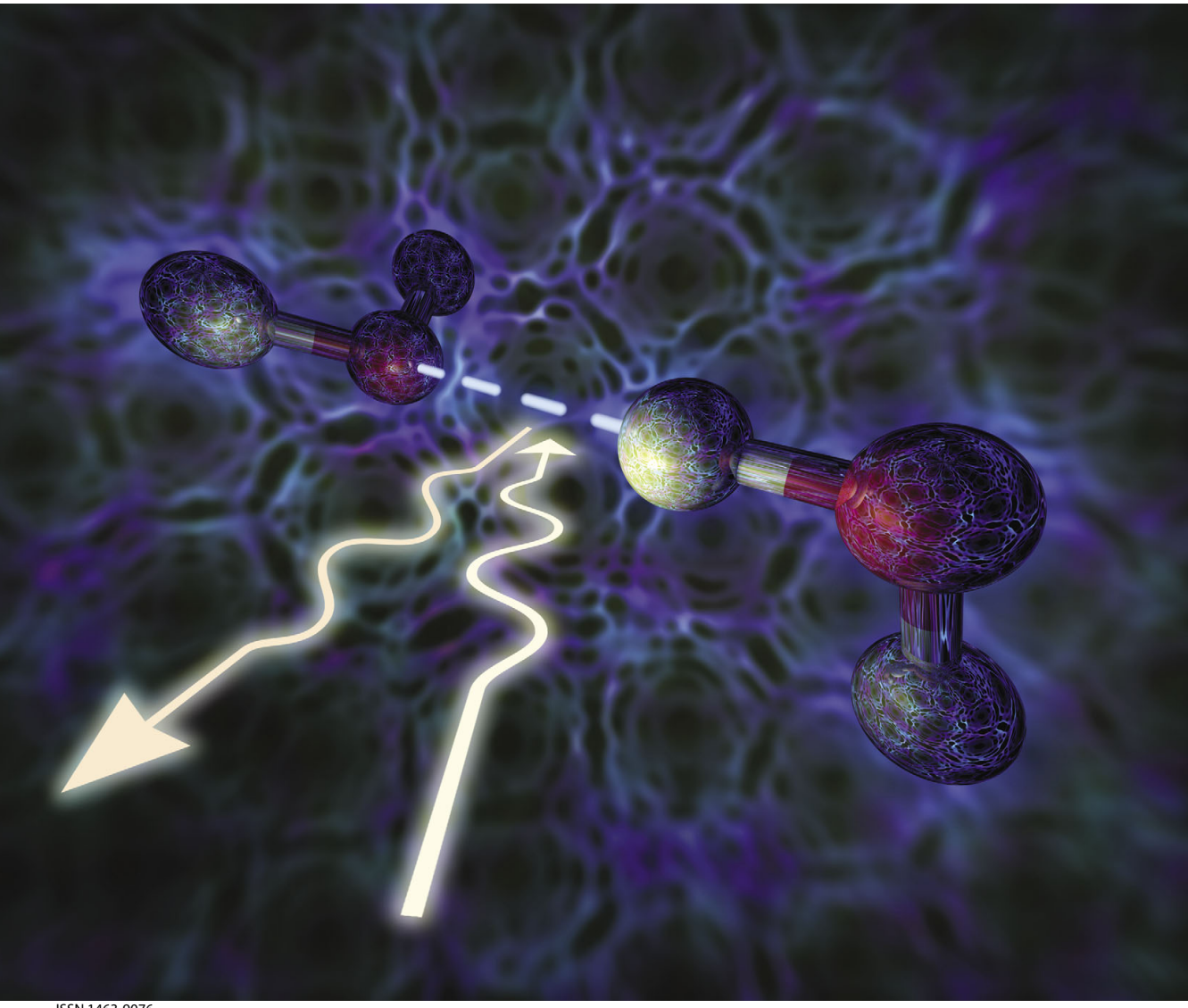
Calculation of isotropic Compton profiles with Gaussian basis sets
Jussi Lehtola, Mikko Hakala, Juha Vaara, and Keijo Hämäläinen
Physical Chemistry Chemical Physics, 2011, 13, 5630–5641
doi:10.1039/C0CP02269A

PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

Volume 13 | Number 13 | 7 April 2011 | Pages 5481–6372



ISSN 1463-9076

PCCP

Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 5630–5641

www.rsc.org/pccp

PAPER

Calculation of isotropic Compton profiles with Gaussian basis sets

Jussi Lehtola,*[a] Mikko Hakala,[a] Juha Vaara[b] and Keijo Hämäläinen[a]

Received 25th October 2010, Accepted 4th January 2011
DOI: 10.1039/c0cp02269a

In this paper we present an adaptive algorithm for calculating the isotropic Compton profile (ICP) for any type of Gaussian basis set. The ICP is a measure of the momentum density of electrons and it can be obtained from inelastic X-ray scattering experiments employing synchrotron radiation. We have performed calculations of the ICP for water and helium monomers and dimers using density-functional theory, Hartree–Fock and post-Hartree–Fock methods, with Dunning-type ((d-)aug-)cc-p(C)VXZ basis sets. We have examined the convergence of the Compton profile as a function of the basis set and the level of theory used for the formation of the density matrix. We demonstrate that diffuse basis functions are of utmost importance to the calculation of Compton profiles. Basis sets of at least triple-ζ quality appended by diffuse functions should be used in Compton profile calculations in order to obtain sufficient convergence with regard to the current, experimentally feasible accuracy for systems consisting of light elements.

1. Introduction

The electron density (ED) is a quantity that is at the heart of materials science, as it is inextricably linked with the structure and behavior of matter. The Hohenberg–Kohn theorems[1] show that the ground-state ED unambiguously defines the quantum mechanical system. Thus, in principle, when the exact ED is known, any property of the system can be calculated exactly. Furthermore, the ED is a quantity that is easy to grasp and intuitive, as it is high near atomic nuclei and covalent chemical bonds. As chemical interactions affect the ED, studies of the ED using, *e.g.*, X-ray diffraction can be used to obtain information about chemical bonds.[2] Experimental information about the ED can be used to verify *ab initio* calculations, which are often necessary for a complete understanding of the behavior of materials. The effect of the inclusion of electron correlation in the calculation of ED has been studied[3] and found to be necessary for the proper description of systems that are poorly described by a single Hartree–Fock (HF) determinant, such as the ozone molecule O_3.

The momentum space counterpart of the ED, the electron momentum density (EMD), $N(\boldsymbol{p})$, can be similarly used to study chemical bonding. Whereas the ED is higher along covalent bonds, the EMD oscillates along the bonding direction (in reciprocal space). A fitting procedure for extracting EMD properties from the ED has been recently suggested.[4] However, EMD properties are more traditionally obtained directly from calculations, or measured with techniques such as $(e, 2e)$ spectroscopy, positron annihilation spectroscopy or X-ray Compton scattering (CS). CS refers to a process where an incident X-ray photon scatters inelastically off an electron from the target sample, exchanging a large amount of energy and momentum.

The theory of CS is well known, for reviews see, *e.g.*, refs. 5 and 6. Assuming that the impulse approximation[7] holds, the double differential cross section, which measures the amount of photons scattered by the sample into the solid angle $\mathrm{d}\Omega$ with energy E_2, can be written as[5,6]

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\,\mathrm{d}E_2} = C(E_1, E_2, \phi)J(p_q). \tag{1}$$

Here E_1 is the energy of the incoming photon, ϕ is the scattering angle and $C(E_1,E_2,\phi)$ only depends on the setup of the experiment. p_q is the projection of the initial momentum of the electron onto the scattering vector. The function $J(p_q)$ is the Compton profile (CP)

$$J(p_q) = \iint N(p_x,p_y,p_z = p_q)\mathrm{d}p_x\mathrm{d}p_y, \tag{2}$$

which measures[5] the projection of the EMD along the direction of the scattering vector $\boldsymbol{q}$.

CS can be used to obtain experimental information about the EMD, and hence on the chemical bonds and the ionic configuration. To circumvent various experimental difficulties with the determination of absolute CPs, it is usual to compare CPs in different conditions, such as the profile along different directions of crystalline samples, or the profile of the same sample in different thermodynamic environments. In crystalline systems the anisotropy of the CP along crystal axes has

[a] *Department of Physics, University of Helsinki, P. O. Box 64, FI-00014 University of Helsinki, Finland. E-mail: jussi.lehtola@helsinki.fi*
[b] *NMR Research Group, Department of Physics, University of Oulu, P. O. Box 3000, FI-90014 University of Oulu, Finland*

been recently used in studying the nature of the hydrogen bond in ice,[8] the hydrogen bond signature in NH_4F[9] and the metal–insulator transition in $La_{2-2x}Sr_{1+2x}Mn_2O_7$.[10]

In isotropic materials, in which no directional information can be obtained, comparisons can be made, for instance, at different temperatures or in various pressures, as well as between isomers. Recently, this method has been used in experiments in the study of the hydrogen bond length and angle correlation,[11] intra- and intermolecular effects,[12] isotope quantum effects[13] and local hydrogen bond geometries[14] in water, configurational energetics in ice[15] and charge localization in alcohols,[16] for example.

Currently the experimental precision of the measurements of light molecular systems (*e.g.*, those composed of carbon, oxygen and hydrogen) is such that CP differences of roughly 0.02% of the peak height of the CP, $J(0)$, can be seen.[12–16] The ongoing development of synchrotron radiation sources and beamline instrumentation will further increase the statistics, and thus the precision, of CS experiments. This progress presents a challenge to the theoretical calculation of the CPs, as the interpretation of the experimental results leans heavily on modeling.

Kohn–Sham density-functional theory (KS-DFT)[1,17] is a widely used method in chemistry and physics, as it is computationally less demanding than high-level *ab initio* wave function methods while it still takes electron correlation into account *via* the use of the exchange–correlation (XC) functional. Lam and Platzman (LP) showed[18] that the expression the EMD acquires an extra term (compared to HF) in the case of KS-DFT, due to the assumption that the fictitious KS one-electron states do not interact. The LP correction to the peak of the absolute CP is of the order of 2% for atomic Ne and Ar[19,20] and 3% in crystalline Li[21] in the local density approximation (LDA).

The LDA correction proposed by LP uses the free-electron Fermi distribution, resulting in an isotropic correction. This can be somewhat dubious in the gas phase or systems where the Fermi sphere can be heavily deformed. Furthermore, the LP correction term is in fact seldom included at all in the analysis, and as far as the authors are aware, it has never been used within the context of the generalized-gradient approximation (GGA). An alternative method for calculating an anisotropic correction to the EMD is discussed in ref. 21.

The momentum-space performance of DFT without LP-type corrections has been investigated in several works. Studies of exact KS-DFT results have been made by examining closed-shell two-electron systems that are exactly solvable, or for which full configuration interaction (FCI) results are available. The exact KS orbital is formed from the ED produced by the reference calculation (exact or FCI), and is then Fourier transformed to form the EMD. Jarzęcki and Davidson investigated[22] the He atom, for which the ED was produced with a Kinoshita-type wave function, and saw that the Fourier transformed KS orbital does not reproduce the correct EMD. Ragot compared[23] the Fourier transforms of exact KS-DFT and HF orbitals for the Moshinsky atom and Hooke's atom, and concluded that the exact KS orbitals produce better momentum-space results than HF as compared to the exact EMD. He noted, however, that in actual DFT calculations of light two-electron ions the momentum space performance strongly depends on the choice of the XC functional.

For molecular systems Ghanty *et al.*[24] studied the CP anisotropy in the water dimer and found that HF and DFT give similar results, whereas Hakala *et al.*[25] found significant deviations between the HF, MP2 and DFT results for the difference CP between the water dimer and two free monomers. Hart and Thakkar calculated[26] moments of the EMD for a large set of closed-shell molecules at HF, various post-HF and DFT levels, also studying the effect of the used basis set. They found that DFT methods provide poor results for the moments; in fact often worse than plain HF when compared to high-level theory (coupled-cluster singles and doubles, CCSD) data. The work of Hart and Thakkar was continued by Miguel and Garcìa de la Vega[27] who analyzed the effect of electron correlation on the radial EMD with similar conclusions. Erba *et al.* studied[28] the performance of the second-order Møller–Plesset (MP2) method *versus* DFT for the calculation of the directional CPs in crystalline urea. They noted that while the DFT results are generally satisfactory, there is some disagreement with the experimental profile anisotropies, which is better accounted for by the MP2 calculations.

In this work we perform a comprehensive study of the basis set requirements for the calculation of isotropic Compton profiles using different levels of theory by studying two systems: the water and the helium dimers. The first experimental and calculated CPs for the systems under study in this work were reported some 40 years ago,[29–31] and they have also been studied more recently.[11,12,24,25,32–34] Our work represents an extension of high-level EMD studies of hydrogen bonded systems, and systems the energetics of which is dominated by dispersive interactions. The study of dimers makes it possible to look at the basis set effects in detail, as, *e.g.*, counterpoise calculations can be performed. Sufficient convergence with respect to the basis set is determined by the current experimental accuracy, which for light systems is of the order of 0.02% of the height of the Compton peak.[11,13–16,35]

The layout of this paper is as follows. First, in section 2 we present the method we use in the calculations of the CPs. In section 3 we discuss the details of the calculations, and present and discuss the results in section 4. Finally, in section 5 we draw the conclusions of this study.

2. Theory

In isotropic systems the CP can be calculated as[36]

$$J(p_q) = \frac{1}{2}\int_{|p_q|}^{\infty} pN(p)\mathrm{d}p \tag{3}$$

$$N(p) = \int N(\boldsymbol{p})\mathrm{d}\Omega_{\mathbf{p}} \tag{4}$$

where $N(p)$ is the EMD integrated over all of the directions of $\boldsymbol{p}$. For the rest of the manuscript, we streamline the notation by denoting p_q in eqn (3) with a plain q. The CP has an intricate connection to the moments of the EMD

$$\langle p^k \rangle = \int_0^{\infty} p^{k+2}N(p)\mathrm{d}p. \tag{5}$$

As can be directly seen from eqn (3)

$$\langle p^{-1} \rangle = 2J(0), \tag{6}$$

and one can also show that[26,37]

$$\langle p^k \rangle = 2(k+1)\int_0^\infty q^k J(q)\mathrm{d}q,\ k \geq 0, \tag{7}$$

and[38]

$$\langle p^{-2} \rangle = 2\int_0^\infty q^{-2}[J(q) - J(0)]\mathrm{d}q. \tag{8}$$

In any given basis the EMD can be obtained by a straightforward Dirac-Fourier transform as[36]

$$N(\boldsymbol{p}) = \sum_{a,b}^{\text{basis}} \overline{\tilde{X}_a(\boldsymbol{p})} G^{ab} \tilde{X}_b(\boldsymbol{p}), \tag{9}$$

where G^{ab} is the one-particle density matrix (OPDM) in the atomic-orbital basis, $\tilde{X}_a(\boldsymbol{p})$ is the Fourier transform of the ath (atomic) basis function, the overline denotes complex conjugation and the sums over a and b run over all basis functions in the basis set. The OPDM G^{ab} can be formed at any level of theory, be it single- (HF, DFT), or multi-determinantal (post-HF methods such as MP2 and CCSD). With DFT, however, a Lam-Platzman-type correction[18] should be added to eqn (9), as discussed in the Introduction.

Inserting eqn (9) in eqn (4) we see that the task of calculating the CP reduces to evaluating the angular integrals over products of basis functions in the momentum space:

$$N(p) = \sum_{a,b}^{\text{basis}} G^{ab} I_{ab}(p), \tag{10}$$

$$I_{ab}(p) = \int \mathrm{d}\Omega_{\mathbf{p}} \overline{\tilde{X}_a(\boldsymbol{p})} \tilde{X}_b(\boldsymbol{p}). \tag{11}$$

The calculation of eqn (11) in a Gaussian basis involves evaluating integrals of the type

$$\int \mathrm{d}\Omega_{\mathbf{p}} p_x^\lambda p_y^\mu p_z^\nu \mathrm{e}^{-\beta p^2 + i\boldsymbol{pr}}. \tag{12}$$

The calculation of these integrals has been examined before by many authors,[36,39,40] most recently by Thakkar and Sharma[41] who showed that evaluating eqn (12) can be reduced to the evaluation of derivatives of spherical Bessel functions. In a Gaussian basis set where the maximum angular momentum is $l_{\max}$ one needs to calculate the integrals of eqn (12) for the set $\lambda, \mu, \nu \geq 0$: $\lambda + \mu + \nu = 0,\ldots,2l_{\max}$. This is because the angular momentum type of a Gaussian basis function does not change in the Fourier transform, although lower order terms may also be produced. Our calculations use up to I-type basis functions ($l_{\max} = 6$) in sextuple-ζ oxygen. Thus, there are 455 integrals that need to be calculated, although by using symmetry considerations the necessary amount is reduced to 102. We decided to approach the problem by writing the product in eqn (11) in terms of spherical harmonics[36] and use their group-theoretical properties to formulate the integration rule. We discuss the specifics in Appendix A and give here the result

$$N(p) = \sum_l c_l p^{n_l} \mathrm{e}^{-\beta_l p^2} + \sum_l d_l j_{n_l}(p\Delta r_l) p^{m_l} \mathrm{e}^{-\alpha_l p^2} \tag{13}$$

in which c_l and d_l are real numbers, Δr_l, α_l, and β_l positive real numbers, m_l and n_l non-negative integers and j_{nl} is a spherical Bessel function. The first term in eqn (13) comes from the one-center integrals and the second from the two-center integrals.

To obtain the CP itself one needs to perform the radial integration in eqn (3). As the integrand contains special functions (see eqn (13)) and the lower limit of the integral is not fixed, we decided to use an adaptive algorithm based on the Simpson rule, which is presented in Appendix B. The results of the algorithm are very satisfactory: numerical integration over the EMD on the radial grid differs from the number of electrons in the system by as little as $\sim 10^{-10}$ electrons, which is of the same order as the numerical error in the density matrix G^{ab}. Furthermore, the estimated integration error for the CP $J(q)$ in the typically studied range with $q \lesssim 10$ a.u. was $< 10^{-9}$ electrons per a.u. To our knowledge this is the best numerical precision ever reported for CPs.

3. Calculations

We have studied the water and helium dimers with fixed geometries, with the internuclear distance in He_2 being $r_{\text{He–He}} = 3.01$ Å and the O–O distance in water $r_{\text{O–O}} = 3.07$ Å (see Appendix C for the full geometry that was used for the water dimer). The density matrices for these systems were calculated with Gaussian 09.[42] Calculations were performed using HF, Møller–Plesset perturbation theory[43] truncated at the second (MP2) and third order (MP3), the coupled-clusted methods QCISD,[44] CCD[45] and CCSD,[46] and also with DFT using the BLYP,[47,48] B3LYP,[48–50] BHandHLYP, BP86,[47,51] PBE,[52,53] PW91[54–58] and SVWN[1,17,59] XC functionals.

Dunning-type correlation consistent cc-p(C)VXZ, aug-cc-p(C)VXZ and d-aug-cc-pVXZ basis sets,[60–64] where the cardinal number X = D,T,Q,5,6 for water and X = D,T,Q,5 for helium, obtained from the ESML basis set exchange,[65,66] were used in the calculations. The sizes of the used basis sets are given in Table 1.

As was already pointed out in the Introduction, difference Compton profiles (ΔCPs) are often used in the analysis of experimental results. In this study, we examine the ΔCP obtained from the relative difference of the dimer and monomer profiles

$$\Delta J(q) = [J_{\text{dimer}}(q) - 2J_{\text{monomer}}(q)]/J_{\text{dimer}}(0), \tag{14}$$

which can be understood as a measure of the change of the CP due to the bonding in the water dimer, for example. This system makes it possible to study basis set superposition errors (BSSE) in detail, as the counterpoise[67] corrected ΔCP can be straightforwardly defined as

$$\Delta J^{\text{dimer basis}}(q) = [J_{\text{dimer}}(q) - 2J^{\text{dimer basis}}_{\text{monomer}}(q)]/J_{\text{dimer}}(0), \tag{15}$$

where the superscript emphasizes that the monomer calculation is performed using the same basis set as for the dimer. As the basis set size is increased the BSSE is expected to vanish, that is, $\Delta J(q) - \Delta J^{\text{dimer basis}}(q) \to 0$. In the case of the water dimer, there are two nonequivalent possibilities for choosing which molecule is removed in the counterpoise calculation.

Table 1 Sizes of the used basis sets for the helium atom and the water monomer. n_{bas} denotes the number of basis functions and n_{prim} the number of primitive Gaussians

	n_{bas}	n_{prim}
(a) Helium atom		
cc-pVDZ	5	7
cc-pVTZ	14	18
cc-pVQZ	30	38
cc-pV5Z	55	73
aug-cc-pVDZ	9	11
d-aug-cc-pVDZ	13	15
aug-cc-pVTZ	23	28
d-aug-cc-pVTZ	32	38
aug-cc-pVQZ	46	58
d-aug-cc-pVQZ	62	78
aug-cc-pV5Z	80	108
d-aug-cc-pV5Z	105	143
(b) Water monomer		
cc-pVDZ	24	47
cc-pVTZ	58	86
cc-pVQZ	115	163
cc-pV5Z	201	295
cc-pV6Z	322	497
aug-cc-pVDZ	41	65
d-aug-cc-pVDZ	58	83
aug-cc-pVTZ	92	126
d-aug-cc-pVTZ	126	166
aug-cc-pVQZ	172	238
d-aug-cc-pVQZ	229	313
aug-cc-pV5Z	287	421
d-aug-cc-pV5Z	373	547
aug-cc-pV6Z	443	693
d-aug-cc-pV6Z	551	861

Being pedantic, one would calculate both possibilities and take their average. However, we examine the counterpoise calculation only as a means to obtain insight about the magnitude of the basis set effect in the CPs. Thus, for simplicity, we chose to calculate only the first molecule, including the basis functions of the second in the calculation.

Before presenting the results, we review a simple system to acquire some insight on what can be seen in the CP. A model wave function for a simple diatomic molecule in a minimal basis can be written as[6]

$$\psi(\boldsymbol{r}) = (2 \pm 2S)^{-1/2}(\psi_a(\boldsymbol{r}) \pm \psi_a(\boldsymbol{r} - \boldsymbol{R})), \qquad (16)$$

where $\psi_a(\boldsymbol{r})$ is the solution to the one-atom Schrödinger equation and S is the overlap between the basis functions centered at the origin and at $\boldsymbol{R}$. Assuming that $\psi_a(\boldsymbol{r})$ is spherically symmetric, the spherically integrated EMD can be found out to be

$$N(p) = 4\pi|\tilde{\psi}_a(p)|^2(1 \pm j_0(pR))(1 \pm S)^{-1}. \qquad (17)$$

A general version of this equation for the EMD of an N_{el}-electron system modeled using any Gaussian basis set was already given above, in eqn (13). As can be seen from the Bessel function in the equation, the EMD of the dimer exhibits a damped oscillation as a function of p, with a period of $2\pi/R$. This is known as the bond oscillation principle[68,69] and it demonstrates the usefulness of CP as a local structural probe: bonding and antibonding orbitals cause oscillations in the CP with a period that is related to the bond length.

4. Results and discussion

We were able to perform the calculations for helium for all the studied basis sets. In the case of the water dimer, results for the most advanced post-HF methods (coupled-cluster and MP3) were attainable only up to quadruple-ζ basis sets due to their computational demands.

Since the sum rule (eqn (7)) fixes the integral of the CP to the number of electrons, the study of the convergence of the CP is a non-trivial task. If the absolute CP increases at some value of q as the basis set is improved, it must decrease at some other value of q: the convergence of the value of the CP at q does not imply convergence at q'. For completeness we will thus study the convergence for all values of q. However, for illustrative purposes, we start by examining the evolution of $J(0)$ as a function of the basis set in Figs. 1 and 2. The horizontal axis in the figures shows the basis set used in the calculations, the sizes of which were given above in Table 1. The size of the basis set increases monotonically from cc-pVDZ to cc-pV6Z (5Z for helium), is then reduced to aug-cc-pVDZ from which it grows again monotonically to d-aug-cc-pV6Z (5Z for helium).

An earlier study showed that[27] the radial EMD is more affected by the correlation part than the exchange part of the used DFT XC functional. Although the values of $J(0)$ produced by different functionals differ, the basis set dependence is similar for each one of them. This behavior is demonstrated by the B3LYP curves in Figs. 1 and 2. The basis set requirements of HF are much like those of DFT methods. A different type of behavior is seen for the set of post-HF methods, as shown by the MP2 and CCSD results in the same figures.

As can be seen from Figs. 1 and 2, the non-augmented cc-pVXZ basis sets are insufficient for modeling the CP accurately; the value of $J(0)$ changes by several percents as the cardinal number of the basis set is increased. Furthermore, there is a large difference in the CP peak value of the monomer and the corresponding counterpoise calculation, which as well implies the basis set is insufficient. We also find that ΔCPs calculated with the cc-pVXZ basis sets change radically as the basis set grows. The results for the cc-pCVXZ basis sets are unsatisfactory in the same manner. The augmented (d-)aug-cc-p(C)VXZ basis sets give better results than the cc-p(C)VXZ basis sets, with a smaller number of basis functions (see Table 1).

Part of this basis set convergence behavior is certainly due to the extreme nature of the comparison (the isolated dimer *versus* the isolated monomer). Additional calculations show, in fact, that the same kind of phenomenon can be seen by comparing water dimers at different oxygen–oxygen distances r_{O-O}. As the average length of the hydrogen bonds in condensed phase is typically temperature dependent, this kind of modeling can be used to understand difference profiles of the same sample at two different temperatures. This type of calculation has been used in the literature, *e.g.*, in the study of bond geometries.[12] When the intermolecular distance is increased by an appreciable amount, the inclusion of diffuse functions is very important. We demonstrate this effect by increasing the intermolecular distance by 0.5 Å to $r_{O-O} = 3.57$ Å and calculating the difference profile

$$\delta J(q) = \frac{J_{\text{dimer}}^{3.07\,\text{Å}}(q) - J_{\text{dimer}}^{3.57\,\text{Å}}(q)}{J_{\text{dimer}}^{3.07\,\text{Å}}(0)}$$

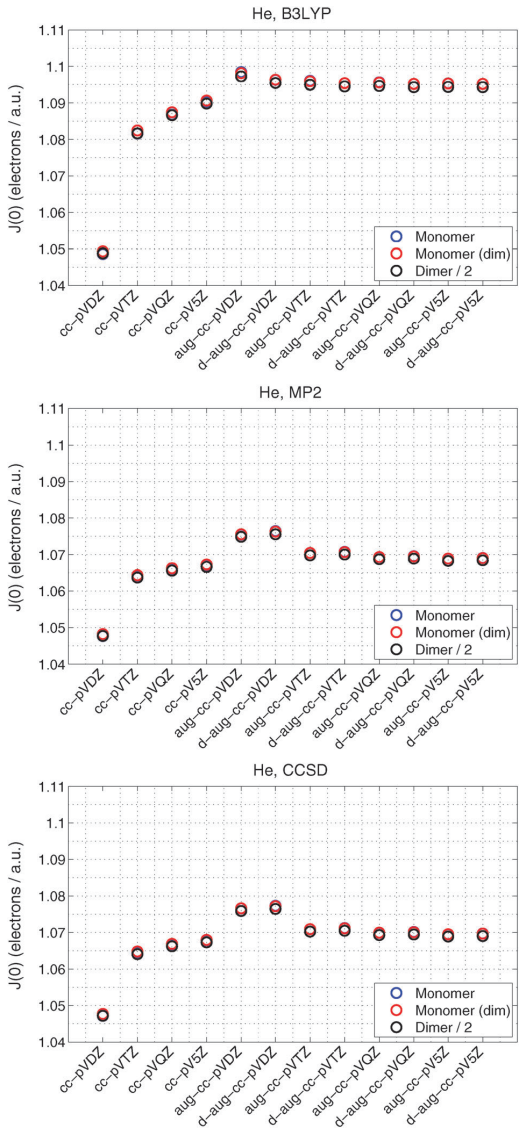


Fig. 1 The evolution of the peak value $J(0)$ of helium for DFT and post-HF methods. "Dimer/2" in the figure indicates that the quantity plotted for the dimer is $J(0)/2$. The notation "(dim)" refers to the dimer basis used in the monomer calculation. The results for the (aug-)cc-pCVXZ are similar to the corresponding (aug-)cc-pVXZ results shown above.

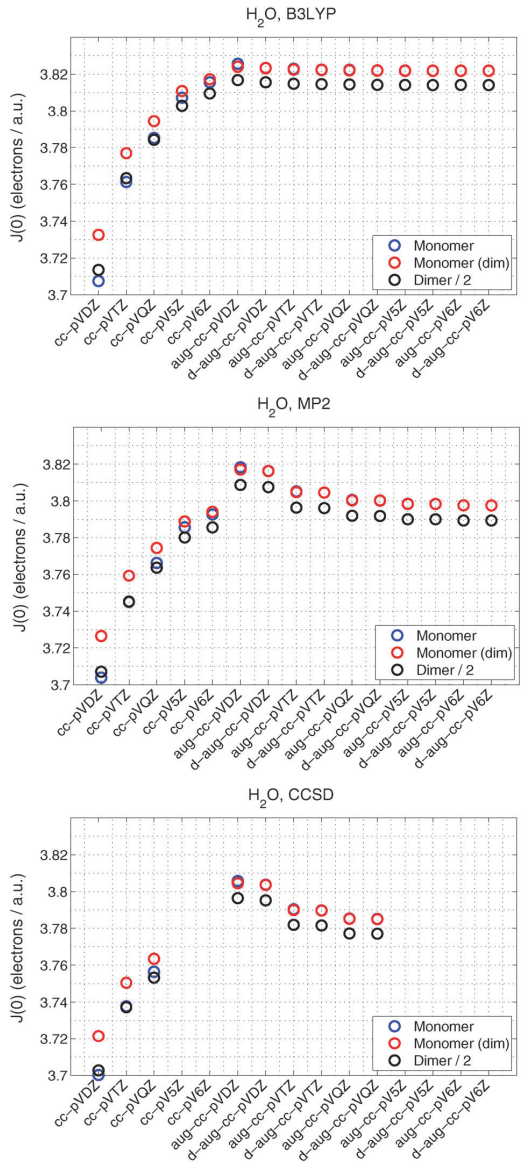


Fig. 2 The evolution of the peak value $J(0)$ of water for DFT and post-HF methods. The notation is the same as in Fig. 1. The results for the (aug-)cc-pCVXZ are similar to the corresponding (aug-)cc-pVXZ results shown above.

as shown by the HF calculation in Fig. 3; DFT and post-HF calculations give similar results. Interestingly, the same kind of basis set behavior as in Fig. 3 is also seen with smaller displacements than the used 0.5 Å, the strength of the signal is just smaller. The maximum absolute value of the ΔCP is 0.17% for the displacement of 0.5 Å, 2.8×10^{-2}% for 0.05 Å and 5.8×10^{-3}% for 0.01 Å. This means that the cc-pVXZ basis sets may yield a closely enough converged result, when the difference between the geometries that are compared is small. What can be furthermore noted is that when gas-phase monomers of different isomers are compared (*e.g.*, *n*-propanol *vs.* isopropanol), the diffuse functions are not as important for the calculation of the ΔCP. Our interpretation for this is that the features of the ΔCP in this case are mainly caused by the differences in intramolecular bonds, which are sufficiently well described by a non-augmented basis set.

