A basis-set error correction based on density-functional theory for strongly correlated molecular systems EP

Emmanuel Giner iD, Anthony Scemama iD, Pierre-François Loos iD, and Julien Toulouse iD

COLLECTIONS

EP   This paper was selected as an Editor's Pick

View Online   Export Citation   CrossMark

A basis-set error correction based on density-functional theory for strongly correlated molecular systems ᴱᴾ

View Online   Export Citation   CrossMark

Emmanuel Giner,[1,a] (iD) Anthony Scemama,[2] (iD) Pierre-François Loos,[2,b] (iD) and Julien Toulouse[1,3,c] (iD)

AFFILIATIONS

[1] Laboratoire de Chimie Théorique (UMR 7616), Sorbonne Université, CNRS, Paris, France
[2] Laboratoire de Chimie et Physique Quantiques (UMR 5626), Université de Toulouse, CNRS, UPS, Toulouse, France
[3] Institut Universitaire de France, Paris, France

[a] Author to whom correspondence should be addressed: emmanuel.giner@lct.jussieu.fr
[b] Electronic mail: loos@irsamc.ups-tlse.fr
[c] Electronic mail: toulouse@lct.jussieu.fr

ABSTRACT

We extend to strongly correlated molecular systems the recently introduced basis-set incompleteness correction based on density-functional theory (DFT) [E. Giner *et al.*, J. Chem. Phys. **149**, 194301 (2018)]. This basis-set correction relies on a mapping between wave-function calculations in a finite basis set and range-separated DFT (RSDFT) through the definition of an effective non-divergent interaction corresponding to the electron–electron Coulomb interaction projected in the finite basis set. This enables the use of RSDFT-type complementary density functionals to recover the dominant part of the short-range correlation effects missing in this finite basis set. To study both weak and strong correlation regimes, we consider the potential energy curves of the H_{10}, N_2, O_2, and F_2 molecules up to the dissociation limit, and we explore various approximations of complementary functionals fulfilling two key properties: spin-multiplet degeneracy (i.e., independence of the energy with respect to the spin projection S_z) and size consistency. Specifically, we investigate the dependence of the functional on different types of on-top pair densities and spin polarizations. The key result of this study is that the explicit dependence on the on-top pair density allows one to completely remove the dependence on any form of spin polarization without any significant loss of accuracy. Quantitatively, we show that the basis-set correction reaches chemical accuracy on atomization energies with triple-ζ quality basis sets for most of the systems studied here. In addition, the present basis-set incompleteness correction provides smooth potential energy curves along the whole range of internuclear distances.

I. INTRODUCTION

The general goal of quantum chemistry is to provide reliable theoretical tools to explore the rich area of chemistry. More specifically, developments in quantum chemistry primarily aim at accurately computing the electronic structure of molecular systems. Despite intense developments, no definitive solution to this problem has been found. The theoretical challenge to tackle belongs to the quantum many-body problem due to the intrinsic quantum nature of the electrons and the Coulomb repulsion between them. This so-called electronic correlation problem corresponds to finding a solution to the Schrödinger equation

for an N-electron system, and two main roads have emerged to approximate this solution: wave-function theory (WFT)[1] and density-functional theory (DFT).[2] Although both WFT and DFT spring from the same Schrödinger equation, they rely on very different formalisms as the former deals with the complicated N-electron wave function, whereas the latter focuses on the much simpler one-electron density. In its Kohn–Sham (KS) formulation,[3] the computational cost of DFT is very appealing since it is a simple mean-field procedure. Therefore, although continued efforts have been made to reduce the computational cost of WFT, DFT still remains the workhorse of quantum computational chemistry.

The difficulty in obtaining a reliable theoretical description of a given chemical system can be roughly categorized by the strength of the electronic correlation. The so-called weakly correlated systems, such as closed-shell organic molecules near their equilibrium geometry, are typically dominated by correlation effects that do not affect the qualitative mean-field picture of the system. These weak-correlation effects can be either short range (near the electron–electron coalescence points)[4] or long range (London dispersion interactions).[5] The theoretical description of weakly correlated systems is one of the most concrete achievements of quantum chemistry, and the main remaining challenge for these systems is to push the limit of the chemical system size that can be treated. The case of the so-called strongly correlated systems, which are ubiquitous in chemistry, is more problematic as they exhibit a much more complex electronic structure. For example, transition metal complexes, low-spin open-shell systems, and covalent bond breaking situations have all in common that they cannot be even qualitatively described by a single electronic configuration. It is now clear that the usual semilocal density-functional approximations of KS DFT fail to accurately describe these situations,[6,7] and WFT is king for the treatment of strongly correlated systems.

In practice, WFT uses a finite one-electron basis set. The exact solution of the Schrödinger equation within this basis set is then provided by full configuration interaction (FCI) that consists in a linear-algebra eigenvalue problem with a dimension scaling exponentially with the system size. Due to this exponential growth of the FCI computational cost, introducing approximations is necessary, with at least two difficulties for strongly correlated systems: (i) the qualitative description of the wave function is determined by a primary set of electronic configurations (whose size can scale exponentially in many cases) among which near degeneracies and/or strong interactions appear in the Hamiltonian matrix and (ii) the quantitative description of the system requires also to account for weak correlation effects, which involve many other electronic configurations with typically much smaller weights in the wave function. Simultaneously addressing these two issues is a rather complicated task for a given approximate WFT method, especially, if one adds the requirement of satisfying formal properties, such as spin-multiplet degeneracy (i.e., independence of the energy with respect to the spin projection S_z) and size consistency.

Beside the difficulties in accurately describing the molecular electronic structure within a given basis set, a crucial limitation of WFT methods is the slow convergence of the energy (and related properties) with respect to the size of the one-electron basis set. As initially shown by the seminal work of Hylleraas[8] and further developed by Kutzelnigg and co-workers,[9–11] the main convergence problem originates from the divergence of the electron–electron Coulomb interaction at the coalescence point, which induces a discontinuity in the first derivative of the exact wave function (the so-called electron–electron cusp). Describing such a discontinuity with an incomplete one-electron basis set is impossible, and as a consequence, the convergence of the computed energies and properties is strongly affected. To alleviate this problem, extrapolation techniques have been developed either based on a partial-wave expansion analysis[12,13] or more recently based on perturbative arguments.[14,15] A more rigorous approach to tackle the basis-set convergence problem is provided by the so-called explicitly correlated F12 (or R12) methods[4,16–20] that introduce a geminal function depending explicitly on the interelectronic distance. This ensures a correct representation of the Coulomb correlation hole around the electron–electron coalescence point and leads to a much faster convergence of the energy than usual WFT methods. For instance, using the explicitly correlated version of the coupled cluster with singles, doubles, and perturbative triples [CCSD(T)] in a triple-ζ basis set is equivalent to using a quintuple-ζ basis set with the usual CCSD(T) method,[21] although a computational overhead is introduced by the auxiliary basis set needed to compute the three-electron integrals involved in F12 theory.[22] In addition to the computational cost, a possible drawback of F12 theory is its rather complex formalism, which requires non-trivial developments for adapting it to a new method. For strongly correlated systems, several multi-reference methods have been extended to explicit correlation (see, for example, Refs. 23–27), including approaches based on the so-called universal F12 theory, which are potentially applicable to any electronic-structure computational methods.[28–31]

An alternative way to improve the convergence toward the complete-basis-set (CBS) limit is to treat the short-range correlation effects within DFT and to use WFT methods to deal only with the long-range and/or strong correlation effects. A rigorous approach achieving this mixing of DFT and WFT is range-separated DFT (RSDFT) (see Ref. 32 and the references therein), which relies on a decomposition of the electron–electron Coulomb interaction in terms of the interelectronic distance, thanks to a range-separation parameter μ. The advantage of this approach is at least twofold: (i) the DFT part deals primarily with the short-range part of the Coulomb interaction, and consequently, the usual semilocal density-functional approximations are more accurate than for standard KS DFT and (ii) the WFT part deals only with a smooth non-divergent interaction, and consequently, the wave function has no electron–electron cusp[33] and the basis-set convergence is much faster.[34] A number of approximate RSDFT schemes have been developed involving single-reference[35–42] and multi-reference[43–48] WFT methods. Nevertheless, there are still some open issues in RSDFT, such as remaining fractional-charge and fractional-spin errors in the short-range density functionals[49] or the dependence of the quality of the results on the value of the range-separation parameter μ.

Building on the development of RSDFT, a possible solution to the basis-set convergence problem has been recently proposed by some of the present authors,[50] in which RSDFT functionals are used to recover only the correlation effects outside a given basis set. The key point here is to realize that a wave function developed in an incomplete basis set is cuspless and could also originate from a Hamiltonian with a non-divergent long-range electron–electron interaction. Therefore, a mapping with RSDFT can be performed through the introduction of an effective non-divergent interaction representing the usual electron–electron Coulomb interaction projected in an incomplete basis set. First, applications to weakly correlated molecular systems have been successfully carried out[51] together with extensions of this approach to the calculations of excitation energies[52] and ionization potentials.[53] The goal of the present work is to further develop this approach for the description of strongly correlated systems.

This paper is organized as follows: In Sec. II, we recall the mathematical framework of the basis-set correction and we present its

extension for strongly correlated systems. In particular, our focus is primarily set on imposing two key formal properties that are highly desirable in the context of strong correlation: spin-multiplet degeneracy and size consistency. To do this, we introduce (i) new functionals using different flavors of spin polarizations and on-top pair densities and (ii) an effective electron–electron interaction based on a multiconfigurational wave function. This generalizes the method used in previous works on weakly correlated systems[50,51] for which it was sufficient to use an effective interaction based on a single-determinant wave function and a functional depending only on the usual density, reduced density gradient, and spin polarization. Then, in Sec. III, we apply the method to the calculation of the potential energy curves of the H_{10}, N_2, O_2, and F_2 molecules up to the dissociation limit. Finally, we conclude in Sec. IV.

II. THEORY

As the theory behind the present basis-set correction has been exposed in detail in Ref. 50, we only briefly recall the main equations and concepts needed for this study in Secs. II A–II C. More specifically, in Sec. II A, we recall the basic mathematical framework of the present theory by introducing the complementary functional to a basis set. Section II B introduces the effective non-divergent interaction in the basis set, which leads us to the definition of the effective *local* range-separation function in Sec. II C. Then, Sec. II D exposes the new approximate RSDFT-based complementary correlation functionals. The generic form of such functionals is exposed in Sec. II D 1, their properties in the context of the basis-set correction are discussed in Sec. II D 2, and the specific requirements for strong correlation are discussed in Sec. II E. Finally, the actual functionals used in this work are introduced in Sec. II F.

A. Basic theory

The exact ground-state energy E_0 of an N-electron system can, in principle, be obtained in DFT by a minimization over N-representable one-electron densities $n(\mathbf{r})$,

$$E_0 = \min_n \left\{ F[n] + \int \mathrm{d}\mathbf{r} \, v_{\mathrm{ne}}(\mathbf{r}) n(\mathbf{r}) \right\}, \quad (1)$$

where $v_{\mathrm{ne}}(\mathbf{r})$ is the nuclei–electron potential and $F[n]$ is the universal Levy–Lieb density functional written with the constrained search formalism as[54,55]

$$F[n] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi \rangle, \quad (2)$$

where $\hat{T}$ and $\hat{W}_{\mathrm{ee}}$ are the kinetic and electron–electron Coulomb operators, and the notation $\Psi \to n$ means that the wave function Ψ yields the density n. The minimizing density n_0 in Eq. (1) is the exact ground-state density. Nevertheless, in practical calculations, the accessible densities are necessarily restricted to the set of densities "representable in a basis set $\mathcal{B}$," i.e., densities coming from wave functions expandable in the N-electron Hilbert space generated by the one-electron basis set $\mathcal{B}$. In the following, we always consider only such representable-in-$\mathcal{B}$ densities. With this restriction, Eq. (1) then gives an upper bound $E_0^{\mathcal{B}}$ of the exact ground-state energy. Since the density has a faster convergence with the size of the basis set

than the wave function, this restriction is a rather weak one, and we can consider that $E_0^{\mathcal{B}}$ is an acceptable approximation to the exact ground-state energy, i.e., $E_0^{\mathcal{B}} \approx E_0$.