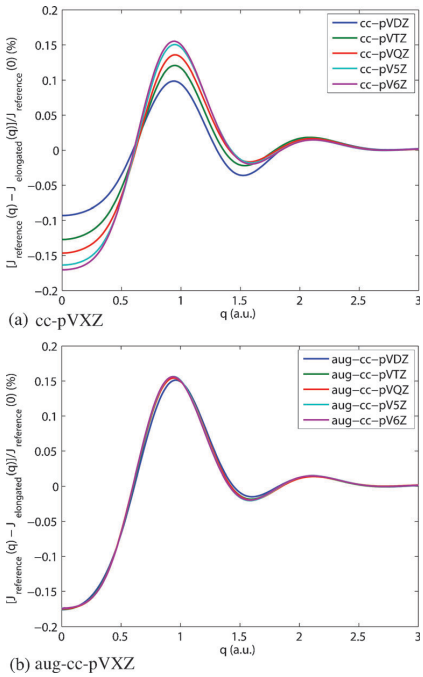


Fig. 3 Difference profile of the water dimer with O–O distance r_{OO} = 3.07 Å *vs.* that with r_{OO} = 3.57 Å, calculated at HF level using different basis sets. The results for the (aug-)cc-pCVXZ are similar to the (aug-)cc-pVXZ results shown above.

Next, we proceed by defining some error metrics for the analysis of the basis set convergence of the absolute and difference CPs. For absolute profiles the error is defined as

$$\sigma_X^J = \max_q |J_X(q) - J_{ref}(q)|/J_{ref}(0) \times 100\%, \qquad (18)$$

where J_X is the profile computed with basis set X and J_{ref} is the reference profile. As reference we use the CP calculated with the best basis set, which is d-aug-cc-pV5Z in the case of helium, and d-aug-cc-pV6Z in that of water with the exception of high-level calculations, for which a d-aug-cc-pVQZ reference is used. For ΔCPs we define the error as

$$\sigma_X^{\Delta J} = \max_q |\Delta J_X(q) - \Delta J_{ref}(q)| \qquad (19)$$

We also define

$$\delta = \max_q |\Delta J_X(q) - \Delta J_{ref}^{\text{dimer basis}}(q)| \qquad (20)$$

as a measure of BSSE in the ΔCP. Note that the error given by eqns (19) and (20) is in %-units, as the ΔCPs are measured in percent. In order to establish a scale for the error in the ΔCP, we use a reference defined by the absolute magnitude of the reference ΔCP

$$\varepsilon = \max_q |\Delta J_{ref}(q)| \qquad (21)$$

which is system-specific and is a limiting factor for both calculations and experiments; the signal must be greater than the noise.

The errors defined by eqns (18) and (19) are shown in Figs. 4 and 5 for the same systems as were shown in Figs. 1 and 2. When diffuse basis sets are used, the absolute Compton profiles converge monotonically as the basis set size is increased. It is also seen from the figures that, as expected, the influence of the counterpoise correction vanishes as the basis set grows. However, it does not diminish monotonically; there is a plateau at triple- and quadruple-ζ level. The sudden reduction

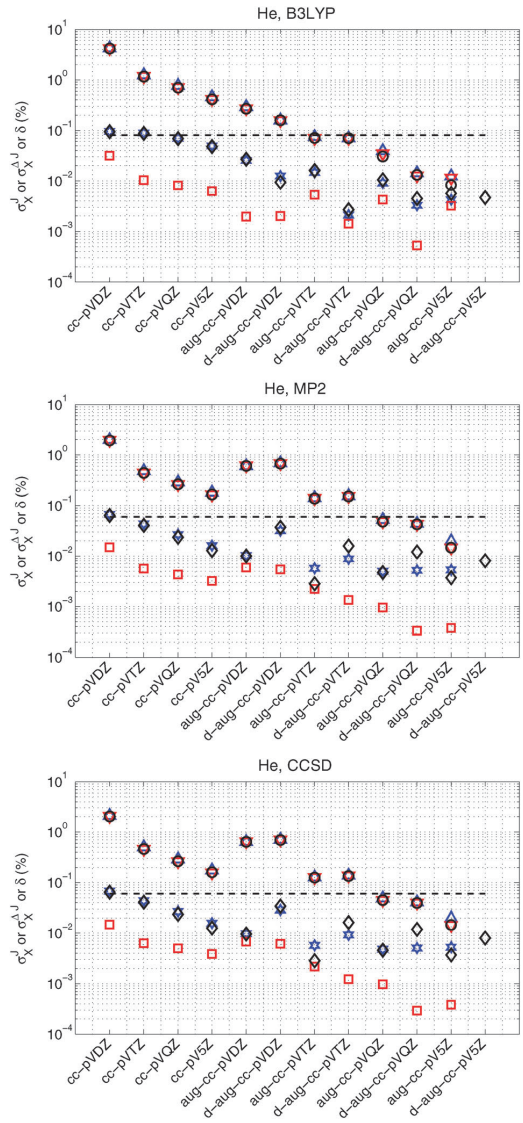


Fig. 4 The convergence of the CPs and ΔCPs of helium. The symbols are: blue upward triangles, red downward triangles and black circles for σ_X^J (eqn (18)) of the monomer, its counterpoise calculation and the dimer, respectively, blue hexagrams and red squares for $\sigma_X^{\Delta J}$ (eqn (19)) and its counterpoise calculation, and black diamonds for δ (eqn (20)). The dashed line represents ε (eqn (21)).

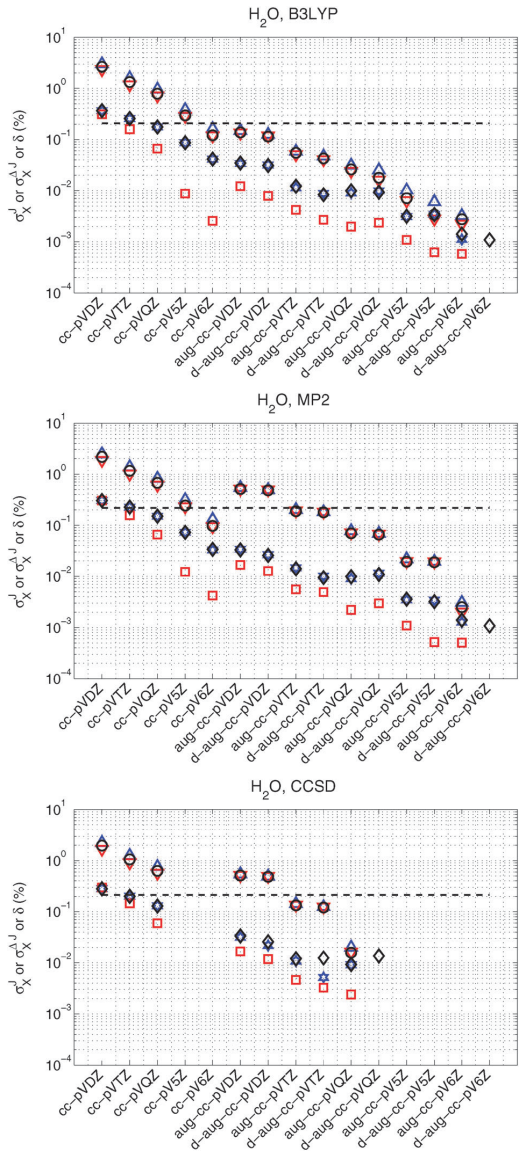


Fig. 5 The convergence of the CPs and ΔCPs of water. The symbols are the same as in Fig. 4.

of the size of the basis set (Table 1) used as the vertical axis can be seen in the post-HF plots, where the error of the absolute profiles jumps up at aug-cc-pVDZ. However, it is rather interesting that the accuracy of the ΔCP stays the same even though the error of the absolute profiles is increased almost an order of magnitude by the reduction of the size of the basis set. In the case of water the aug-cc-pVDZ ΔCP is of similar quality as that of cc-pV6Z, even for post-HF calculations.

A polarized triple-ζ basis has been found sufficient[26] to converge $\langle p^k \rangle$ within 0.1% of the HF limit for $k = -1,1,2,3$, where $\langle p^{-1} \rangle = 2J(0)$ as discussed in section 2. It was found in the present work that the accuracy of the ΔCPs of approximately 0.01%-units is obtained with the aug-cc-pVTZ basis set, with and without counterpoise corrections.

For the simple case of the helium atom it is possible to obtain an estimate of how far one is from the Gaussian basis set limit by comparing our best results (CCSD (d-)aug-cc-pV5Z) to a highly accurate Gaussian geminal calculation.[32] As CCSD corresponds to FCI for two-electron systems such as the helium atom, the difference between the CCSD and the geminal calculation arises purely from basis set effects. We find that the CPs differ by approximately 4.5×10^{-4} electrons per a.u., *i.e.*, roughly 0.04% of $J(0)$ (see Fig. 6). This is a rather large difference as compared to the experimental accuracy. Thus, if absolute profiles are used in future work, one must be extremely careful in choosing the basis set.

In the case of water the basis set requirement for the computation of difference profiles is not completely clear, as high-level calculations with large basis sets were not computationally feasible. Nonetheless, the MP2 calculation does suggest that the aug-cc-pVTZ basis set is sufficient also in the case of water.

Now, as we have determined sufficient convergence with respect to the basis set, we can proceed with the comparison between the results obtained using different levels of theory. We determine the deviation from the CCSD results using the error metrics defined by

$$\delta_{\mathrm{CP}} = \max_q |J^{\mathrm{method}}_{\mathrm{dimer}}(q) - J^{\mathrm{CCSD}}_{\mathrm{dimer}}(q)| / J^{\mathrm{CCSD}}_{\mathrm{dimer}}(0) \qquad (22)$$

for absolute CPs and

$$\delta_{\Delta\mathrm{CP}} = \max_q |\Delta J^{\mathrm{method}}(q) - \Delta J^{\mathrm{CCSD}}(q)| \qquad (23)$$

for difference profiles, both errors being measured in %-units. The notation $J^{\mathrm{method}}(q)$ and $\Delta J^{\mathrm{method}}(q)$ in eqns (22) and (23) stresses that the CP $J(q)$ and ΔCP $\Delta J(q)$ are calculated using DFT, HF and post-HF methods, and the result is compared to the CCSD one. The relevant numbers are shown in Table 2, whereas some of the corresponding ΔCPs are shown in Fig. 7.

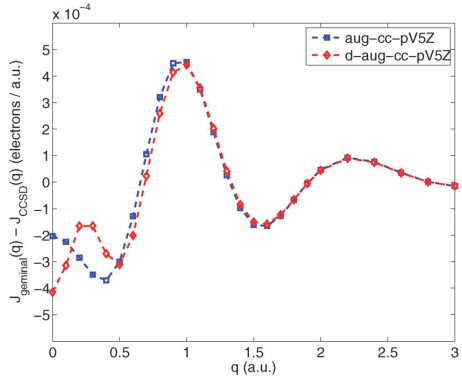


Fig. 6 Deviations of the CCSD (d-)aug-cc-pV5Z helium monomer Compton profile calculations from the accurate Gaussian geminal result. The data points are denoted by circles, with the dashed lines drawn as aids for the eye.

Table 2 Differences of CPs and ΔCPs calculated with DFT, HF and post-HF methods from the CCSD results. The notation is $A(n) = A \times 10^n$. The error measures δ_{CP} and $\delta_{\Delta CP}$ have been defined in eqn (22) and (23), both quantities being given in %-units

	δ_{CP} (dimer)	$\delta_{\Delta CP}$
(a) d-aug-cc-pV5Z helium		
SVWN	4.25	3.89(−2)
PBE	3.17	2.78(−2)
PW91	3.16	5.51(−2)
BP86	2.55	5.80(−2)
BLYP	3.18	3.73(−2)
B3LYP	2.36	1.84(−2)
BHandHLYP	1.40	5.16(−3)
HF	2.22(−1)	4.33(−3)
MP2	6.16(−2)	8.26(−4)
MP3	3.06(−2)	6.15(−4)
CCD	7.54(−3)	1.04(−4)
QCISD	2.13(−6)	5.28(−3)
(b) d-aug-cc-pVQZ water		
SVWN	1.51	9.72(−2)
PBE	1.28	3.86(−2)
PW91	1.25	4.04(−2)
BP86	1.14	3.98(−2)
BLYP	1.45	1.67(−2)
B3LYP	9.86(−1)	1.43(−2)
BHandHLYP	5.00(−1)	7.17(−3)
HF	4.61(−1)	1.98(−2)
MP2	3.91(−1)	1.11(−2)
MP3	8.25(−2)	4.41(−3)
CCD	2.50(−2)	2.12(−3)
QCISD	1.61(−2)	4.26(−4)

First of all, it can be seen that absolute profiles are not reproduced well by DFT. This was expected, as earlier studies have shown that EMD properties are not reproduced by DFT.[23–28] The systematic behavior of the sequence BLYP → B3LYP → BHandHLYP → HF makes it rather tempting to attribute a major part of the DFT error to the lack of a LP-type correction: the more weight is given to HF exchange the smaller the error of the absolute CP becomes, both in the case of water and helium. The error is also of the same order of magnitude as the LP correction in Li, Ne and Ar.[19–21] However, the same kind of systematics with respect to the weight attributed to HF exchange is often seen also in the calculation of other kinds of properties. CCD and QCISD are seen to produce practically CCSD-level results for the ΔCPs, as compared to the experimental precision. A similar observation has been made before in relation to the moments of the EMD.[26]

The helium dimer is clearly problematic for DFT as the magnitude of the DFT ΔCPs is almost twice as big as that produced with CCSD. BHandHLYP is much closer to the reference ΔCP. What is rather striking is that the form of the ΔCPs for the He dimer produced by the DFT functionals PBE, PW91, BLYP and B3LYP is very much different from the *ab initio* result; similar differences have been noticed for radial EMDs.[27] Still, the numerical value of the difference is of the order of 0.02%; differences of the same magnitude can be seen in the case of water, where the magnitude of the ΔCP signal is just larger. HF outperforms all DFT methods (in the absence of an LP-type correction), both for absolute CPs and ΔCPs of He. In the case of water, HF is still superior to DFT methods when absolute profiles are considered. However, in the case of

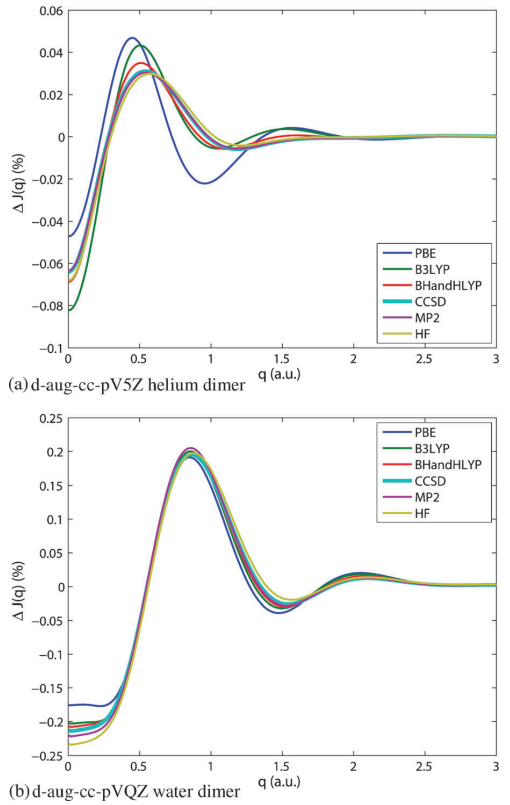


Fig. 7 Difference profiles for the helium and water dimers using eqn (14).

difference profiles BLYP, B3LYP and BHandHLYP fare better than HF. What is rather interesting, is that BHandHLYP outperforms MP2, as well.

It is found that the core region is well enough described by the aug-cc-pVXZ basis sets, as the addition of more functions to the core region of oxygen as in the aug-cc-pCVXZ basis sets did not result in any notable changes in the profiles. The differences in the absolute profiles were of the order 0.01% of the Compton peak height. In the ΔCPs the difference between the results was smaller by an order of magnitude (0.001%-units) and thus of little relevance regarding the analysis of experimental data (although taking core correlation into account may be more important for systems containing second row or heavier atoms). Furthermore, we found that the removal of the higher-than d-type ($l > 2$) basis functions from the basis sets changes the ΔCPs in the case of water by less than 0.01%, which is extremely advantageous as larger systems can be calculated with smaller computational effort.

In this work we have used Dunning-type correlation consistent basis sets for DFT calculations in order to be able to compare them to high level wave function based methods on an equal footing. However, as the basis set requirements of DFT are less demanding than those of high level methods, smaller basis sets, such as the minimally augmented maug-cc-pVXZ[70–72] or

Jensen's aug-pc-N DFT basis sets,[73–78] should be tested in future work for the calculations of ΔCPs using DFT.

5. Conclusions

In this paper we have implemented an adaptive method for the calculation of isotropic CPs and applied it to calculations of the helium and water dimers. We have studied the convergence of the CPs as a function of the basis set used in the calculation, and the level of theory used to form the density matrix.

We have found that the CPs are extremely sensitive to the presence of diffuse basis functions and that basis sets of at least aug-cc-pVTZ quality are necessary to calculate ΔCPs to obtain an accuracy of circa 0.01% of $J(0)$, in comparison to the results obtained with the largest basis set. Hybrid XC functionals were found to produce better results for the CPs and ΔCPs than pure XC functionals, as compared to the CCSD results. We tentatively attribute this behavior to the lack of an LP-type correction in the treatment.

It has been found out in previous work that DFT does not provide an accurate description of EMD properties,[23–28] but also that differences in the EMD may be approximately reproduced by DFT.[24,25,28] Our results demonstrate that it is possible to benchmark DFT XC functionals against post-HF methods by using a sufficiently small system, so that the XC functional that best describes the system can be chosen.

Appendix A

A general Gaussian basis function can be written in the form

$$X_a(\boldsymbol{r}) = \left[\sum_\alpha c_\alpha (x-x_a)^{l_\alpha}(y-y_a)^{m_\alpha}(z-z_a)^{n_\alpha}\right] \times \left[\sum_\beta d_\beta \mathrm{e}^{-\zeta_\beta \boldsymbol{r}^2}\right],$$

in which $\boldsymbol{r}_a = (x_a, y_a, z_a)$ are the coordinates of the center of the a:th basis function. The first bracket contains the angular momentum type of the basis function (a single term when a cartesian basis is used, many terms when spherical harmonics are used) and the second bracket contains the exponential contraction. Because of the Gaussian exponent, the Fourier transform factorizes to

$$\begin{aligned}\tilde{X}_a(\boldsymbol{p}) &= \frac{1}{(2\pi)^{3/2}}\int X_a(\boldsymbol{r})\mathrm{e}^{-i\boldsymbol{p}\boldsymbol{r}}d^3r \\ &= \mathrm{e}^{-i\boldsymbol{p}\boldsymbol{r}_a}\sum_{\alpha,\beta} c_\alpha d_\beta (2\zeta_\beta)^{-(l_\alpha+m_\alpha+n_\alpha+3/2)}\mathscr{R}_{l_\alpha}(p_x;\zeta_\beta) \\ &\quad \times \mathscr{R}_{m_\alpha}(p_y;\zeta_\beta)\mathscr{R}_{n_\alpha}(p_z;\zeta_\beta)\mathrm{e}^{-p^2/4\zeta_\beta}\end{aligned}$$

and the one-dimensional transforms can easily be calculated using the recursion relation[31]

$$\begin{aligned}\mathscr{R}_0(p_i;\zeta) &= 1 \\ \mathscr{R}_1(p_i;\zeta) &= -ip_i \\ \mathscr{R}_l(p_i;\zeta) &= -ip_i\mathscr{R}_{l-1}(p_i;\zeta) + 2\zeta(l-1)\mathscr{R}_{l-2}(p_i;\zeta)\end{aligned}$$

where $p_i = p_x, p_y, p_z$. Now that we have the expression for the Fourier transforms of the basis functions, we can proceed by the evaluation of the integral in eqn (11), which is trivial to do when spherical harmonics are used. In the Condon–Shortley phase convention

$$1 = \sqrt{4\pi}\,Y_0^0 \tag{24}$$

$$p_x = p\sqrt{\frac{2\pi}{3}}\left(Y_1^{-1}(\hat{\boldsymbol{p}}) - Y_1^1(\hat{\boldsymbol{p}})\right) \tag{25}$$

$$p_y = ip\sqrt{\frac{2\pi}{3}}\left(Y_1^{-1}(\hat{\boldsymbol{p}}) + Y_1^1(\hat{\boldsymbol{p}})\right) \tag{26}$$

$$p_z = p\sqrt{\frac{4\pi}{3}}Y_1^0(\hat{\boldsymbol{p}}) \tag{27}$$

where Y_l^m are (complex) spherical harmonics and $\hat{\boldsymbol{p}}$ is the unit vector in the direction of $\boldsymbol{p}$. The spherical harmonics expansions of the polynomials $p_x^\lambda p_y^\mu p_z^\nu$ (12) can be calculated using the group-theoretical properties:[79]

$$Y_{l_1}^{m_1}(\hat{\boldsymbol{p}})\,Y_{l_2}^{m2}(\hat{\boldsymbol{p}}) = \sum_{l=|l_1-l_2|}^{l_1+l_2}\sum_{m=-1}^{1} G_{l_1 l_2, m}^{m_1 m_2, l}\,Y_l^m(\hat{\boldsymbol{p}}),$$

in which $G_{l_1 l_2, m}^{m_1 m_2, l}$ is the Gaunt coefficient

$$\begin{aligned}G_{l_1 l_2, m}^{m_1 m_2, l} &= (-1)^{-m}\sqrt{\frac{(2l+1)(2l_1+1)(2l_2+1)}{4\pi}} \\ &\quad \times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & -m \end{pmatrix}\begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix}\end{aligned}$$

and $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ are Wigner 3-j symbols. Now we are able to expand the product in eqn (11)

$$\overline{\tilde{X}_a}(\boldsymbol{p})\tilde{X}_b(\boldsymbol{p}) = \mathrm{e}^{i\boldsymbol{p}\cdot(\boldsymbol{r}_a-\boldsymbol{r}_b)}\sum_{l,m} C_{ab,m}^l(p)\,Y_l^m(\hat{\boldsymbol{p}}),$$

where $C_{ab,m}^l$ is a complex expansion coefficient. Using the spherical wave expansion of plane waves for the case $\boldsymbol{p}\cdot\boldsymbol{x} \neq 0$

$$\mathrm{e}^{i\boldsymbol{p}\cdot\boldsymbol{x}} = 4\pi\sum_{l=0}^{\infty} i^l j_l(px)\sum_{m=-l}^{l}\overline{Y_l^m(\hat{\boldsymbol{p}})}\,Y_l^m(\hat{\boldsymbol{x}})$$

we can integrate over the angles in eqn (11), obtaining the result

$$I_{ab}(p) = \begin{cases} \dfrac{\sqrt{\pi}}{2}C_{ab,0}^0(p), \text{ when } \Delta\boldsymbol{r} = \boldsymbol{0} \text{ or } p = 0 \\ 4\pi\sum_l i^l j_l(p\Delta r)\sum_m C_{ab,m}^l(p)\,Y_l^m(\widehat{\Delta\boldsymbol{r}}) \end{cases}$$

where $\Delta\boldsymbol{r} = \boldsymbol{r}_a - \boldsymbol{r}_b$. Performing the contractions in eqn (10) we end up with the final result as was already given in section 2

$$N(p) = \sum_l c_l j_{n_l}(p\Delta r_l)p^{m_l}\mathrm{e}^{-\alpha_l p^2} + \sum_l d_l p^{n_l}\mathrm{e}^{-\beta_l p^2},\ p>0 \tag{28}$$

in which c_l and d_l are real numbers, Δr_l, α_l, and β_l positive numbers and m_l and n_l non-negative integers. Although it was assumed in the calculation that $p \neq 0$, it is clear from eqn (28) that it is also valid for $p = 0$.

The important thing to note here is that the integration rule of eqn (28) to be calculated only once. Also, the expansion coefficients only depend on the type of the basis function (not on the coordinates of its center). Since in typical calculations the number of elements (types of basis functions) is much

smaller than the amount of atoms, the formation of the integration rule is rather fast. This method is applicable to any Gaussian basis set. The only requirement for the implementation is that the values of the Bessel functions and the 3*j* symbols can be calculated. The method can also be straightforwardly extended to calculate the integrals in a Slater-type orbital basis.

We have used the GNU Scientific Library[80] for calculating the values of the Bessel functions and the Wigner coefficients and C++ to implement the algorithm. Copies of the program are available upon request.

Appendix B

We have used an adaptive method for calculating the radial integral in eqn (3), which minimizes the necessary number of radial points (angular integrals) needed to obtain the wanted precision for the Compton profile. The method is based on a combined use of the 3-point and 5-point Simpson rules

$$I_3 = \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \tag{29}$$

$$I_5 = \frac{b-a}{12}\left[f(a) + 4f\left(\frac{3a+b}{4}\right) + 2f\left(\frac{a+b}{2}\right) + 4f\left(\frac{a+3b}{4}\right) + f(b)\right] \tag{30}$$

in which $f(x)$ is the integrand and a and b are the lower and upper limits of integration, respectively. The integration error of the 3-point rule is[81]

$$\left|\int_a^b f(x)\mathrm{d}x - I_3\right| \leq \frac{(b-a)^5}{2880}|f^{(4)}(\eta)|,\ \eta \in [a,b],$$

where $f^{(4)}$ denotes the fourth derivative of f. As the 5-point rule (eqn (30)) is in fact the 3-point rule (eqn (29)) on a finer grid, the integration error can be estimated as

$$\left|\int_a^b f(x)dx - I_5\right| \leq |I_3 - I_5|/15 = \varepsilon. \tag{31}$$

This estimate can be used for adapting the integration grid. The algorithm we have used for the calculation of the Compton profile is as follows:

(1) Fill out an initial grid from $p = 0$ to p: $I(p) = 0$ by adding 4 points each time to the end of the grid, using a constant spacing.

(2) Evaluate the charge in the system by calculating

$$Q = \langle p^0 \rangle = \int_0^\infty p^2 N(p)\mathrm{d}p \tag{32}$$

on the grid, using eqn (29)–(31).

(3) If the charge obtained from eqn (32) differs by more than the tolerance from the number of electrons in the system, add 4 more points to the interval with the largest error estimate ε and return to step 2.

(4) To make sure that the integration grid is fine enough in regions where the EMD changes, repeat the procedure to converge the physical moments[82,83] of the EMD

$$\int_0^\infty p^{n+2}N(p)dp,\ n = -2,\ldots,4$$

by adding more points to the area of maximum estimated error, until the relative error is smaller than the tolerance.

(5) The calculation has converged. Calculate the CP.

(6) Finally, to make the comparison of profiles possible in spite of the use of an adaptive grid, interpolate the CP and its error on a uniform grid.

In this paper we have used spline interpolation, which seems to be well suited for the purpose, as the difference profiles obtained with it are smooth. Linear interpolation was found not to be enough, as unphysical spikes were seen in the differences. The spline interpolation error can be assumed small, as the grid produced by the adaptive method above is quite dense in the interval used for the interpolation.

Table 3 The used geometry for the water dimer

Atom	x/Å	y/Å	z/Å
O	−1.4838	0.0065	0.0
H	−2.0682	0.7438	0.0
H	−2.0214	−0.7659	0.0
O	1.5849	−0.0550	0.0
H	1.8451	0.8481	0.0
H	0.6404	−0.0570	0.0

Appendix C

The used geometry for the water dimer is given in Table 3.

Acknowledgements

J.L. thanks the Jenny and Antti Wihuri foundation for financial support. The authors thank the Center for Scientific Computing (Espoo, Finland) for providing computational resources. This work has been supported by the Academy of Finland [Contract 1127462 and the Centers of Excellence Program (2006-2011)] and the Research Funds of the University of Helsinki.

References

1 P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev. Sect. B*, 1964, **136**, B864–B871.
2 P. Coppens, *X-Ray Charge Densities and Chemical Bonding*, Oxford University Press, 1997.
3 K. B. Wiberg, C. M. Hadad, T. J. LePage, C. M. Breneman and M. J. Frisch, Analysis of the effect of electron correlation on charge density distributions, *J. Phys. Chem.*, 1992, **96**, 671–679.
4 M. K. Harbola, R. R. Zope, A. Kshirsagar and R. K. Pathak, Momentum-space properties from coordinate-space electron density, *J. Chem. Phys.*, 2005, **122**, 204110.
5 M. J. Cooper, P. E. Mijnarends, N. Shiotani, N. Sakai and A. Bansil, *X-ray Compton Scattering*, volume 5 of Oxford Series on Synchrotron Radiation, Oxford University Press, 2004.
6 M. J. Cooper, Compton scattering and electron momentum determination, *Rep. Prog. Phys.*, 1985, **48**, 415–482.
7 P. Eisenberger and P. M. Platzman, Compton scattering of X rays from bound electrons, *Phys. Rev. A*, 1970, **2**, 415–423.
8 E. D. Isaacs, A. Shukla, P. M. Platzman, D. R. Hamann, B. Barbiellini and C. A. Tulk, Compton scattering evidence for covalency of the hydrogen bond in ice, *J. Phys. Chem. Solids*, 2000, **61**, 403–406.
9 B. Barbiellini, Ch. Bellin, G. Loupias, T. Buslaps and A. Shukla, How the hydrogen bond in NH_4F is revealed with Compton scattering, *Phys. Rev. B: Condens. Matter*, 2009, **79**, 155115.

10 B. Barbiellini, A. Koizumi, P. E. Mijnarends, W. Al-Sawai, H. Lin, T. Nagao, K. Hirota, M. Itou, Y. Sakurai and A. Bansil, Role of oxygen electrons in the metal-insulator transition in the magnetoresistive oxide $La_{2-2x}Sr_{1+2x}Mn_2O_7$ probed by Compton scattering, *Phys. Rev. Lett.*, 2009, **102**(20), 206402.
11 M. Hakala, K. Nygård, S. Manninen, S. Huotari, T. Buslaps, A. Nilsson, L. G. M. Pettersson and K. Hämäläinen, Correlation of hydrogen bond lengths and angles in liquid water based on Compton scattering, *J. Chem. Phys.*, 2006, **125**, 084504.
12 M. Hakala, K. Nygård, S. Manninen, L. G. M. Pettersson and K. Hämäläinen, Intra- and intermolecular effects in the Compton profile of water, *Phys. Rev. B: Condens. Matter*, 2006, **73**, 035432.
13 K. Nygård, M. Hakala, T. Pylkkänen, S. Manninen, T. Buslaps, M. Itou, A. Andrejczuk, Y. Sakurai, M. Odelius and K. Hämäläinen, Isotope quantum effects in the electron momentum density of water, *J. Chem. Phys.*, 2007, **126**, 154508.
14 K. Nygård, M. Hakala, S. Manninen, A. Andrejczuk, M. Itou, Y. Sakurai, L. G. M. Pettersson and K. Hämäläinen, Compton scattering study of water *versus* ice I*h*: Intra- and intermolecular structure, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 031503.
15 K. Nygård, M. Hakala, S. Manninen, M. Itou, Y. Sakurai and K. Hämäläinen, Configurational energetics in ice I*h* probed by Compton scattering, *Phys. Rev. Lett.*, 2007, **99**, 197401.
16 M. Hakala, K. Nygård, J. Vaara, M. Itou, Y. Sakurai and K. Hämäläinen, Charge localization in alcohol isomers studied by Compton scattering, *J. Chem. Phys.*, 2009, **130**, 034506.
17 W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev. Sect. A*, 1965, **140**, A1133–A1138.
18 L. Lam and P. M. Platzman, Momentum density and Compton profile of the inhomogeneous interacting electronic system. I. Formalism, *Phys. Rev. B*, 1974, **9**, 5122–5127.
19 B. Y. Tong and L. Lam, Compton profiles of Ne, Ar, and Kr, *Phys. Rev. A*, 1978, **18**, 552–558.
20 A. Harmalkar, P. V. Panat and D. G. Kanhere, Compton profiles for neon and argon from Xα wavefunctions, *J. Phys. B: At. Mol. Phys.*, 1980, **13**, 3075–3080.
21 B. Barbiellini and A. Bansil, Treatment of correlation effects in electron momentum density: density functional theory and beyond, *J. Phys. Chem. Solids*, 2001, **62**, 2181–2189.
22 A. A. Jarzęcki and E. R. Davidson, Kinetic and potential energy of isoelectronic atomic ions from density functional theory compared with exact values, *Mol. Phys.*, 2000, **98**, 1089–1097.
23 S. Ragot, Exact Kohn-Sham *versus* Hartree-Fock in momentum space: Examples of two-fermion systems, *J. Chem. Phys.*, 2006, **125**, 014106.
24 T. K. Ghanty, V. N. Staroverov, P. R. Koren and E. R. Davidson, Is the hydrogen bond in water dimer and ice covalent?, *J. Am. Chem. Soc.*, 2000, **122**, 1210–1214.
25 M. Hakala, S. Huotari, K. Hämäläinen, S. Manninen, Ph. Wernet, A. Nilsson and L. G. M. Pettersson, Compton profiles for water and mixed water-neon clusters: a measure of coordination, *Phys. Rev. B: Condens. Matter*, 2004, **70**, 125413.
26 J. R. Hart and A. J. Thakkar, Moments of the electron momentum density: Requirements for *ab initio* and density functional theory calculations, *Int. J. Quantum Chem.*, 2005, **102**, 673–683.
27 B. Miguel and J. M. García de la Vega, Influence of electronic correlation in monoelectronic density in p-space, *Theor. Chem. Acc.*, 2007, **118**, 723–732.
28 A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz and D. Usvyat, MP2 *versus* density-functional theory study of the Compton profiles of crystalline urea, *Phys. Rev. B: Condens. Matter*, 2010, **81**, 165108.
29 P. Eisenberger, Electron momentum density of He and H_2; Compton X-ray scattering, *Phys. Rev. A*, 1970, **2**, 1678–1686.
30 M. J. Cooper, M. Roux, M. Cornille and B. Tsapline, The Compton profile of water, *Philos. Mag.*, 1968, **18**, 309–312.
31 L. C. Snyder and T. A. Weber, The Compton profile of water: Computed from an SCF–MO wavefunction in a double-zeta Gaussian basis set, *J. Chem. Phys.*, 1975, **63**, 113–114.
32 P. E. Regier and A. J. Thakkar, Momentum space properties of two-electron atoms, *J. Phys. B: At. Mol. Phys.*, 1985, **18**, 3061.
33 P. H.-L. Sit, C. Bellin, B. Barbiellini, D. Testemale, J.-L. Hazemann, T. Buslaps, N. Marzari and A. Shukla, Hydrogen bonding and coordination in normal and supercritical water from X-ray inelastic scattering, *Phys. Rev. B: Condens. Matter*, 2007, **76**, 245413.
34 S. Huotari, *Electronic structure of matter studied by Compton scattering*, PhD thesis, University of Helsinki, 2002.
35 K. Nygård, M. Hakala, S. Manninen, K. Hämäläinen, M. Itou, A. Andrejczuk and Y. Sakurai, Ion hydration studied by X-ray Compton scattering, *Phys. Rev. B: Condens. Matter*, 2006, **73**, 024208.
36 P. Kaijser and V. H. Smith, Evaluation of momentum distributions and Compton profiles for atomic and molecular systems, *Adv. Quantum Chem.*, 1977, **10**, 37–75.
37 I. R. Epstein, Calculation of atomic and molecular momentum expectation values and total energies from Compton-scattering data, *Phys. Rev. A*, 1973, **8**, 160–168.
38 A. J. Thakkar, A. M. Simas and V. H. Smith, Extraction of momentum expectation values from Compton profiles, *Mol. Phys.*, 1980, **41**, 1153–1162.
39 J. Wang and V. H. Smith, Evaluation of cross sections for X-ray and high-energy electron scattering from molecular systems, *Int. J. Quantum Chem.*, 1994, **52**, 1145–1151.
40 J. Wang and V. H. Smith, Spherically averaged molecular electron densities and radial moments in position and momentum spaces, *J. Phys. B: At., Mol. Opt. Phys.*, 1994, **27**, 5159–5173.
41 A. J. Thakkar and B. S. Sharma, A fresh look at the computation of spherically averaged electron momentum densities for wave functions built from Gaussian-type functions, *Int. J. Quantum Chem.*, 2001, **85**, 258–262.
42 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven Jr., J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN 09 (Revision A.1.)*, Gaussian, Inc., Wallingford, CT, 2009.
43 C. Møller and M. S. Plesset, Note on an approximation treatment for many-electron systems, *Phys. Rev.*, 1934, **46**, 618–622.
44 J. A. Pople, M. Head-Gordon and K. Raghavachari, Quadratic configuration interaction, A general technique for determining electron correlation energies, *J. Chem. Phys.*, 1987, **87**, 5968–5975.
45 J. Cizek, On the correlation problem in atomic, molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods, *J. Chem. Phys.*, 1966, **45**, 4256–4266.
46 G. D. Purvis III and R. J. Bartlett, A full coupled-cluster singles and doubles model: The inclusion of disconnected triples, *J. Chem. Phys.*, 1982, **76**, 1910–1918.
47 A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A*, 1988, **38**, 3098–3100.
48 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: Condens. Matter*, 1988, **37**, 785–789.
49 A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
50 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *Ab initio* calculation of vibrational absorption and circular dichroism spectra using density functional force fields, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
51 J. P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas, *Phys. Rev. B: Condens. Matter*, 1986, **33**, 8822–8824.
52 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

53 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1997, **78**, 1396.
54 J. P. Perdew, *Electronic Structure of Solids '91*, Akademie Verlag, Berlin, 1991.
55 J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B: Condens. Matter*, 1992, **46**, 6671–6687.
56 J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B: Condens. Matter*, 1993, **48**, 4978.
57 J. P. Perdew, K. Burke and Y. Wang, Generalized gradient approximation for the exchange-correlation hole of a many-electron system, *Phys. Rev. B: Condens. Matter*, 1996, **54**, 16533–16539.
58 K. Burke, J. P. Perdew and Y. Wang, *Electronic Density Functional Theory: Recent Progress and New Directions*, Plenum, 1998.
59 J. C. Slater, *The Self-Consistent Field for Molecules and Solids*, volume 4 of Quantum Theory of Molecules and Solids, McGraw-Hill, New York, 1974.
60 T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
61 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.*, 1992, **96**, 6796–6806.
62 K. A. Peterson, D. E. Woon and T. H. Dunning, Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $H + H_2 \rightarrow H_2 + H$ reaction, *J. Chem. Phys.*, 1994, **100**, 7410–7415.
63 A. K. Wilson, T. van Mourik and T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets for boron through neon, *THEOCHEM*, 1996, **388**, 339–349.
64 D. E. Woon and T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon, *J. Chem. Phys.*, 1995, **103**, 4572–4585.
65 D. Feller, The role of databases in support of computational chemistry calculations, *J. Comput. Chem.*, 1996, **17**, 1571–1586.
66 K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T.L. Windus, Basis set exchange: A community database for computational sciences, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.
67 S. F. Boys and F. Bernardi, The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors, *Mol. Phys.*, 1970, **19**, 553–566.
68 I. R. Epstein and A. C. Tanner, *Chemistry* in *Compton scattering*, ed. B. Williams, McGraw-Hill, London, 1977, p. 209.
69 A. C. Tanner, The bond directional principle for momentum space wavefunctions: comments and cautions, *Chem. Phys.*, 1988, **123**, 241–247.
70 E. Papajak, H. R. Leverentz, J. Zheng and D. G. Truhlar, Efficient diffuse basis sets: cc-pVxZ + and maug-cc-pVxZ, *J. Chem. Theory Comput.*, 2009, **5**, 1197–1202.
71 E. Papajak, H. R. Leverentz, J. Zheng and D. G. Truhlar, Erratum: Efficient diffuse basis sets: cc-pVxZ + and maug-cc-pVxZ, *J. Chem. Theory Comput.*, 2009, **5**, 3330.
72 E. Papajak and D. G. Truhlar, Efficient diffuse basis sets for density functional theory, *J. Chem. Theory Comput.*, 2010, **6**, 597–601.
73 F. Jensen, Polarization consistent basis sets: Principles, *J. Chem. Phys.*, 2001, **115**, 9113–9125.
74 F. Jensen, Erratum: "Polarization consistent basis sets: Principles" [*J. Chem. Phys.*, 2001, **115**, 9113], *J. Chem. Phys.*, 2002, **116**, 3502.
75 F. Jensen, Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit, *J. Chem. Phys.*, 2002, **116**, 7372–7379.
76 F. Jensen, Polarization consistent basis sets. III. The importance of diffuse functions, *J. Chem. Phys.*, 2002, **117**, 9234–9240.
77 F. Jensen and T. Helgaker, Polarization consistent basis sets. V. The elements Si–Cl, *J. Chem. Phys.*, 2004, **121**, 3463–3470.
78 F. Jensen, Polarization consistent basis sets 4: The elements He, Li, Be, B, Ne, Na, Mg, Al, and Ar, *J. Phys. Chem. A*, 2007, **111**, 11198–11204, PMID: 17439196.
79 J. J. Sakurai, *Modern Quantum Mechanics*, Addison-Wesley Publishing Company, Inc., revised edn, 1994.
80 M. Galassi, *et al.*, *GNU Scientific Library Reference Manual*, 3rd edn, 2009.
81 K. A. Atkinson, *An Introduction to Numerical Analysis*, John Wiley & Sons, 2nd edn, 1989.
82 A. J. Thakkar, The momentum density perspective of the electronic structure of atoms and molecules, *Adv. Chem. Phys.*, 2004, **128**, 303–352.
83 A. J. Thakkar, Asymptotic behavior of atomic momentals, *J. Chem. Phys.*, 1987, **86**, 5060–5062.

Paper II

Completeness-optimized basis sets: Application to ground-state electron momentum densities
Jussi Lehtola, Pekka Manninen, Mikko Hakala, and Keijo Hämäläinen
Journal of Chemical Physics 137, 104105 (2012)
doi:10.1063/1.4749272

Completeness-optimized basis sets: Application to ground-state electron momentum densities

Jussi Lehtola,[1,a)] Pekka Manninen,[2] Mikko Hakala,[1] and Keijo Hämäläinen[1]
[1]*Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[2]*CSC, IT Center for Science Ltd., P.O. Box 405, FI-02101 Espoo, Finland*

(Received 4 June 2012; accepted 17 August 2012; published online 12 September 2012)

In the current work we apply the completeness-optimization paradigm [P. Manninen and J. Vaara, J. Comput. Chem. **27**, 434 (2006)] to investigate the basis set convergence of the moments of the ground-state electron momentum density at the self-consistent field level of theory. We present a black-box completeness-optimization algorithm that can be used to generate computationally efficient basis sets for computing any property at any level of theory. We show that the complete basis set (CBS) limit of the moments of the electron momentum density can be reached more cost effectively using completeness-optimized basis sets than using conventional, energy-optimized Gaussian basis sets. By using the established CBS limits, we generate a series of smaller basis sets which can be used to systematically approach the CBS and to perform calculations on larger, experimentally interesting systems. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4749272]

I. INTRODUCTION

Conventional, commonly used energy-optimized basis sets, such as the Dunning-style correlation consistent *cc-pVXZ*[1–4] or Jensen's polarization-consistent *pc-N* series,[5–9] may not be optimal for computing other properties than the energy. For instance, although extremely important for the dipole moment, diffuse functions have very little effect on the energy and thus do not appear automatically in energy-optimized basis sets (unless extremely large basis sets are used). Instead, the necessary additional diffuse functions need to be determined manually for forming efficient basis sets capable of accurately computing properties such as the dipole moment.

Furthermore, whereas the error in the energy is second order in the error in wave function, errors in one-electron properties such as the dipole moment are only first order and are thus much more sensitive to deficiencies in the wave function. Also because of this, in contrast to, e.g., the Hartree-Fock (HF) energy, many experimentally relevant properties are not variational with respect to the basis set, making the application of conventional basis set optimization methods unattractive for generating basis sets tuned for particular properties.

Completeness optimization[10] was introduced in order to approach the complete basis set (CBS) limit of also non-variational properties in a systematic fashion. In contrast to other basis sets where the exponents are typically fine-tuned to produce optimal energies for the element in question, in the completeness-optimization paradigm the exact values of the exponents are of no interest—the only thing that matters is that they span the relevant part of the exponent space, in the necessary precision, to accurately procure the property being modeled. Although so far only applied for magnetic properties,[10–13] completeness optimization is a completely general, black-box method that can be used for any property at any level of theory. Current applications have shown the resulting basis sets to be computationally more efficient than conventional, energy-optimized basis sets.[10–13]

Despite the extending use of ground-state (GS) electron momentum density (EMD)[14] methods in materials science, there has been surprisingly little research into the basis set requirements for its accurate modeling. In the current work we employ completeness-optimization to calculations of the GS-EMD, which is a first-order property highly sensitive to diffuse functions. Introducing a novel, computerized optimization algorithm, we investigate the basis set convergence of the moments of the GS-EMD at the self-consistent field (SCF) level of theory. We demonstrate that the CBS limit can be easily reached using completeness-optimized Gaussian type orbital (GTO) basis sets, as shown by the comparison to near HF values produced with Slater type orbital (STO) basis sets. Furthermore, we generate cost efficient basis sets for the computation of the moments of the GS-EMD at the SCF level of theory for the first three rows of the periodic table, i.e., the elements H–Ar.

Classically equivalent to the velocity distribution of the electrons, the GS-EMD is a computationally interesting quantity as it can be measured by using $(e, 2e)$ spectroscopy,[15] x-ray Compton scattering,[16,17] or positron annihilation spectroscopy. The EMD $n(\mathbf{p})$ can be calculated in any atom-centered basis set as[18]

$$n(\mathbf{p}) = \sum_{\mu\nu} P_{\mu\nu} \overline{\tilde{\chi}_\mu(\mathbf{p})} \tilde{\chi}_\nu(\mathbf{p}), \tag{1}$$

where $\mathbf{p}$ is the momentum, $P_{\mu\nu}$ is the one-particle density matrix, $\tilde{\chi}_\mu$ is the Fourier transform of the μth basis function, and the overline denotes complex conjugation. A similar expression to Eq. (1) applies in periodic systems.[19]

X-ray Compton scattering experiments are usually performed using highly energetic x-rays at modern synchrotron radiation sources. When the impulse approximation[20] holds,

a)Electronic mail: jussi.lehtola@helsinki.fi.

the resulting differential cross section for Compton scattered photons is found out to be[16,17]

$$\frac{d^2\sigma}{d\Omega dE_2} = C(E_1, E_2, \phi)J(p_q), \tag{2}$$

where $d^2\sigma/d\Omega dE_2$ measures the amount of photons detected in the solid angle $d\Omega$ with the energy in the range $E_2, \ldots, E_2 + dE_2$, E_1 is the energy of the incoming photon, ϕ is the scattering angle, and $C(E_1, E_2, \phi)$ only depends on the setup of the experiment. The function $J(p_q)$ is the Compton profile, which measures the projection of the EMD along the scattering vector $\mathbf{q}$,[16]

$$J\left(p_q\right) = \iint n\left(p_x, p_y, p_z \equiv p_q\right) dp_x dp_y. \tag{3}$$

By measuring the Compton profiles from single crystal samples and aligning the scattering vector along certain crystallographic directions it is possible to obtain anisotropic information about the EMD.[21,22] However, in the rest of the paper, we will discuss only isotropic systems, in which the relevant quantities are the radial EMD

$$n\left(p\right) = \int n\left(\mathbf{p}\right) d\Omega_{\mathbf{p}} \tag{4}$$

and the isotropic Compton profile

$$J(p_q) = \frac{1}{2}\int_{|p_q|}^{\infty} pn\left(p\right) dp. \tag{5}$$

The radial EMD can be characterized using its moments

$$\langle p^k \rangle = \int_0^{\infty} p^{k+2} n\left(p\right) dp, \ -2 \leq k \leq 4. \tag{6}$$

The k range is limited as the moments with other values of k diverge with the exact momentum density.[14,23]

Even though absolute Compton profiles can be measured,[24–26] differences of Compton profiles (so called difference profiles) are often used for studying the structure of materials. Recent studies include covalent bonding in liquid silicon,[27] oxygen disorder in ice,[28] doped holes in cuprate superconductors,[29] solvation in water-ethanol solutions,[30] and energetics in tetrahydrofuran clathrate.[31] While x-ray Compton scattering within the impulse approximation yields collective GS-EMD properties, $(e, 2e)$ scattering can be used to extract the GS-EMD of individual orbitals.[15] Recent $(e, 2e)$ studies include the orbital momentum profiles of naphthalene,[32] isobutane[33] and n-butane,[34] as well as the study of the conformers of n-propanol.[35]

Near HF results using large STO basis sets have been reported in the literature for diatomic and linear molecules.[36,37] For diatomics a numerical solution of the HF equations is possible as well, allowing one to obtain results directly at the HF limit.[38] As usual, calculations on more complicated molecules and post-HF studies for investigating the effect of correlation have been performed using GTO basis sets.[39–41] Also crystalline systems can be studied at the HF and post-HF levels using GTOs.[42–45] While correlation effects have been shown to be important, relativistic effects are insignificant for the first three rows of the periodic table.[26,46]

TABLE I. Deviation of the CCSD *d-aug-cc-pV5Z* moments of the helium atom[47] from the 40 Gaussian geminal (40 GG) calculation from Ref. 49.

	CCSD	40 GG	Error (%)
$\langle p^{-2} \rangle$	4.1056	4.0986	0.17
$\langle p^{-1} \rangle$	2.1394	2.1386	0.04
$\langle p \rangle$	2.8140	2.8146	− 0.02
$\langle p^2 \rangle$	5.8050	5.8074	− 0.04
$\langle p^3 \rangle$	18.387	18.406	− 0.10
$\langle p^4 \rangle$	105.06	106.56	− 1.41

GTO basis set requirements for the calculation of GS-EMD properties of gas-phase molecules have been discussed recently by Hart and Thakkar[41] and Lehtola *et al.*[47] The former studied basis set convergence of the moments of the GS-EMD at HF level of theory using a pruned version of Jensen's augmented triple-ζ polarization-consistent basis set (*aug-pc-3*)[5,7] as the reference, and drew conclusions about the accuracy and necessary basis sets by comparing the results obtained with the reference basis set to those obtained with smaller basis sets. In the latter study the basis set convergence of the isotropic Compton profile (Eq. (5)) was investigated at HF, density-functional theory (DFT), Møller-Plesset perturbation theory[48] truncated at the second order (MP2), and coupled-cluster singles and doubles (CCSD) levels of theory using the Dunning-style basis set series. A notable difference was found in the Compton profiles for the helium atom produced by the CCSD calculation using the doubly augmented quintuple-ζ correlation consistent (*d-aug-cc-pV5Z*) basis set[1,3] and a highly accurate Gaussian geminal based calculation.[49] The difference was found to be of the order of 5×10^{-4} a.u. (corresponding to ∼0.05% of $J(0)$), while an accuracy of 0.02% of $J(0)$ can often be achieved for difference Compton profiles in experiment.[50–54] Rather large differences can also be seen in the moments of the GS-EMD, as shown in Table I; this was the original motivation for this work.

The organization of the paper is the following. In Sec. II, we discuss completeness optimization in more detail and suggest an automatized algorithm for the generation of novel basis sets. In Sec. III we present the algorithm applied in the current work, and describe how the calculations were performed. Then, in Sec. IV we investigate the basis set convergence of the moments of the GS-EMD and present efficient basis sets, concluding in Sec. V.

II. METHODS

A. Completeness profiles

The idea of completeness optimization is founded on the concept of completeness profiles, which were introduced by Chong[55] as a graphical tool for evaluating the completeness of one-electron basis sets. The completeness profile of an atomic basis set can be calculated as[55]

$$Y\left(\alpha\right) = \langle G(\alpha)| \sum_{\mu\nu} |\mu\rangle S^{-1}_{\mu\nu} \langle \nu|G(\alpha)\rangle, \tag{7}$$

where $|\mu\rangle$ and $|\nu\rangle$ are functions in the studied basis set and $S^{-1}_{\mu\nu}$ denotes the (μ, ν) element of the inverse of the overlap matrix. $G(\alpha)$ is a scanning function, α being a parameter used to scan the basis set. The profile is (nearly) unity when the basis set is flexible enough to faithfully represent the scanning function, whereas $Y(\alpha)$ vanishes if the basis set is unable to do so.

In principle any normalizable function can be used as a scanning function. However, only Gaussian primitives have been used as probes so far. This choice is advantageous when GTO basis sets are used, as the overlap between a primitive function in the probed basis set and the scanning function is unity when the exponents match and decays rapidly when their difference grows. As a result, assuming no contractions and no linear dependencies in the basis set, a well-defined region where $Y(\alpha) \approx 1$ is seen in the profile; this is called the completeness plateau. For the reasons stated above, the scanning function used in the current work is

$$\langle \mathbf{r} | G(\alpha) \rangle = N r^l Y_{lm}(\hat{\mathbf{r}}) \exp(-\alpha r^2), \tag{8}$$

where N is a normalization constant, Y_{lm} are real spherical harmonics, $\hat{\mathbf{r}}$ denotes the unit vector in the direction of r, and the scanning parameter α has been identified with the Gaussian exponent. When pure spherical functions are used in the basis set, it is easily seen that a different profile is obtained for every value of the angular momentum l, whereas all values of the z component m yield an identical profile.

B. Completeness optimization

As the completeness profile visualizes the flexibility of the basis set, it can also be used to maximize it: the idea is to place the exponents on every angular momentum shell so that the completeness profile becomes as flat as possible, i.e., so that the ripples in the plateau are minimized.[10] Technically, the measure used for the optimization is[56]

$$\tau_n = \left(\frac{1}{\lg \alpha_{\max} - \lg \alpha_{\min}} \int_{\lg \alpha_{\min}}^{\lg \alpha_{\max}} [1 - Y(\alpha)]^n \, d \lg \alpha \right)^{1/n}, \tag{9}$$

where lg is the ISO standard notation for the 10-base logarithm, $\alpha_{\min}$ and $\alpha_{\max}$ are the lower and upper limits of the plateau, and $n \geq 1$ is a parameter. It is easily seen that by letting $\tau_n \to 0$, $\alpha_{\min} \to -\infty$, and $\alpha_{\max} \to +\infty$ a truly l-complete[57] basis set is obtained, proving that the method is mathematically rigorous.

While the ranges $\lg \alpha \in [\alpha_{\min}, \alpha_{\max}]$ for all values of l that are necessary to accurately calculate the wanted property are *a priori* unknown, they can be determined by varying the limits $\alpha_{\min}$ and $\alpha_{\max}$ for each value of l while examining the evolution of the wanted property.[10] An arbitrary minimal basis set (for instance a small s set for H–Be and a small sp set for B–Ar) is taken as the starting point. Now, by fixing the value of τ_n to a small (positive) number one is able to systematically examine the effect of the completeness of the basis set on the studied property, for which no variationality is required. This is done by extending the range of completeness of the basis set to the steep (larger $\alpha_{\max}$) and/or to the diffuse end (smaller $\alpha_{\min}$). Once the property does not change any more when the limits of existing shells are extended or shells with higher angular momentum are added, the CBS limit has been achieved.

The special case $n = 1$ in Eq. (9) corresponds to maximization of the area of completeness, and $n = 2$ to minimization of the root mean square deviation from completeness. Qualitative analysis of the completeness profiles produced by the two metrics reveals that $n = 2$ is stricter on the deviation from completeness at the edges of the plateau, leading to exponents placed closer by near the edges. However, as in the limit $\tau_n \to 0$ both cases lead to a CBS, in analogy to earlier work,[10–13] we will use maximization of the area and denote τ_1 with a plain τ in the rest of the paper.

C. Expansion of completeness

As the overlap of Gaussian functions is invariant under scaling of the exponents (which corresponds to translation in $\log \alpha$), the completeness-optimized exponents for an angular momentum shell corresponding to an interval of length $\lambda := \lg \alpha_{\max} - \lg \alpha_{\min}$ are equivalent up to a scaling factor. Thus, for a fixed value of τ, the width of the completeness plateau attainable with N exponents for the angular momentum l is uniquely defined by Eq. (9). The width λ_N – and the corresponding values of the exponents – can be easily found from Eq. (9), e.g., by using the bisection method. As a consequence of the uniquely defined width, one is able to define a step size σ related to the addition or to the removal of a single exponent on the angular momentum shell as $\sigma := \lambda_{N+1} - \lambda_N$. Correspondingly, a set criteria for the addition or removal of a single exponent can be easily placed for the formation of the basis set.

Trials are made to extend the completeness plateau of each value of angular momentum l in the basis set to the steep or the diffuse end by σ. Another trial is to make the existing plateau even flatter by adding a single exponent to the current interval, thereby reducing the value of τ on the current shell. The trial that leads to the largest change Δ in the computed property is then accepted. With this method, a (quasi)monotonically reducing change in the property is observed as the basis set grows more complete. The necessity of further correlation/polarization shells is evaluated by adding a new shell with a single exponent, and maximizing Δ with respect to the additional exponent. Once the change induced in the property by the addition of more functions is small enough, the CBS limit can be said to have been effectively reached.

III. IMPLEMENTATION

Although the method presented above in Sec. II for the formation of completeness-optimized basis sets can be applied to any property at any level of theory, for simplicity we will consider the restricted open-shell Hartree-Fock (ROHF) level of theory[58] in this paper. The completeness-optimization procedure outlined in Sec. II has been summarized in Algorithm I. The measure of the change Δ used in the current

ALGORITHM I. Generation of completeness-optimized basis sets.

1. Pick the wanted target tolerance ϵ_t and the initial tolerance $\epsilon_i > \epsilon_t$.
2. Pick an initial completeness tolerance $\tau \lesssim 10^{-3}$.
3. Choose an initial range of completeness: $Y(\alpha) \approx 1$ (within τ) for $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ for all occupied values of the angular momentum l.
4. Set the current tolerance value $\epsilon = \epsilon_i$.
5. Refine the atomic basis set(s).
   Loop over values of angular momentum in the basis set $l = 0, \ldots, l_{\max}$:
   (a) Compute the step size σ corresponding to the addition of an exponent to the shell l.
   (b) Form trials by expanding the range of completeness and reoptimizing the exponents by minimizing τ in Eq. (9):
      i. $\alpha_{\min} = 10^{-\sigma}\alpha_{\min}$.
      ii. $\alpha_{\max} = 10^{\sigma}\alpha_{\max}$.
      iii. Keep limits, but add one function to the current shell (reducing τ for the current value of l).
   (c) Compute the quantity in the trial basis sets.
   (d) Compute the maximum change Δ from the current reference value.
   (e) If $\Delta \geq \epsilon$, accept the trial corresponding to Δ, update the current reference value, and go back to step 5a. Otherwise, continue to next value of l.
6. Refine the polarization shells using the dimer(s).
   (a) Check that all existing polarization shells are sufficiently filled by looping over the polarization shells in the basis set analogously to step 5.
   (b) All existing shells have been refined. Check if the addition of a further polarization shell is necessary:
   Compute the step size σ corresponding to $l = l_{\max} + 1$ and set the range of completeness of the polarization shell to $\alpha_{\min} = 10^{x}$, $\alpha_{\max} = 10^{x+\sigma}$, where x is an (arbitrary) offset. Determine x by maximizing Δ.
   If the resulting $\Delta \geq \epsilon$, add the shell and return to step 6(a).
7. If $\epsilon > \epsilon_t$, set $\epsilon \to \epsilon/10$ and go back to step 5. Otherwise, end the loop.

work is

$$\Delta := \max_{k \in \mathcal{K}} \left| \frac{\langle p^k_{\text{new}} \rangle - \langle p^k_{\text{ref}} \rangle}{\langle p^k_{\text{ref}} \rangle} \right|, \tag{10}$$

where $\langle p^k_{\text{new}} \rangle$ is the new value and $\langle p^k_{\text{ref}} \rangle$ the current reference value, $\mathcal{K}$ being the set of moments considered in the optimization. The tolerance ϵ for the change in the computed property Δ is tightened slowly towards its final value in Algorithm I in order to guarantee that the shells with different l values are balanced, as done in Jensen's basis sets.[5]

When the CBS limit has been determined using Algorithm I, the completeness-optimization method can be used in reverse, as presented in Algorithm II, to prune unimportant functions in order to obtain computationally efficient basis sets for practical calculations on large systems. As it is now available, the CBS value is used as the reference during the reduction of the basis set. Since the estimated accuracy of the CBS value is ϵ_t, we start the reduction by forming a set with a similar accuracy, after which sets of decreasing accuracy are formed. Although presented here in the current scope of SCF calculations, Algorithms I and II can be straighforwardly generalized for post-HF calculations.

The recently published ERKALE program[56,59] is used to perform the fully automatized completeness-optimization procedure, including the necessary Hartree-Fock calculations. ERKALE uses the Nelder-Mead simplex algorithm[60] for the completeness optimization of the exponents in Eq. (9).[56] The algorithm used by ERKALE to compute the radial momentum density and the moments of the GS-EMD has been discussed extensively in Ref. 47.

ALGORITHM II. Reduction of completeness-optimized basis sets.

1. Pick the wanted final tolerance ϵ_f.
2. Set the current tolerance value $\epsilon = \epsilon_t$.
3. Reduce the completeness of the shells.
   (a) Loop over the values of angular momentum $l = l_{\max}, \ldots, 0$:
      i. Compute the step size σ corresponding to the removal of an exponent.
      ii. Form trials by reducing the range of completeness and reoptimizing the exponents by minimizing τ in Eq. (9):
         A. $\alpha_{\min} = 10^{\sigma}\alpha_{\min}$.
         B. $\alpha_{\max} = 10^{-\sigma}\alpha_{\max}$.
         C. Keep limits, but remove a function, increasing τ for the current value of l.
      iii. Compute the quantity in the trial basis sets.
      iv. Compute the minimal change δ_l from the CBS reference value.
   (b) Find out the minimal change $\delta = \min_l \delta_l$. If $\delta < \epsilon$, accept the trial corresponding to δ and go back to step 5.
4. If $\epsilon < \epsilon_f$, set $\epsilon \to 10\epsilon$ and go back to step 3. Otherwise, end the loop.

We form basis sets of uncontracted primitives for the first three rows of the periodic table. Since the completeness optimization of primitives contains no chemical information in itself – it is just a mathematical method – we group the elements of the first three rows into five groups with similar atomic electronic structures and use the same basis set for each of the elements in the group; this grouping is given in Table II.

In analogy to the polarization-consistent basis sets of Jensen[5] we obtain polarization functions by studying homoatomic molecules. The completeness-optimized basis set is then required to reproduce the moments of the GS-EMD both of the individual atoms and the homoatomic dimers. As the dimers represent rather different chemical environments, one can be assured of the transferability of the resulting basis sets. Experimental geometries for the dimer ground states were used,[61] their details are given in Table III.

Canonical orthonormalization[62] with a cutoff of 10^{-5} for the eigenvalues of the overlap matrix was used both in the completeness optimization and in the SCF calculations. Closed-shell species were computed using restricted Hartree-Fock during the completeness optimization of the basis sets, whereas a recently introduced ROHF formalism[63]

TABLE II. The grouping of electronic structures.

	Atomic configuration	Elements
Group I	$1s^n$	H–He
Group II	$[\text{He}]2s^n$	Li–Be
Group III	$[\text{He}]2s^2 2p^n$	B–Ne
Group IV	$[\text{Ne}]3s^n$	Na–Mg
Group V	$[\text{Ne}]3s^2 3p^n$	Al–Ar

TABLE III. The dimers used in the parametrization of the basis set. The used bond lengths, given in atomic units, are r_e values from the NIST database.[61] The Be_2 molecule is not stable, thus the (fictitious) bond length was adopted from Ref. 37.