In the present context, it is important to note that the wave functions Ψ defined in Eq. (2) are not restricted to a finite basis set, i.e., they should be expanded in a complete basis set. In Ref. 50, it was then proposed to decompose $F[n]$ as, for a representable-in-$\mathcal{B}$ density n,

$$F[n] = \min_{\Psi^{\mathcal{B}} \to n} \langle \Psi^{\mathcal{B}} | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi^{\mathcal{B}} \rangle + \bar{E}^{\mathcal{B}}[n], \quad (3)$$

where $\Psi^{\mathcal{B}}$ are wave functions expandable in the N-electron Hilbert space generated by $\mathcal{B}$ and

$$\bar{E}^{\mathcal{B}}[n] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi \rangle - \min_{\Psi^{\mathcal{B}} \to n} \langle \Psi^{\mathcal{B}} | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi^{\mathcal{B}} \rangle \quad (4)$$

is the complementary density functional to the basis set $\mathcal{B}$. Introducing the decomposition in Eq. (3) back into Eq. (1) yields

$$E_0^{\mathcal{B}} = \min_{\Psi^{\mathcal{B}}} \left\{ \langle \Psi^{\mathcal{B}} | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi^{\mathcal{B}} \rangle + \bar{E}^{\mathcal{B}}[n_{\Psi^{\mathcal{B}}}] + \int \mathrm{d}\mathbf{r} \, v_{\mathrm{ne}}(\mathbf{r}) n_{\Psi^{\mathcal{B}}}(\mathbf{r}) \right\}, \quad (5)$$

where the minimization is only over wave functions $\Psi^{\mathcal{B}}$ restricted to the basis set $\mathcal{B}$ and $n_{\Psi^{\mathcal{B}}}(\mathbf{r})$ refers to the density generated from $\Psi^{\mathcal{B}}$. Therefore, thanks to Eq. (5), one can properly combine a WFT calculation in a finite basis set with a density functional (hereafter referred to as complementary functional) accounting for the correlation effects that are not included in the basis set.

As a simple non-self-consistent version of this approach, we can approximate the minimizing wave function $\Psi_0^{\mathcal{B}}$ in Eq. (5) by the ground-state FCI wave function $\Psi_{\mathrm{FCI}}^{\mathcal{B}}$ within $\mathcal{B}$, and we then obtain the following approximation for the exact ground-state energy [see Eqs. (12)–(15) of Ref. 50]:

$$E_0 \approx E_0^{\mathcal{B}} \approx E_{\mathrm{FCI}}^{\mathcal{B}} + \bar{E}^{\mathcal{B}}[n_{\mathrm{FCI}}^{\mathcal{B}}], \quad (6)$$

where $E_{\mathrm{FCI}}^{\mathcal{B}}$ and $n_{\mathrm{FCI}}^{\mathcal{B}}$ are the ground-state FCI energy and density, respectively. As was originally shown in Ref. 50 and further emphasized in Refs. 51 and 52, the main role of $\bar{E}^{\mathcal{B}}[n_{\mathrm{FCI}}^{\mathcal{B}}]$ is to correct for the basis-set incompleteness error, a large part of which originating from the lack of electron–electron cusp in the wave function expanded in an incomplete basis set. The whole purpose of this work is to determine approximations for $\bar{E}^{\mathcal{B}}[n_{\mathrm{FCI}}^{\mathcal{B}}]$, which are suitable for strongly correlated molecular systems. Two key requirements for this purpose are (i) spin-multiplet degeneracy and (ii) size consistency.

B. Effective interaction in a finite basis

As originally shown by Kato,[56] the electron–electron cusp of the exact wave function originates from the divergence of the Coulomb interaction at the coalescence point. Therefore, a cuspless wave function $\Psi^{\mathcal{B}}$ could also be obtained from a Hamiltonian with a non-divergent electron–electron interaction. In other words, the impact of the basis set incompleteness can be understood as the removal of the divergence of the usual electron–electron Coulomb interaction.

As originally derived in Ref. 50 (see Sec. II D and Appendixes A and B), one can obtain an effective non-divergent electron–electron interaction, here referred to as $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)$, which reproduces the expectation value of the electron–electron Coulomb interaction operator over a given wave function $\Psi^{\mathcal{B}}$. As we are interested

in the behavior at the coalescence point, we focus on the opposite-spin part of the electron–electron interaction. More specifically, the effective electron–electron interaction associated with a given wave function $\Psi^{\mathcal{B}}$ is defined as

$$W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} f_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)/n_{2,\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) & \text{if } n_{2,\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) \neq 0 \\ \infty, & \text{otherwise,} \end{cases} \quad (7)$$

where

$$n_{2,\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pqrs \in \mathcal{B}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2)\Gamma_{pq}^{rs}\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2) \quad (8)$$

is the opposite-spin pair density associated with $\Psi^{\mathcal{B}}$, and $\Gamma_{pq}^{rs} = 2\langle \Psi^{\mathcal{B}}|\hat{a}_{r_\uparrow}^\dagger \hat{a}_{s_\uparrow}^\dagger \hat{a}_{q_\uparrow} \hat{a}_{p_\downarrow}|\Psi^{\mathcal{B}}\rangle$ is its associated tensor in a basis of spatial orthonormal orbitals $\{\phi_p(\mathbf{r})\}$,

$$f_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pqrstu \in \mathcal{B}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2) V_{pq}^{rs}\Gamma_{rs}^{tu}\phi_t(\mathbf{r}_1)\phi_u(\mathbf{r}_2) \quad (9)$$

and $V_{pq}^{rs} = \langle pq|rs \rangle$ are the usual two-electron Coulomb integrals. With such a definition, one can show that $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)$ satisfies

$$\frac{1}{2}\iint d\mathbf{r}_1 d\mathbf{r}_2 \, W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) n_{2,\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2}\iint d\mathbf{r}_1 d\mathbf{r}_2 \frac{n_{2,\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \quad (10)$$

As shown in Ref. 50, the effective interaction $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)$ is necessarily finite at coalescence for an incomplete basis set and tends to the usual Coulomb interaction in the CBS limit for any choice of wave function $\Psi^{\mathcal{B}}$, i.e.,

$$\lim_{\mathcal{B} \to \text{CBS}} W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad \forall \, \Psi^{\mathcal{B}}. \quad (11)$$

The condition in Eq. (11) is fundamental as it guarantees the correct behavior of the theory in the CBS limit.

C. Local range-separation function

1. General definition

The effective interaction within a finite basis, $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)$, is bounded and resembles the long-range interaction used in RSDFT,

$$w_{ee}^{lr}(\mu; r_{12}) = \frac{\text{erf}(\mu \, r_{12})}{r_{12}}, \quad (12)$$

where μ is the range-separation parameter. As originally proposed in Ref. 50, we make the correspondence between these two interactions by using the local range-separation function

$$\mu_{\Psi^{\mathcal{B}}}(\mathbf{r}) = \frac{\sqrt{\pi}}{2} W_{\Psi^{\mathcal{B}}}(\mathbf{r}, \mathbf{r}) \quad (13)$$

such that the two interactions coincide at the electron–electron coalescence point for each $\mathbf{r}$,

$$w_{ee}^{lr}(\mu_{\Psi^{\mathcal{B}}}(\mathbf{r}); 0) = W_{\Psi^{\mathcal{B}}}(\mathbf{r}, \mathbf{r}), \quad \forall \, \mathbf{r}. \quad (14)$$

Because of the very definition of $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2)$, one has the following property in the CBS limit [see Eq. (11)]:

$$\lim_{\mathcal{B} \to \text{CBS}} \mu_{\Psi^{\mathcal{B}}}(\mathbf{r}) = \infty, \quad \forall \, \Psi^{\mathcal{B}}, \quad (15)$$

which is again fundamental to guarantee the correct behavior of the theory in the CBS limit.

2. Frozen-core approximation

As all WFT calculations in this work are performed within the frozen-core approximation, we use a "valence-only" (or no-core) version of the various quantities needed for the complementary functional introduced in Ref. 51. We partition the basis set as $\mathcal{B} = \mathcal{C} \cup \mathcal{V}$, where $\mathcal{C}$ and $\mathcal{V}$ are the sets of core and "valence" (i.e., non-core) orbitals, respectively, and define the valence-only local range-separation function as

$$\mu_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}) = \frac{\sqrt{\pi}}{2} W_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}, \mathbf{r}), \quad (16)$$

where

$$W_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} f_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2)/n_{2,\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2) & \text{if } n_{2,\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2) \neq 0 \\ \infty & \text{otherwise} \end{cases} \quad (17)$$

is the valence-only effective interaction and

$$f_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pq \in \mathcal{B}} \sum_{rstu \in \mathcal{V}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2) V_{pq}^{rs}\Gamma_{rs}^{tu}\phi_t(\mathbf{r}_1)\phi_u(\mathbf{r}_2), \quad (18)$$

$$n_{2,\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pqrs \in \mathcal{V}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2)\Gamma_{pq}^{rs}\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2). \quad (19)$$

One would note the restrictions of the sums to the set $\mathcal{V}$ in Eqs. (18) and (19). It is also noteworthy that, with the present definition, $W_{\Psi^{\mathcal{B}}}^{\text{val}}(\mathbf{r}_1, \mathbf{r}_2)$ still tends to the usual Coulomb interaction as $\mathcal{B} \to \text{CBS}$. For simplicity, we will drop the indication "val" in the notation for the rest of the paper.

D. General form of the complementary functional

1. Generic approximate form

As originally proposed and motivated in Ref. 50, we approximate the complementary functional $\bar{E}^{\mathcal{B}}[n]$ by using the so-called correlation energy functional with multideterminant reference (ECMD) introduced by Toulouse *et al.*[57,58] Following the recent work in Ref. 51, we propose to consider a Perdew–Burke–Ernzerhof (PBE)-like functional that uses the one-electron density $n(\mathbf{r})$, the spin polarization $\zeta(\mathbf{r}) = [n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r})]/n(\mathbf{r})$ [where $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$ are the spin-up and spin-down densities], the reduced density gradient $s(\mathbf{r}) = \nabla n(\mathbf{r})/n(\mathbf{r})^{4/3}$, and the on-top pair density $n_2(\mathbf{r}) \equiv n_2(\mathbf{r}, \mathbf{r})$. In the present work, all these quantities are computed with the same wave function $\Psi^{\mathcal{B}}$ used to define $\mu(\mathbf{r}) \equiv \mu_{\Psi^{\mathcal{B}}}(\mathbf{r})$. Therefore, $\bar{E}^{\mathcal{B}}[n]$ has the following generic form:

$$\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu] = \int d\mathbf{r} \, n(\mathbf{r}) \bar{\varepsilon}_{\text{c,md}}^{\text{sr,PBE}}(n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r})), \quad (20)$$

where

$$\bar{\varepsilon}_{\text{c,md}}^{\text{sr,PBE}}(n, \zeta, s, n_2, \mu) = \frac{\varepsilon_c^{\text{PBE}}(n, \zeta, s)}{1 + \beta(n, \zeta, s, n_2)\,\mu^3} \quad (21)$$

is the correlation energy per particle with

$$\beta(n, \zeta, s, n_2) = \frac{3}{2\sqrt{\pi}(1 - \sqrt{2})} \frac{\varepsilon_c^{\text{PBE}}(n, \zeta, s)}{n_2/n}, \quad (22)$$

where $\varepsilon_c^{\mathrm{PBE}}(n, \zeta, s)$ is the usual PBE correlation energy per particle.[59] Before introducing the different flavors of approximate functionals that we will use here (see Sec. II F), we would like to give some motivations for this choice of functional form.