		Bond length	State
Group I	H_2	1.401	$^1\Sigma_g$
	He_2	5.612	$^1\Sigma_g$
Group II	Li_2	5.051	$^1\Sigma_g$
	Be_2	4.000	$^1\Sigma_g$
Group III	B_2	3.005	$^3\Sigma_g$
	C_2	2.348	$^1\Sigma_g$
	N_2	2.074	$^1\Sigma_g$
	O_2	2.282	$^3\Sigma_g$
	F_2	2.668	$^1\Sigma_g$
	Ne_2	5.858	$^1\Sigma_g$
Group IV	Na_2	5.818	$^1\Sigma_g$
	Mg_2	7.351	$^1\Sigma_g$
Group V	Al_2	4.660	$^3\Sigma_g$
	Si_2	4.244	$^3\Sigma_g$
	P_2	3.578	$^1\Sigma_g$
	S_2	3.570	$^3\Sigma_g$
	Cl_2	3.755	$^1\Sigma_g$
	Ar_2	7.102	$^1\Sigma_g$

was used for open-shell species. The Broyden method[64] was used for Fock matrix updates. In cases where the Broyden method failed to converge, the calculation was performed with the trust-region Roothaan-Hall (TRRH) method.[65] The direct SCF scheme[66] was used with Schwarz prescreening of electron repulsion integrals with a threshold of 10^{-15} for the upper limit of the absolute value of the integrals. The used relative precision for the numerical integration of the GS-EMD moments (Eq. (6)) was 10^{-8}.

IV. RESULTS

Initial test calculations on the helium atom demonstrated that modeling $\langle p^{-2}\rangle$ and $\langle p^4\rangle$ with the accuracy aimed at in the current work using a Gaussian basis set is unfeasible due to numerical problems. This is not surprising, as these moments probe exactly the asymptotic regions where Gaussian type basis functions are inherently deficient, notably the nuclear cusp and the asymptotic decay.[41] However, this limitation is not a theoretical one, since STOs can be expanded exactly using GTOs as[67]

$$\exp(-\zeta r) = \frac{2}{\sqrt{\pi}} \int_0^\infty \exp(-t^2) \exp\left(-\left[\frac{\zeta r}{2t}\right]^2\right) dt \equiv \int_0^\infty w(t) \exp\left[-\alpha(t) r^2\right] dt, \quad (11)$$

where $w(t) = 2\exp(-t^2)/\sqrt{\pi}$ has been identified as the weighting factor and $\alpha(t) = \zeta^2/4t^2$ as the Gaussian exponent. Instead, the limitation is caused by the use of a finite number of functions and finite precision arithmetic, and it been observed before in the case of helium with energy-optimized GTOs;[68] with energy-optimized STOs all of the moments converge rapidly.[68]

Therefore, we decided to concentrate on the moments $\langle p^k\rangle$ for $k \in \mathcal{K} = \{-1, 1, 2, 3\}$,[69] of which the experimentally interesting quantities are the height of the Compton peak $J(0) = \langle p^{-1}\rangle/2$ and the kinetic energy of electrons $T = \langle p^2\rangle/2$.[70] Accordingly, in the rest of the paper we will refer to $\langle p^{-1}\rangle$, $\langle p\rangle$, $\langle p^2\rangle$, and $\langle p^3\rangle$ simply as *the moments*.

We started the completeness optimization from a set with a completeness plateau similar to that of the uncontracted *pc-0* basis set.[5] Algorithm I was then applied with the parameter values $\epsilon_i = 10^{-2}$, $\epsilon_t = 10^{-4}$, and $\tau = 10^{-4}$ to find the CBS limits, yielding the *coemd-ref* set. Finally, the production basis sets were formed by reducing the full basis sets using Algorithm II to the accuracies $\epsilon_f \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, compared to the results obtained with the *coemd-ref* basis set. We label the resulting production basis sets as *coemd-4*, *coemd-3*, *coemd-2*, and *coemd-1*, respectively. The *coemd*

TABLE IV. Estimated HF limits for the moments of the EMD of the molecules studied in the current work.

Molecule	Bond length (a.u.)	$\langle p^{-1}\rangle$	$\langle p\rangle$	$\langle p^2\rangle$	$\langle p^3\rangle$
H_2	1.400	3.115 (0)	1.820 (0)	2.252 (0)	3.964 (0)
	Near HF[a]	3.115 (0)	1.820 (0)	2.252 (0)	3.964 (0)
	Near HF[b]	3.115 (0)	1.821 (0)	2.252 (0)	3.964 (0)
Li_2	5.051	1.059 (1)	9.822 (1)	2.978 (1)	1.421 (2)
	Near HF[a,b]	1.053 (1)	9.825 (1)	2.978 (1)	1.421 (2)
Be_2	4.000	1.086 (1)	1.513 (1)	5.853 (1)	3.708 (2)
	Near HF[a,b]	1.077 (1)	1.514 (1)	5.855 (1)	3.707 (2)
B_2	3.005	1.062 (1)	2.158 (1)	9.829 (1)	7.657 (2)
	Near HF[a,b]	1.063 (1)	2.158 (1)	9.829 (1)	7.660 (2)
C_2	2.348	1.041 (1)	2.922 (1)	1.508 (2)	1.379 (3)
	Near HF[a]	1.042 (1)	2.922 (1)	1.508 (2)	1.379 (3)
	Near HF[b]	1.042 (1)	2.922 (1)	1.508 (2)	1.380 (3)
N_2	2.068	1.070 (1)	3.804 (1)	2.176 (2)	2.264 (3)
	Near HF[a,b]	1.069 (1)	3.804 (1)	2.176 (2)	2.265 (3)
O_2	2.282	1.070 (1)	4.767 (1)	2.988 (2)	3.482 (3)
	Near HF[a]	1.069 (1)	4.767 (1)	2.988 (2)	3.483 (3)
	Near HF[b]	1.069 (1)	4.767 (1)	2.989 (2)	3.483 (3)
F_2	2.680	1.076 (1)	5.839 (1)	3.971 (2)	5.099 (3)
	Near HF[a,b]	1.075 (1)	5.839 (1)	3.972 (2)	5.101 (3)
LiH	3.015	5.660 (0)	5.902 (0)	1.598 (1)	7.264 (1)
	Near HF[a,b]	5.660 (0)	5.902 (0)	1.598 (1)	7.266 (1)
HF	1.7328	6.470 (0)	3.037 (1)	2.000 (2)	2.546 (3)
	Near HF[a,b]	6.469 (0)	3.037 (1)	2.001 (2)	2.547 (3)
NaH	3.566	~9.3 (0)[c]	4.169 (1)	3.248 (2)	4.894 (3)
	Near HF[a,b]	9.305 (0)	4.169 (1)	3.248 (2)	4.893 (3)
HCl	2.4087	1.134 (1)	8.188 (1)	9.202 (2)	2.022 (4)
	Near HF[a,b]	1.134 (1)	8.188 (1)	9.202 (2)	2.022 (4)
LiF	2.955	7.941 (0)	3.468 (1)	2.140 (2)	2.614 (3)
	Near HF[a,b]	7.897 (0)	3.472 (1)	2.140 (2)	2.615 (3)
LiCl	3.825	1.301 (1)	8.613 (1)	9.341 (2)	2.029 (4)
	Near HF[a,b]	1.294 (1)	8.615 (1)	9.341 (2)	2.029 (4)
NaF	3.62883	1.128 (1)	7.050 (1)	5.228 (2)	7.434 (3)
	Near HF[a,b]	1.127 (1)	7.051 (1)	5.227 (2)	7.433 (3)
AlF	3.126	1.485 (1)	8.238 (1)	6.829 (2)	1.099 (4)
	Near HF[a,b]	1.475 (1)	8.239 (1)	6.829 (2)	1.099 (4)
NaCl	4.4609	1.638 (1)	1.220 (1)	1.243 (3)	2.511 (4)
	Near HF[a,b]	1.630 (1)	1.220 (1)	1.243 (3)	2.511 (4)

[a]The near HF values are from Ref. 36.
[b]The near HF values are from Ref. 37.
[c]The HF limit value for $\langle p^{-1}\rangle$ could not be determined.

basis sets, along with their completeness profiles, are available in the supplementary material.[71]

To demonstrate the computational efficiency of the generated basis sets, for comparison we have also performed calculations using conventional energy-optimized sets obtained from the EMSL basis set exchange,[72,73] i.e., the Dunning-style (*d-*)*aug-cc-p*(*C*)*VXZ* basis sets[1–4] and Jensen's augmented polarization-consistent *aug-pc-N* basis sets[5–9] of similar size to the *coemd* sets. We also included the *atpc3* set used by Hart and Thakkar, which is obtained from *aug-pc-3* by deleting *f* functions from hydrogen and *g* functions from all atoms.[41] Calculations with the energy-optimized *cc* and *pc* basis sets were performed not only with the original sets but also by decontracting them completely, i.e., using only the corresponding primitive set. These uncontracted sets are denoted with the prefix "*un-*." The large set of calculations performed with different style basis sets approaching the CBS limit enabled us to estimate the HF limits of the moments studied in the current work with an accuracy of four digits.

Demonstration calculations were performed for homoatomic dimers for which near HF values have been presented in the literature. As the basis sets were parametrized for homoatomic molecules, we also examined the transferability of the *coemd* basis sets by calculating the heteroatomic molecules LiH [I–II], HF [I–III], NaH [I–IV], HCl [I–V], LiF [II–III], LiCl [II–V], NaF [III–IV], AlF [III–V], and NaCl [IV–V], for which near HF values are available as well. For the sake of clarity, in the following we will present only a summary of the full set of results, which is available as a part of the supplementary material.[71] The HF limits estimated for the molecules along with near HF values from Refs. 36 and 37 are shown in Table IV. As seen from the table, we propose new HF limit values for $\langle p^{-1} \rangle$ and $\langle p \rangle$ of Li_2, LiF, and LiCl; $\langle p^{-1} \rangle$ of Be_2, AlF, and NaCl; and $\langle p^3 \rangle$ of B_2.

Double augmentation with diffuse functions (*d-aug-cc-pVXZ*) or the addition of more core functions (*aug-cc-pCVXZ*) is not necessary, as the results are qualitatively the same as in the *aug-cc-pVXZ* set. The moments produced by the *atpc3* and the *aug-pc-3* are practically identical, pointing to the small importance of higher order polarization functions to the EMD at the SCF level of theory. More interesting is the comparison of the effect of the contraction of the polarization-consistent basis sets. In most cases the moments produced by the normal, contracted forms of the *aug-pc* sets converge to incorrect values, the difference being most clearly seen for $\langle p^3 \rangle$. The decontraction of the basis sets restores the compatibility of the results with those obtained with the *coemd* and *aug-cc* sets. Thus, we conclude that moments calculated at the HF level of theory using the *aug-pc* sets[41] are not accurate.

The accuracy of the results obtained with the *aug-cc* sets suffers slightly from the contractions as well, but to a smaller extent than with the *aug-pc* sets. Obviously, this feature is caused by an incomplete description of the HF core orbitals in the contracted *pc* sets, which have been formed with DFT calculations,[6] whereas the less problematic Dunning sets have been contracted using HF.[1]

The *coemd-ref* and *coemd-4* sets correctly reproduce the estimated CBS limits for all molecules studied.[74] A similar amount of systematical accuracy can be achieved using the *un-atpc3* set or the *aug-cc-pVQZ* set, in the latter case with the exception of the hydrogen molecule for which a decontracted *aug-cc-pV6Z* set is necessary. However, in the best case the *un-atpc3* set contains the same amount of functions as the *coemd-ref* set, whereas often even the *coemd-ref* set is much smaller.

On a molecule by molecule basis the *coemd* sets compare even more favorably to the *aug-pc* and *aug-cc* sets. The minimal size of the basis sets required for reproducing the HF limits of the moments (Table IV) within an accuracy of one unit in the last decimal are shown in Table V. With the exception of the molecules containing a chlorine atom, the *coemd* sets reproduce the CBS limit with a smaller amount of

TABLE V. Minimal necessary basis sets and the resulting amounts of basis functions for reproducing the HF limits shown in Table IV. The second column shows the number of basis functions in the *coemd-ref* set. [a]The HF limit value for $\langle p^{-1} \rangle$ of NaH could not be determined and thus it was not used in determining the basis set convergence.

Molecule	*coemd-ref*	Minimal *coemd*		Minimal *aug-cc*		Minimal *aug-pc*	
H_2	74	*coemd-4*	54	*un-aug-cc-pV6Z*	262	*un-atpc3*	80
Li_2	114	*coemd-3*	52	*aug-cc-pVTZ*	92	*un-aug-pc-2*	72
Be_2	114	*coemd-3*	52	*aug-cc-pVTZ*	92	*un-atpc3*	130
B_2	182	*coemd-4*	138	*aug-cc-pVQZ*	160	*un-atpc3*	182
C_2	182	*coemd-3*	86	*aug-cc-pVTZ*	92	*un-atpc3*	182
N_2	182	*coemd-3*	86	*un-aug-cc-pVTZ*	116	*un-aug-pc-2*	122
O_2	182	*coemd-3*	86	*aug-cc-pVQZ*	160	*un-atpc3*	182
F_2	182	*coemd-3*	86	*aug-cc-pVQZ*	160	*un-atpc3*	182
LiH	94	*coemd-4*	74	*un-aug-cc-pVQZ*	141	*un-atpc3*	105
HF	128	*coemd-3*	61	*aug-cc-pVQZ*	126	*un-atpc3*	131
NaH[a]	101	*coemd-4*	85	*aug-cc-pVTZ*	73	*un-aptc3*	126
HCl	143	*coemd-4*	105	*aug-cc-pVDZ*	36	*un-aug-pc-2*	102
LiF	148	*coemd-3*	116	*aug-cc-pVQZ*	160	*un-atpc3*	156
LiCl	163	*coemd-4*	125	*aug-cc-pVDZ*	50	*un-aug-pc-2*	112
NaF	155	*coemd-4*	127	*aug-cc-pVQZ*	164	*un-atpc3*	177
AlF	197	*coemd-4*	147	*un-aug-cc-pVTZ*	133	*un-aug-pc-2*	137
NaCl	170	*coemd-3*	93	*aug-cc-pVDZ*	54	*un-aug-pc-2*	130

functions than the *aug-pc* or *aug-cc* sets in either contracted or decontracted form. Whereas normally a quadruple-ζ basis set (*aug-cc-pVQZ* or decontracted *atpc3*) is required, surprisingly for the chlorine systems the *aug-cc-pVDZ* set already yields a converged result. This is likely caused by the closely ionic nature of HCl, LiCl, and NaCl, and as a special case does not raise concerns about the performance of the *coemd* basis sets; the *aug-pc* sets of comparable performance are of similar size to the *coemd* sets.

V. CONCLUSIONS

We have shown that the CBS limit of the moments of the GS-EMD at the SCF level of theory can be achieved at a modest computational cost by using completeness-optimized GTO basis sets. Using the established CBS limits, we have generated cost efficient, completeness-optimized basis sets for the calculation of GS-EMD properties at the wanted accuracy at the SCF level of theory. The generated basis sets are much more compact than conventional energy-optimized sets, while producing more accurate results (at the SCF level of theory), hence enabling calculations to be performed near the CBS limit even in large systems.

We summarize this and previous studies on the suitable level of theory for computing EMDs as follows.

i. MP2 accounts for most correlation effects on the EMD,[41] as compared to a CCSD reference.
ii. A basis set equivalent to the Dunning-style *aug-cc-pVQZ* basis set is generally capable of reproducing the moments of the EMD at the SCF level of theory with a four-digit accuracy.

The possible contraction of the completeness-optimized primitive sets, along with the straightforward extension to generating basis sets suitable for post-HF calculations, are left for future work.

ACKNOWLEDGMENTS

This work has been supported by the Jenny and Antti Wihuri foundation, the Academy of Finland through its Centers of Excellence program and Project No. 1127462, and by the research funds of the University of Helsinki (No. 490064). J.L. thanks Lars Pettersson for help in implementing the TRRH algorithm. The computational resources of CSC - IT Center for Science Ltd. (Espoo, Finland) are gratefully acknowledged.

[1] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[2] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[3] K. A. Peterson, D. E. Woon, and T. H. Dunning, J. Chem. Phys. **100**, 7410 (1994).
[4] D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).
[5] F. Jensen, J. Chem. Phys. **115**, 9113 (2001).
[6] F. Jensen, J. Chem. Phys. **116**, 7372 (2002).
[7] F. Jensen, J. Chem. Phys. **117**, 9234 (2002).
[8] F. Jensen and T. Helgaker, J. Chem. Phys. **121**, 3463 (2004).
[9] F. Jensen, J. Phys. Chem. A **111**, 11198 (2007).
[10] P. Manninen and J. Vaara, J. Comput. Chem. **27**, 434 (2006).
[11] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara, J. Chem. Phys. **129**, 124102 (2008).
[12] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara, Phys. Chem. Chem. Phys. **11**, 11404 (2009).
[13] S. Ikäläinen, M. Romalis, P. Lantto, and J. Vaara, Phys. Rev. Lett. **105**, 153001 (2010).
[14] A. J. Thakkar, Adv. Chem. Phys. **128**, 303 (2004).
[15] M. Coplan, J. Moore, and J. Doering, Rev. Mod. Phys. **66**, 985 (1994).
[16] M. J. Cooper, P. E. Mijnarends, N. Shiotani, N. Sakai, and A. Bansil, *X-Ray Compton Scattering*, Oxford Series on Synchrotron Radiation Vol. 5 (Oxford University Press, 2004).
[17] M. J. Cooper, Rep. Prog. Phys. **48**, 415 (1985).
[18] P. Kaijser and V. H. Smith, Adv. Quantum Chem. **10**, 37 (1977).
[19] A. Erba and C. Pisani, J. Comput. Chem. **33**, 822 (2012).
[20] P. Eisenberger and P. M. Platzman, Phys. Rev. A **2**, 415 (1970).
[21] C. Blaas, J. Redinger, S. Manninen, V. Honkimäki, K. Hämäläinen, and P. Suortti, Phys. Rev. Lett. **75**, 1984 (1995).
[22] K. Hämäläinen, S. Manninen, C.-C. Kao, W. Caliebe, J. B. Hastings, A. Bansil, S. Kaprzyk, and P. M. Platzman, Phys. Rev. B **54**, 5453 (1996).
[23] A. J. Thakkar, J. Chem. Phys. **86**, 5060 (1987).
[24] S. Huotari, B. Boldrini, V. Honkimäki, P. Suortti, and W. Weyrich, J. Synchrotron Radiat. **16**, 672 (2009).
[25] S. Huotari, J. A. Soininen, T. Pylkkänen, K. Hämäläinen, A. Issolah, A. Titov, J. McMinis, J. Kim, K. Esler, D. M. Ceperley, M. Holzmann, and V. Olevano, Phys. Rev. Lett. **105**, 086403 (2010).
[26] H. Sakurai, H. Ota, N. Tsuji, M. Itou, and Y. Sakurai, J. Phys. B **44**, 065001 (2011).
[27] J. Okada, P. Sit, Y. Watanabe, Y. Wang, B. Barbiellini, T. Ishikawa, M. Itou, Y. Sakurai, A. Bansil, R. Ishikawa, M. Hamaishi, T. Masaki, P.-F. Paradis, K. Kimura, and S. Nanao, Phys. Rev. Lett. **108**, 067402 (2012).
[28] Ch. Bellin, B. Barbiellini, S. Klotz, T. Buslaps, G. Rousse, Th. Strässle, and A. Shukla, Phys. Rev. B **83**, 094117 (2011).
[29] Y. Sakurai, M. Itou, B. Barbiellini, P. E. Mijnarends, R. S. Markiewicz, S. Kaprzyk, J.-M. Gillet, S. Wakimoto, M. Fujita, S. Basak, Y. J. Wang, W. Al-Sawai, H. Lin, A. Bansil, and K. Yamada, Science **332**, 698 (2011).
[30] I. Juurinen, K. Nakahara, N. Ando, T. Nishiumi, H. Seta, N. Yoshida, T. Morinaga, M. Itou, T. Ninomiya, Y. Sakurai, E. Salonen, K. Nordlund, K. Hämäläinen, and M. Hakala, Phys. Rev. Lett. **107**, 197401 (2011).
[31] F. Lehmkühler, A. Sakko, C. Sternemann, M. Hakala, K. Nygård, C. J. Sahle, S. Galambosi, I. Steinke, S. Tiemeyer, A. Nyrow, T. Buslaps, D. Pontoni, M. Tolan, and K. Hämäläinen, J. Phys. Chem. Lett. **1**, 2832 (2010).
[32] L. Shi, K. Liu, C. Ning, and J. Deng, Sci. China, Ser. G **54**, 1981 (2011).
[33] J. K. Deng, G. Q. Li, Y. He, J. D. Huang, H. Deng, X. D. Wang, F. Wang, Y. A. Zhang, C. G. Ning, N. F. Gao, Y. Wang, X. J. Chen, and Y. Zheng, J. Chem. Phys. **114**, 882 (2001).
[34] W. N. Pang, J. F. Gao, C. J. Ruan, R. C. Shang, A. B. Trofimov, and M. S. Deleuze, J. Chem. Phys. **112**, 8043 (2000).
[35] Z. H. Luo, C. G. Ning, K. Liu, Y. R. Huang, and J. K. Deng, J. Phys. B **42**, 165205 (2009).
[36] R. K. Pathak, B. S. Sharma, and A. J. Thakkar, J. Chem. Phys. **85**, 958 (1986).
[37] A. J. Thakkar and W. A. Pedersen, Int. J. Quantum Chem., Symp. **38**, 327 (1990).
[38] A. K. Roy and A. J. Thakkar, Chem. Phys. Lett. **362**, 428 (2002).
[39] J. Wang and V. H. Smith, J. Phys. B **27**, 5159 (1994).
[40] B. Miguel and J. M. García de la Vega, Theor. Chem. Acc. **118**, 723 (2007).
[41] J. R. Hart and A. J. Thakkar, Int. J. Quantum Chem. **102**, 673 (2005).
[42] C. Pisani, R. Dovesi, and R. Orlando, Int. J. Quantum Chem. **42**, 5 (1992).
[43] A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz, and D. Usvyat, Phys. Rev. B **81**, 165108 (2010).
[44] A. Erba, M. Itou, Y. Sakurai, R. Yamaki, M. Ito, S. Casassa, L. Maschio, A. Terentjes, and C. Pisani, Phys. Rev. B **83**, 125208 (2011).
[45] C. Pisani, A. Erba, S. Casassa, M. Itou, and Y. Sakurai, Phys. Rev. B **84**, 245102 (2011).
[46] P. Jaiswal and A. Shukla, Phys. Rev. A **75**, 022504 (2007).
[47] J. Lehtola, M. Hakala, J. Vaara, and K. Hämäläinen, Phys. Chem. Chem. Phys. **13**, 5630 (2011).
[48] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[49] P. E. Regier and A. J. Thakkar, J. Phys. B **18**, 3061 (1985).
[50] M. Hakala, K. Nygård, S. Manninen, L. G. M. Pettersson, and K. Hämäläinen, Phys. Rev. B **73**, 035432 (2006).
[51] K. Nygård, M. Hakala, S. Manninen, A. Andrejczuk, M. Itou, Y. Sakurai, L. G. M. Pettersson, and K. Hämäläinen, Phys. Rev. E **74**, 031503 (2006).

[52]K. Nygård, M. Hakala, T. Pylkkänen, S. Manninen, T. Buslaps, M. Itou, A. Andrejczuk, Y. Sakurai, M. Odelius, and K. Hämäläinen, J. Chem. Phys. **126**, 154508 (2007).
[53]K. Nygård, M. Hakala, S. Manninen, M. Itou, Y. Sakurai, and K. Hämäläinen, Phys. Rev. Lett. **99**, 197401 (2007).
[54]M. Hakala, K. Nygård, J. Vaara, M. Itou, Y. Sakurai, and K. Hämäläinen, J. Chem. Phys. **130**, 034506 (2009).
[55]D. P. Chong, Can. J. Chem. **73**, 79 (1995).
[56]J. Lehtola, M. Hakala, A. Sakko, and K. Hämäläinen, J. Comput. Chem. **33**, 1572 (2012).
[57]As all values of angular momentum l yield a separate profile, a complete basis set must be complete for all values of l.
[58]C. Roothaan, Rev. Mod. Phys. **32**, 179 (1960).
[59]J. Lehtola, ERKALE – HF/DFT from Hel, 2012, see http://erkale.googlecode.com.
[60]J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[61]K. P. Huber, G. Herzberg, J. W. Gallagher, and R. D. Johnson III, "Constants of diatomic molecules," in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, edited by P. J. Linstrom and W. G. Mallard (National Institute of Standards and Technology, Gaithersburg, MD), see http://webbook.nist.gov.
[62]P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).
[63]T. Tsuchimochi and G. E. Scuseria, J. Chem. Phys. **133**, 141102 (2010).
[64]K. Baarman, T. Eirola, and V. Havu, J. Chem. Phys. **134**, 134109 (2011).
[65]L. Thøgersen, J. Olsen, A. Köhn, P. Jørgensen, P. Sałek, and T. Helgaker, J. Chem. Phys. **123**, 074103 (2005).
[66]J. Almlöf, K. Faegri, and K. Korsell, J. Comput. Chem. **3**, 385 (1982).
[67]R. Kikuchi, J. Chem. Phys. **22**, 148 (1954).
[68]A. M. Simas, A. J. Thakkar, and V. H. Smith, Int. J. Quantum Chem. **21**, 419 (1982).
[69]$\langle p^0 \rangle$ is simply the normalization condition, giving the number of electrons.
[70]According to the virial theorem this is, in turn, the negative of the total energy if the system is in its ground-state geometry.
[71]See supplementary material at http://dx.doi.org/10.1063/1.4749272 for the *coemd* basis sets, their completeness profiles, and the values of the moments of the EMD of the molecules in Tables IV and V computed in the *coemd*, *aug-cc*, and *aug-pc* basis sets.
[72]K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Model. **47**, 1045 (2007).
[73]See http://bse.pnl.gov for EMSL basis set exchange.
[74]To allow for rounding errors, a deviation of one unit in the last decimal is allowed in the values of the moments.

Paper III

Contraction of completeness-optimized basis sets: Application to ground-state electron momentum densities
Susi Lehtola, Pekka Manninen, Mikko Hakala, and Keijo Hämäläinen
Journal of Chemical Physics 138, 044109 (2013)
doi:10.1063/1.4788635

Contraction of completeness-optimized basis sets: Application to ground-state electron momentum densities

Susi Lehtola,[1,a)] Pekka Manninen,[2,b)] Mikko Hakala,[1] and Keijo Hämäläinen[1]
[1]*Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[2]*CSC – IT Center for Science Ltd., P.O. Box 405, FI-02101 Espoo, Finland*

(Received 12 November 2012; accepted 6 January 2013; published online 28 January 2013)

Completeness-optimization is a novel method for the formation of one-electron basis sets. Contrary to conventional methods of basis set generation that optimize the basis set with respect to ground-state energy, completeness-optimization is a completely general, black-box method that can be used to form cost-effective basis sets for any wanted property at any level of theory. In our recent work [J. Lehtola, P. Manninen, M. Hakala, and K. Hämäläinen, J. Chem. Phys. **137**, 104105 (2012)] we applied the completeness-optimization approach to forming primitive basis sets tuned for calculations of the electron momentum density at the Hartree-Fock (HF) level of theory. The current work extends the discussion to contracted basis sets and to the post-HF level of theory. Contractions are found to yield significant reductions in the amount of functions without compromising the accuracy. We suggest polarization-consistent and correlation-consistent basis sets for the first three rows of the periodic table, which are completeness-optimized for electron momentum density calculations.
 [http://dx.doi.org/10.1063/1.4788635]

I. INTRODUCTION

Completeness-optimization[1] is a mathematical tool for spanning the physically and chemically relevant atom-centered one-electron basis set in the best possible way. As the method does not rely on any specific type of calculation (single *vs* multiple determinants, variational *vs* non-variational method) or even the use of the basis set (orbital *vs* auxiliary basis), it provides a systematic, black box approach for reaching the complete basis set (CBS) limit for any property at any level of theory.

In our previous work[2] we presented an automated algorithm for performing completeness-optimization of primitive Gaussian basis sets at the self-consistent field (SCF) level of theory and applied it to calculations of the moments of the ground-state electron momentum density[3] $n(\mathbf{p})$ (EMD)

$$\langle p^k \rangle = \int_0^\infty p^{k+2} n(p) dp, \quad -2 \leq k \leq 4, \tag{1}$$

where the radial EMD $n(p)$ can be obtained as[4]

$$n(p) = \sum_{\mu\nu} P_{\mu\nu} \int \overline{\tilde{\chi}_\mu(\mathbf{p})} \tilde{\chi}_\nu(\mathbf{p}) d\Omega_{\mathbf{p}}. \tag{2}$$

Here $\boldsymbol{P}$ is the one-electron density matrix, $\tilde{\chi}_\mu(\mathbf{p})$ is the Fourier transform of the μth basis function and the overline denotes complex conjugation. Similarly to previous works that have applied completeness-optimization to magnetic properties,[1,5–7] completeness-optimization was found to produce basis sets that are computationally more efficient than conventional basis sets also in the case of the EMD.

All current applications of completeness-optimization have been restricted to the SCF level of theory and have used only basis sets of uncontracted primitives. In the current work we discuss the contraction of completeness-optimized basis sets and the straightforward application of the completeness-optimization algorithms introduced in Ref. 2 to the post-Hartree–Fock (post-HF) level of theory, still focusing on the EMD as the studied property.

Analogously to the polarization-consistent basis sets of Jensen[8–11] and the correlation-consistent basis sets of Dunning and co-workers[12–15] that produce systematic convergence with respect to the energy, we present novel polarization-consistent *pcemd* basis sets (parametrized at the HF level of theory) and correlation-consistent *ccemd* basis sets (parametrized at the post-HF level of theory) that produce systematic convergence with respect to the EMD. The basis sets are available as part of the supplementary information.[16]

The EMD was chosen for the study, because as a first-order property it is very sensitive to the deficiencies in the wave function, and because EMD-based structural studies using synchrotron radiation have recently become more popular.[17–29] Therefore devising accurate but computationally feasible approaches for *ab initio* calculations of the EMD is of interest and practical importance.

The structure of the manuscript is the following. In Sec. II we briefly review the completeness-optimization scheme. Next, in Sec. III, we discuss the implementation of the scheme in practice. The results of the study are presented in Sec. IV, ending in discussion and conclusions in Sec. V. Atomic units are used throughout the manuscript.

a)Electronic mail: susi.lehtola@alumni.helsinki.fi.
b)Electronic mail: pekka.manninen@csc.fi.

II. METHOD

The method of completeness-optimization is based on Chong's concept of completeness profiles[30,31]

$$Y(\alpha) = \langle G(\alpha)| \sum_{\mu\nu} |\mu\rangle S^{-1}_{\mu\nu} \langle\nu||G(\alpha)\rangle, \tag{3}$$

which measure the flexibility of the basis set to represent the used scanning function $|G(\alpha)\rangle$. $|\mu\rangle$ and $|\nu\rangle$ are functions in the probed basis set and $S^{-1}_{\mu\nu}$ denotes the (μ, ν) element of the inverse overlap matrix. The scanning function used in the current work is a Gaussian primitive

$$\langle \mathbf{r}|G(\alpha)\rangle = N r^l \exp(-\alpha r^2) Y_{lm}(\hat{\mathbf{r}}), \tag{4}$$

where l is the angular momentum, Y_{lm} are spherical harmonics in the real form, and the parameter α has been identified with the Gaussian exponent. A separate profile is obtained for all values of l present in a basis set, while all values of m yield an identical profile. When the basis set is capable of representing the scanning function, $Y(\alpha) \approx 1$.

In contrast to the conventional method of basis set generation, i.e., optimizing a fixed amount of exponents in order to produce the minimal possible atomic (or molecular) energy, in the completeness-optimization scheme the exponents for each angular momentum shell are obtained by minimizing the deviation from unity of the completeness profile.[1] In the current work the exponents are determined by minimization of the measure[2,32]

$$\tau = \frac{1}{\lg \alpha_{\max} - \lg \alpha_{\min}} \int_{\lg \alpha_{\min}}^{\lg \alpha_{\max}} [1 - Y(\alpha)]\, d \lg \alpha, \tag{5}$$

where lg is the 10-base logarithm. While the limits $\alpha_{\min}$ and $\alpha_{\max}$ in equation (5) that describe the relevant exponent space for the examined property are *a priori* unknown, they can be determined by a trial and error approach.[1,2,5–7]

As the values of the exponents do not have any direct connection to the chemistry of the element, completeness-optimized basis sets have an universal[33,34] nature, since adjacent elements often have similar electronic structures; that is, basis set requirements.[1]

The method of completeness-optimization of the exponents bears resemblance to the solution of the Griffin–Hill–Wheeler (GHW) version of the Hartree–Fock (HF) equations, which leads to integration over the exponent space.[35] This integral can be accurately discretized, while the exact set of values of the used exponents do not matter as in the completeness-optimization scheme. Even though the GHW method is very efficient for generating atomic basis sets at the HF level,[35–39] its applicability beyond the HF atom is unclear. In contrast, the completeness-optimization scheme does not rely on the used level of theory or the property calculated.

The scheme for forming the completeness-optimized basis sets is the following. Following the procedure of Ref. 2, we collect the elements of the first three rows into five groups, shown in Table I, and use the same primitive basis set and contraction pattern for every element in the group. The primitive set yielding the CBS limit for the moments of the EMD of the atoms and homoatomic dimers of Table I is formed by expanding the basis set in a trial-and-error approach until convergence is achieved as measured by[2]

TABLE I. The grouping of the elements into similar shell structures and the dimers used in the parametrization of the basis sets.[2] The primitive set is expanded until the CBS limits of the moments of the atoms and homoatomic dimers in the group are found. The pruned (and contracted) basis sets are then required to reproduce these limits within the wanted accuracy.

		r (bohr)	State
Group I	H_2	1.401	$^1\Sigma_g$
	He_2	5.612	$^1\Sigma_g$
Group II	Li_2	5.051	$^1\Sigma_g$
	Be_2	4.000	$^1\Sigma_g$
Group III	B_2	3.005	$^3\Sigma_g$
	C_2	2.348	$^1\Sigma_g$
	N_2	2.074	$^1\Sigma_g$
	O_2	2.282	$^3\Sigma_g$
	F_2	2.668	$^1\Sigma_g$
	Ne_2	5.858	$^1\Sigma_g$
Group IV	Na_2	5.818	$^1\Sigma_g$
	Mg_2	7.351	$^1\Sigma_g$
Group V	Al_2	4.660	$^3\Sigma_g$
	Si_2	4.244	$^3\Sigma_g$
	P_2	3.578	$^1\Sigma_g$
	S_2	3.570	$^3\Sigma_g$
	Cl_2	3.755	$^1\Sigma_g$
	Ar_2	7.102	$^1\Sigma_g$

$$\Delta = \max_{k \in \mathcal{K}} \left| \frac{\langle p^k \rangle_{\text{new}} - \langle p^k \rangle_{\text{ref}}}{\langle p^k \rangle_{\text{ref}}} \right|, \tag{6}$$

where the set of studied moments is $\mathcal{K} = \{-1, 1, 2, 3\}$, to which we will refer simply as the moments.

Next, still following Ref. 2, "production-level" primitive basis sets are formed by pruning unnecessarily thick sampling of the exponent space. Finally, the pruned exponents are contracted to form the final basis sets, the transferability of which is verified by calculations on a set of heteroatomic molecules.

In the current work we use slightly revised versions of the expansion and pruning algorithms, which are available as part of the supplementary material.[16] Furthermore, as the result of the completeness-optimization procedure is rather insensitive to its starting point, in the current work we have introduced Algorithm I which initializes the atomic basis set at the SCF level of theory, one electronic shell at a time, hence quickly providing a sensible starting point for the full completeness-optimization procedure.

ALGORITHM I. Generation of the initial basis set for the group with maximum atomic number Z_0.

1. Initialize the basis set with 2 s functions, starting from $\alpha_{\min} = 0$ and ranging to the value of $\alpha_{\max}$ attainable[2] with the set value of τ.
2. Loop over the noble gases with $Z \leq Z_0$, from lightest to heaviest:
   (a) If the atom has electrons on a higher angular momentum shell than is currently contained in the basis set, add the shell with two functions on it analogously to step 1.
   (b) Expand the basis set until the CBS limit for the gas-phase atom at the SCF level of theory is attained within $\epsilon_t = 10^{-1}$.

ALGORITHM II. Contraction of completeness-optimized basis sets.

1. Compute the atomic one-electron density matrices for the atoms in the group at the wanted level of theory.
2. For every value of angular momentum l occupied at the SCF level of theory (s for H-Be and sp for groups B-Ar), diagonalize the density matrix in the (ll) sublock to obtain the natural orbitals.
3. If $l = 0$ the MO coefficients give the exponential contraction straight away, whereas if $l > 0$ an average is taken over all of the coefficients of the orbitals with different values of the z projection m, weighed by their respective occupation numbers.
4. Compute measures of goodness Δ_l corresponding to the contraction of an additional exponent in the steep end, considering both the monomers and dimers.
5. If the minimal change is smaller than the used threshold ϵ, $\min_l \Delta_l \leq \epsilon$, accept the contraction corresponding to the minimal value of Δ and return to step 4. Otherwise, end the algorithm.

In order to make the contraction method independent of the level of theory used, we use atomic natural orbitals (NOs) to form generally contracted basis sets,[40] as summarized in Algorithm II. In contrast to the previously published ANO sets of Almlöf, Roos and co-workers,[41–43] not all primitives are contracted. Furthermore, while in the former scheme the first (contracted) basis functions are simply the occupied HF orbitals, in our scheme all orbitals are NOs, thus including correlation effects already in the minimal basis set. At the SCF level of theory our scheme is equivalent to the conventional HF orbital contraction scheme.

Finally, we note that since $\langle p^2 \rangle/2$ gives the kinetic energy of the electrons, which at equilibrium is equal to the negative of the total energy of the system by virtue of the quantum mechanical virial theorem, the resulting completeness-optimized basis sets are polarization-consistent (and also correlation-consistent in the post-HF case) in the sense of the energy, as well. However, while the energy holds a special place in the Schrödinger equation and is thus variational in many models of quantum chemistry, such as HF, multi-configurational SCF, and full configuration-interaction, this is not true for the kinetic energy which is non-variational. In consequence, the completeness-optimization with respect to the kinetic energy is more stringent in terms of the basis set requirements, as the fulfilment of the virial theorem is implicitly required as well.

III. IMPLEMENTATION

We chose to use Møller-Plesset perturbation theory[44] truncated at the second order (MP2) in the current work, as it has been found to account for most correlation effects on the moments of the EMD.[45] The truncation of the virtual orbital space with respect to MP2 natural orbitals has been recently found to significantly speed up coupled-cluster calculations, without compromising their accuracy.[46–49] Thus, although the actual values of the moments produced at the MP2 level of theory may be far from correct for systems with significant multideterminental nature such as[50] N_2, we expect the parametrization at the MP2 level of theory to produce basis sets suitable also for calculations at higher levels of theory, such as multi-reference configuration interaction or coupled-cluster methods.

TABLE II. Compositions of the completeness-optimized polarization-consistent CBS limit sets.

	coemd-ref		*un-pcemd-ref*	
Group	n_{bf}	Composition	n_{bf}	Composition
I	37	17s5p1d	31	17s3p1d
II	57	20s9p2d	83	21s10p5d1f
III	91	20s13p5d1f	87	21s13p4d1f
IV	64	23s12p1d	82	24s12p3d1f
V	111	21s16p7d1f	111	22s16p5d1f1g

The ERKALE program[32,51] was used to perform the automatical completeness-optimization procedure and to calculate the moments of the EMD during the optimization.[52] The necessary density matrices were calculated using GAUSSIAN 09.[53] All calculations during the optimization were performed in the spin-unrestricted formalism. However, the calculations of closed-shell molecules reported in Sec. IV were performed in the restricted spin formalism. In both cases convergence to saddle point solutions was prohibited via internal stability analysis[54] of the SCF wave function. All electrons were correlated in the MP2 calculations.

In addition to the calculations performed with the completeness-optimized basis sets, we also performed calculations using Jensen's polarization-consistent *aug-pc-N* basis set series,[8–11] the Dunning-style correlation-consistent *aug-cc-pVXZ* basis set series,[12–15] and the *coemd* basis sets[2] obtained from the EMSL basis set exchange.[55,56] Calculations with the *aug-pc-N* and *aug-cc-pVXZ* series were performed both in contracted and uncontracted form, the latter being denoted with the *un-* prefix. Finally, convergence with respect to diffuse functions was checked by even-tempered extrapolation of the *aug-cc-pVXZ* series to the doubly and triply augmented *d-aug-cc-pVXZ* and *t-aug-cc-pVXZ* sets.[57]

IV. RESULTS

A. Polarization-consistent basis sets

Due to the slight improvements in the completeness-optimization algorithms, the use of point group symmetry and the use of a stability-analyzed unrestricted Hartree–Fock (UHF) reference during the completeness-optimization, we began by revisiting the parametrization of the polarization-consistent primitive sets for the first three rows of the periodic table at the SCF level of theory, analogously to the *coemd* basis sets of Ref. 2.

The CBS limit is obtained with the parameter values $\tau = 10^{-4}$, $\epsilon_i = 10^{-1}$, and $\epsilon_t = 10^{-4}$ in the notation of Ref. 2 (meaning that the addition of a further function at any angular momentum to the final basis set causes a relative change in the moments smaller than $\epsilon_t = 10^{-4}$). The composition of the resulting *un-pcemd-ref* basis set, shown in Table II, differs slightly from that of the *coemd-ref* set of Ref. 2: the amount of primitives is significantly increased in groups II and IV, but slightly reduced in groups I and III. In the *un-pcemd-ref* basis set groups II, III, and IV are much more similar to each other than in the *coemd-ref* set.

TABLE III. Contraction schemes of the *pcemd* and *ccemd* basis sets. The notation is [primitive | contracted].

	Group I	Group II	Group III	Group IV	Group V
pcemd-2:	[6s1p|3s]	[10s3p1d|3s]	[10s5p|3s2p]	[10s4p|4s1p]	[10s7p|3s3p]
pcemd-3:	[12s1p|6s]	[16s5p2d|7s]	[13s7p1d|6s4p]	[16s7p1d1f|12s2p]	[15s9p2d|5s4p]
pcemd-4:	[16s3p1d|9s]	[20s8p3d1f|10s]	[18s11p3d1f|10s7p]	[22s10p2d1f|17s3p]	[19s12p3d1f|16s8p]
ccemd-2:	[7s1p|3s]	[10s4p1d|3s]	[9s5p1d|3s4p]	[12s4p1d|3s2p]	[10s7p2d|3s3p]
ccemd-3:	[12s3p1d1f|7s]	[14s6p3d1f|8s]	[13s7p3d1f1g|8s]	[14s8p3d|11s]	[15s9p4d2f1g|10s7p]

TABLE IV. Molecules used for the benchmark.

Groups	HF	MP2	Molecule	r (bohr)	State
I-I	✓	✓	H_2	1.400	$^1\Sigma$
			He_2	5.628	$^1\Sigma$
I-II	✓		LiH	3.015	$^1\Sigma$
			BeH	2.538	$^2\Sigma$
I-III	✓	✓	BH	2.336	$^1\Sigma$
			CH	2.124	$^2\Sigma$
			NH	1.9614	$^3\Sigma$
			OH	1.8342	$^2\Sigma$
			HF	1.7328	$^1\Sigma$
I-IV	✓		NaH	3.566	$^1\Sigma$
			MgH	3.271	$^2\Sigma$
I-V	✓	✓	AlH	3.114	$^1\Sigma$
			SiH	2.874	$^2\Sigma$
			PH	2.708	$^3\Sigma$
			SH	2.551	$^2\Sigma$
			HCl	2.4087	$^1\Sigma$
II-II	✓		Li_2	5.051	$^1\Sigma$
			Be_2	4.000	$^1\Sigma$
II-III	✓		LiO	3.184	$^2\Sigma$
			LiF	2.955	$^1\Sigma$
			BeO	2.5149	$^1\Sigma$
			BeF	2.572	$^2\Sigma$
II-V	✓		LiCl	3.825	$^1\Sigma$
			BeS	3.291	$^1\Sigma$
III-III	✓	✓	B_2	3.005	$^3\Sigma$
			BN	2.421	$^3\Sigma$
			BO	2.275	$^2\Sigma$
			BF	2.391	$^1\Sigma$
			C_2	2.3481	$^1\Sigma$
			CN	2.214	$^2\Sigma$
			CO	2.132	$^1\Sigma$
			CF	2.402	$^2\Sigma$
			N_2	2.068	$^1\Sigma$
			NO	2.1747	$^2\Sigma$
			NF	2.489	$^3\Sigma$
			O_2	2.282	$^3\Sigma$
			OF	2.4958	$^2\Sigma$
			F_2	2.680	$^1\Sigma$
III-IV	✓		NaF	3.62883	$^1\Sigma$
			MgO	3.3052	$^1\Sigma$
III-V	✓	✓	PN	2.818	$^1\Sigma$
			SiO	2.854	$^1\Sigma$
			AlF	3.126	$^1\Sigma$
IV-V	✓		NaCl	4.4609	$^1\Sigma$
			MgS	4.049	$^1\Sigma$

The *un-pcemd-ref* basis set was then pruned to form the primitive *un-pcemd-4*, *un-pcemd-3* and *un-pcemd-2* basis sets, similarly to Ref. 2. The sets reproduce the moments of the EMD at the UHF level of theory with maximum relative errors of 10^{-4}, 10^{-3}, and 10^{-2}, respectively, as compared to the CBS limit given by the *un-pcemd-ref* set.

Finally, while the speed of SCF level calculations is bound by the cost of integral evaluation, contractions do offer some speed benefits also on the SCF level of theory: the amount of degrees of freedom is reduced, and so is the number of two-electron integrals. The primitive *un-pcemd* sets

TABLE V. The amount of basis functions for elements in groups I-V, and the mean amount of functions in the calculations of the molecules in Table IV.

Basis	I	II	III	IV	V
ccemd-2	6	20	20	14	22
un-ccemd-2	10	27	29	29	41
ccemd-3	28	48	60	50	74
un-ccemd-3	33	54	65	53	85
un-ccemd-ref	53	86	106	97	133
pcemd-2	6	17	9	7	12
un-pcemd-2	9	24	25	22	31
pcemd-3	9	32	23	30	27
un-pcemd-3	15	41	39	49	52
pcemd-4	23	56	53	43	62
un-pcemd-4	30	66	73	69	77
un-pcemd-ref	31	83	87	82	111
coemd-2	7	15	27	23	34
coemd-3	17	26	50	43	50
coemd-4	27	42	69	58	83
coemd-ref	37	57	91	64	111
aug-cc-pVDZ	9	23	23	27	27
un-aug-cc-pVDZ	11	35	35	50	50
aug-cc-pVTZ	23	46	46	50	50
un-aug-cc-pVTZ	25	59	58	79	75
aug-cc-pVQZ	46	80	80	84	84
un-aug-cc-pVQZ	48	93	93	118	112
un-aug-cc-pV5Z[a]	83	...	144	...	162
aug-pc-2	23	24	46	28	50
un-aug-pc-2	26	36	61	54	76
aug-pc-3	50	48	89	51	89
un-aug-pc-3	54	65	109	86	124
aug-pc-4	88	83	145	85	141
un-aug-pc-4	92	106	167	129	185

[a]A parametrization of the *aug-cc-pV5Z* basis set has not been published for groups II and IV.

were contracted to produce the *pcemd* sets, while retaining the same upper limit for the error as used in the reduction phase.[61] The resulting contraction schemes are shown in Table III.

B. Correlation-consistent basis sets

The target accuracy in the polarization-consistent basis sets $\epsilon_t = 10^{-4}$ leads to inconveniently large basis sets when the post-HF level of theory is considered – unsurprisingly, obtaining the CBS limit at the MP2 level of theory is much harder than at the SCF level of theory due to the slow convergence of the description of the electronic cusp. For this reason, the convergence threshold of the *un-ccemd-ref* set was increased to $\epsilon_t = 5 \times 10^{-4}$. The resulting basis set was then pruned and contracted analogously as above, yielding the *(un-)ccemd-3* and *(un-)ccemd-2* sets that reproduce the moments of the EMD with maximum relative errors of 10^{-3} and 10^{-2} compared to the *un-ccemd-ref* reference value. The composition and contraction schemes of the *ccemd* sets are shown in Table III.

It was found that the use of the NO-based contraction method described in Sec. II yields slightly smaller contraction errors compared to the use of SCF orbital coefficients. Still, unlike the polarization-consistent *pcemd* basis sets, the contraction of the *ccemd* sets does not result in a significant decrease in the number of functions for two reasons. First, while an atomic calculation in a minimal basis set (all functions contracted) yields the same result at the SCF level as with the full uncontracted set of primitives, at the post-HF level the additional virtual orbitals are necessary for obtaining a proper description of the system. Second, due to the more stringent basis set requirements, the correlation-consistent *ccemd* basis sets feature much more high angular momentum functions. Thus, the reduction of a few *s* and *p* functions does not play as big a role as at the SCF level of theory. We also note that the contraction scheme produced by Algorithm II in the case of the post-HF basis is unconventional, as in groups II and III of the *ccemd-2* basis set there are more *p* functions than *s* functions.

C. Benchmarks

In order to perform a thorough benchmark of the basis sets, calculations were performed for 45 diatomic molecules with bond lengths adapted from References 59 and 60, shown in Table IV. While near-HF calculations for the moments of the EMD for most of these molecules have been performed in the literature,[58,59] the results have been found to be unsatisfactory.[2] Most of the molecules are heteroatomic, and thus provide a good probe of the transferability of the basis sets.

In order to establish a compact representation, we examine the distribution of the errors in the values of the moments reproduced by different basis sets to the CBS estimate, which for the HF level was taken to be the *un-aug-pc-4* result, and for the MP2 level the *un-aug-cc-pV5Z* result. As the convergence of $\langle p^{-2} \rangle$ and $\langle p^4 \rangle$ with Gaussian basis sets is not certain,[2] we only examine $\langle p^k \rangle$ for $k \in \{-1, 1, 2, 3\}$.

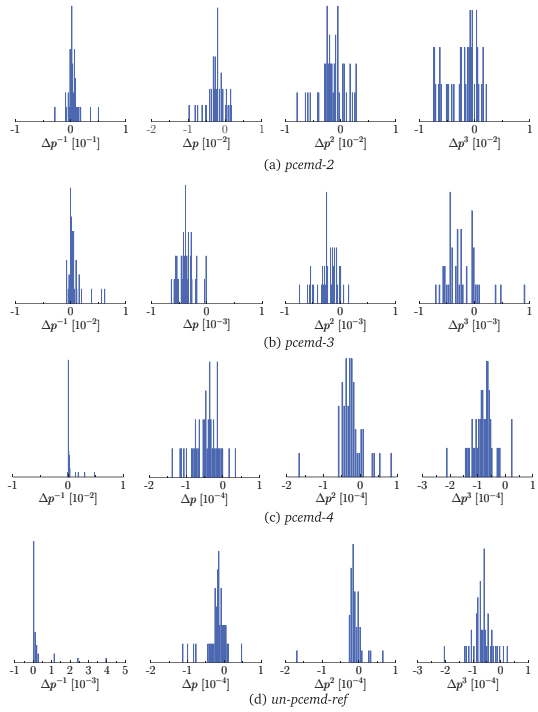


FIG. 1. *pcemd* series, errors at HF level of theory.

The amount of functions in the used basis sets is shown in Table V.

1. Hartree–Fock level of theory

The mean and mean absolute errors of the moments at the HF level of theory are given in Table VI. The error histogram for the *pcemd-N* series is shown in Figure 1, the histograms for the other basis sets are available in the supplementary material.[16]

As can be seen by comparing the errors in the *un-aug-cc-pVQZ*, *un-d-aug-cc-pVQZ,* and *un-t-aug-cc-pVQZ* results in Table VI, the reference calculation is converged with respect to the diffuse exponents. The accuracy of the chosen reference values can be estimated from the *un-aug-pc-3* errors.

Previous work[1,2,5–7] has shown completeness-optimized basis sets to be much more computationally efficient in reproducing CBS limit results than conventional basis sets, when the property one is interested in is not the energy. The data in Table VI is conclusive – the completeness-optimized sets reproduce the moments of the EMD more accurately with a smaller amount of functions than conventional basis sets.

The reference calculations show that the completeness-optimized (*co-*) basis sets systematically overestimate $\langle p^k \rangle$ for $k = -1$ and underestimate it for $k \in \{1, 2, 3\}$. This occurs because the energy is more sensitive to steep functions than the EMD at the chosen convergence threshold, and thus the energy-optimized basis sets contain tighter functions. Still, the estimated error in the *un-pcemd-ref* results is

TABLE VI. Mean errors (ME) and mean absolute errors (MAE) at Hartree–Fock level of theory in units of 10^{-4}, compared to the *un-aug-pc-4* reference values. The $\langle n_{\mathrm{bf}} \rangle$ column gives the mean amount of basis functions in the molecular calculations using the given basis set.

	$\langle p^{-1} \rangle$		$\langle p \rangle$		$\langle p^2 \rangle$		$\langle p^3 \rangle$				
	ME	MAE	ME	MAE	ME	MAE	ME	MAE	Total ME	Total MAE	$\langle n_{\mathrm{bf}} \rangle$
ccemd-2	22.802	45.287	− 22.877	22.877	− 12.802	14.370	− 35.289	36.851	− 12.041	29.846	34
un-ccemd-2	18.645	21.327	− 24.364	24.364	− 18.267	18.267	− 43.499	43.499	− 16.871	26.864	53
ccemd-3	− 0.493	2.194	− 2.488	2.503	− 2.055	2.078	− 6.645	6.668	− 2.920	3.361	106
un-ccemd-3	− 0.496	2.191	− 2.493	2.508	− 2.096	2.120	− 7.142	7.142	− 3.057	3.490	118
un-ccemd-ref	0.222	0.418	− 0.344	0.358	− 0.657	0.664	− 4.975	4.975	− 1.438	1.604	191
pcemd-2	36.767	69.273	− 24.068	27.329	− 14.344	21.838	− 22.951	27.018	− 6.149	36.364	19
un-pcemd-2	24.772	30.575	− 22.194	22.194	− 17.396	17.564	− 28.412	29.190	− 10.807	24.881	44
pcemd-3	6.254	7.748	− 3.717	3.717	− 2.266	2.373	− 2.284	3.155	− 0.503	4.248	44
un-pcemd-3	5.584	6.996	− 3.998	3.998	− 2.540	2.675	− 2.574	3.476	− 0.882	4.286	73
pcemd-4	3.164	3.239	− 0.459	0.482	− 0.244	0.345	− 0.797	0.819	0.416	1.221	95
un-pcemd-4	3.054	3.129	− 0.462	0.485	− 0.231	0.304	− 0.716	0.745	0.411	1.166	127
un-pcemd-ref	2.346	2.403	− 0.205	0.243	− 0.119	0.183	− 0.654	0.671	0.342	0.875	156
coemd-2	42.989	46.577	− 29.108	29.108	− 19.418	21.983	− 16.054	37.518	− 5.398	33.797	44
coemd-3	3.378	4.513	− 2.280	2.280	− 0.900	1.039	− 3.162	3.256	− 0.741	2.772	79
coemd-4	1.324	1.477	− 0.261	0.321	− 0.134	0.197	− 0.788	0.791	0.035	0.696	116
coemd-ref	0.908	1.085	− 0.173	0.200	− 0.113	0.134	− 0.781	0.783	− 0.040	0.550	153
aug-pc-2	6.626	7.052	− 0.677	8.620	6.635	12.649	20.499	28.245	8.271	14.141	75
un-aug-pc-2	2.775	3.401	− 0.887	0.887	− 0.991	0.991	− 7.562	7.562	− 1.666	3.210	104
aug-pc-3	1.621	1.985	3.033	5.555	11.340	11.400	30.049	31.367	11.511	12.577	146
un-aug-pc-3	0.214	0.409	− 0.042	0.042	− 0.037	0.038	− 0.577	0.577	− 0.110	0.266	185
aug-pc-4	0.945	1.298	4.066	4.456	8.663	8.699	12.714	12.884	6.597	6.834	241
aug-cc-pVDZ	11.611	12.100	− 7.365	7.854	− 11.765	13.116	− 35.161	36.046	− 10.670	17.279	42
un-aug-cc-pVDZ	10.336	10.697	− 6.578	6.578	− 7.772	7.772	− 21.222	21.222	− 6.309	11.567	66
aug-cc-pVTZ	3.710	3.968	− 1.845	1.948	− 3.192	3.346	− 12.100	12.102	− 3.357	5.341	84
un-aug-cc-pVTZ	3.379	3.592	− 1.540	1.540	− 1.666	1.666	− 7.725	7.725	− 1.888	3.631	110
aug-cc-pVQZ	1.602	1.679	− 0.260	0.464	− 0.288	0.720	− 2.760	2.872	− 0.427	1.434	148
un-aug-cc-pVQZ	1.506	1.602	− 0.432	0.432	− 0.424	0.424	− 2.412	2.412	− 0.440	1.218	176
un-d-aug-cc-pVQZ	1.535	1.837	− 0.419	0.419	− 0.445	0.445	− 2.424	2.424	− 0.438	1.281	222
un-t-aug-cc-pVQZ	1.575	1.586	− 0.386	0.386	− 0.428	0.428	− 2.421	2.421	− 0.415	1.206	269

extremely satisfactory compared to the used CBS convergence threshold.

As has been previously noted,[2] the properties of the core electrons are not correctly reproduced by the *pc-N* series, as the error in $\langle p^k \rangle$ tends to increase with k. However, decontracting the sets leads to much better results. Contractions do not have a significant effect on the results of the *ccemd*, *pcemd* and the *cc-pVXZ* calculations.

2. MP2 level of theory

The mean and mean absolute errors of the moments at the MP2 level of theory are given in Table VII. The error histogram for the *ccemd-N* series is shown in Figure 2, the histograms for the other basis sets are available in the supplementary material.[16] Surprisingly, the *aug-pc-3* and *aug-pc-4* basis sets reproduce better results at the MP2 level of theory than at the HF level of theory, although the sets have not been parametrized for post-HF calculations. In contrast, the performance of the completeness-optimized *pcemd* and *coemd* basis sets is much worse at the post-HF level – as the basis set

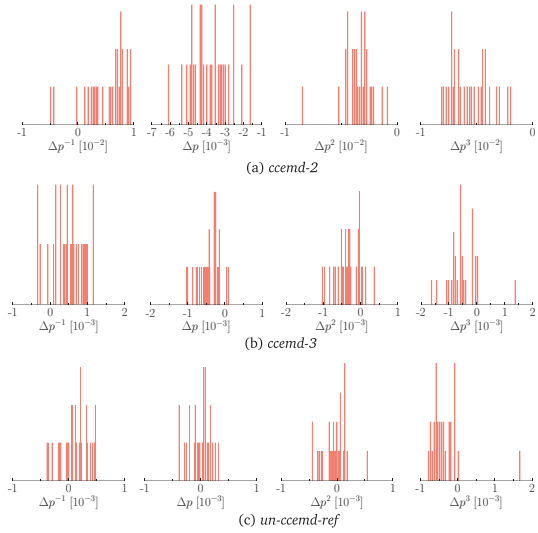


FIG. 2. *ccemd* series, errors at MP2 level of theory.

TABLE VII. Mean errors (ME) and mean absolute errors (MAE) at MP2 level of theory in units of 10^{-4}, compared to the *un-aug-cc-pV5Z* reference values. The $\langle n_{\mathrm{bf}} \rangle$ column gives the mean amount of basis functions in the molecular calculations using the given basis set.

	$\langle p^{-1} \rangle$		$\langle p \rangle$		$\langle p^2 \rangle$		$\langle p^3 \rangle$				
	ME	MAE	ME	MAE	ME	MAE	ME	MAE	Total ME	Total MAE	$\langle n_{\mathrm{bf}} \rangle$
ccemd-2	51.766	58.531	− 37.865	37.865	− 34.987	34.987	− 56.149	56.149	− 19.309	46.883	33
un-ccemd-2	41.982	49.033	− 30.185	30.185	− 24.720	24.795	− 45.145	45.714	− 14.517	37.432	52
ccemd-3	4.926	5.637	− 4.301	4.415	− 3.299	3.713	− 5.137	6.137	− 1.953	4.976	108
un-ccemd-3	4.819	5.570	− 4.069	4.190	− 2.818	3.264	− 5.632	6.601	− 1.925	4.906	120
un-ccemd-ref	1.108	2.232	− 0.068	1.443	− 0.409	1.568	− 3.429	4.605	− 0.700	2.462	193
pcemd-2	45.144	146.520	− 78.655	78.655	− 64.880	64.880	− 45.245	48.763	− 35.909	84.705	17
un-pcemd-2	81.975	81.975	− 37.967	37.967	− 24.813	24.813	− 24.302	26.059	− 1.277	42.703	43
pcemd-3	40.080	45.925	− 23.058	23.186	− 22.944	23.490	− 17.958	20.910	− 5.970	28.378	40
un-pcemd-3	32.833	39.309	− 16.217	16.582	− 9.570	10.324	− 1.358	5.869	1.422	18.021	70
pcemd-4	18.175	18.257	− 9.394	9.394	− 10.131	10.131	− 7.756	7.852	− 2.276	11.408	94
un-pcemd-4	17.201	17.327	− 7.153	7.254	− 4.808	5.411	− 1.237	2.016	1.001	8.002	126
un-pcemd-ref	14.383	14.405	− 6.015	6.092	− 4.135	4.537	− 1.162	1.667	0.768	6.675	153
coemd-2	99.512	100.618	− 51.892	51.892	− 37.956	37.956	− 15.525	34.713	− 1.465	56.295	46
coemd-3	25.666	32.889	− 13.732	14.199	− 10.077	10.870	− 7.038	7.838	− 1.295	16.449	84
coemd-4	16.816	16.816	− 7.606	7.787	− 5.743	6.431	− 2.433	3.244	0.259	8.569	121
coemd-ref	12.869	12.869	− 4.575	4.600	− 3.129	3.508	− 1.173	1.647	0.998	5.656	161
aug-pc-2	22.654	30.652	− 14.657	16.251	− 16.547	21.610	− 9.852	17.807	− 4.600	21.580	82
un-aug-pc-2	14.771	14.771	− 5.865	5.865	− 5.437	5.465	− 11.034	11.034	− 1.891	9.284	109
aug-pc-3	5.640	5.640	− 3.193	6.806	− 2.325	9.834	11.653	13.059	2.944	8.835	159
un-aug-pc-3	0.962	2.287	0.614	1.827	0.494	1.687	− 0.132	0.545	0.485	1.586	195
aug-pc-4	− 0.311	0.808	1.010	3.341	− 0.083	7.586	0.392	10.167	0.252	5.476	261
un-aug-pc-4	− 2.842	2.886	2.363	2.363	2.105	2.139	1.300	1.300	0.732	2.172	302
aug-cc-pVDZ	57.118	62.230	− 34.132	34.132	− 48.133	48.133	− 80.021	80.021	− 26.292	56.129	40
un-aug-cc-pVDZ	46.047	51.582	− 22.984	23.206	− 25.512	25.637	− 40.846	40.846	− 10.824	35.318	62
aug-cc-pVTZ	24.518	24.518	− 15.161	15.161	− 22.280	22.280	− 34.657	34.657	− 11.895	24.154	82
un-aug-cc-pVTZ	18.877	18.877	− 7.509	7.509	− 7.331	7.914	− 14.240	14.303	− 2.551	12.151	104
aug-cc-pVQZ	10.369	10.369	− 7.213	7.213	− 11.971	11.971	− 15.925	15.925	− 6.185	11.370	144
un-aug-cc-pVQZ	7.205	7.205	− 2.243	2.314	− 1.858	2.600	− 3.599	3.941	− 0.124	4.015	169
un-d-aug-cc-pVQZ	7.465	7.465	− 2.236	2.307	− 1.900	2.626	− 3.638	3.975	− 0.077	4.093	215
un-t-aug-cc-pVQZ	7.276	7.276	− 2.221	2.291	− 1.906	2.628	− 3.647	3.982	− 0.125	4.044	260

requirements of the level of theory change, so, too, must the basis set.