The form of $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$ in Eq. (21) has been originally proposed in Ref. 48 in the context of RSDFT. In the $\mu \to 0$ limit, it reduces to the usual PBE correlation functional, i.e.,

$$\lim_{\mu \to 0} \bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu) = \varepsilon_c^{\mathrm{PBE}}(n, \zeta, s), \qquad (23)$$

which is relevant in the weak-correlation (or high-density) limit. In the large-μ limit, it behaves as

$$\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu) \underset{\mu \to \infty}{\sim} \frac{2\sqrt{\pi}(1 - \sqrt{2})}{3\,\mu^3} \frac{n_2}{n}, \qquad (24)$$

which is the exact large-μ behavior of the exact ECMD correlation energy.[48,60] Of course, for a specific system, the large-μ behavior will be exact only if one uses for n_2 the *exact* on-top pair density of this system. This large-μ limit in Eq. (24) is relevant in the strong-correlation (or low-density) limit. In the context of RSDFT, some of the present authors have illustrated in Ref. 48 that the on-top pair density involved in Eq. (21) plays, indeed, a crucial role when reaching the strong-correlation regime. The importance of the on-top pair density in the strong-correlation regime has been also recently acknowledged by Gagliardi and co-workers[61] and Pernal and co-workers.[62]

Note also that $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$ vanishes when n_2 vanishes, i.e.,

$$\lim_{n_2 \to 0} \bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu) = 0, \qquad (25)$$

which is expected for systems with a vanishing on-top pair density. Finally, the function $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$ vanishes when $\mu \to \infty$ like all RSDFT short-range functionals, i.e.,

$$\lim_{\mu \to \infty} \bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu) = 0. \qquad (26)$$

2. Two limits where the complementary functional vanishes

Within the definitions of Eqs. (13) and (20), any approximate complementary functional $\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu]$ satisfies two important properties.

First, thanks to the properties in Eqs. (15) and (26), $\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu]$ vanishes in the CBS limit, independently of the type of wave function $\Psi^{\mathcal{B}}$ used to define the local range-separation function $\mu(\mathbf{r})$ in a given basis set $\mathcal{B}$,

$$\lim_{\mathcal{B} \to \mathrm{CBS}} \bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu] = 0, \quad \forall\, \Psi^{\mathcal{B}}. \qquad (27)$$

Second, $\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu]$ correctly vanishes for systems with uniformly vanishing on-top pair density, such as one-electron systems, and for the stretched H_2 molecule,

$$\lim_{n_2 \to 0} \bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu] = 0. \qquad (28)$$

This property is doubly guaranteed by (i) the choice of setting $W_{\Psi^{\mathcal{B}}}(\mathbf{r}_1, \mathbf{r}_2) = \infty$ for a vanishing pair density [see Eq. (7)], which leads to $\mu(\mathbf{r}) \to \infty$ and thus a vanishing $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$ [see

Eq. (26)], and (ii) the fact that $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$ vanishes anyway when the on-top pair density vanishes [see Eq. (25)].

E. Requirements on the complementary functional for strong correlation

An important requirement for any electronic-structure method is size consistency, i.e., the additivity of the energies of non-interacting fragments, which is mandatory to avoid any ambiguity in computing interaction energies. When two subsystems A and B dissociate in closed-shell systems, as in the case of weak intermolecular interactions, for instance, spin-restricted Hartree–Fock (RHF) is size-consistent. When the two subsystems dissociate in open-shell systems, such as in covalent bond breaking, it is well known that the RHF approach fails and an alternative is to use a complete-active-space self-consistent-field (CASSCF) wave function that, provided that the active space has been properly chosen, leads to additive energies.

Another important requirement is spin-multiplet degeneracy, i.e., the independence of the energy with respect to the S_z component of a given spin state, which is also a property of any exact wave function. Such a property is also important in the context of covalent bond breaking, where the ground state of the supersystem A + B is generally of lower spin than the corresponding ground states of the fragments (A and B) that can have multiple S_z components.

1. Spin-multiplet degeneracy

A sufficient condition to achieve spin-multiplet degeneracy is to eliminate all dependencies on S_z. In the case of the function $\bar{\varepsilon}_{c,\mathrm{md}}^{\mathrm{sr,PBE}}(n, \zeta, s, n_2, \mu)$, this means removing the dependence on the spin polarization $\zeta(\mathbf{r})$ originating from the PBE correlation functional $\varepsilon_c^{\mathrm{PBE}}(n, \zeta, s)$ [see Eq. (21)].

To do so, it has been proposed to replace the dependence on the spin polarization by the dependence on the on-top pair density. Most often, it is done by introducing an effective spin polarization[7,63–75] (see, also, Refs. 76 and 77)

$$\tilde{\zeta}(n, n_2) = \sqrt{1 - 2\,n_2/n^2} \qquad (29)$$

expressed as a function of the density n and the on-top pair density n_2 calculated from a given wave function. The advantage of this approach is that this effective spin polarization $\tilde{\zeta}$ is independent from S_z since the on-top pair density is S_z-independent. Nevertheless, the use of $\tilde{\zeta}$ in Eq. (29) presents some disadvantages since this expression was derived for a single-determinant wave function. Hence, it does not appear justified to use it for a multideterminant wave function. More particularly, it may happen in the multideterminant case that $1 - 2\,n_2/n^2 < 0$, which results in a complex-valued effective spin polarization.[64] Therefore, following other authors,[67,72,73] we use the following definition:

$$\tilde{\zeta}(n, n_2) = \begin{cases} \sqrt{1 - 2\,n_2/n^2} & \text{if } n^2 \geq 2n_2 \\ 0 & \text{otherwise.} \end{cases} \qquad (30)$$

An alternative way to eliminate the S_z dependence is to simply set $\zeta = 0$, i.e., to resort to the spin-unpolarized functional. This lowers the accuracy for open-shell systems at $\mu = 0$, i.e., for the usual PBE correlation functional $\varepsilon_c^{\mathrm{PBE}}(n, \zeta, s)$. Nevertheless, we argue that

for sufficiently large μ, it is a viable option. Indeed, the purpose of introducing the spin polarization in semilocal density-functional approximations is to mimic the exact on-top pair density,[76] but our functional $\bar{\varepsilon}_{c,md}^{sr,PBE}(n, \zeta, s, n_2, \mu)$ already explicitly depends on the on-top pair density [see Eqs. (21) and (22)]. The dependencies on ζ and n_2 can thus be expected to be largely redundant. Consequently, we propose here to test the use of $\bar{\varepsilon}_{c,md}^{sr,PBE}$ with *a zero spin polarization*. This ensures its S_z independence and, as will be numerically demonstrated, very weakly affects the complementary functional accuracy.

2. Size consistency

Since $\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu]$ is computed via a single integral over $\mathbb{R}^3$ [see Eq. (20)] that involves only local quantities [$n(\mathbf{r})$, $\zeta(\mathbf{r})$, $s(\mathbf{r})$, $n_2(\mathbf{r})$, and $\mu(\mathbf{r})$], in the case of non-overlapping fragments A + B, it can be written as the sum of two local contributions: one coming from the integration over the region of subsystem A and the other from the region of subsystem B. Therefore, a sufficient condition for size consistency is that these quantities locally coincide in the isolated fragments and in the supersystem A + B. Since these local quantities are calculated from the wave function $\Psi^{\mathcal{B}}$, a sufficient condition is that the wave function is multiplicatively separable in the limit of non-interacting fragments, i.e., $|\Psi_{A+B}^{\mathcal{B}}\rangle = |\Psi_A^{\mathcal{B}}\rangle \otimes |\Psi_B^{\mathcal{B}}\rangle$. We refer the interested reader to Appendix A for a detailed proof and discussion of the latter statement. In the case where the two subsystems A and B dissociate in closed-shell systems, a simple RHF wave function ensures this property, but when one or several covalent bonds are broken, a properly chosen CASSCF wave function can be used to recover this property. The underlying active space must however be chosen in such a way that it leads to size-consistent energies in the limit of dissociated fragments.

F. Actual approximations used for the complementary functional

As the present work focuses on the strong-correlation regime, we propose here to investigate only approximate functionals that are S_z independent and size-consistent in the case of covalent bond breaking. Therefore, the wave functions $\Psi^{\mathcal{B}}$ used throughout this paper are CASSCF wave functions in order to ensure size consistency of all local quantities. The difference between the different flavors of functionals is only due to the types of spin polarization and on-top pair density used.

Regarding the spin polarization that enters into the function $\varepsilon_c^{PBE}(n, \zeta, s)$, two different types of S_z-independent formulations are considered: (i) the *effective* spin polarization $\tilde{\zeta}$ defined in Eq. (30) and calculated from the CASSCF wave function and (ii) a *zero* spin polarization. In the latter case, the functional is referred to as "SU," which stands for "spin unpolarized."

Regarding the on-top pair density entering in Eq. (22), we use two different approximations. The first one is based on the uniform electron gas (UEG) and reads

$$n_2^{UEG}(n, \zeta) \approx n^2(1 - \zeta^2)g_0(n),\qquad(31)$$

where the pair-distribution function $g_0(n)$ is taken from Eq. (46) of Ref. 33. As the spin polarization appears in Eq. (31), we use the

effective spin polarization $\tilde{\zeta}$ of Eq. (30) in order to ensure S_z independence. Thus, n_2^{UEG} will depend indirectly on the on-top pair density of the CASSCF wave function through $\tilde{\zeta}$. When using $n_2^{UEG}(\mathbf{r}) \equiv n_2^{UEG}(n(\mathbf{r}), \tilde{\zeta}(\mathbf{r}))$ in a functional, we will refer to it as "UEG."

The second approach to approximate the exact on-top pair density consists in using directly the on-top pair density of the CASSCF wave function. Following the work of some of the present authors,[48,52] we introduce the extrapolated on-top pair density

$$\mathring{n}_2(n_2, \mu) = \left(1 + \frac{2}{\sqrt{\pi}\mu}\right)^{-1} n_2,\qquad(32)$$

which directly follows from the large-μ extrapolation of the exact on-top pair density derived by Gori-Giorgi and Savin[33] in the context of RSDFT. Thus, the extrapolated on-top pair density $\mathring{n}_2$ depends on the local range-separation function μ. When using $\mathring{n}_2(\mathbf{r}) \equiv \mathring{n}_2(n_2(\mathbf{r}), \mu(\mathbf{r}))$ in a functional, we will simply refer to it as "OT," which stands for "on-top."

We then define three complementary functionals:

(i) PBE-UEG that combines the effective spin polarization of Eq. (30) and the UEG on-top pair density defined in Eq. (31),

$$\bar{E}_{PBE-UEG}^{\mathcal{B}} = \int d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), \tilde{\zeta}(\mathbf{r}), s(\mathbf{r}), n_2^{UEG}(\mathbf{r}), \mu(\mathbf{r})),$$

$$(33)$$

(ii) PBE-OT that combines the effective spin polarization of Eq. (30) and the on-top pair density of Eq. (32),

$$\bar{E}_{PBE-OT}^{\mathcal{B}} = \int d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), \tilde{\zeta}(\mathbf{r}), s(\mathbf{r}), \mathring{n}_2(\mathbf{r}), \mu(\mathbf{r})),$$

$$(34)$$

(iii) SU-PBE-OT that combines a zero spin polarization and the on-top pair density of Eq. (32),

$$\bar{E}_{SU-PBE-OT}^{\mathcal{B}} = \int d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), 0, s(\mathbf{r}), \mathring{n}_2(\mathbf{r}), \mu(\mathbf{r})).$$

$$(35)$$

The performance of each of these functionals is tested in the following. Note that we did not define a spin-unpolarized version of the PBE-UEG functional because it would have been significantly inferior (in terms of performance) compared to the three other functionals. Indeed, because of the lack of knowledge on the spin polarization or on the accurate on-top pair density, such a functional would be inaccurate, in particular, for open-shell systems. This assumption has been numerically confirmed by calculations.

III. RESULTS

A. Computational details

We present potential energy curves of small molecules up to the dissociation limit to investigate the performance of the basis-set correction in regimes of both weak and strong correlations. The considered systems are the H_{10} linear chain with equally spaced atoms and the N_2, O_2, and F_2 diatomics.