The performance of the completeness-optimized basis sets is insuperable also at the post-HF level of theory. While the *ccemd-3* set is much smaller than the *aug-cc-pVQZ* set, it is more than twice as accurate. Although the same overestimation-underestimation behavior is seen for the *co-* sets also at the MP2 level of theory, the errors in the *un-ccemd-ref* set compare very favorably to the used CBS limit convergence threshold.

V. DISCUSSION AND CONCLUSIONS

Using the completeness-optimization scheme, we parametrized contracted polarization-consistent *pcemd* basis sets that are completeness-optimized for the modeling of the electron momentum density. The introduction of the use of stability analysis of the SCF wave function during the optimization resulted in minor changes in the compositions of the primitive basis sets, as compared to the *coemd* sets introduced in Ref. 2. In spite of these differences, the two sets were found to reproduce results of a similar accuracy. The contraction of the primitive *un-pcemd-n* sets to the *pcemd-n* sets was found to yield significant reductions in the amount of basis functions.

Next, the completeness-optimization approach was shown to work also at the post-HF level of theory, and to converge towards the corresponding CBS limit. However, due to more the stringent basis set requirements at the post-HF level of theory the convergence with respect to the size of the basis set is much slower and contractions are less cost-effective than at the SCF level of theory.

The deviation of the results computed using the CBS limit *un-pcemd-ref* and *un-ccemd-ref* sets from the CBS limits estimated with the *un-aug-pc-4* and *un-aug-cc-pV5Z* basis sets at HF and MP2 levels of theory, respectively, were found to be similar to the used convergence threshold of the completeness-optimization algorithm.

The *pcemd* and *ccemd* basis sets, pruned and contracted from the CBS limit *un-pcemd-ref* and *un-ccemd-ref* sets, respectively, were found to outperform conventional basis sets in terms of accuracy and computational performance at the HF and MP2 levels of theory, respectively. The contraction of the basis sets was shown not to compromise the accuracy of the results.

In the current and our preceding article we have demonstrated the applicability of the completeness-optimization paradigm to the electron momentum density. We have introduced automatic algorithms that can be used for *black-box optimization of contracted basis sets adapted for the computation of any property at any level of theory*. Due to the straightforward applicability of the algorithms, we expect them to be of great practical importance for the generation of compact, application-specific basis sets giving near-CBS limit results. The used completeness-optimization algorithms are freely available as part of the ERKALE program suite.[51]

ACKNOWLEDGMENTS

This work has been supported by the Jenny and Antti Wihuri foundation, the Academy of Finland through its Centers of Excellence program and Project Nos. 1127462 and 1259526, and by the research funds of the University of Helsinki (Grant No. 490064). The computational resources of CSC – IT Center for Science Ltd. (Espoo, Finland) are gratefully acknowledged.

[1] P. Manninen and J. Vaara, J. Comput. Chem. **27**, 434 (2006).
[2] J. Lehtola, P. Manninen, M. Hakala, and K. Hämäläinen, J. Chem. Phys. **137**, 104105 (2012).
[3] A. J. Thakkar, Adv. Chem. Phys. **128**, 303 (2004).
[4] P. Kaijser and V. H. Smith, Adv. Quantum Chem. **10**, 37 (1977).
[5] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara, J. Chem. Phys. **129**, 124102 (2008).
[6] S. Ikäläinen, P. Lantto, P. Manninen, and J. Vaara, Phys. Chem. Chem. Phys. **11**, 11404 (2009).
[7] S. Ikäläinen, M. Romalis, P. Lantto, and J. Vaara, Phys. Rev. Lett. **105**, 153001 (2010).
[8] F. Jensen, J. Chem. Phys. **115**, 9113 (2001).
[9] F. Jensen, J. Chem. Phys. **116**, 7372 (2002).
[10] F. Jensen, J. Chem. Phys. **117**, 9234 (2002).
[11] F. Jensen, J. Phys. Chem. A **111**, 11198 (2007).
[12] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[13] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[14] D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358 (1993).
[15] A. K. Wilson, T. van Mourik, and T. H. Dunning, J. Mol. Struct.: THEOCHEM **388**, 339 (1996).
[16] See supplementary material at http://dx.doi.org/10.1063/1.4788635 for the revised completeness-optimization algorithms, the *pcemd* and *ccemd* basis sets, and the error histograms for the calculations at the HF and MP2 levels of theory for the studied basis sets.
[17] K. Nygård, M. Hakala, S. Manninen, K. Hämäläinen, M. Itou, A. Andrejczuk, and Y. Sakurai, Phys. Rev. B **73**, 024208 (2006).
[18] K. Nygård, M. Hakala, T. Pylkkänen, S. Manninen, T. Buslaps, M. Itou, A. Andrejczuk, Y. Sakurai, M. Odelius, and K. Hämäläinen, J. Chem. Phys. **126**, 154508 (2007).
[19] K. Nygård, M. Hakala, S. Manninen, M. Itou, Y. Sakurai, and K. Hämäläinen, Phys. Rev. Lett. **99**, 197401 (2007).
[20] K. Nygård, M. Hakala, S. Manninen, A. Andrejczuk, M. Itou, Y. Sakurai, L. G. M. Pettersson, and K. Hämäläinen, Phys. Rev. E **74**, 031503 (2006).
[21] M. Hakala, K. Nygård, S. Manninen, L. G. M. Pettersson, and K. Hämäläinen, Phys. Rev. B **73**, 035432 (2006).
[22] M. Hakala, K. Nygård, S. Manninen, S. Huotari, T. Buslaps, A. Nilsson, L. G. M. Pettersson, and K. Hämäläinen, J. Chem. Phys. **125**, 084504 (2006).
[23] M. Hakala, K. Nygård, J. Vaara, M. Itou, Y. Sakurai, and K. Hämäläinen, J. Chem. Phys. **130**, 034506 (2009).
[24] P. Sit, Ch. Bellin, B. Barbiellini, D. Testemale, J.-L. Hazemann, T. Buslaps, N. Marzari, and A. Shukla, Phys. Rev. B **76**, 245413 (2007).
[25] B. Barbiellini, Ch. Bellin, G. Loupias, T. Buslaps, and A. Shukla, Phys. Rev. B **79**, 155115 (2009).
[26] B. Barbiellini, A. Koizumi, P. Mijnarends, W. Al-Sawai, H. Lin, T. Nagao, K. Hirota, M. Itou, Y. Sakurai, and A. Bansil, Phys. Rev. Lett. **102**, 206402 (2009).
[27] A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz, and D. Usvyat, Phys. Rev. B **81**, 165108 (2010).
[28] A. Erba, M. Itou, Y. Sakurai, R. Yamaki, M. Ito, S. Casassa, L. Maschio, A. Terentjevs, and C. Pisani, Phys. Rev. B **83**, 125208 (2011).
[29] I. Juurinen, K. Nakahara, N. Ando, T. Nishiumi, H. Seta, N. Yoshida, T. Morinaga, M. Itou, T. Ninomiya, Y. Sakurai, E. Salonen, K. Nordlund, K. Hämäläinen, and M. Hakala, Phys. Rev. Lett. **107**, 197401 (2011).
[30] D. P. Chong and S. R. Langhoff, J. Chem. Phys. **93**, 570 (1990).
[31] D. P. Chong, Can. J. Chem. **73**, 79 (1995).
[32] J. Lehtola, M. Hakala, A. Sakko, and K. Hämäläinen, J. Comput. Chem. **33**, 1572 (2012).
[33] D. M. Silver and W. C. Nieuwpoort, Chem. Phys. Lett. **57**, 421 (1978).
[34] D. M. Silver, S. Wilson, and W. C. Nieuwpoort, Int. J. Quantum Chem. **14**, 635 (1978).
[35] J. R. Mohallem, R. M. Dreizler, and M. Trsic, Int. J. Quantum. Chem., Symp. **30**, 45 (1986).
[36] H. F. M. da Costa, M. Trsic, and J. R. Mohallem, Mol. Phys. **62**, 91 (1987).
[37] J. R. Mohallem and M. Trsic, J. Chem. Phys. **86**, 5043 (1987).
[38] A. B. F. da Silva, H. F. M. da Costa, and M. Trsic, Mol. Phys **68**, 433 (1989).
[39] E. V. R. de Castro and F. E. Jorge, J. Chem. Phys. **108**, 5225 (1998).
[40] R. C. Raffenetti, J. Chem. Phys. **58**, 4452 (1973).
[41] J. Almlöf and P. R. Taylor, J. Chem. Phys. **86**, 4070 (1987).
[42] P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, Theor. Chim. Acta **77**, 291 (1990).
[43] P.-O. Widmark, B. J. Persson, and B. O. Roos, Theor. Chim. Acta **79**, 419 (1991).
[44] Chr. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[45] J. R. Hart and A. J. Thakkar, Int. J. Quantum. Chem. **102**, 673 (2005).
[46] A. G. Taube and R. J. Bartlett, Collect. Czech. Chem. Commun. **70**, 837 (2005).
[47] A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 164101 (2008).
[48] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theory Comput. **7**, 2780 (2011).
[49] Z. Rolik and M. Kállay, J. Chem. Phys. **134**, 124111 (2011).
[50] W. D. Laidig, P. Saxe, and R. J. Bartlett, J. Chem. Phys. **86**, 887 (1987).
[51] S. Lehtola, ERKALE - HF/DFT from Hel, 2012. See http://erkale.googlecode.com.
[52] J. Lehtola, M. Hakala, J. Vaara, and K. Hämäläinen, Phys. Chem. Chem. Phys. **13**, 5630 (2011).
[53] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 09, Revision C.01, Gaussian, Inc., Wallingford, CT, 2009.
[54] R. Seeger and J. A. Pople, J. Chem. Phys. **66**, 3045 (1977).
[55] See http://bse.pnl.gov for EMSL basis set exchange.
[56] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Model. **47**, 1045 (2007).
[57] The additional diffuse exponents for every angular momentum shell are obtained from the second-most and most diffuse exponents α_0 and α_1, respectively, in the *aug-cc-pVXZ* basis set of the element as $\alpha_n = \alpha_0/(\alpha_0/\alpha_1)^n$ with $n = 2$ (*d-aug-cc-pVXZ*) and $n = 2, 3$ (*t-aug-cc-pVXZ*).
[58] R. K. Pathak, B. S. Sharma, and A. J. Thakkar, J. Chem. Phys. **85**, 958 (1986).
[59] A. J. Thakkar and W. A. Pedersen, Int. J. Quantum. Chem., Symp. **38**, 327 (1990).
[60] K. P. Huber and G. Herzberg, "Constants of diatomic molecules" in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, edited by P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, see http://webbook.nist.gov.
[61] Both the *pcemd-n* and *un-pcemd-n* basis sets reproduce the CBS value of the *un-pcemd-ref* set with a maximum relative error of 10^{-n}.

Paper IV

ERKALE — A Flexible Program Package for X-ray Properties of Atoms and Molecules
Jussi Lehtola, Mikko Hakala, Arto Sakko, and Keijo Hämäläinen
Journal of Computational Chemistry 2012, 33, 1572–1585
doi:10.1002/jcc.22987

ERKALE—A Flexible Program Package for X-ray Properties of Atoms and Molecules

Jussi Lehtola,*[a] Mikko Hakala,[a] Arto Sakko,†[a] and Keijo Hämäläinen[a]

ERKALE is a novel software program for computing X-ray properties, such as ground-state electron momentum densities, Compton profiles, and core and valence electron excitation spectra of atoms and molecules. The program operates at Hartree–Fock or density-functional level of theory and supports Gaussian basis sets of arbitrary angular momentum and a wide variety of exchange-correlation functionals. ERKALE includes modern convergence accelerators such as Broyden and ADIIS and it is suitable for general use, as calculations with thousands of basis functions can routinely be performed on desktop computers. Furthermore, ERKALE is written in an object oriented manner, making the code easy to understand and to extend to new properties while being ideal also for teaching purposes.

DOI: 10.1002/jcc.22987

Introduction

Checking the reproducibility of published results is sometimes troublesome in computational science, as the path from an algorithm to a working program may be long and winding. Verifying or improving on the correct functionality of existing software packages is often not possible due to the lack of access to the source code or its restrictive licensing. On the other hand, developing a completely new code may be prohibitively expensive. The easy availability of free software thus not only increases the productivity of scientific research but also benefits society on a larger scale, as parts of existing programs can be reused for new purposes.

In the field of quantum chemistry, there is a multitude of free density-functional theory[1,2] (DFT) codes using a wide variety of approaches to represent the molecular orbitals (MOs), such as plane-waves (e.g., ABINIT,[3,4] NWChem,[5,6] and Quantum ESPRESSO[7,8]), wavelets (BigDFT,[9,10] DFT++[11,12] and M-A-D-N-E-S-S[13]), numerical atomic orbitals (OpenMX[14] and GPAW[15,16]), and numerical grids (GPAW and Octopus[17,18]), for example. However, these approaches (with the exception of multiresolution grids) become computationally problematic when hybrid DFT functionals[19] are used, due to the need to compute the exact exchange (see Hartree-Fock section).

Using a basis set consisting of Gaussian type orbitals (GTOs)[20] greatly simplifies the implementation of Hartree–Fock (HF) and hybrid DFT, as most of the necessary integrals can be computed efficiently and analytically using recursion relations.[21,22] The main drawback of the Gaussian basis set is its incorrect asymptotic behavior, notably the violation of the nuclear cusp condition and the asymptotic decay that is too fast.[23] Slater-type orbital (STO) basis sets do not have these shortcomings and can also be used to perform HF and hybrid DFT calculations when suitable fitting algorithms are used,[24,25] but a greater number of integrals need to be performed using purely numerical quadrature. This approach has been chosen in the commercial ADF code.[26,27]

However, global basis sets (e.g., STOs and GTOs) often suffer from linear dependence problems and nonsmooth basis set convergence, making it hard to approach the complete basis set limit and thus to verify the numerical accuracy of the results. Comparing properties in different molecular geometries is also nontrivial when a localized basis is used, as the basis set changes with the geometry. This is known as the basis set superposition error, which can be partly remedied by performing counterpoise calculations.[28] Even so, the complexity of the necessary counterpoise procedure increases rapidly with system size, absorbing some of the computational gains obtained with the use of a localized basis set.

The problems related to global basis sets can be solved by using multiresolution grids that grant uniform convergence and guaranteed precision, as is done, for example, in the M-A-D-N-E-S-S code.[13] Although HF and (pure and hybrid) DFT calculations using multiresolution grids have been reported in the literature[29–31] and the approach holds great promise, it is still in its infancy and has larger computational requirements than basis set calculations performed at moderate accuracy.

Despite all their shortcomings, Gaussian basis sets predominate quantum chemical calculations, as the simplicity of integral evaluation not only allows fast HF and DFT calculations but also makes it straightforward to perform post-HF calculations at, for example, Møller–Plesset[32] or coupled-cluster[33] level of theory. Free GTO codes include the linear-scaling codes Ergo,[34,35] FreeON,[36] and CP2K,[37,38] and the orbital-based ACES II[39,40] and ACES III,[40,41] MPQC,[42] NWChem,[5,6] and PSI3[43,44] codes. These programs constantly provide the usual functionalities related to energetics and static response functions, but often

[a] J. Lehtola, M. Hakala, A. Sakko, K. Hämäläinen
Department of Physics, University of Helsinki, Finland
E-mail: jussi.lehtola@helsinki.fi

†Present address: Department of Applied Physics, Aalto University, Finland.

it is difficult to find software for more specific calculations, for example, for various spectroscopic applications.

In this article, we present a novel software program—ERKALE[45]—that has been developed to remedy the situation, especially in the domain of nonresonant inelastic X-ray scattering (NRIXS) spectroscopy. The program shares substantial amounts of code with, for example, PSI3, MPQC, and CP2K through its use of the libint library[46] for the computation of two-electron integrals, and on the other hand with ABINIT, BigDFT, and GPAW through the use of the libxc library[47] for evaluating exchange-correlation (XC) energy densities and their functional derivatives in DFT calculations. The program is freely available under the terms of the GNU General Public License[48][‡].

The organization of this article is as follows. First, we provide a brief review of the theoretical methods used in the program (Theory section), and then we discuss their implementation in practice (Implementation section). Next, we present applications of the code (Applications section) to computing ground-state electron momentum density (EMD) properties and the Compton profile (Electron momentum density and Compton profile section), completeness-optimization of basis sets (Completeness-optimized basis sets section) as well as the modeling of core and valence electron excitations (Core electron excitation and Valence electron excitations sections, respectively) in a simple model system: the water dimer. Finally, we conclude with an outlook on the future prospects of the code (Conclusion and Outlook section). All equations in the current work are given in atomic units. Furthermore, the coordinate representations of the basis functions and MOs are assumed to have real values.

Theory

Basis set

ERKALE operates with a basis set consisting of segmented contractions of GTOs[20,23] in either cartesian [Eqs. (1) and (2)] or spherical [Eqs. (1) and (3)] form.

$$\phi_\mu(\mathbf{r}) = \Omega_{\boldsymbol{\lambda}_\mu}(\mathbf{r}) \sum_p d_{\mu p} e^{-\zeta_{\mu p}(\mathbf{r}-\mathbf{r}_\mu)^2} \tag{1}$$

$$\Omega^{\text{cart}}_{\boldsymbol{\lambda}_\mu} = N^{\text{cart}}_\mu (x-x_\mu)^{l_\mu}(y-y_\mu)^{m_\mu}(z-z_\mu)^{n_\mu} \tag{2}$$

$$\Omega^{\text{sph}}_{\boldsymbol{\lambda}_\mu} = N^{\text{sph}}_\mu |\mathbf{r}-\mathbf{r}_\mu|^{\lambda_\mu} Y_{\lambda_\mu M}(\widehat{\mathbf{r}-\mathbf{r}_\mu}). \tag{3}$$

Here, $\phi_\mu(\mathbf{r}) = \langle \mathbf{r} | \mu \rangle$ is the coordinate representation of the μth basis function, centered at $\mathbf{r}_\mu = (x_\mu, y_\mu, z_\mu)$. $\Omega_{\boldsymbol{\lambda}_\mu}$ contains the angular part of the basis function and $\boldsymbol{\lambda}_\mu = (l_\mu, m_\mu, n_\mu)$ is its angular momentum. $d_{\mu p}$ and $\zeta_{\mu p}$ are the contraction coefficient and exponent, respectively, of the pth primitive Gaussian function in the μth basis function. N^{cart}_μ and N^{sph}_μ are normalization constants, such that $\langle \mu | \mu \rangle = 1$.

Basis functions, which are located on the same center $\mathbf{r}_\mu$, share the same primitive contraction and have the same angular momentum $\lambda_\mu = l_\mu + m_\mu + n_\mu$ [see Eqs. (2) and (3)], form a shell of functions. The functions $Y_{LM}(\hat{\mathbf{r}})$ in Eq. (3) are spherical harmonics in the real form, $\hat{\mathbf{r}}$ denoting the unit vector in the direction of $\mathbf{r}$. The functions Y_{LM} can be expressed as a sum of the cartesian functions of the same shell.[23] All operations with basis functions, such as evaluation of the basis functions at a point, are performed on a shell basis in ERKALE.

Self-consistent field equations

Expressing the spatial part of the MOs in the basis set as

$$\psi_i^\sigma(\mathbf{r}) = \sum_\mu C^\sigma_{\mu i} \phi_\mu(\mathbf{r}), \tag{4}$$

where ψ_i^σ is the ith MO of spin σ and $C^\sigma_{\mu i}$ is the coefficient of the μth basis function in its expansion, the sum running over all basis functions μ in the basis set, one famously obtains the Roothaan–Hall self-consistent field (SCF) equation for the MO coefficients $\mathbf{C}$[49]

$$\mathbf{FC} = \mathbf{ESC} \tag{5}$$

or its spin-polarized counterpart, the Pople–Nesbet equations[50]

$$\begin{cases} \mathbf{F}^\alpha \mathbf{C}^\alpha &= \mathbf{E}^\alpha \mathbf{S} \mathbf{C}^\alpha, \\ \mathbf{F}^\beta \mathbf{C}^\beta &= \mathbf{E}^\beta \mathbf{S} \mathbf{C}^\beta. \end{cases} \tag{6}$$

α and β denote the states for spin up and spin down electrons, respectively, and $\mathbf{S}$ is the overlap matrix with elements $S_{\mu\nu} = \langle \mu | \nu \rangle$. The Fock matrix $\mathbf{F}$ is defined as

$$\mathbf{F}^\sigma = \mathbf{H}^{\text{core}} + \mathbf{J} + \mathbf{K}^\sigma, \tag{7}$$

where $\mathbf{H}^{\text{core}}$ contains the kinetic energy of the electrons $\mathbf{T}$ and the Coulombic attraction of the nuclei $\mathbf{V}^{\text{nuc}}$, with matrix elements

$$\mathbf{H}^{\text{core}} = \mathbf{T} + \mathbf{V}^{\text{nuc}}, \tag{8}$$

$$T_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle, \tag{9}$$

$$V^{\text{nuc}}_{\mu\nu} = \left\langle \mu \left| -\sum_N \frac{Z_N}{|\mathbf{r}-\mathbf{R}_N|} \right| \nu \right\rangle. \tag{10}$$

The matrix $\mathbf{J}$ in Eq. (7) describes the static Coulomb repulsion of the electrons with themselves

$$J_{\mu\nu} = \sum_{\rho\tau} P_{\rho\tau} \langle \mu\nu | \rho\tau \rangle, \tag{11}$$

where the sums over ρ and τ are taken over all basis functions. $\mathbf{P}$ is the total electron density matrix

$$\mathbf{P} = \mathbf{P}^\alpha + \mathbf{P}^\beta, \tag{12}$$

where the spin-density matrices $\mathbf{P}^\alpha$ and $\mathbf{P}^\beta$ are given by

$$P^\sigma_{\mu\nu} = \sum_i n_i^\sigma C^\sigma_{\mu i} C^\sigma_{\nu i}, \tag{13}$$

[‡](ERKALE is licensed under the GNU General Public License version 2, or (at your option) any later version)

in which n_i^σ is the occupation number of orbital i of spin σ, the sum over i running over all MOs. $(\mu\nu\,|\rho\tau\,)$ in Eq. (11) denotes an electron repulsion integral (ERI)

$$(\mu\nu\,|\rho\tau\,) = \int d^3r \int d^3r' \frac{\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\phi_\rho(\mathbf{r}')\phi_\tau(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \quad (14)$$

Next, we will discuss the form of the XC matrix $\mathbf{K}^\sigma$, which is different in HF and DFT.

Hartree–Fock

In HF, the matrix $\mathbf{K}^\sigma$ in Eq. (7) contains the exchange contribution to the Fock matrix, and its elements are

$$K_{\mu\nu}^{\mathrm{HF},\sigma} = -\sum_{\rho\tau} P_{\rho\tau}^\sigma\,(\mu\rho\,|\tau\nu\,). \quad (15)$$

In the closed-shell case, $\mathbf{P}^\sigma$ is replaced with $\mathbf{P}/2$ in Eq. (15).

The energy expressions for restricted (RHF) and unrestricted HF (UHF) are[51]

$$E^{\mathrm{RHF}} = \mathrm{Tr}\left[\mathbf{P}\left(\mathbf{H}^{\mathrm{core}} + \frac{1}{4}\left(2\mathbf{J} + \mathbf{K}^{\mathrm{HF}}\right)\right)\right], \quad (16)$$

$$E^{\mathrm{UHF}} = \mathrm{Tr}\left[\mathbf{P}\left(\mathbf{H}^{\mathrm{core}} + \frac{1}{2}\left(\mathbf{J} + \mathbf{K}^{\mathrm{HF},\alpha} + \mathbf{K}^{\mathrm{HF},\beta}\right)\right)\right]. \quad (17)$$

Restricted open-shell HF (ROHF) is also available in ERKALE by the constraint of spin-contamination in the UHF Fock operators.[52] In this approach, the energy expression retains its UHF form.

Kohn–Sham DFT

In DFT calculations, $\mathbf{K}^\sigma$ contains both exchange and correlation. For unrestricted DFT (UDFT) using functionals in the generalized-gradient approximation[53] (GGA), the elements of $\mathbf{K}^\sigma$ are[54]

$$K_{\mu\nu}^{\mathrm{UDFT},\sigma} = \int \left[\frac{\delta f^{\mathrm{xc}}}{\delta\rho^\sigma(\mathbf{r})}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) + \left(2\frac{\delta f^{\mathrm{xc}}}{\delta\gamma^{\sigma\sigma}(\mathbf{r})}\nabla\rho^\sigma(\mathbf{r}) + \frac{\delta f^{\mathrm{xc}}}{\delta\gamma^{\sigma\sigma'}(\mathbf{r})}\nabla\rho^{\sigma'}(\mathbf{r})\right)\cdot\nabla(\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}))\right] d^3r \quad (18)$$

where $f^{\mathrm{xc}} = f^{\mathrm{xc}}(\rho^\alpha(\mathbf{r}), \rho^\beta(\mathbf{r}), \nabla\rho^\alpha(\mathbf{r}), \nabla\rho^\beta(\mathbf{r}))$ is the GGA XC energy density and $\delta f/\delta\rho$ denotes the functional derivative. The spin variable σ' is defined in Eq. (18) by $\sigma' \neq \sigma$. The quantity $\gamma^{\sigma\sigma'}$ is given by

$$\gamma^{\sigma\sigma'}(\mathbf{r}) = \nabla\rho^\sigma(\mathbf{r})\cdot\nabla\rho^{\sigma'}(\mathbf{r}), \quad (19)$$

in which ρ^σ is the spin density

$$\rho^\sigma(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}^\sigma \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}), \quad (20)$$

where the sums over μ and ν run over all basis functions.

In the case of restricted DFT (RDFT), the matrix $\mathbf{K}$ is simply

$$K_{\mu\nu}^{\mathrm{RDFT}} = \int \left[\frac{\delta f^{\mathrm{xc}}}{\delta\rho(\mathbf{r})}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) + 2\frac{\delta f^{\mathrm{xc}}}{\delta\gamma(\mathbf{r})}\nabla\rho(\mathbf{r})\cdot\nabla(\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}))\right] d^3r, \quad (21)$$

where $\rho(\mathbf{r}) = \rho^\alpha(\mathbf{r}) + \rho^\beta(\mathbf{r})$ is the total electron density and $\gamma(\mathbf{r}) = \nabla\rho(\mathbf{r})\cdot\nabla\rho(\mathbf{r})$. The energy expressions for DFT are[54]

$$E^{\mathrm{DFT}} = \mathrm{Tr}\left[\mathbf{P}\left(\mathbf{H}^{\mathrm{core}} + \frac{1}{2}\mathbf{J}\right)\right] + E^{\mathrm{xc}}, \quad (22)$$

$$E^{\mathrm{xc,RDFT}} = \int f^{\mathrm{xc}}(\rho(\mathbf{r}), \gamma(\mathbf{r}))d^3r, \quad (23)$$

$$E^{\mathrm{xc,UDFT}} = \int f^{\mathrm{xc}}(\rho_\alpha(\mathbf{r}), \rho_\beta(\mathbf{r}), \gamma_{\alpha\alpha}(\mathbf{r}), \gamma_{\alpha\beta}(\mathbf{r}), \gamma_{\beta\beta}(\mathbf{r}))d^3r. \quad (24)$$

Although Eqs. (18), (21), (22), (23), and (24) were presented for reasons of simplicity only for the case of GGA functionals, local spin density[2] (LDA), hybrid[19] and meta-GGA[55] functionals are supported in ERKALE as well. When hybrid functionals such as B3LYP[56] are used, a part of HF exchange [Eq. (15)] is added to Eq. (18) or (21). On the other hand, when hybrid functionals are not used, there is no need to compute Eq. (15), so the Coulomb term [Eq. (11)] can be approximated using density fitting[57,58] as[59]

$$J_{\mu\nu} \approx \sum_{AB} (\mu\nu\,|A\,)\,(A\,|B\,)^{-1} \sum_{\rho\tau} P_{\rho\tau}\,(B\,|\rho\tau\,), \quad (25)$$

where $(\mu\nu\,|A\,)$ and $(A\,|B\,)$ are three- and two-center ERIs[§] and $(A\,|B\,)^{-1}$ denotes the (A,B)-element of the inverse matrix. The sums over A and B run over the functions in the auxiliary basis set, which is by default formed automatically[60] in ERKALE.

Implementation

Formation of matrices

The matrix elements of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{V}^{\mathrm{nuc}}$ are computed in ERKALE on a shell basis by using Obara–Saika recursion routines.[21] ERIs. are computed with the libint library[46] that uses Obara–Saika[21] and Head–Gordon–Pople recursions.[22] All of these integrals are performed in the Cartesian basis and then transformed into the spherical basis if necessary. The computation of matrix elements over the Coulomb operator $1/r$ (i.e., for $\mathbf{V}^{\mathrm{nuc}}$ and ERIs) requires the evaluation of the Boys' function[20]

$$F_m(x) = \int_0^1 t^{2m} e^{-xt^2} dt, \quad (26)$$

which is computed in ERKALE using a Taylor series when $x \ll 1$. Otherwise, the Boys' function is computed as

$$F_m(x) = \frac{1}{2}x^{-m-\frac{1}{2}}\Gamma\left(m+\frac{1}{2}\right)\gamma\left(m+\frac{1}{2};x\right), \quad (27)$$

[§]The two- and three-center ERIs can be written as usual four-center ERIs as, for example, $(A\,|B\,) = (As\,|Bs\,)$, where s denotes the s-type dummy GTO given by $1 \times \exp(-0 \cdot r^2)$, respectively.

where Γ and γ are Euler's complete and incomplete Gamma functions,[61] respectively, the values of which are computed using the GNU Scientific Library.[62]

In DFT calculations, the elements of the XC potential matrix $\mathbf{K}^{\mathrm{DFT}}$ need to be integrated numerically. When pure DFT methods are used, the most computational time in large systems is spent in forming $\mathbf{K}^{\mathrm{DFT}}$. The formation of the XC potential matrix in ERKALE is based on a modern variant[63] of Becke's original atom-centric approach,[64] in which the XC integration grid is formed adaptively to converge the diagonal elements of $\mathbf{K}^{\mathrm{DFT}}$ within the wanted tolerance τ. This procedure guarantees accuracy with the minimal amount of computational time. The XC grid formation in large systems is typically found to take less time than a single SCF iteration.

Furthermore, as the initial guess density is often far from the converged density, inspired by the work of Köster *et al.*[63] ERKALE first performs a complete SCF cycle with a modest integration grid (tolerance $\tau_{\mathrm{initial}} = 10^{-4}$ a.u.) and convergence criteria. When the calculation has reached this modest stage of convergence, the grid is reformed with an increased precision ($\tau_{\mathrm{final}} = 10^{-5}$ a.u.), and the iteration is continued until full self-consistency is achieved. We have found that this choice saves computational time, while reproducing the same result as when using $\tau_{\mathrm{initial}} = \tau_{\mathrm{final}}$. The default values of τ_{initial} and τ_{final} correspond to energy errors of less than 2×10^{-4} and 2×10^{-5} a.u., respectively, for systems containing first and second row elements[63] (after the grid has been reformed with the correct SCF density). All of these control parameters can be freely modified in the input file.