The computation of the ground-state energy in Eq. (6) in a given basis set requires approximations to the FCI energy $E_{FCI}^{\mathcal{B}}$ and

to the basis-set correction $\bar{E}^{\mathcal{B}}[n_{\text{FCI}}^{\mathcal{B}}]$. For diatomics with the aug-cc-pVDZ and aug-cc-pVTZ basis sets,[78] energies are obtained using frozen-core selected-CI calculations (using the CIPSI algorithm), followed by the extrapolation scheme proposed by Holmes *et al.* (see Refs. 79–84 for more details). All these calculations are performed with the latest version of QUANTUM PACKAGE[84] and will be labeled exFCI in the following. In the case of F_2, we also use the correlation energy extrapolated by intrinsic scaling (CEEIS)[85] as an estimate of the FCI correlation energy with the cc-pVXZ (X = D, T, and Q) basis sets.[86] The estimated exact potential energy curves are obtained from experimental data[87] for the N_2 and O_2 molecules and from CEEIS calculations in the case of F_2. For all geometries and basis sets, the error with respect to the exact FCI energies is estimated to be of the order of 0.5 mHa. For the three diatomics, we performed an additional exFCI calculation with the aug-cc-pVQZ basis set at the equilibrium geometry to obtain reliable estimates of the FCI/CBS dissociation energy. In the case of the H_{10} chain, the approximation to the FCI energies together with the estimated exact potential energy curves is obtained from the data of Ref. 88, where the authors performed MRCI+Q calculations with a minimal valence active space as reference (see below for the description of the active space).

We note that even though we use near-FCI energies in this work, the DFT-based basis-set correction could also be applied to any approximation to FCI such as multireference perturbation theory, similar to what was done for weakly correlated systems for which the basis-set correction was applied to CCSD(T) calculations.[51]

Regarding the complementary functional, we first perform full-valence CASSCF calculations with the GAMESS-US software[89] to obtain the wave function $\Psi^{\mathcal{B}}$. Then, all density-related quantities involved in the functional [density $n(\mathbf{r})$, effective spin polarization $\tilde{\zeta}(\mathbf{r})$, reduced density gradient $s(\mathbf{r})$, and on-top pair density $n_2(\mathbf{r})$] together with the local range-separation function $\mu(\mathbf{r})$ are calculated with this full-valence CASSCF wave function. The CASSCF calculations are performed with the following active spaces: (10e, 10o) for H_{10}, (10e, 8o) for N_2, (12e, 8o) for O_2, and (14e, 8o) for F_2. We note that instead of using CASSCF wave functions for $\Psi^{\mathcal{B}}$, one could of course use the same selected-CI wave functions used for calculating the energy, but the calculations of $n_2(\mathbf{r})$ and $\mu(\mathbf{r})$ would then be more costly.

In addition, as the frozen-core approximation is used in all our selected-CI calculations, we use the corresponding valence-only complementary functionals (see Subsection II C 2). Therefore, all density-related quantities exclude any contribution from the 1s core orbitals, and the range-separation function follows the definition given in Eq. (16).

It should be stressed that the computational cost of the basis-set correction (see Appendix B) is negligible compared to the cost of the selected-CI calculations.

B. H_{10} chain

The H_{10} chain with equally spaced atoms is a prototype of strongly correlated systems as it consists in the simultaneous breaking of 10 interacting covalent σ bonds. As it is a relatively small system, benchmark calculations at near-CBS values are available (see Ref. 88 for a detailed study of this system).

We report in Fig. 1 the potential energy curves computed using the cc-pVXZ (X = D, T, and Q) basis sets for different levels of approximation, and the corresponding atomization energies are reported in Table I. As a general trend, the addition of the basis-set correction globally improves the quality of the potential energy curves, independently of the approximation level of $\bar{E}^{\mathcal{B}}[n]$. In addition, no erratic behavior is found when stretching the bonds, which shows that the present procedure (i.e., the determination of the range-separation function and the definition of the functionals) is robust when reaching the strong-correlation regime. In other words, smooth potential energy curves are obtained with the present basis-set correction. More quantitatively, the values of the atomization energies are within chemical accuracy (i.e., an error below 1.4 mHa) with the cc-pVTZ basis set when using the PBE-OT and SU-PBE-OT functionals, whereas such an accuracy is not yet reached at the standard MRCI+Q/cc-pVQZ level of theory.

Analyzing more carefully the performance of the different types of approximate functionals, the results show that PBE-OT and SU-PBE-OT are very similar (the maximal difference on the atomization energy being 0.3 mHa) and that they give slightly more accurate results than PBE-UEG. These findings provide two important clues on the role of the different physical ingredients included in these functionals: (i) the explicit use of the on-top pair density originating from the CASSCF wave function [see Eq. (32)] is preferable over the use of the UEG on-top pair density [see Eq. (31)], which is somewhat understandable, and (ii) removing the dependence on any kind of spin polarization does not lead to a significant loss of accuracy, providing that one employs a qualitatively correct on-top pair density. The latter point is crucial as it confirms that the spin polarization in density-functional approximations essentially plays the same role as the on-top pair density. This could have significant implications for the construction of more robust families of density-functional approximations within DFT.

C. Dissociation of diatomics

The N_2, O_2, and F_2 molecules are complementary to the H_{10} system for the present study. The level of strong correlation in these diatomics also increases while stretching the bonds, similar to the case of H_{10}, but in addition, these molecules exhibit more important and versatile types of weak correlations due to the larger number of electrons. Indeed, the short-range correlation effects are known to play a strong differential effect on the computation of the atomization energy at equilibrium, while the shape of the curve far from the equilibrium geometry is governed by dispersion interactions that are medium to long-range weak-correlation effects.[5] The dispersion interactions in H_{10} play a minor role on the potential energy curve due to the much smaller number of near-neighbor electron pairs compared to N_2, O_2, or F_2. In addition, O_2 has a triplet ground state and is therefore a good candidate for checking the spin-polarization dependence of the various functionals proposed here.

We report in Figs. 2–4 the potential energy curves of N_2, O_2, and F_2 computed at various approximation levels using the aug-cc-pVDZ and aug-cc-pVTZ basis sets. The atomization energies for each level of theory with different basis sets are reported in Table I.

Just as in H_{10}, the accuracy of the atomization energies is globally improved by adding the basis-set correction, and it is remarkable that PBE-OT and SU-PBE-OT provide again very similar results.
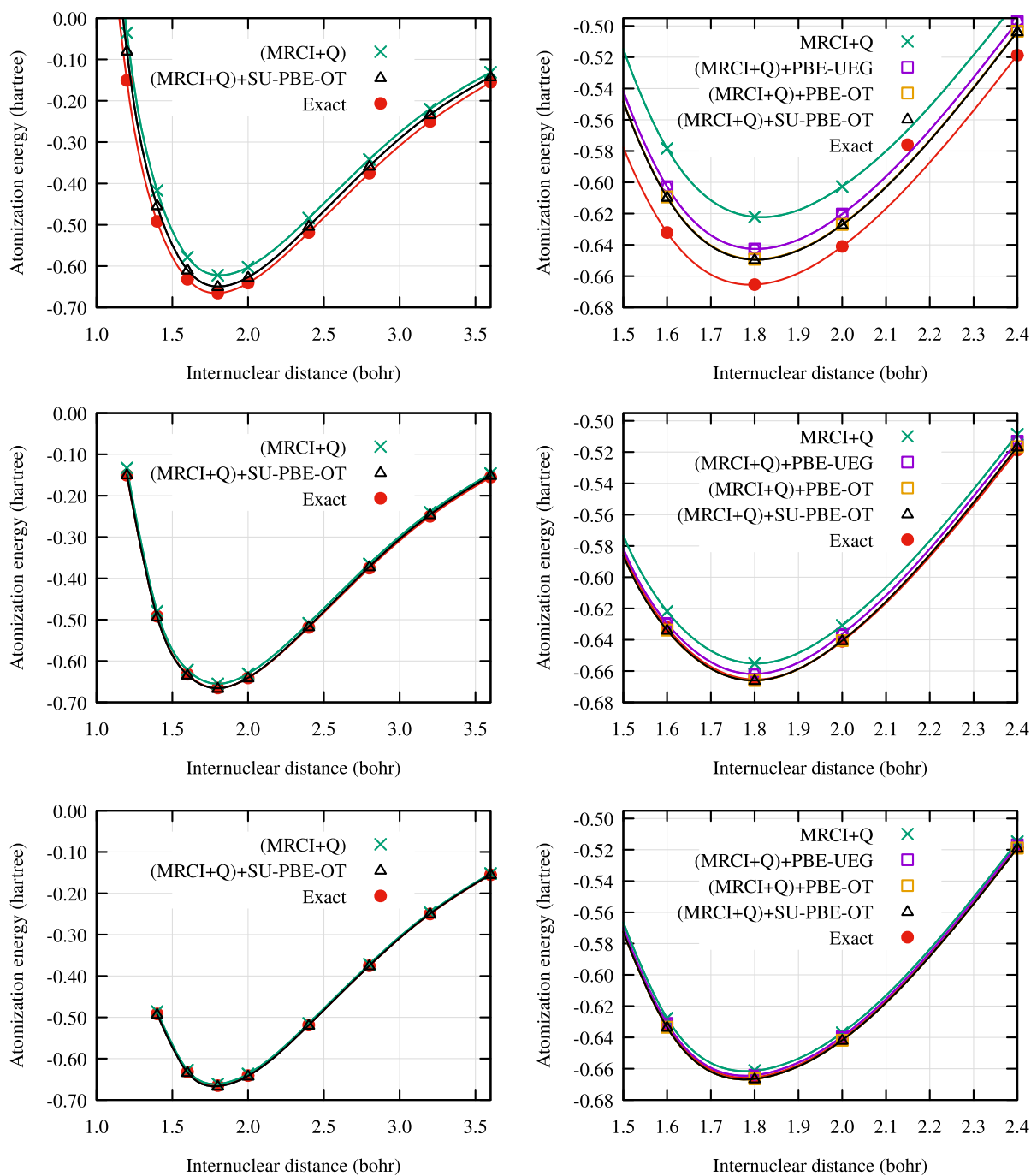
FIG. 1. Potential energy curves of the H_{10} chain with equally spaced atoms calculated with MRCI+Q and basis-set corrected MRCI+Q using the cc-pVDZ (top), cc-pVTZ (middle), and cc-pVQZ (bottom) basis sets. The MRCI+Q energies and the estimated exact energies have been extracted from Ref. 88.

The latter observation confirms that the dependence on the on-top pair density allows one to remove the dependence of any kind of spin polarization for a quite wide range of covalent bonds and also for an open-shell system such as O_2. More quantitatively, an error below 1.0 mHa compared to the estimated exact valence-only atomization energy is found for N_2, O_2, and F_2 with the aug-cc-pVTZ basis set using the SU-PBE-OT functional, whereas such a feat is far from being reached within the same basis set at the near-FCI level.

TABLE I. Atomization energies (in mHa) and associated errors (in square brackets) with respect to the estimated exact values computed at different levels of theory with various basis sets.

System	Basis set	MRCI+Q[a]	(MRCI+Q)+PBE-UEG	(MRCI+Q)+PBE-OT	(MRCI+Q)+SU-PBE-OT
H_{10}	cc-pVDZ	622.1 [43.3]	642.6 [22.8]	649.2 [16.2]	649.5 [15.9]
	cc-pVTZ	655.2 [10.2]	661.9 [3.5]	666.0 [−0.6]	666.0 [−0.6]
	cc-pVQZ	661.2 [4.2]	664.1 [1.3]	666.4 [−1.0]	666.5 [−1.1]
			Estimated exact:[a] 665.4		
		exFCI	exFCI+PBE-UEG	exFCI+PBE-OT	exFCI+SU-PBE-OT
N_2	aug-cc-pVDZ	321.9 [40.8]	356.2 [6.5]	355.5 [7.2]	354.6 [8.1]
	aug-cc-pVTZ	348.5 [14.2]	361.5 [1.2]	363.5 [−0.5]	363.2 [−0.3]
	aug-cc-pVQZ	356.6 [6.1]	362.8 [−0.1]	364.2 [−1.5]	364.3 [−1.6]
			Estimated exact:[b] 362.7		
		exFCI	exFCI+PBE-UEG	exFCI+PBE-OT	exFCI+SU-PBE-OT
O_2	aug-cc-pVDZ	171.4 [20.5]	187.6 [4.3]	187.6 [4.3]	187.1 [4.8]
	aug-cc-pVTZ	184.5 [7.4]	190.3 [1.6]	191.2 [0.7]	191.0 [0.9]
	aug-cc-pVQZ	188.3 [3.6]	190.3 [1.6]	191.0 [0.9]	190.9 [1.0]
			Estimated exact:[b] 191.9		
		exFCI	exFCI+PBE-UEG	exFCI+PBE-OT	exFCI+SU-PBE-OT
F_2	aug-cc-pVDZ	49.6 [12.6]	54.8 [7.4]	54.9 [7.3]	54.8 [7.4]
	aug-cc-pVTZ	59.3 [2.9]	61.2 [1.0]	61.5 [0.7]	61.5 [0.7]
	aug-cc-pVQZ	60.1 [2.1]	61.0 [1.2]	61.3 [0.9]	61.3 [0.9]
		CEEIS[c]	CEEIS[c]+PBE-UEG	CEEIS[c]+PBE-OT	CEEIS[c]+SU-PBE-OT
	cc-pVDZ	43.7 [18.5]	51.0 [11.2]	51.0 [11.2]	50.7 [11.5]
	cc-pVTZ	56.3 [5.9]	59.2 [3.0]	59.6 [2.6]	59.5 [2.7]
	cc-pVQZ	59.9 [2.3]	61.3 [0.9]	61.6 [0.6]	61.6 [0.6]
			Estimated exact:[b] 62.2		

[a] From Ref. 88.
[b] From the CEEIS valence-only non-relativistic calculations of Ref. 90.
[c] From the CEEIS valence-only non-relativistic calculations of Ref. 85.

In the case of F_2, it is clear that the addition of diffuse functions in the double- and triple-ζ basis sets strongly improves the accuracy of the results, which could have been anticipated due to the strong breathing-orbital effect induced by the ionic valence-bond forms in this molecule.[91] It should also be noticed that when reaching the aug-cc-pVQZ basis set for N_2, the accuracy of the atomization energy slightly deteriorates for the PBE-OT and SU-PBE-OT functionals, but it remains, nevertheless, more accurate than the estimated FCI atomization energy and very close to chemical accuracy.

The overestimation of the basis-set-corrected atomization energy observed for N_2 in large basis sets reveals an unbalanced treatment between the molecule and the atom in favor of the molecular system. Since the integral over $\mathbf{r}$ of the on-top pair density $n_2(\mathbf{r})$ is proportional to the short-range correlation energy in the large-μ limit[48,60] [see Eq. (24)], the accuracy of a given approximation of the exact on-top pair density will have a direct influence on the accuracy of the related basis-set correction energy $\bar{E}^{\mathcal{B}}$. To quantify the quality

of different flavors of on-top pair densities for a given system and a given basis set $\mathcal{B}$, we define the system-averaged CASSCF on-top pair density and extrapolated on-top pair density as

$$\langle n_{2,\text{CASSCF}} \rangle = \int d\mathbf{r}\, n_{2,\text{CASSCF}}(\mathbf{r}), \quad (36a)$$

$$\langle \mathring{n}_{2,\text{CASSCF}} \rangle = \int d\mathbf{r}\, \mathring{n}_{2,\text{CASSCF}}(\mathbf{r}), \quad (36b)$$

where $\mathring{n}_{2,\text{CASSCF}}(\mathbf{r}) = \mathring{n}_2(n_{2,\text{CASSCF}}(\mathbf{r}), \mu_{\text{CASSCF}}(\mathbf{r}))$ [see Eq. (32)] and $\mu_{\text{CASSCF}}(\mathbf{r})$ is the local range-separation function calculated with the CASSCF wave function, and similarly, we define the system-averaged CIPSI on-top pair density and extrapolated on-top pair density as

$$\langle n_{2,\text{CIPSI}} \rangle = \int d\mathbf{r}\, n_{2,\text{CIPSI}}(\mathbf{r}), \quad (37a)$$

$$\langle \mathring{n}_{2,\text{CIPSI}} \rangle = \int d\mathbf{r}\, \mathring{n}_{2,\text{CIPSI}}(\mathbf{r}), \quad (37b)$$
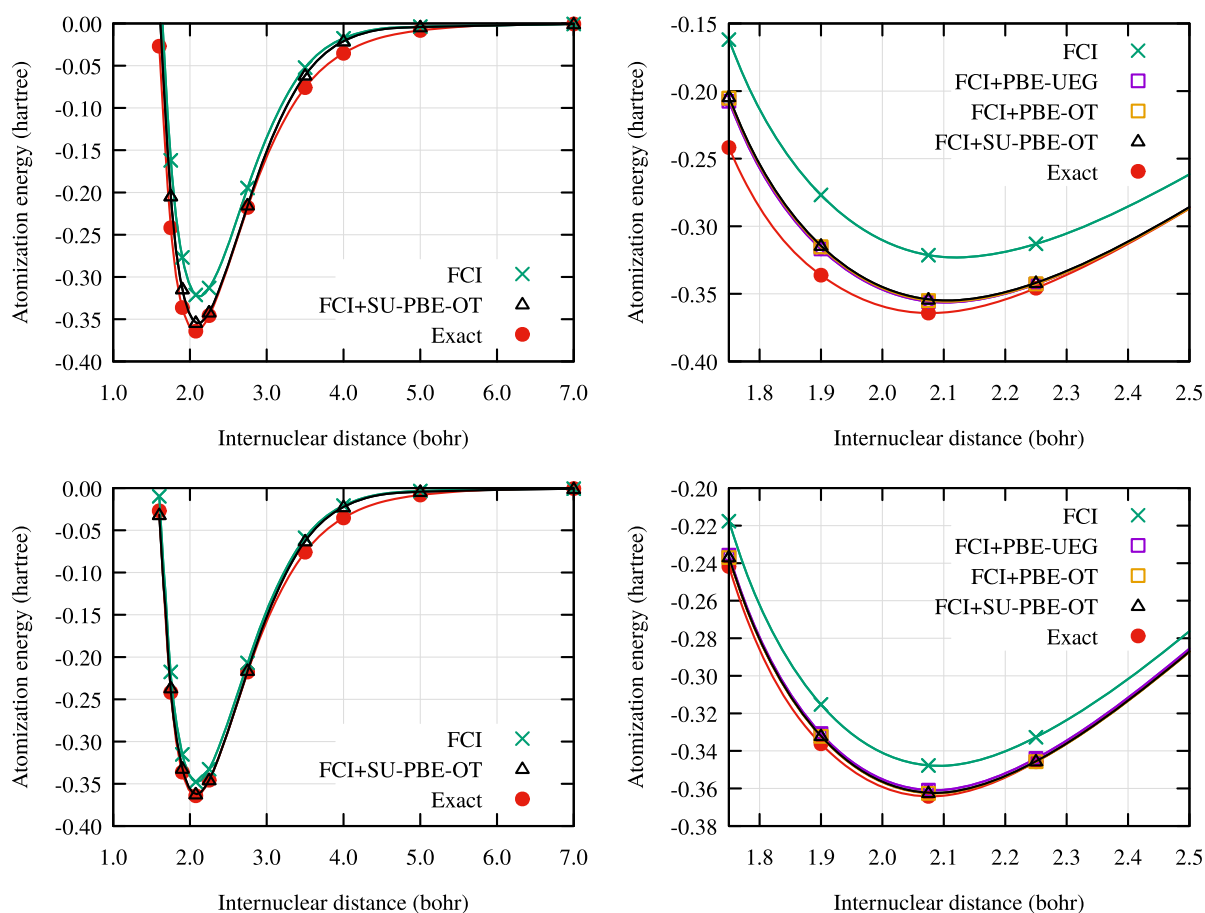
FIG. 2. Potential energy curves of the N_2 molecule calculated with exFCI and basis-set corrected exFCI using the aug-cc-pVDZ (top) and aug-cc-pVTZ (bottom) basis sets. The estimated exact energies are based on a fit of experimental data and obtained from Ref. 87.

where $\mathring{n}_{2,\text{CISPI}}(\mathbf{r}) = \mathring{n}_2(n_{2,\text{CIPSI}}(\mathbf{r}), \mu_{\text{CIPSI}}(\mathbf{r}))$ and $\mu_{\text{CIPSI}}(\mathbf{r})$ is the local range-separation function calculated with the CIPSI wave function. We also define the system-averaged range-separation parameters as

$$\langle \mu_{\text{CASSCF}} \rangle = \frac{1}{N} \int d\mathbf{r} \, n_{\text{CASSCF}}(\mathbf{r}) \, \mu_{\text{CASSCF}}(\mathbf{r}), \qquad (38a)$$

$$\langle \mu_{\text{CIPSI}} \rangle = \frac{1}{N} \int d\mathbf{r} \, n_{\text{CIPSI}}(\mathbf{r}) \, \mu_{\text{CIPSI}}(\mathbf{r}), \qquad (38b)$$

where $n_{\text{CASSCF}}(\mathbf{r})$ and $n_{\text{CIPSI}}(\mathbf{r})$ are the CASSCF and CIPSI densities, respectively. All the CIPSI quantities have been calculated with the largest variational wave function computed in the CIPSI calculation with a given basis, which contains here at least 10^7 Slater determinants. In particular, $\mu_{\text{CIPSI}}(\mathbf{r})$ has been calculated from Eqs. (16)–(19) with the opposite-spin two-body density matrix Γ_{pq}^{rs} of the largest variational CIPSI wave function for a given basis. All quantities in Eqs. (36a)–(38a) were computed excluding all contributions from the 1s orbitals, i.e., they are "valence-only" quantities.

We report in Table II these quantities for N_2 and N for various basis sets. One notes that the system-averaged on-top pair density