The XC energy density and its functional derivatives are evaluated with the libxc library.[47] The atom-centric radial integrals in the Becke scheme are computed using Gauss–Chebyshev quadrature of the second kind.[63] The angular integrals on each radial shell are performed by default with Lebedev quadrature,[65–70] but Lobatto quadrature[71] is available as well. The local nature of the basis functions is used to screen and select only the numerically significant functions in the formation of $\mathbf{K}^{\mathrm{DFT}}$[72] by requiring that the absolute value of the basis function is at least $\epsilon = 10^{-10}$ a.u.[73] in the grid point¶. The quadrature in Eq. (18) and (21) is performed directly in the spherical basis, if applicable.

None of the aforementioned integrals have any limitation with respect to the maximum angular momentum of the used basis set, thus ERKALE supports basis sets of arbitrary angular momentum‖. Furthermore, all of the integrals are parallellized,** which is especially important when large systems are studied.

When many basis functions are included in the calculation, the necessary ERIs [Eqs. (11), (15), and (25)] might not fit in the memory. Therefore, direct SCF calculations[74] with Schwarz prescreening of ERIs are supported both for HF and DFT methods. Because of the parallellization and the possibility of computing integrals on-the-fly, calculations with thousands of basis functions can be routinely performed on desktop computers.

Self-consistent solution

Equations (5) and (6) are solved iteratively until self-consistency is attained. The iteration starts by default from an atomic ROHF guess[75]; the core guess (eigenvectors of $\mathbf{H}^{\mathrm{core}}$) is also available. The solution is obtained in a canonically orthonormalized[76] basis, with eigenvectors of $\mathbf{S}$ with eigenvalues smaller than ϵ (by default $\epsilon = 10^{-5}$) removed from the basis set.

Accelerators are used to improve the convergence of the fixed point iteration of Eqs. (5) and (6). ERKALE performs Pulay's direct inversion in the iterative subspace (DIIS),[77,78] which is a well-established tool for improving SCF convergence. As a new feature, DIIS combined with energy estimation (ADIIS),[79] which is based on the augmented Roothaan–Hall energy expression,[80] can be used to speed up the initial convergence. A novel Broyden-type secant method[81] is available as well for Fock matrix updates. We have found the Broyden method to outperform DIIS and ADIIS on many occasions, especially in calculations using the transition potential approximation (TP; see Core electron excitations section).

Applications

In the following section, we present applications of ERKALE to different X-ray spectroscopic methods, using a simple model system: the water dimer, the geometry of which is given in Table 1.

EMD and Compton profile

The EMD is analogous to the electron density in momentum space. The EMD can be computed in general in localized basis sets as[82–84]

$$n(\mathbf{p}) = \sum_{\mu\nu} P_{\mu\nu} \overline{\tilde{\phi}_\mu(\mathbf{p})} \tilde{\phi}_\nu(\mathbf{p}), \tag{28}$$

where $\tilde{\phi}_\mu$ is the Fourier transform of the μth basis function and the overline denotes complex conjugation. The Fourier transforms of the Gaussian basis functions are evaluated in ERKALE with a recursion relation.[85] In isotropic or amorphous systems, such as liquids and gases, the relevant quantity to compute is the radial EMD

$$n(p) = \int n(\mathbf{p}) d\Omega_{\mathbf{p}}. \tag{29}$$

The ground-state EMD can be measured, for example, with Compton scattering (CS) experiments at modern synchrotron

¶The screening of basis functions is performed in ERKALE shell by shell. For each shell of basis functions (centered at $\mathbf{R}_{\mathrm{sh}}$) an effective radius R_ϵ is determined by the requirement $|\phi_i(|\mathbf{r} - \mathbf{R}_{\mathrm{sh}}|)| \leq \epsilon$, when $|\mathbf{r} - \mathbf{R}_{\mathrm{sh}}| > R_\epsilon$, where ϕ_i are basis functions belonging to the studied shell of basis functions. Basis functions are prescreened on the radial grid before point-by-point screening on the angular mesh. This is done by including the shell at $\mathbf{R}_{\mathrm{sh}}$ for screening on the angular points of the radial shell (radius r) of the atom at $\mathbf{R}_0$ only if $||\mathbf{R}_{\mathrm{sh}} - \mathbf{R}_0| - r| \leq R_\epsilon$.

‖However, when extremely high values of angular momentum are used, the numerical accuracy of the recursion routines is not guaranteed.

**The parallellization is implemented using OpenMP, which can be used to speed up the computation on shared-memory computers, such as modern multicore workstations.

radiation facilities.[83,84,86] CS experiments on isotropic samples yield the isotropic Compton profile

$$J(q) = \frac{1}{2}\int_{|q|}^{\infty} pn(p)dp. \quad (30)$$

Difference Compton profiles are often studied in experiments. Although DFT does not reproduce the absolute momentum density correctly,[87–91] it often reproduces these differences well,[90,92,93] hybrid functionals likely faring better than pure functionals.[90]

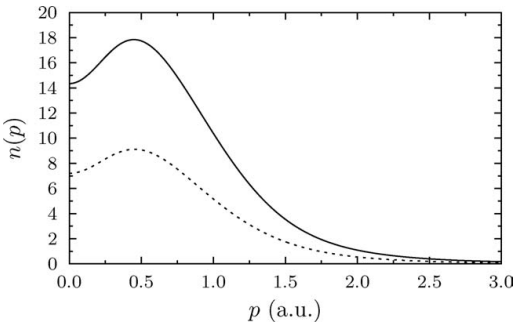


Figure 1. Radial EMDs of the water dimer (solid line) and the monomer (dashed line), calculated at HF level using the aug-cc-pVTZ basis set.

Table 1. The used water dimer geometry.

	x/Å	y/Å	z/Å
O	−1.464	0.099	0.300
H	−1.956	0.624	−0.340
H	−1.797	−0.799	0.206
O	1.369	0.146	−0.395
H	1.894	0.486	0.335
H	0.451	0.165	−0.008

ERKALE can currently compute the EMD [Eq. (28)], the radial EMD [Eq. (29)] and the isotropic Compton profile [Eq. (30)] as described in reference Ref. [90] using HF or any of the DFT methods that are implemented. As an illustration, Figure 1 shows the radial EMDs of the water dimer and of the monomer (taken as the first molecule in Table 1), computed using HF and the Dunning-style aug-cc-pVTZ basis set[94–96] obtained from the ESML basis set exchange.[97,98] Figure 2 shows the difference Compton profile between these systems, which we define as in Ref. [90] as

$$\Delta J(q) = \frac{J^{\text{dimer}}(q) - 2\,J^{\text{monomer}}(q)}{J^{\text{dimer}}(0)}. \quad (31)$$

The difference profile can be understood as a measure of the change induced on the EMD by the chemical environment.

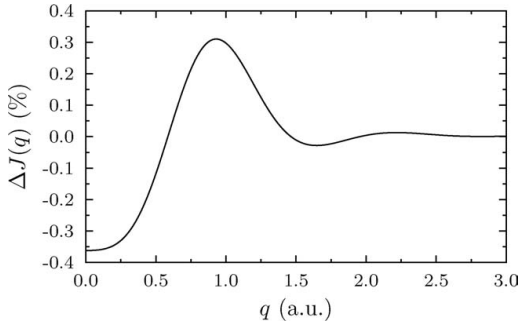


Figure 2. Difference Compton profile of the water dimer versus the monomer, calculated at HF level using the aug-cc-pVTZ basis set.

Difference profiles have been successfully used in studying, for example, temperature-dependent hydrogen bond networks in water,[99,100] configurational energetics in ice,[101] and solvation in water–ethanol mixtures.[102]

Completeness-optimized basis sets

Completeness profiles were introduced by Chong as a graphical means to study the completeness of one-electron basis sets.[103,104] The completeness profile for an atomic basis set is given by

$$Y(\alpha) = \left\langle \alpha \left| \sum_{\mu\nu} |\mu\rangle\, S^{-1}_{\mu\nu}\, \langle\nu| \right| \alpha \right\rangle, \quad (32)$$

where $|\mu\rangle$ and $|\nu\rangle$ are the functions centered on the atom, the basis set of which is investigated, and $|\alpha\rangle$ is the scanning function. When the basis set is flexible enough to represent the scanning function $Y(\alpha) \approx 1$. Although, in principle, any normalizable function can be used to probe the basis set, Gaussian scanning functions (also centered on the same atom) are conventionally used:

$$\langle \mathbf{r}\,|\alpha\,\rangle = Nr^l Y_{lm}(\hat{\mathbf{r}})e^{-\alpha r^2}, \quad (33)$$

where N is the normalization constant. In the computation of completeness profiles, it is assumed that also the basis set uses pure spherical functions [Eq. (3)]. A separate profile is then obtained for each value of l present in the basis set, while all values of m yield an identical profile, as is easily seen from Eq. (32) by using the orthogonality properties of spherical harmonics.

Conventional basis sets are energy optimized, which tends to overemphasize the tightly bound region and neglect the diffuse area.[104] Manninen and Vaara introduced completeness-optimization of basis sets[105] as a systematical as well as practical approach to the complete basis set limit. Indeed, many molecular properties have been found to converge faster with completeness-optimized basis sets than with conventional, energy optimized basis sets.[105–108] As well as the Kruununhaka basis set tool kit,[109] ERKALE can be used to plot completeness profiles and generate completeness-optimized primitive sets.

In the completeness optimization paradigm, the primitives in each angular momentum shell in the basis set are optimized so that the resulting completeness profile is as close to unity as possible in the exponent range $\alpha \in [\alpha_1, \alpha_2]$ that is relevant for the computed property[105]††. The optimization in ERKALE is performed by minimizing the measure

$$\tau_n = \left[\frac{1}{\log\alpha_2 - \log\alpha_1} \int_{\log\alpha_1}^{\log\alpha_2} [1 - Y(\alpha)]^n d\log\alpha \right]^{1/n} \quad (34)$$

where $n = 1$ corresponds to maximization of the area of completeness (as done in Refs. [105–108]) and $n = 2$ to minimization of the root mean square deviation from completeness. The completeness-optimization of the basis set using Eq. (34) is a highly nonlinear task. However, it is greatly simplified by the fact that the solution is independent of the ordering of the exponents. ERKALE performs the minimization using the Nelder–Mead simplex algorithm[110] as implemented in GSL,[62] which in our calculations systematically provided better optimized basis sets (smaller τ_n) than the stochastic differential evolution algorithm[111] used in previous work.[105]

The modeling of excitation spectra discussed in the following subsections requires rather large basis sets to be used to obtain satisfactorily converged results. X-ray absorption and X-ray Raman calculations (see Core electron excitations section) often use the IGLO-III basis set[112] for describing the excited atom, whereas Casida calculations for valence excitations often use the Sadlej basis set.[113,114] The contraction scheme for IGLO-III is 6*s*2*p*/4*s*2*p* (H) and 11*s*7*p*2*d*/7*s*6*p*2*d* (O), and for the Sadlej basis, it is 6*s*4*p*/3*s*2*p* (H) and 10*s*6*p*4*d*/5*s*3*p*2*d* (O). For comparison, the aforementioned aug-cc-pVTZ basis is 6*s*3*p*2*d*/4*s*3*p*2*d* (O) and 11*s*6*p*3*d*2*f*/5*s*4*p*3*d*2*f* (H). For this work, we have generated a 7*s*6*p* basis for hydrogen and a 10*s*10*p*5*d* basis for oxygen using the root mean square metric in Eq. (34). The generated bases are more complete than IGLO-III and Sadlej basis sets, and more *sp*- and *spd*-complete than the aforementioned aug-cc-pVTZ basis set for hydrogen and oxygen, respectively. As an illustration, the completeness profiles of the basis sets taken from Ref. [98] are shown in Figure 3. The completeness-optimized basis sets are available in Supporting Information.

Core electron excitations

In X-ray absorption spectroscopy (XAS), the energy of the incoming photon beam is in the vicinity of a core orbital binding energy of the investigated sample. Because of this, XAS is an element-selective probe of the electronic structure. When XAS is performed in the soft X-ray regime (e.g., *K*-edges of second-row elements), it becomes highly sensitive to the sample surface.[115] Because of the high absorption, the experiments are often performed in high-vacuum environments. In contrast, in X-ray Raman scattering (XRS), only the energy transfer is close to the core electron binding energy, making XRS a bulk probe also for excitations in the soft X-ray energy range and enabling, for example, measurements of the oxygen *K*-edge in complicated experimental setups, such as high-pressure measurements in diamond anvil cells.[115,116] The information contents of XRS is analogous to that obtained through XAS. However, the XRS spectrum is also dependent on the momentum transfer, which can be used to extract supporting information not available in XAS experiments.[117–120]

XAS and XRS spectra are computed within ERKALE in the TP approximation,[121] which is based on Slater's transition state approximation.[122,123] The same approach is also available in, for example, the StoBe-deMon,[124] CP2K,[37,38] and the GPAW[15] codes for modeling XAS. Although XRS spectra have been computed with TP in Gaussian basis sets before[125] using a post-analysis tool to StoBe-deMon, ERKALE is to our knowledge the first published code to perform this task as a standard feature. We note that XAS/XRS can also be computed using the Bethe–Salpeter equation,[126] for example, with the OCEAN code,[127] or in the real-space multiple scattering approach[120,128,129] with the FEFF code.[130]

In the TP approximation, an occupation of 0.5 is set on the core orbital with spin σ, which is localized on the atom being excited. The system is left with a net charge $+0.5$‡‡. For systems with a closed-shell ground state, the excitation is conventionally performed for the spin α core orbital ($\sigma = \alpha$). On the other hand, when the ground state is spin polarized, a separate calculation needs to be performed for both α and β excitations. The main advantage of TP is that all excitation energies and transition rates are obtained with a single calculation. TP has also been shown to reproduce experimental XAS spectra well, see Ref. [131] for a review in the case of water.

However, TP does not produce a reliable energy scale,[131] as it neglects the relaxation of the electrons in the presence of the full core hole, the magnitude of which depends on the excited center. When aggregate spectra are computed from cluster snapshots or from nonequivalent sites in a molecule, a uniform energy scale needs to be established. This can be done by computing the difference in the total energy of the first core-excited state (XCH)§§ and the ground state for each configuration or nonequivalent center.[132] The first TP transition is then set to occur at this energy.[132,133] After an absolute energy scale has been established, the resulting spectra need still to be shifted to account for, for example, deficiencies in the approximate DFT functional as well as relativistic and basis set incompleteness effects.[134]

The determination of the excited core state is performed dynamically in ERKALE during the SCF iteration process by computing the centers $\langle \mathbf{r}_i \rangle$ and rms widths Δr_i of each occupied orbital ψ_i^σ of the excited spin σ

$$\langle \mathbf{r}_i \rangle = \left\langle \psi_i^\sigma \left| \mathbf{r} \right| \psi_i^\sigma \right\rangle \quad (35)$$

$$\Delta r_i = \sqrt{\left\langle \psi_i^\sigma \left| (\mathbf{r} - \langle \mathbf{r}_i \rangle)^2 \right| \psi_i^\sigma \right\rangle} \quad (36)$$

†† See Refs. [105–107] on how to choose the relevant values for α_1 and α_2.

‡‡ This refers to gas-phase calculations (i.e., molecules and clusters) as computed by ERKALE.

§§ By the first core-excited state, we mean a full core hole with the excited electron placed on the lowest unoccupied state.

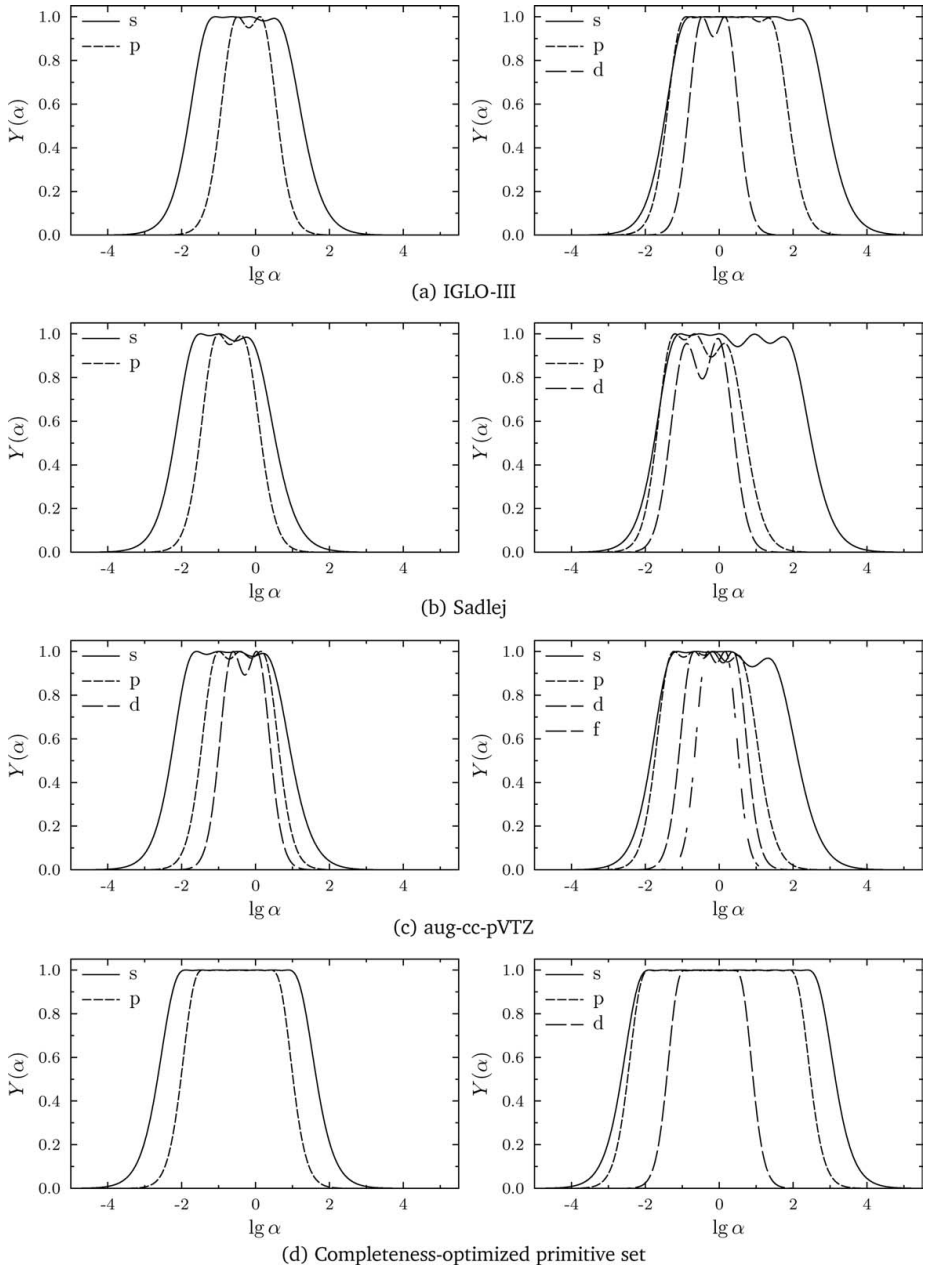


Figure 3. Completeness profiles for the studied basis sets. The profile for hydrogen is on the left, whereas oxygen is on the right.

and then by choosing the orbital i that is the most localized on the excited atom at $\mathbf{R}_0$, determined by the minimal value of the product $\Delta r_i |\mathbf{r}_i - \mathbf{R}_0|$.

When calculating systems with several symmetry equivalent sites such as benzene, localization of the excited core state (here carbon 1*s*) can be enforced by freezing the other similar (carbon 1*s*) states in the system.[131] The frozen core for each atom is determined in ERKALE from a ground-state calculation by computing the matrix elements of $\mathbf{r}$ around the atom (located at $\mathbf{R}_I$)

$$r_{ij}^2 = \sqrt{\left\langle \psi_i^\sigma \left| (\mathbf{r} - \mathbf{R}_l)^2 \right| \psi_j^\sigma \right\rangle}, \quad (37)$$

in the space of occupied orbitals i,j and diagonalizing the resulting matrix to obtain the corresponding eigenvectors and eigenvalues[¶¶]. The actual freezing of the orbitals is performed in ERKALE in the SCF iteration after the formation of the Kohn–Sham Fock matrix by transforming it into the MO basis

$$\mathbf{F}_{\mathrm{MO}}^\sigma = (\mathbf{C}^\sigma)^{\mathrm{T}} \mathbf{F}_{\mathrm{AO}}^\sigma \mathbf{C}^\sigma \quad (38)$$

and setting the (n,f) and (f,n) elements to zero, where f and n stand for indices of frozen and nonfrozen orbitals, respectively. Then the Fock operator is transformed back into the AO basis with

$$\mathbf{F}_{\mathrm{AO}}^\sigma = \mathbf{S}\mathbf{C}^\sigma \mathbf{F}_{\mathrm{MO}}^\sigma (\mathbf{C}^\sigma)^{\mathrm{T}} \mathbf{S}. \quad (39)$$

The frozen core approximation presented above has been found to have little effect on the total and orbital energies in molecular systems.[135] A commonly used approximation to the frozen core is the effective core potential (ECP) approach, which has been shown to have little effect on the ground state geometry, spectroscopic parameters, and the dissociation energy.[136] Although ECPs can be used to reduce the computational demands by removing the core electrons from the calculation and also incorporate relativistic effects, for light elements, the first is not a substantial gain (as there are but few core electrons and their integrals are effectively screened out) and the second are not notable. Thus, in this case using ECPs just introduces an unnecessary approximation.

Furthermore, most ECPs reported in the literature for Gaussian basis sets have been parameterized against HF (or relativistic Dirac–Fock) calculations. Although core orbitals are similar in HF and DFT, being able to use (frozen) core orbitals corresponding exactly to the used XC functional can be deemed important, especially when developing new XC functionals. The method summarized above is more flexible also in that the used (core and valence) basis sets can be easily changed.

In the dipole approximation, the XAS spectrum is proportional to the dipole oscillator strength

$$f_i(\omega) = \sum_f \left(\epsilon_f^\sigma - \epsilon_i^\sigma\right) \left| \left\langle \psi_f^\sigma \left| \boldsymbol{\varepsilon} \cdot \mathbf{r} \right| \psi_i^\sigma \right\rangle \right|^2 \delta\left(\epsilon_i^\sigma + \omega - \epsilon_f^\sigma\right), \quad (40)$$

where $\boldsymbol{\varepsilon}$ is the polarization of the incoming photon beam, $\left|\psi_i^\sigma\right\rangle$ and $\left|\psi_f^\sigma\right\rangle$ are the excited initial one-particle state and the final one-particle state with energies ϵ_i^σ and ϵ_f^σ, respectively, and the sum runs over all unoccupied states f of the TP calculation. The XRS spectrum, on the other hand, is proportional to the dynamic structure factor

$$S(\mathbf{q},\omega) = \sum_f \left| \left\langle \psi_f^\sigma \left| \exp(i\mathbf{q}\cdot\mathbf{r}) \right| \psi_i^\sigma \right\rangle \right|^2 \delta\left(\epsilon_i^\sigma + \omega - \epsilon_f^\sigma\right), \quad (41)$$

where $\mathbf{q}$ is the momentum transfer.

The TP ground state, that is, the occupied orbitals in the TP calculation, can be accurately modelled using a much smaller basis set than what is required for obtaining the virtual orbitals relevant for XAS and XRS calculations [the final states in Eqs. (40) and (41)]. For this reason, the double-basis set procedure[137,138] is often used to improve the description of the unoccupied states after SCF convergence (of the occupied orbitals) has been achieved. In this method, the basis set is augmented with extremely diffuse functions on the excited atom, keeping the occupied orbitals constant. The Kohn–Sham Fock operator is then diagonalized in the unoccupied space in order to obtain the virtual orbitals and their energies.

Equations (35), (36), and (40) are easily computed with moment integrals.[21] However, due to its $\mathbf{q}$ dependence, Eq. (41) is harder to evaluate. Writing out the troublesome matrix element in Eq. (41) more explicitly, we see that

$$\left\langle \psi_f^\sigma \right| \exp(i\mathbf{q}\cdot\mathbf{r}) \left| \psi_i^\sigma \right\rangle = \sum_{\mu\nu} C_{\mu f}^\sigma C_{\nu i}^\sigma S_{\mu\nu}^{\mathbf{q}}, \quad (42)$$

$$S_{\mu\nu}^{\mathbf{q}} = \int \phi_\mu(\mathbf{r}) \exp(i\mathbf{q}\cdot\mathbf{r}) \phi_\nu(\mathbf{r}) d^3r, \quad (43)$$

where $\mathbf{S}^{\mathbf{q}}$ is the momentum transfer matrix. Evaluation of Eq. (43) through a series expansion of the exponential is practical for $q \ll 1$; this leads once again to moment integrals, which are readily computed. However, for larger values of q, the series converges very slowly. Instead, the matrix in Eq. (43) can also be formed by using the closure property of Gaussians and directly computing the values of the Fourier transforms of the products $\phi_\mu\phi_\nu$. Indeed, this is the default method in ERKALE of evaluating Eq. (41), as it is both fast and accurate.

When the initial state is sufficiently localized around a nucleus at $\mathbf{R}_0$, the integral in Eq. (41) can also be calculated by using a single-center expansion for the orbitals[125]

$$\psi_i^\sigma(\mathbf{r}) \approx \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{m} c_{il}^{\sigma m}(\rho) Y_l^m(\hat{\boldsymbol{\rho}}), \quad (44)$$

$$\boldsymbol{\rho} = \mathbf{r} - \mathbf{R}_0. \quad (45)$$

Here, Y_l^m are spherical harmonics in the complex form and $l_{\max}$ is the truncation order, $c_{il}^{\sigma m}(\rho)$ being complex radial expansion coefficients given by

$$c_{il}^{\sigma m}(\rho) = \int \overline{Y_l^m(\hat{\boldsymbol{\rho}})} \psi_i^\sigma(\boldsymbol{\rho} + \mathbf{R}_0) d\Omega_\rho. \quad (46)$$

This method allows extraction of the angular momentum distribution of the final states.[125] In contrast to Ref. [125] in which $l_{\max} = 2$, ERKALE has no limitation to the order of truncation of the expansion in Eq. (44).

[¶¶] In the benzene example, there are 42 electrons, that is, 21 occupied orbitals in the (restricted) ground state. Localizing the 1s states of the excited spin σ on each of the carbons, for the first atom, the size of the diagonalized r_{ij}^2 matrix is 21×21. The occupied spin σ orbitals are updated to correspond to the eigenvectors, and the one corresponding to the lowest eigenvalue is assigned to be frozen and dropped from further iterations. The iteration continues over the other centers, and the diagonalizations are then perfomed on matrices of size $20 \times 20, 19 \times 19, \ldots$, and 16×16.

When isotropic or amorphous systems are studied, an average must be taken over all directions of the scattering vector **q**. Although the Fourier method of Eq. (43) would allow analytical integration along the lines of Ref. [90], it was found that the scaling of the method is far too restrictive to allow its use already in systems of small size. Instead, the directional average is computed numerically using Lobatto quadrature (as introduced in Formation of matrices section).

We demonstrate XAS/XRS calculations with ERKALE by the spherically averaged oxygen *K*-edge XRS spectra of the water monomer and the molecules in the water dimer, shown in Figure 4. The spectra have been computed with the PBE XC functional[139, 140] with density fitting and the (orbital) basis set presented above in Completeness-optimized basis sets section. The double-basis set procedure was used with a large set of diffuse functions added on the excited atom to improve the description of the continuum states. An absolute energy scale correction was performed, shifting the monomer spectrum by −3.4 eV, and the acceptor and donor spectra by −0.07 and −0.24 eV, respectively, with respect to the monomer spectrum. Finally, the stick spectra were convoluted with a Lorentzian with a full width at half maximum (FWHM) of 0.15 eV, and the area of the convoluted spectrum was normalized to unity in the range $\hbar\omega \in [527, 536.5]$ eV. The spectra display the typical pre-edge behavior[141] at the excitation energy of roughly 529eV. As is well known, the acceptor spectrum bears a close resemblance to the monomer spectrum, whereas the donor spectrum is substantially affected by the formation of the hydrogen bond[141]; the same feature is seen also in the case of alcohols.[142] The experimental XAS edge of water in the gas phase is roughly at 534 eV,[143] thus the XCH method (without corrections for functional dependence, *etc.*) underestimates the transition energy by some 5.5 eV.

As a further demonstration, we have reproduced the calculation in Ref. [125] for the dipole limit of the dynamic structure factor of benzene, using the geometry obtained from the NIST Standard Reference Database[144] and the revised PBE functional.[139, 140, 145] The IGLO-III basis set[112] was used on all atoms, and the double-basis set method was used as above. The localization of the core hole on the excited carbon atom was enforced by using frozen cores for all other carbons in the calculation. An absolute energy correction was performed, shifting the spectrum by −2.19 eV. For comparison with experimental data taken from Ref. [146], the spectrum was then convoluted with a Lorentzian with an increasing FWHM Γ_ω given by

$$\Gamma_\omega = \begin{cases} \Gamma_0, & \omega < \omega_0 \\ \Gamma_0 + \gamma\,(\omega - \omega_0)\,, & \omega_0 \le \omega \le \omega_1 \\ \Gamma_0 + \gamma\,(\omega_1 - \omega_0)\,, & \omega_1 < \omega \end{cases} \tag{47}$$

where the parameters are $\Gamma_0 = 0.347$ eV, $\gamma = 0.122$, $\omega_0 = 284.7$ eV and $\omega_1 = 296.3$ eV, to account for finite lifetime effects, and then a Gaussian with an FWHM of 1.1 eV to account for instrumental broadening. Finally, the spectrum was shifted by 0.37 eV so that the computed maximum coincides with the experimental absorption edge[146] at 285.10 eV, and the area of the computed and experimental spectra were normalized to unity in the region $\hbar\omega \in [280\text{ eV}, 302\text{ eV}]$.

For comparison, we repeated the calculation using StoBe-deMon[124] with the IGLO-III basis on all hydrogen atoms and the excited carbon atom, while ECPs and corresponding valence basis sets[‖‖] were used on the other carbon atoms. The first transition in the TP approximation occurs at 286.75 eV in the StoBe-deMon calculation, compared to 286.93 eV in the calculation performed with ERKALE. In XCH, the values become

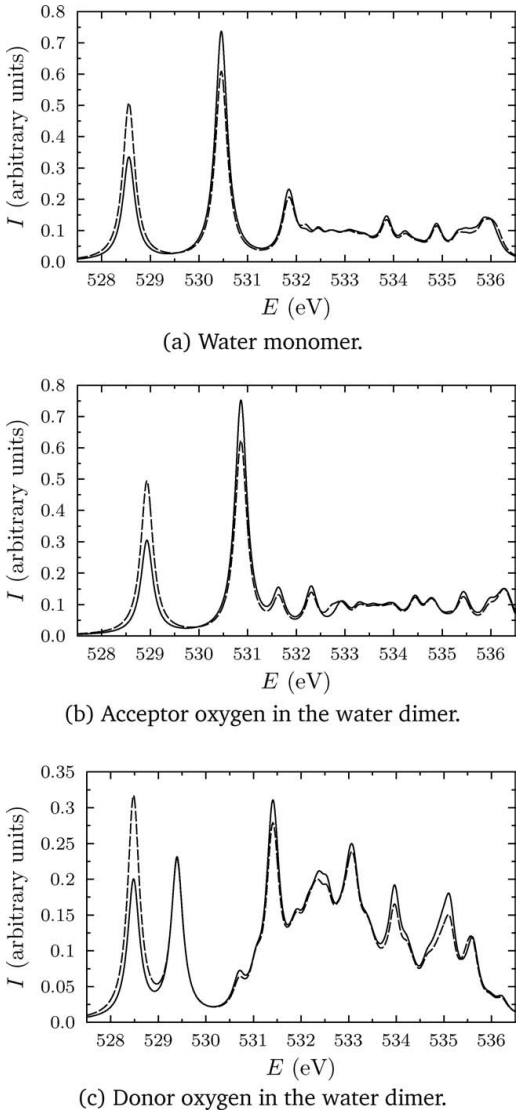


Figure 4. The oxygen *K*-edge XRS spectra computed for the water monomer and the acceptor and the donor in the water dimer in the TP approximation. The solid line is the spectrum in the dipole approximation (corresponding to $q = 0$), the dashed line represents $q = 5.0$ a.u.