at the CIPSI level $\langle n_{2,\text{CIPSI}} \rangle$ is systematically lower than its CASSCF analog $\langle n_{2,\text{CASSCF}} \rangle$, which is expected since short-range correlation, i.e., digging the correlation hole in a given basis set at the near FCI level, is missing from the valence CASSCF wave function. In addition, $\langle n_{2,\text{CIPSI}} \rangle$ decreases in a monotonous way as the size of the basis set increases, leading to roughly a 20% decrease from the aug-cc-pVDZ to the aug-cc-pVQZ basis sets, whereas $\langle n_{2,\text{CASSCF}} \rangle$ is almost constant with respect to the basis set. Regarding the extrapolated on-top pair densities, $\langle \mathring{n}_{2,\text{CASSCF}} \rangle$ and $\langle \mathring{n}_{2,\text{CIPSI}} \rangle$, it is interesting to note that they are substantially lower than their non-extrapolated counterparts, $\langle n_{2,\text{CASSCF}} \rangle$ and $\langle n_{2,\text{CIPSI}} \rangle$. Nevertheless, the behaviors of $\langle \mathring{n}_{2,\text{CASSCF}} \rangle$ and $\langle \mathring{n}_{2,\text{CIPSI}} \rangle$ are qualitatively different: $\langle \mathring{n}_{2,\text{CASSCF}} \rangle$ clearly increases when enlarging the basis set, whereas $\langle \mathring{n}_{2,\text{CIPSI}} \rangle$ remains almost constant. More precisely, in the case of N_2, the value of $\langle \mathring{n}_{2,\text{CASSCF}} \rangle$ increases by about 30% from the aug-cc-pVDZ to the aug-cc-pVQZ basis sets, whereas the value of $\langle \mathring{n}_{2,\text{CIPSI}} \rangle$ only fluctuates within 5% for the same basis sets. The behavior of $\langle \mathring{n}_{2,\text{CASSCF}} \rangle$ can be understood by noting that (i) the value of $\mu_{\text{CASSCF}}(\mathbf{r})$ globally increases when enlarging the basis set (as evidenced by $\langle \mu_{\text{CASSCF}} \rangle$) and (ii) $\lim_{\mu \to \infty} \mathring{n}_2(n_2, \mu) = n_2$ [see Eq. (32)]. Therefore, in the CBS
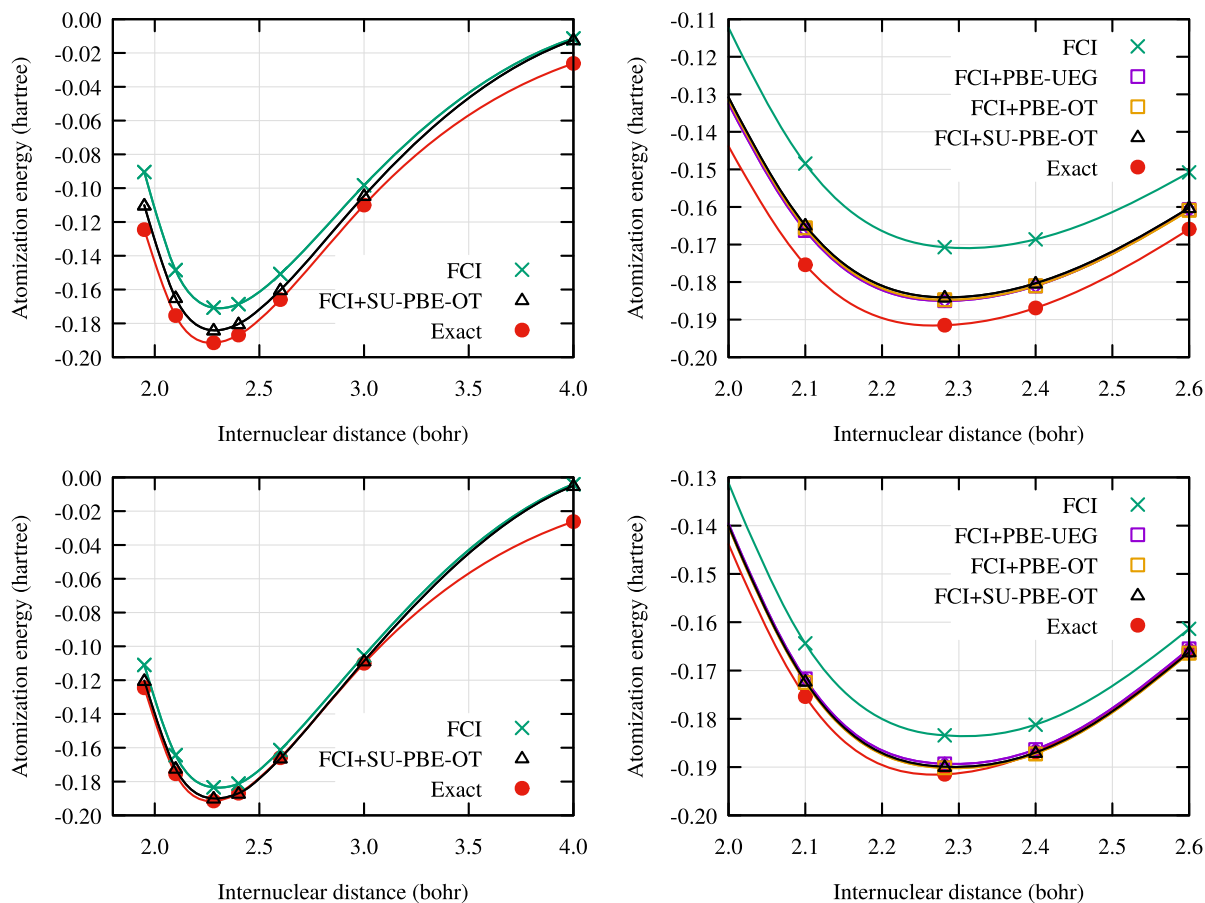
FIG. 3. Potential energy curves of the O_2 molecule calculated with exFCI and basis-set corrected exFCI using the aug-cc-pVDZ (top) and aug-cc-pVTZ (bottom) basis sets. The estimated exact energies are based on a fit of experimental data and obtained from Ref. 87.

limit, $\mu_{CASSCF}(\mathbf{r}) \to \infty$ and one obtains

$$\lim_{\mathcal{B} \to CBS} \langle \mathring{n}_{2,CASSCF} \rangle = \lim_{\mathcal{B} \to CBS} \langle n_{2,CASSCF} \rangle, \tag{39}$$

i.e., $\langle \mathring{n}_{2,CASSCF} \rangle$ must increase with the size of the basis set $\mathcal{B}$ to eventually converge to $\lim_{\mathcal{B} \to CBS} \langle n_{2,CASSCF} \rangle$, the latter limit being essentially reached with the present basis sets. On the other hand, the stability of $\langle \mathring{n}_{2,CIPSI} \rangle$ with respect to the basis set is quite remarkable and must come from the fact that (i) $\langle n_{2,CIPSI} \rangle$ is a good approximation to the corresponding FCI value within the considered basis sets and (ii) the extrapolation formula in Eq. (32) together with the choice of $\mu_{CIPSI}(\mathbf{r})$ is quantitatively correct. Therefore, we expect the calculated values of $\langle \mathring{n}_{2,CIPSI} \rangle$ to be nearly converged with respect to the basis set, and we will take the value of $\langle \mathring{n}_{2,CIPSI} \rangle$ in the aug-cc-pVQZ basis set as an estimate of the exact system-averaged on-top pair density.

For the present work, it is important to keep in mind that $\langle \mathring{n}_{2,CASSCF} \rangle$ directly determines the basis-set correction in the large-μ limit. More precisely, the correlation energy contribution associated with the basis-set correction is (in absolute value)

an increasing function of $\langle \mathring{n}_{2,CASSCF} \rangle$. Therefore, the error on $\langle \mathring{n}_{2,CASSCF} \rangle$ with respect to the estimated exact system-averaged on-top pair density provides an indication of the error made by the basis-set correction for a given system and basis set. With the aug-cc-pVQZ basis set, we have $\langle \mathring{n}_{2,CASSCF} \rangle - \langle \mathring{n}_{2,CIPSI} \rangle = 0.240$ for the N_2 molecule, while $2(\langle \mathring{n}_{2,CASSCF} \rangle - \langle \mathring{n}_{2,CIPSI} \rangle) = 0.190$ for two isolated N atoms. We can then conclude that the overestimation of the system-averaged on-top pair density, and therefore of the basis-set correction, is more important for the N_2 molecule at the equilibrium distance than for the isolated N atoms. This probably explains the observed overestimation of the atomization energy. To confirm this statement, we computed the basis-set correction for both the N_2 molecule at the equilibrium distance and the isolated atoms using $\mu_{CIPSI}(\mathbf{r})$ and $\mathring{n}_{2,CIPSI}(\mathbf{r})$ with the aug-cc-pVTZ and aug-cc-pVQZ basis sets. We obtained the following values for the atomization energies: 362.12 mH with aug-cc-pVTZ and 362.15 mH with aug-cc-pVQZ, which are indeed more accurate values than those obtained using $\mu_{CASSCF}(\mathbf{r})$ and $\mathring{n}_{2,CASSCF}(\mathbf{r})$.

Finally, regarding now the performance of the basis-set correction along the whole potential energy curves reported in Figs. 2–4,
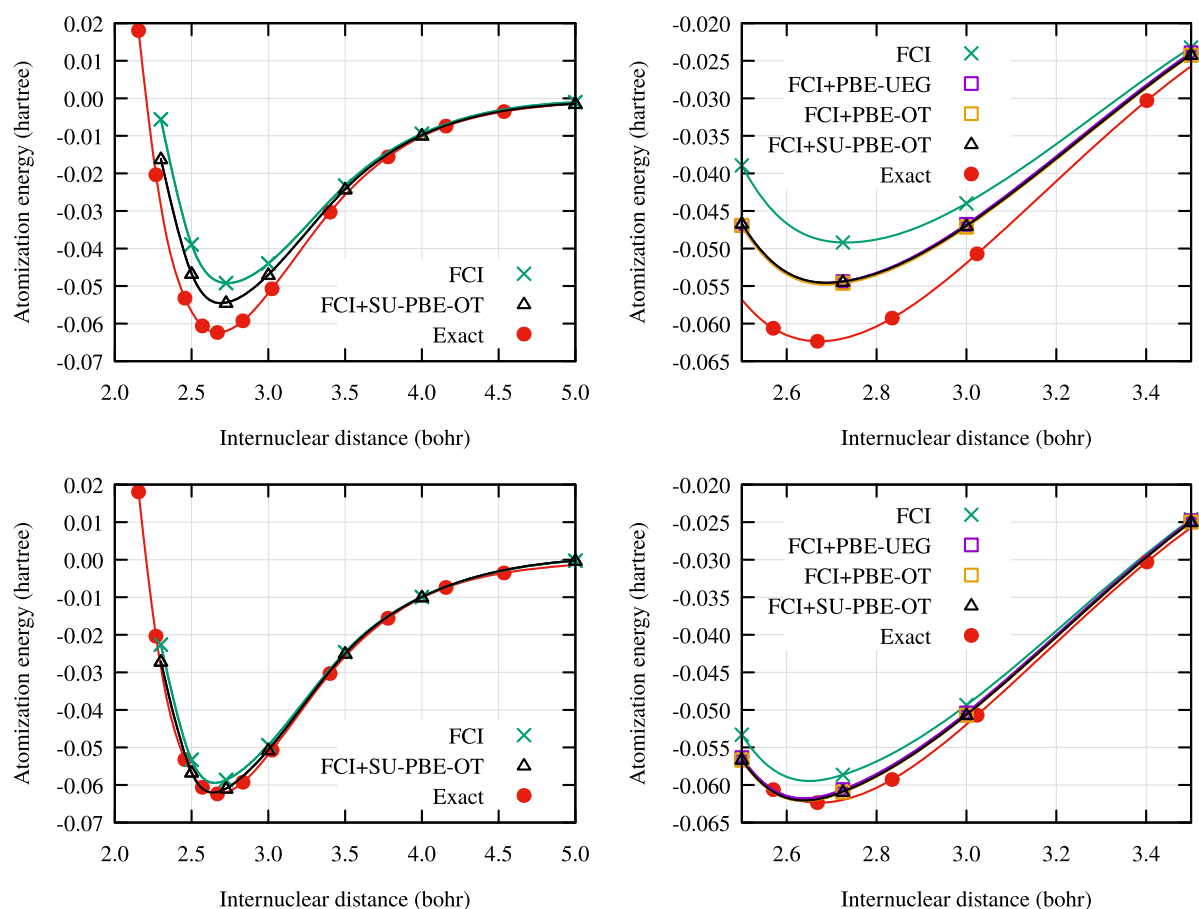
FIG. 4. Potential energy curves of the F_2 molecule calculated with exFCI and basis-set corrected exFCI using the aug-cc-pVDZ (top) and aug-cc-pVTZ (bottom) basis sets. The estimated exact energies are based on a fit of experimental data and obtained from Ref. 87.

it is interesting to note that it fails to provide a noticeable improvement far from the equilibrium geometry. Acknowledging that the weak-correlation effects in these regions are dominated by dispersion interactions that are long-range effects, the failure of the present approximations for the complementary functional can be

understood easily. Indeed, the whole scheme designed here is based on the physics of correlation near the electron–electron coalescence point: the local range-separation function $\mu(\mathbf{r})$ is based on the value of the effective electron–electron interaction at coalescence and the ECMD functionals are suited for short-range correlation effects.

TABLE II. System-averaged on-top pair density $\langle n_2 \rangle$, extrapolated on-top pair density $\langle \mathring{n}_2 \rangle$, and range-separation parameter $\langle \mu \rangle$ (all in atomic units) calculated with full-valence CASSCF and CIPSI wave functions (see the text for details) for N_2 and N in the aug-cc-pVXZ basis sets (X = D, T, and Q). All quantities were computed within the frozen-core approximation, i.e., excluding all contributions from the 1s orbitals.

System	Basis set	$\langle n_{2,\text{CASSCF}} \rangle$	$\langle \mathring{n}_{2,\text{CASSCF}} \rangle$	$\langle n_{2,\text{CIPSI}} \rangle$	$\langle \mathring{n}_{2,\text{CIPSI}} \rangle$	$\langle \mu_{\text{CASSCF}} \rangle$	$\langle \mu_{\text{CIPSI}} \rangle$
N_2	aug-cc-pVDZ	1.175 42	0.659 66	1.027 92	0.582 28	0.946	0.962
	aug-cc-pVTZ	1.183 24	0.770 12	0.922 76	0.610 74	1.328	1.364
	aug-cc-pVQZ	1.184 84	0.840 12	0.838 66	0.599 82	1.706	1.746
N	aug-cc-pVDZ	0.344 64	0.196 22	0.254 84	0.146 86	0.910	0.922
	aug-cc-pVTZ	0.346 04	0.226 30	0.223 44	0.148 28	1.263	1.299
	aug-cc-pVQZ	0.346 14	0.246 66	0.212 24	0.151 64	1.601	1.653

Therefore, the failure of the present basis-set correction to describe dispersion interactions is theoretically expected. We hope to report further on this in the near future.

IV. CONCLUSION

In the present paper, we have extended the recently proposed DFT-based basis-set correction to strongly correlated systems. We have applied the method to the H_{10}, N_2, O_2, and F_2 molecules up to the dissociation limit at the near-FCI level in increasingly large basis sets and investigated how the basis-set correction affects the convergence toward the CBS limit of the potential energy curves of these molecular systems.

The density-based basis-set correction relies on three aspects: (i) the definition of an effective non-divergent electron–electron interaction obtained from the expectation value over a wave function $\Psi^{\mathcal{B}}$ of the Coulomb electron–electron interaction projected into an incomplete basis set $\mathcal{B}$, (ii) the fit of this effective interaction with the long-range interaction used in RSDFT, and (iii) the use of a short-range, complementary functional borrowed from RSDFT. In the present paper, we investigated (i) and (iii) in the context of strong correlation and focused on potential energy curves and atomization energies. More precisely, we proposed a new scheme to design functionals fulfilling spin-multiplet degeneracy and size consistency. To fulfill such requirements, we proposed to use CASSCF wave functions leading to size-consistent energies, and we developed functionals using only S_z-independent density-like quantities.

The development of new S_z-independent and size-consistent functionals has led us to investigate the role of two related quantities: the spin polarization and the on-top pair density. One important result of the present study is that by using functionals *explicitly* depending on the on-top pair density, one can eschew its spin-polarization dependence without loss of accuracy. This avoids the commonly used effective spin polarization originally proposed in Ref. 64, which has the disadvantage of possibly becoming complex-valued in the multideterminant case. From a more fundamental aspect, this confirms that, in a DFT framework, the spin polarization mimics the role of the on-top pair density. Consequently, we believe that one could potentially develop new families of density-functional approximations, where the spin polarization is abandoned and replaced by the on-top pair density.

Regarding the results of the present approach, the basis-set correction systematically improves the near-FCI calculations in a given basis set. More quantitatively, it is shown that with only triple-ζ quality basis sets, chemically accurate atomization energies are obtained for all systems, whereas the uncorrected near-FCI results are far from this accuracy within the same basis set.

In addition, it is shown that the basis-set correction gives substantial differential contribution to potential energy curves close to the equilibrium geometries, but at long internuclear distances, it cannot recover the dispersion interaction energy missing because of the basis-set incompleteness. This behavior is actually expected as dispersion interactions are of long-range nature, and the present approach is designed to recover only short-range correlation effects.

APPENDIX A: SIZE CONSISTENCY OF THE BASIS-SET CORRECTION

1. Sufficient condition for size consistency

The basis-set correction is expressed as an integral in real space,

$$\bar{E}^{\mathcal{B}}[n, \zeta, n_2, \mu] = \int d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r})),$$

(A1)

where all the local quantities $n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r})$ are obtained from the same wave function Ψ. In the limit of two non-overlapping and non-interacting dissociated fragments A + B, this integral can be rewritten as the sum of the integral over the region Ω_A and the integral over the region Ω_B,

$$\bar{E}_{A+B}^{\mathcal{B}}[n, \zeta, n_2, \mu] = \int_{\Omega_A} d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r}))$$

$$+ \int_{\Omega_B} d\mathbf{r}\, n(\mathbf{r}) \bar{\varepsilon}_{c,md}^{sr,PBE}(n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r})).$$

(A2)

Therefore, a sufficient condition to obtain size consistency is that all the local quantities $n(\mathbf{r}), \zeta(\mathbf{r}), s(\mathbf{r}), n_2(\mathbf{r}), \mu(\mathbf{r})$ are *intensive*, i.e., they *locally* coincide in the supersystem A + B and in each isolated fragment X = A or B. Hence, we must have, for $\mathbf{r} \in \Omega_X$,

$$n_{A+B}(\mathbf{r}) = n_X(\mathbf{r}), \quad \text{(A3a)}$$

$$\zeta_{A+B}(\mathbf{r}) = \zeta_X(\mathbf{r}), \quad \text{(A3b)}$$

$$s_{A+B}(\mathbf{r}) = s_X(\mathbf{r}), \quad \text{(A3c)}$$

$$n_{2,A+B}(\mathbf{r}) = n_{2,X}(\mathbf{r}), \quad \text{(A3d)}$$

$$\mu_{A+B}(\mathbf{r}) = \mu_X(\mathbf{r}), \quad \text{(A3e)}$$

where the left-hand-side quantities are for the supersystem and the right-hand-side quantities are for an isolated fragment. Such conditions can be difficult to fulfill in the presence of degeneracies since system X can be in a different mixed state (i.e., ensemble) in the supersystem A + B and in the isolated fragment.[92] Here, we will consider the simple case, where the wave function of the supersystem is multiplicatively separable, i.e.,

$$|\Psi_{A+B}\rangle = |\Psi_A\rangle \otimes |\Psi_B\rangle, \quad \text{(A4)}$$

where $\otimes$ is the antisymmetric tensor product. In this case, it is easy to show that Eqs. (A3a)–(A3c) are valid, as well known, and it remains to show that Eqs. (A3d) and (A3e) are also valid. Before showing this, we note that even though we do not explicitly consider the case of degeneracies, the lack of size consistency that could arise from spin-multiplet degeneracies can be avoided by the same strategy used for imposing the energy independence on S_z, i.e., by using the effective spin polarization $\tilde{\zeta}(n(\mathbf{r}), n_2(\mathbf{r}))$ or a zero spin polarization $\zeta(\mathbf{r}) = 0$. Moreover, for the systems treated in this work, the lack of size consistency that could arise from spatial degeneracies (coming from atomic p states) can also be avoided by selecting the same state in the supersystem and in the isolated fragment. For example, for the F_2 molecule, the CASSCF wave function dissociates into the atomic

configuration $p_x^2 p_y^2 p_z^1$ for each fragment, and we thus choose the same configuration for the calculation of the isolated atom. The same argument applies to the N_2 and O_2 molecules. For other systems, it may not be always possible to do so.

2. Intensity of the on-top pair density and the local range-separation function

The on-top pair density can be written in an orthonormal spatial orbital basis set $\{\phi_p(\mathbf{r})\}$ as

$$n_2(\mathbf{r}) = \sum_{pqrs \in \mathcal{B}} \phi_p(\mathbf{r})\phi_q(\mathbf{r})\Gamma_{pq}^{rs}\phi_r(\mathbf{r})\phi_s(\mathbf{r}), \quad (A5)$$

with $\Gamma_{pq}^{rs} = 2\langle\Psi|\hat{a}_{r_\downarrow}^\dagger \hat{a}_{s_\uparrow}^\dagger \hat{a}_{q_\uparrow} \hat{a}_{p_\downarrow}|\Psi\rangle$. As the summations run over all orbitals in the basis set $\mathcal{B}$, $n_2(\mathbf{r})$ is invariant to orbital rotations and can thus be expressed in terms of localized orbitals. For two non-overlapping fragments A + B, the basis set can then be partitioned into orbitals localized on fragment A and orbitals localized on B, i.e., $\mathcal{B} = \mathcal{B}_A \cup \mathcal{B}_B$. Therefore, we see that the on-top pair density of the supersystem A + B is additively separable,

$$n_{2,A+B}(\mathbf{r}) = n_{2,A}(\mathbf{r}) + n_{2,B}(\mathbf{r}), \quad (A6)$$

where $n_{2,X}(\mathbf{r})$ is the on-top pair density of fragment X,

$$n_{2,X}(\mathbf{r}) = \sum_{pqrs \in \mathcal{B}_X} \phi_p(\mathbf{r})\phi_q(\mathbf{r})\Gamma_{pq}^{rs}\phi_r(\mathbf{r})\phi_s(\mathbf{r}), \quad (A7)$$

in which the elements Γ_{pq}^{rs} with orbital indices restricted to fragment X are $\Gamma_{pq}^{rs} = 2\langle\Psi_{A+B}|\hat{a}_{r_\downarrow}^\dagger \hat{a}_{s_\uparrow}^\dagger \hat{a}_{q_\uparrow} \hat{a}_{p_\downarrow}|\Psi_{A+B}\rangle = 2\langle\Psi_X|\hat{a}_{r_\downarrow}^\dagger \hat{a}_{s_\uparrow}^\dagger \hat{a}_{q_\uparrow} \hat{a}_{p_\downarrow}|\Psi_X\rangle$, owing to the multiplicative structure of the wave function [see Eq. (A4)]. This shows that the on-top pair density is a local intensive quantity.

The local range-separation function is defined as, for $n_2(\mathbf{r}) \neq 0$,

$$\mu(\mathbf{r}) = \frac{\sqrt{\pi}}{2}\frac{f(\mathbf{r},\mathbf{r})}{n_2(\mathbf{r})}, \quad (A8)$$

where

$$f(\mathbf{r},\mathbf{r}) = \sum_{pqrstu \in \mathcal{B}} \phi_p(\mathbf{r})\phi_q(\mathbf{r})V_{pq}^{rs}\Gamma_{rs}^{tu}\phi_t(\mathbf{r})\phi_u(\mathbf{r}). \quad (A9)$$

Again, $f(\mathbf{r},\mathbf{r})$ is invariant to orbital rotations and can be expressed in terms of orbitals localized on fragments A and B. In the limit of infinitely separated fragments, the Coulomb interaction vanishes between A and B, and therefore, any two-electron integral V_{pq}^{rs} involving orbitals on both A and B vanishes. We thus see that the quantity $f(\mathbf{r},\mathbf{r})$ of the supersystem A + B is additively separable,

$$f_{A+B}(\mathbf{r},\mathbf{r}) = f_A(\mathbf{r},\mathbf{r}) + f_B(\mathbf{r},\mathbf{r}), \quad (A10)$$

with

$$f_X(\mathbf{r},\mathbf{r}) = \sum_{pqrstu \in \mathcal{B}_X} \phi_p(\mathbf{r})\phi_q(\mathbf{r})V_{pq}^{rs}\Gamma_{rs}^{tu}\phi_t(\mathbf{r})\phi_u(\mathbf{r}). \quad (A11)$$

So, $f(\mathbf{r},\mathbf{r})$ is a local intensive quantity. As a consequence, the local range-separation function of the supersystem A + B is

$$\mu_{A+B}(\mathbf{r}) = \frac{\sqrt{\pi}}{2}\frac{f_A(\mathbf{r},\mathbf{r}) + f_B(\mathbf{r},\mathbf{r})}{n_{2,A}(\mathbf{r}) + n_{2,B}(\mathbf{r})}, \quad (A12)$$

which implies

$$\mu_{A+B}(\mathbf{r}) = \mu_X(\mathbf{r}) \text{ if } \mathbf{r} \in \Omega_X, \quad (A13)$$

where $\mu_X(\mathbf{r}) = (\sqrt{\pi}/2)f_X(\mathbf{r},\mathbf{r})/n_{2,X}(\mathbf{r})$. The local range-separation function is thus a local intensive quantity.

We can therefore conclude that if the wave function of the supersystem A + B is multiplicative separable, all local quantities used in the basis-set correction functional are intensive, and therefore, the basis-set correction is size consistent.

APPENDIX B: COMPUTATIONAL COST OF THE BASIS-SET CORRECTION FOR A CASSCF WAVE FUNCTION

The computational cost of the basis-set correction is determined by the calculation of the on-top pair density $n_2(\mathbf{r})$ and the local range-separation function $\mu(\mathbf{r})$ on the real-space grid. For a general multideterminant wave function, the computational cost is of order $O(N_{grid}N_\mathcal{B}^4)$, where N_{grid} is the number of grid points and $N_\mathcal{B}$ is the number of basis functions.[51] For a CASSCF wave function, a significant reduction in the scaling of the computational cost can be achieved.