[‖‖] L. G. M. Pettersson, unpublished. However, the used ECP and valence basis set are part of the standard StoBe distribution.

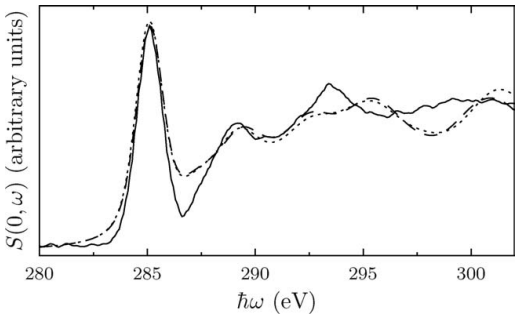


Figure 5. The dipole limits of the dynamic structure factor of benzene. The solid line represents experiment, whereas the dashed and dotted lines represent the TP calculations with ERKALE and StoBe, respectively.

284.72 eV and 284.74 eV, respectively. The StoBe-deMon spectrum was convoluted, normalized, and shifted to coincide with the experimental spectrum by using the same procedure as above. These two calculations are shown in Figure 5 accompanied with experimental data taken from Ref. [146]. As expected, the use of ECPs in the StoBe-deMon calculation has no notable effect on the convoluted spectrum (compared to the frozen-core calculation with ERKALE).

Valence electron excitations

ERKALE can compute electron excitation energies and oscillator strengths using linear response time-dependent DFT[147,148] through the Casida method,[149] which is especially useful in simulating discrete, bound-to-bound state transitions in moderately sized systems. Indeed, it has been shown to give excellent results for valence electron excitations.[150] The Casida method can be formulated as a matrix eigenvalue equation[151]

$$\mathbf{\Omega F}_I = \omega_I^2 \mathbf{F}_I, \tag{48}$$

where ω_I are the vertical excitation energies and $\mathbf{F}_I$ are the corresponding eigenmodes. The coupling matrix $\mathbf{\Omega}$ has the elements

$$\Omega_{ij\sigma,kl\sigma'} = \delta_{\sigma\sigma'}\delta_{ik}\delta_{jl}\left(\epsilon_l^{\sigma'} - \epsilon_k^{\sigma'}\right)^2 + \tilde{f}_{ij\sigma}K_{ij\sigma,kl\sigma'}\tilde{f}_{kl\sigma'}, \tag{49}$$

$$\tilde{f}_{ij\sigma} = \sqrt{\left(n_i^\sigma - n_j^\sigma\right)\left(\epsilon_j^\sigma - \epsilon_i^\sigma\right)}, \tag{50}$$

$$K_{ij\sigma,kl\sigma'} = \int d^3r \int d^3r' \psi_i^\sigma(\mathbf{r})\psi_j^\sigma(\mathbf{r})\psi_k^{\sigma'}(\mathbf{r}')\psi_l^{\sigma'}(\mathbf{r}') \times \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 f^{\mathrm{xc}}}{\delta\rho_\sigma(\mathbf{r})\delta\rho_{\sigma'}(\mathbf{r}')}\right). \tag{51}$$

Occupied orbitals are denoted here with i and k, whereas j and l denote virtual orbitals. Because of this rather surprisingly simple computational formulation, the Casida method forms a standard part of modern quantum chemistry programs (e.g., CP2K[38] and NWChem[6]). As the size of the matrices in Eq. (48) grows as N_{pairs}^2***, but often only a small subset of the orbitals significantly affect the response properties of the system in the investigated energy range, it is common to make the method computationally less demanding by omitting insignificant orbitals from the Casida treatment.

The Casida formalism has been recently generalized[152] to yield NRIXS spectra or more specifically the momentum transfer dependent dynamic structure factor

$$S(\mathbf{q},\omega) = \sum_I \delta(\omega - \omega_I)\left|\sum_{ij\sigma}\left\langle\psi_i^\sigma\left|e^{i\mathbf{q}\cdot\mathbf{r}}\right|\psi_j^\sigma\right\rangle\tilde{f}_{ij\sigma}F_{I,ij\sigma}\right|^2, \tag{52}$$

which can be computed by ERKALE as well. The evaluation of the matrix element in Eq. (52) is performed using the same Fourier method described before in Core electron excitations section. As for XRS spectra, directionally averaged, momentum transfer dependent Casida spectra are computed in ERKALE using Lobatto quadrature (Formation of matrices section). The calculations in Ref. [152] were done using Octopus,[17,18] which uses a numerical real space grid to represent the Kohn–Sham orbitals. While real space grids offer a straightforward way for improving the accuracy of the calculation by reducing the grid spacing, the benefit of a localized basis set (as in ERKALE) is the significantly smaller number of degrees of freedom that makes spectrum calculations very fast. Localized basis sets also easily provide explicit description of core orbitals and thereby automatically enable calculations of core excitations as well (i.e., XAS/XRS spectra). Despite some promising results (for instance, see Ref. [153] for a recent application to XAS calculation on liquid water), this approach does not provide very reliable spectra for deeper excitations with conventional DFT XC functionals; TDDFT heavily underestimates core excitation energies†††, which has been shown to be caused by the self-interaction error[154] at short distances.[155] As a result, contrary to TP calculations,[134] the XAS spectra predicted by TDDFT with conventional functionals show large dependence on the functional used.[153] A perspective on TDDFT calculations for core electron spectroscopies has recently been published, see Ref. [156].

ERKALE currently supports calculations only at the time-dependent LDA (TDLDA) level‡‡‡. However, the orbitals used for the calculation can be obtained with HF or any supported DFT method§§§. The Coulombic term in Eq. (51) is computed using Eq. (25) as in the work of Jamorski *et al.*[151] However, contrary to their work, fitting of the XC term in Eq. (51) is not performed in ERKALE, as the Coulomb part is often found to dominate the computational cost. We demonstrate the Casida method by computing the momentum transfer-dependent NRIXS spectrum

***N_{pairs} is the amount of pairs of occupied and virtual orbitals. For restricted calculations $N_{\mathrm{pairs}} = N_{\mathrm{occ}}N_{\mathrm{virt}}$, for unrestricted calculations, $N_{\mathrm{pairs}} = N_{\mathrm{occ}}^{\alpha}N_{\mathrm{virt}}^{\alpha} + N_{\mathrm{occ}}^{\beta}N_{\mathrm{virt}}^{\beta}$. N_{occ} and N_{virt} denote the number of occupied and virtual orbitals, respectively.

†††According to a commonly used rule of thumb, when the excitation energy approaches the negative of the orbital energy of the highest occupied MO, TDLDA will likely underestimate the excitation energy.

‡‡‡Calculations can also be done in ERKALE in the random phase approximation (RPA) and the independent particle approximation (IPA), which are a subset of TDLDA.

§§§It must be noted that a time-dependent formulation of Hartree–Fock theory (TDHF) exists as well. TDLDA performed with Hartree–Fock orbitals is not equivalent to TDHF.

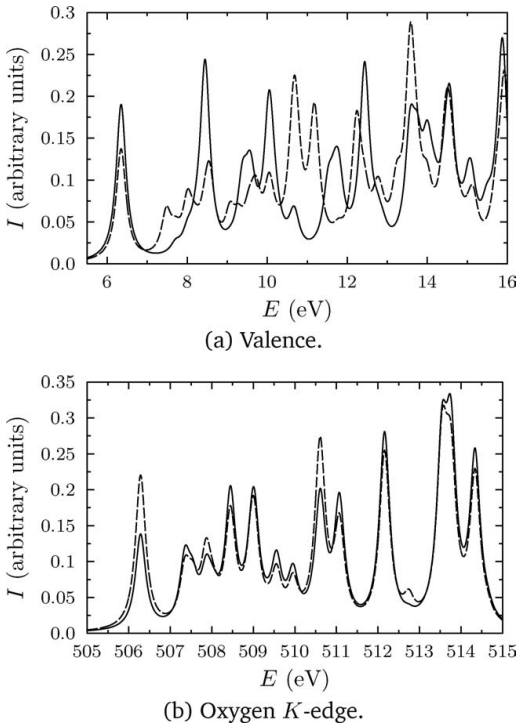


Figure 6. NRIXS spectra of the water monomer, calculated using Casida's equation. The solid line is the spectrum in the dipole approximation (corresponding to $q = 0$), the dashed line represents $q = 5.0$ a.u.

corresponding to valence electron excitations in Figure 6a. As a further illustration, an XRS spectrum of the water monomer at the oxygen *K*-edge, produced by the same Casida calculation, is shown in Figure 6b. The MOs were computed using the PBE exchange-correction functionals[139,140] using density fitting, after which the Casida equation was solved using the first 75 orbitals, with LDA exchange[157,158] and VWN correlation.[159] Finally, the spectrum was convoluted and normalized analogously to the procedure in Core electron excitations section. From the discussion above, it is not surprising that TDLDA underestimates the energy of the first *K* edge transition by almost 30 eV. Another issue to be analyzed is the incompatibility of the TP and TDLDA spectra. However, specially designed functionals (e.g., Refs. [154, 155]) have been shown to produce good compatibility of TDLDA near-edge X-ray absorption fine structure spectra with experiment.[155]

Conclusion and Outlook

We have presented a novel code, ERKALE,[45] for electronic structure calculations of atoms and molecules using HF or DFT level of theory with Gaussian basis sets of arbitrary angular momentum. In addition to a large selection of supported XC functionals, the program features modern convergence accelerators and is thoroughly parallellized, making calculations with thousands of basis functions routine on desktop computers.

The main advantage of the program is its direct applicability to computing X-ray properties, such as Compton profiles and NRIXS spectra of core and valence electronic excitations. We have demonstrated the code by performing sample calculations of the electronic structure and excitations of a simple, well-known model system, the water molecule and the water dimer. Thanks to the free availability and simple organization of the program, it can easily be modified to suit the user's needs. Because of its clarity, the code is also ideal for teaching purposes.

We have introduced a new way to localize the core hole for XAS and XRS calculations in the TP approximation. We have also introduced a fast and accurate method of computing elements of the momentum transfer matrix $S^{\mathbf{q}}_{\mu\nu} = \langle \mu | \exp(i\mathbf{q} \cdot \mathbf{r}) | \nu \rangle$ in a Gaussian basis. The method is based on the direct Fourier transform of the products of basis functions, and we have used it to implement the first readily available momentum transfer dependent NRIXS spectrum calculators for Gaussian basis sets, both in the TP approximation and Casida method.

The development of ERKALE will further continue to include new features, mainly focusing on X-ray spectroscopic applications. One of the future goals is the implementation of self-interaction error corrected XC functionals for more accurate TDDFT calculations of core electron spectroscopies. ERKALE will also be used in shortly upcoming X-ray studies and basis set work.

Acknowledgments

JL thanks Miguel Marques, Edward Valeev, and Justin T. Fermann for publishing the libxc and libint libraries, the existence of which was one of the original motivations for this work, and Takashi Tsuchimochi for support in the implementation of the ROHF algorithm. We thank Pekka Manninen, Dage Sundholm, and Juha Vaara for comments on the manuscript, and Lars Pettersson and Tuomas Pylkkänen for discussions. This work has been supported by the Jenny and Antti Wihuri foundation, the Academy of Finland through its Centers of Excellence program and project number 1127462, and by the research funds of the University of Helsinki (number 490064). The resources of the Center for Scientific Computing (Espoo, Finland) are gratefully acknowledged.

Keywords: electronic structure • density-functional theory • Hardree–Fock • Compton scattering • electron momentum density • X-ray absorption • X-ray Raman Scattering • time-dependent density-junctional theory • completeness optimization

How to cite this article: J. Lehtola, M. Hakala, A. Sakko, K. Hämäläinen, *J. Comput. Chem.* **2012**, *33*, 1572–1585. DOI: 10.1002/jcc.22987

Additional Supporting Information may be found in the online version of this article.

[1] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.

[2] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[3] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas,, M. Côt, *Comput. Phys. Commun.* **2009**, *180*, 2582.

[4] ABINIT, http://www.abinit.org/.

[5] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, *Comput. Phys. Commun.* **2010**, *181*, 1477.

[6] NWChem, http://www.nwchem-sw.org/.

[7] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, R. M. Wentzcovitch, *J. Phys. Condens. Mater* **2009**, *21*, 395502.

[8] Quantum ESPRESSO, http://www.quantum-espresso.org/.

[9] L. Genovese, B. Videau, M. Ospici, T. Deutsch, S. Goedecker, J.-F. Méhaut, *C. R. Méc.* **2011**, *339*, 149.

[10] BigDFT, http://inac.cea.fr/L_Sim/BigDFT/.

[11] S. Ismail-Beigi, T. A. Arias, *Comput. Phys. Commun.* **2000**, *128*, 1.

[12] *DFT++*, http://dft.physics.cornell.edu/.

[13] M-A-D-N-E-S-S, http://m-a-d-n-e-s-s.googlecode.com.

[14] OpenMX, http://www.openmx-square.org.

[15] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, K. W. Jacobsen, *J. Phys. Condens. Matter* **2010**, *22*, 253202.

[16] GPAW, https://wiki.fysik.dtu.dk/gpaw/.

[17] A. Castro, H. Appel, M. Oliveira, C. A. Rozzi, X. Andrade, F. Lorenzen, M. A. L. Marques, E. K. U. Gross, A. Rubio, *Phys. Status Solidi* (b) **2006**, *243*, 2465.

[18] Octopus, http://www.tddft.org/programs/octopus/.

[19] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[20] S. F. Boys, *Proc. R. Soc. A* **1950**, *200*, 542.

[21] S. Obara, A. Saika, *J. Chem. Phys.* **1986**, *84*, 3963.

[22] M. Head-Gordon, J. A. Pople, *J. Chem. Phys.* **1988**, *89*, 5777.

[23] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory*; Wiley, Chichester **2000**.

[24] O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*, 514.

[25] M. A. Watson, N. C. Handy, A. J. Cohen, *J. Chem. Phys.* **2003**, *119*, 6475.

[26] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.

[27] ADF, http://www.scm.com/.

[28] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.

[29] T. Yanai, G. I. Fann, Z. Gan, R. J. Harrison, G. Beylkin, *J. Chem. Phys.* **2004**, *121*, 6680.

[30] R. J. Harrison, G. I. Fann, T. Yanai, Z. Gan, G. Beylkin, *J. Chem. Phys.* **2004**, *121*, 11587.

[31] F. A. Bischoff, E. F. Valeev, *J. Chem. Phys.* **2011**, *134*, 104104.

[32] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.

[33] J. Čížek, *J. Chem. Phys.* **1966**, *45*, 4256.

[34] E. Rudberg, E. H. Rubensson, P. Sałek, *J. Chem. Theor. Comp.* **2011**, *7*, 340.

[35] Ergo, http://www.ergoscf.org/.

[36] N. Bock, M. Challacombe, C. K. Gan, G. Henkelman, K. Nemeth, A. M. N. Niklasson, A. Odell, E. Schwegler, C. J. Tymczak, V. Weber, FreeON. Available at: http://www.freeon.org/, **2011**.

[37] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Comput. Phys. Commun.* **2005**, *167*, 103.

[38] CP2K, http://www.cp2k.org/.

[39] J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale, R. J. Bartlett, *Int. J. Quantum. Chem.* **1992**, *44*, 879.

[40] ACES, http://www.qtp.ufl.edu/ACES/.

[41] V. Lotrich, N. Flocke, M. Ponton, A. D. Yau, A. Perera, E. Deumens, R. J. Bartlett, *J. Chem. Phys.* **2008**, *128*, 194104.

[42] C. L. Janssen, I. B. Nielsen, M. L. Leininger, E. F. Valeev, J. P. Kenny, E. T. Seidl, *The Massively Parallel Quantum Chemistry Program (MPQC)*, http://www.mpqc.org.

[43] T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny, W. D. Allen, *J. Comput. Chem.* **2007**, *28*, 1610.

[44] PSI3, http://psicode.org/.

[45] J. Lehtola, *ERKALE - HF/DFT from Hel* (2011), http://erkale.googlecode.com.

[46] E. F. Valeev, J. T. Fermann, *Libint—efficient computation of quantum mechanical matrix elements over Gaussian basis sets*, http://sourceforge.net/p/libint/.

[47] M. A. L. Marques, M. Oliveira, T. Burnus, G. Madsen, X. Andrade, *Libxc–A Library of Exchange-Correlation Functionals for Density-Functional Theory*, http://www.tddft.org/programs/octopus/wiki/index.php/Libxc.

[48] http://www.gnu.org/licenses/old-licenses/gpl-2.0.txt

[49] C. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69.

[50] J. A. Pople, R. K. Nesbet, *J. Chem. Phys.* **1954**, *22*, 571.

[51] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc., New York **1996**.

[52] T. Tsuchimochi, G. E. Scuseria, *J. Chem. Phys.* **2010**, *133*, 141102.

[53] D. Langreth, J. Perdew, *Phys. Rev. B* **1980**, *21*, 5469.

[54] J. A. Pople, P. M. Gill, B. G. Johnson, *Chem. Phys. Lett.* **1992**, *199*, 557.

[55] J. Perdew, S. Kurth, A. Zupan, P. Blaha, *Phys. Rev. Lett.* **1999**, *82*, 2544.

[56] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

[57] E. Baerends, *Chem. Phys.* **1973**, *2*, 41.

[58] H. Sambe, R. H. Felton, *J. Chem. Phys.* **1975**, *62*, 1122.

[59] K. Eichkorn, O. Treutler, H. Ohm, M. Haser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283.

[60] R. Yang, A. P. Rendell, M. J. Frisch, *J. Chem. Phys.* **2007**, *127*, 074102.

[61] G. B. Arfken, H. J. Weber, F. Harris, *Mathematical Methods for Physicists*; 5th ed. Academic Press: New York, **2000**.

[62] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, *GNU Scientific Library Reference Manual*, 3rd ed., Network Theory Ltd, United Kingdom, **2009** ISBN: 0-9546120-7-8.

[63] A. M. Köster, R. Flores-Moreno, J. U. Reveles, *J. Chem. Phys.* **2004**, *121*, 681.

[64] A. D. Becke, *J. Chem. Phys.* **1988**, *88*, 2547.

[65] V. I. Lebedev, *Zh. Vychisl. Mat. Mat. Fiz.* **1975**, *15*, 48.

[66] V. I. Lebedev, *USSR Comput. Math. Math. Phys* **1976**, *16*, 10.

[67] V. I. Lebedev, *Sib. Math. J.* **1977**, *18*, 99.

[68] V. I. Lebedev, A. L. Skorokhodov, *Russ. Acad. Sci. Dokl. Math.* **1992**, *45*, 587.

[69] V. I. Lebedev, *Russ Acad. Sci. Dokl. Math.* **1995**, *50*, 283.

[70] V. I. Lebedev, D. N. Laikov, *Dokl. Math.* **1999**, *59*, 477.

[71] O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346.

[72] B. G. Johnson, Ph.D. Thesis, Chapter 6, Carnegie Mellon University, **1993**.

[73] R. E. Stratmann, G. E. Scuseria, M. J. Frisch, *Chem. Phys. Lett.* **1996**, *257*, 213.

[74] J. Almlöf, K. Faegri, K. Korsell, *J. Comput. Chem.* **1982**, *3*, 385.

[75] J. H. van Lenthe, R. Zwaans, H. J. J. van Dam, M. F. Guest, *J. Comput. Chem.* **2006**, *27*, 926.

[76] P.-O. Löwdin, *J. Chem. Phys.* **1950**, *18*, 365.

[77] P. Pulay, *Chem. Phys. Lett.* **1980**, *73*, 393.

[78] H. Sellers, *Int. J. Quantum. Chem.* **1993**, *45*, 31.

[79] X. Hu, W. Yang, *J. Chem. Phys.* **2010**, *132*, 054109.

[80] S. Høst, J. Olsen, B. Jansík, L. Thøgersen, P. Jørgensen, T. Helgaker, *J. Chem. Phys.* **2008**, *129*, 124106.

[81] K. Baarman, T. Eirola, V. Havu, *J. Chem. Phys.* **2011**, *134*, 134109.

[82] P. Kaijser, V. H. Smith, *Adv. Quantum Chem.* **1977**, *10*, 37.

[83] M. J. Cooper, P. E. Mijnarends, N. Shiotani, N. Sakai, A. Bansil, *X-ray Compton Scattering*, *Oxford Series on Synchrotron Radiation*, Vol. 5 Oxford University Press, Inc., New York **2004**.

[84] M. J. Cooper, *Rep. Prog. Phys.* **1985**, *48*, 415.

[85] L. C. Snyder, T. A. Weber, *J. Chem. Phys.* **1975**, *63*, 113.

[86] S. Huotari, J. Soininen, T. Pylkkänen, K. Hämäläinen, A. Issolah, A. Titov, J. McMinis, J. Kim, K. Esler, D. Ceperley, M. Holzmann, V. Olevano, *Phys. Rev. Lett.* **2010**, *105*, 20.

[87] A. Jarzecki, E. Davidson, *Mol. Phys.* **2000**, *98*, 1089.

[88] S. Ragot, *J. Chem. Phys.* **2006**, *125*, 014106.

[89] B. Miguel, J. M. García de la Vega, *Theor. Chem. Acc.* **2007**, *118*, 723.

[90] J. Lehtola, M. Hakala, J. Vaara, K. Hämäläinen, *Phys. Chem. Chem. Phys.* **2011**, *13*, 5630.

[91] J. R. Hart, A. J. Thakkar, *Int. J. Quantum. Chem.* **2005**, *102*, 673.

[92] T. K. Ghanty, V. N. Staroverov, P. R. Koren, E. R. Davidson, *J. Am. Chem. Soc.* **2000**, *122*, 1210.

[93] A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz, D. Usvyat, *Phys. Rev. B* **2010**, *81*, 165108.

[94] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.

[95] R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796.

[96] K. A. Peterson, D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1994**, *100*, 7410.

[97] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. L. Windus, *J. Chem. Inf. Mod.* **2007**, *47*, 1045.

[98] ESML Basis Set Exchange, http://bse.pnl.gov.

[99] M. Hakala, K. Nygård, S. Manninen, S. Huotari, T. Buslaps, A. Nilsson, L. G. M. Pettersson, K. Hämäläinen, *J. Chem. Phys.* **2006**, *125*, 084504.

[100] P. Sit, C. Bellin, B. Barbiellini, D. Testemale, J.-L. Hazemann, T. Buslaps, N. Marzari, A. Shukla, *Phys. Rev. B* **2007**, *76*, 245413.

[101] K. Nygård, M. Hakala, S. Manninen, M. Itou, Y. Sakurai, K. Hämäläinen, *Phys. Rev. Lett.* **2007**, *99*, 197401.

[102] I. Juurinen, K. Nakahara, N. Ando, T. Nishiumi, H. Seta, N. Yoshida, T. Morinaga, M. Itou, T. Ninomiya, Y. Sakurai, E. Salonen, K. Nordlund, K. Hämäläinen, M. Hakala, *Phys. Rev. Lett.* **2011**, *107*, 197401.

[103] D. P. Chong, S. R. Langhoff, *J. Chem. Phys.* **1990**, *93*, 570.

[104] D. P. Chong, *Can. J. Chem.* **1995**, *73*, 79.

[105] P. Manninen, J. Vaara, *J. Comput. Chem.* **2006**, *27*, 434.

[106] S. Ikäläinen, P. Lantto, P. Manninen, J. Vaara, *J. Chem. Phys.* **2008**, *129*, 124102.

[107] S. Ikäläinen, P. Lantto, P. Manninen, J. Vaara, *Phys. Chem. Chem. Phys.* **2009**, *11*, 11404.

[108] S. Ikäläinen, M. Romalis, P. Lantto, J. Vaara, *Phys. Rev. Lett.* **2010**, *105*, 153001.

[109] *Kruununhaka basis set tool kit, written by P. Manninen and J. Lehtola*, http://www.chem.helsinki.fi/ manninen/kruununhaka/.

[110] J. A. Nelder, R. Mead, *Comput J.* **1965**, *7*, 308.

[111] L. Davis, *Handbook of Genetic Algorithms*; Van Nostrand Reinhold, **1991**.

[112] W. Kutzelnigg, U. Fleischer, M. Schindler, *The IGLO-Method: Ab Initio Calculation and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities* Vol. 23; Springer Verlag: Berlin/Heidelberg, **1991**, pp. 165–262.

[113] A. J. Sadlej, *Collect Czech. Chem. Commun.* **1988**, *53*, 1995.

[114] A. J. Sadlej, *Theor. Chim. Acta* **1991**, *81*, 45.

[115] J.-P. Rueff, *Rev. Mod. Phys.* **2010**, *82*, 847.

[116] T. Pylkkänen, V. M. Giordano, J.-C. Chervin, A. Sakko, M. Hakala, J. A. Soininen, K. Hämäläinen, G. Monaco, S. Huotari, *J. Phys. Chem. B* **2010**, *114*, 3804.

[117] S. Doniach, P. Platzman, J. Yue, *Phys. Rev. B* **1971**, *4*, 3345.

[118] W. Schülke, *Electron Dynamics by Inelastic X-Ray Scattering*, *Oxford Series on Synchrotron Radiation*; Vol. 7 Oxford University Press, Inc., New York **2007**.

[119] K. Hämäläinen, S. Galambosi, J. Soininen, E. Shirley, J.-P. Rueff, A. Shukla, *Phys. Rev. B* **2002**, *65*, 155111.

[120] J. A. Soininen, A. L. Ankudinov, J. J. Rehr, *Phys. Rev. B* **2005**, *72*, 045136.

[121] L. Triguero, L. G. M. Pettersson, H. Ågren, *Phys. Rev. B* **1998**, *58*, 8097.

[122] J. C. Slater, vol. 6 of *Advances Quantinum Chemistry*; Wiley: New York, **1972**.

[123] J. C. Slater, *J. Chem. Phys.* **1972**, *57*, 2389.

[124] K. Hermann, L. G. M. Pettersson, M. E. Casida, C. Daul, A. Goursot, A. Köster, E. Proynov, A. St-Amant, D. R. Salahub, *StoBe-deMon, a Density Functional Theory molecule/cluster package*, http://www.fhi-berlin.mpg.de/KHsoftware/StoBe.

[125] A. Sakko, M. Hakala, J. A. Soininen, K. Hämäläinen, *Phys. Rev. B* **2007**, *76*, 205115.

[126] J. A. Soininen, E. L. Shirley, *Phys. Rev. B* **2001**, *64*, 165112.

[127] J. Vinson, J. J. Rehr, J. J. Kas, E. L. Shirley, *Phys. Rev. B* **2011**, *83*, 115106.

[128] A. L. Ankudinov, B. Ravel, J. J. Rehr, S. D. Conradson, *Phys. Rev. B* **1998**, *58*, 7565.

[129] J. J. Rehr, R. C. Albers, *Rev. Mod. Phys.* **2000**, *72*, 621.

[130] The FEFF code, http://leonardo.phys.washington.edu/feff/.

[131] M. Leetmaa, M. P. Ljungberg, A. Lyubartsev, A. Nilsson, L. G. M. Pettersson, *J. El. Spec.* **2010**, *177*, 135.

[132] C. Kolczewski, R. Püttner, O. Plashkevych, H. Ågren, V. Staemmler, M. Martins, G. Snell, A. S. Schlachter, M. Sant'Anna, G. Kaindl, L. G. M. Pettersson, *J. Chem. Phys.* **2001**, *115*, 6426.

[133] M. Odelius, M. Cavalleri, A. Nilsson, L. G. M. Pettersson, *Phys. Rev. B* **2006**, *73*, 024205.

[134] O. Takahashi, L. G. M. Pettersson, *J. Chem. Phys.* **2004**, *121*, 10339.

[135] L. Pettersson, U. Wahlgren, *Chem. Phys.* **69**, 185-192 (1982).

[136] L. G. M. Pettersson, U. Wahlgren, O. Gropen, *J. Chem. Phys.* **1987**, *86*, 2176.

[137] H. Ågren, V. Carravetta, O. Vahtras, L. G. M. Pettersson, *Chem. Phys. Lett.* **1994**, *222*, 75.

[138] H. Ågren, V. Carravetta, O. Vahtras, L. G. M. Pettersson, *Theor. Chim. Acta* **1997**, *97*, 14.

[139] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.

[140] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1997**, *78*, 1396.

[141] M. Cavalleri, H. Ogasawara, L. G. M. Pettersson, A. Nilsson, *Chem. Phys. Lett.* **2002**, *364*, 363.

[142] T. Pylkkänen, J. Lehtola, M. Hakala, A. Sakko, G. Monaco, S. Huotari, K. Hämäläinen, *J. Phys. Chem. B* **2010**, *114*, 13076.

[143] S. Myneni, Y. Luo, L. Å. Näslund, M. Cavalleri, L. Ojamäe, H. Ogasawara, A. Pelmenschikov, Ph. Wernet, P. Väterlein, C. Heske, Z. Hussain, L. G. M. Pettersson, A. Nilsson, *J. Phys. Condens Matter* **2002**, *14*, L213.

[144] P. J. Linstrom, W. G. Mallard, Eds., *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; National Institute of Standards and Technology, Gaithersburg: MD, p. 20899; Available at: http://webbook.nist.gov.

[145] B. Hammer, L. B. Hansen, J. K. Nørskov, *Phys. Rev. B* **1999**, *59*, 7413.

[146] M. L. Gordon, D. Tulumello, G. Cooper, A. P. Hitchcock, P. Glatzel, O. C. Mullins, S. P. Cramer, U. Bergmann, *J. Phys. Chem. A* **2003**, *107*, 8512.

[147] E. Runge, E. K. U. Gross, *Phys. Rev. Lett.* **1984**, *52*, 997.

[148] M. A. L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, E. K. U. Gross, Eds., *Time-Dependent Density Functional Theory*; Springer Verlag, **2006**.

[149] M. E. Casida, In *Recent Advances in Computational Chemistry*; D. P. Chong, Ed.; World Scientific, Singapore, **1995**.

[150] M. E. Casida, C. Jamorski, K. C. Casida, D. R. Salahub, *J. Chem. Phys.* **1998**, *108*, 4439.
[151] C. Jamorski, M. E. Casida, D. R. Salahub, *J. Chem. Phys.* **1996**, *104*, 5134.
[152] A. Sakko, A. Rubio, M. Hakala, K. Hämäläinen, *J. Chem. Phys.* **2010**, *133*, 174111.
[153] G. Brancato, N. Rega, V. Barone, *Phys. Rev. Lett.* **2008**, *100*, 107401.
[154] A. Nakata, Y. Imamura, H. Nakai, *J. Chem. Phys.* **2006**, *125*, 064109.
[155] N. A. Besley, M. J. G. Peach, D. J. Tozer, *Phys. Chem. Chem. Phys.* **2009**, *11*, 10350.
[156] N. A. Besley, F. A. Asmuruf, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12024.
[157] F. Bloch, *Z. Phys. A* **1929**, *57*, 545.
[158] P. A. M. Dirac, *Proc. Cambridge Philas. Soc.* **1930**, *26*, 376.
[159] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.

Received: 15 November 2011
Revised: 2 February 2012
Accepted: 16 March 2012
Published online on 24 April 2012