1. Computation of the on-top pair density

For a CASSCF wave function Ψ, the occupied orbitals can be partitioned into a set of active orbitals $\mathcal{A}$ and a set of inactive (doubly occupied) orbitals $\mathcal{I}$. The CASSCF on-top pair density can then be written as

$$n_2(\mathbf{r}) = n_{2,\mathcal{A}}(\mathbf{r}) + n_\mathcal{A}(\mathbf{r})n_\mathcal{I}(\mathbf{r}) + \frac{n_\mathcal{I}(\mathbf{r})^2}{2}, \quad (B1)$$

where

$$n_{2,\mathcal{A}}(\mathbf{r}) = \sum_{pqrs \in \mathcal{A}} \phi_p(\mathbf{r})\phi_q(\mathbf{r})\Gamma_{pq}^{rs}\phi_r(\mathbf{r})\phi_s(\mathbf{r}), \quad (B2a)$$

$$n_\mathcal{A}(\mathbf{r}) = \sum_{pq \in \mathcal{A}} \phi_p(\mathbf{r})\phi_q(\mathbf{r})\langle\Psi|\hat{a}_{p_\uparrow}^\dagger \hat{a}_{q_\uparrow} + \hat{a}_{p_\downarrow}^\dagger \hat{a}_{q_\downarrow}|\Psi\rangle, \quad (B2b)$$

$$n_\mathcal{I}(\mathbf{r}) = 2\sum_{p \in \mathcal{I}} \phi_p(\mathbf{r})^2 \quad (B2c)$$

are the purely active part of the on-top pair density, the active part of the density, and the inactive part of the density, respectively. The leading computational cost is the evaluation of $n_{2,\mathcal{A}}(\mathbf{r})$ on the grid, which, according to Eq. (B2a), scales as $O(N_{grid}N_\mathcal{A}^4)$, where $N_\mathcal{A}$ is the number of active orbitals that is much smaller than the number of basis functions $N_\mathcal{B}$.

2. Computation of the local range-separation function

In addition to the on-top pair density, the computation of $\mu(\mathbf{r})$ needs the computation of $f(\mathbf{r},\mathbf{r})$ [see Eq. (A9)] at each grid point. It can be factorized as

$$f(\mathbf{r},\mathbf{r}) = \sum_{rs \in \mathcal{B}} V^{rs}(\mathbf{r})\Gamma_{rs}(\mathbf{r}), \quad (B3)$$

where

$$V^{rs}(\mathbf{r}) = \sum_{pq \in \mathcal{B}} V_{pq}^{rs}\phi_p(\mathbf{r})\phi_q(\mathbf{r}), \quad (B4a)$$

$$\Gamma_{rs}(\mathbf{r}) = \sum_{pq \in \mathcal{B}} \Gamma_{rs}^{pq}\phi_p(\mathbf{r})\phi_q(\mathbf{r}). \quad (B4b)$$

For a general multideterminant wave function, the computational cost of $f(\mathbf{r}, \mathbf{r})$ thus scales as $O(N_{\text{grid}} N_{\mathcal{B}}^4)$.

In the case of a CASSCF wave function, Γ_{rs}^{pq} vanishes if one index p, q, r, s does not belong to the set of inactive or active occupied orbitals $\mathcal{I} \cup \mathcal{A}$. Therefore, at a given grid point, the number of non-zero elements $\Gamma_{rs}(\mathbf{r})$ is only at most $(N_{\mathcal{I}} + N_{\mathcal{A}})^2$, which is usually much smaller than $N_{\mathcal{B}}^2$. As a consequence, one can also restrict the sum in the calculation of

$$f(\mathbf{r}, \mathbf{r}) = \sum_{rs \in \mathcal{I} \cup \mathcal{A}} V^{rs}(\mathbf{r}) \Gamma_{rs}(\mathbf{r}). \quad (B5)$$

The overall computational cost is dominated by that of $V^{rs}(\mathbf{r})$, which scales as $O(N_{\text{grid}}(N_{\mathcal{I}} + N_{\mathcal{A}})^2 N_{\mathcal{B}}^2)$, which is much smaller than $O(N_{\text{grid}} N_{\mathcal{B}}^4)$.

REFERENCES

[1] J. A. Pople, Rev. Mod. Phys. **71**, 1267 (1999).

[2] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999).

[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[4] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2012).

[5] J. Ángyán, J. Dobson, G. Jansen, and T. Gould, *London Dispersion Forces in Molecules, Solids and Nano-Structures: An Introduction to Physical Models and Computational Methods*, Theoretical and Computational Chemistry Series (Royal Society of Chemistry, 2020).

[6] P. Gori-Giorgi, M. Seidl, and A. Savin, Phys. Chem. Chem. Phys. **10**, 3440 (2008).

[7] L. Gagliardi, D. G. Truhlar, G. Li Manni, R. K. Carlson, C. E. Hoyer, and J. L. Bao, Acc. Chem. Res. **50**, 66 (2017).

[8] E. A. Hylleraas, Z. Phys. **54**, 347 (1929).

[9] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).

[10] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).

[11] J. Noga and W. Kutzelnigg, J. Chem. Phys. **101**, 7738 (1994).

[12] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[13] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[14] A. Irmler, F. Hummel, and A. Grüneis, Phys. Rev. Lett. **123**, 156401 (2019).

[15] A. Irmler and A. Grüneis, J. Chem. Phys. **151**, 104107 (2019).

[16] S. Ten-no, Theor. Chem. Acc. **131**, 1070 (2012).

[17] S. Ten-no and J. Noga, WIREs Comput. Mol. Sci. **2**, 114 (2012).

[18] L. Kong, F. A. Bischoff, and E. F. Valeev, Chem. Rev. **112**, 75 (2012).

[19] A. Grüneis, S. Hirata, Y.-Y. Ohnishi, and S. Ten-no, J. Chem. Phys. **146**, 080901 (2017).

[20] Q. Ma and H.-J. Werner, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1371 (2018).

[21] D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).

[22] G. M. J. Barca and P.-F. Loos, J. Chem. Phys. **147**, 024103 (2017).

[23] S. Ten-no, Chem. Phys. Lett. **447**, 175 (2007).

[24] T. Shiozaki and H.-J. Werner, J. Chem. Phys. **133**, 141103 (2010).

[25] T. Shiozaki, G. Knizia, and H.-J. Werner, J. Chem. Phys. **134**, 034113 (2011).

[26] O. Demel, S. Kedžuch, M. Švaňa, S. Ten-no, J. Pittner, and J. Noga, Phys. Chem. Chem. Phys. **14**, 4753 (2012).

[27] Y. Guo, K. Sivalingam, E. F. Valeev, and F. Neese, J. Chem. Phys. **147**, 064110 (2017).

[28] M. Torheyden and E. F. Valeev, J. Chem. Phys. **131**, 171103 (2009).

[29] L. Kong and E. F. Valeev, J. Chem. Phys. **135**, 214105 (2011).

[30] R. Haunschild, S. Mao, D. Mukherjee, and W. Klopper, Chem. Phys. Lett. **531**, 247 (2012).

[31] G. H. Booth, D. Cleland, A. Alavi, and D. P. Tew, J. Chem. Phys. **137**, 164112 (2012).

[32] J. Toulouse, F. Colonna, and A. Savin, Phys. Rev. A **70**, 062505 (2004).

[33] P. Gori-Giorgi and A. Savin, Phys. Rev. A **73**, 032506 (2006).

[34] O. Franck, B. Mussard, E. Luppi, and J. Toulouse, J. Chem. Phys. **142**, 074107 (2015).

[35] J. G. Ángyán, I. C. Gerber, A. Savin, and J. Toulouse, Phys. Rev. A **72**, 012510 (2005).

[36] E. Goll, H.-J. Werner, and H. Stoll, Phys. Chem. Chem. Phys. **7**, 3917 (2005).

[37] J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán, Phys. Rev. Lett. **102**, 096404 (2009).

[38] B. G. Janesko, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **130**, 081105 (2009).

[39] J. Toulouse, W. Zhu, A. Savin, G. Jansen, and J. G. Ángyán, J. Chem. Phys. **135**, 084119 (2011).

[40] B. Mussard, P. Reinhardt, J. G. Ángyán, and J. Toulouse, J. Chem. Phys. **142**, 154123 (2015); Erratum, **142**, 219901 (2015).

[41] C. Kalai and J. Toulouse, J. Chem. Phys. **148**, 164105 (2018).

[42] C. Kalai, B. Mussard, and J. Toulouse, J. Chem. Phys. **151**, 074102 (2019).

[43] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, Chem. Phys. Lett. **275**, 151 (1997).

[44] E. Fromager, J. Toulouse, and H. J. A. Jensen, J. Chem. Phys. **126**, 074111 (2007).

[45] E. Fromager, R. Cimiraglia, and H. J. A. Jensen, Phys. Rev. A **81**, 024502 (2010).

[46] E. D. Hedegård, S. Knecht, J. S. Kielberg, H. J. A. Jensen, and M. Reiher, J. Chem. Phys. **142**, 224108 (2015).

[47] E. D. Hedegård, J. Toulouse, and H. J. A. Jensen, J. Chem. Phys. **148**, 214103 (2018).

[48] A. Ferté, E. Giner, and J. Toulouse, J. Chem. Phys. **150**, 084103 (2019).

[49] B. Mussard and J. Toulouse, Mol. Phys. **115**, 161 (2017).

[50] E. Giner, B. Pradines, A. Ferté, R. Assaraf, A. Savin, and J. Toulouse, J. Chem. Phys. **149**, 194301 (2018).

[51] P.-F. Loos, B. Pradines, A. Scemama, J. Toulouse, and E. Giner, J. Phys. Chem. Lett. **10**, 2931 (2019).

[52] E. Giner, A. Scemama, J. Toulouse, and P.-F. Loos, J. Chem. Phys. **151**, 144118 (2019).

[53] P.-F. Loos, B. Pradines, A. Scemama, E. Giner, and J. Toulouse, J. Chem. Theory Comput. **16**, 1018 (2020).

[54] M. Levy, Proc. Natl. Acad. Sci. U. S. A. **76**, 6062 (1979).

[55] E. H. Lieb, Int. J. Quantum Chem. **24**, 24 (1983).

[56] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).

[57] J. Toulouse, P. Gori-Giorgi, and A. Savin, Theor. Chem. Acc. **114**, 305 (2005).

[58] J. Toulouse, Ph.D. thesis, Université Pierre et Marie Curie (Paris 6), 2005, tel.archives-ouvertes.fr/tel-00550772.

[59] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[60] S. Paziani, S. Moroni, P. Gori-Giorgi, and G. B. Bachelet, Phys. Rev. B **73**, 155111 (2006).

[61] R. K. Carlson, D. G. Truhlar, and L. Gagliardi, J. Phys. Chem. A **121**, 5540 (2017).

[62] O. V. Gritsenko, R. van Meer, and K. Pernal, Phys. Rev. A **98**, 062510 (2018).

[63] F. Moscardó and E. San-Fabián, Phys. Rev. A **44**, 1549 (1991).

[64] A. D. Becke, A. Savin, and H. Stoll, Theoret. Chim. Acta **91**, 147 (1995).

[65] A. Savin, in *Recent Advances in Density Functional Theory*, edited by D. P. Chong (World Scientific, 1995), pp. 129–153.

[66] A. Savin, in *Recent Developments of Modern Density Functional Theory*, edited by J. M. Seminario (Elsevier, Amsterdam, 1996), pp. 327–357.

[67] B. B. Miehlich, H. Stoll, and A. Savin, Mol. Phys. **91**, 527 (1997).

[68] R. Takeda, S. Yamanaka, and K. Yamaguchi, Chem. Phys. Lett. **366**, 321 (2002).

[69] R. Takeda, S. Yamanaka, and K. Yamaguchi, Int. J. Quantum Chem. **96**, 463 (2004).

[70] J. Gräfenstein and D. Cremer, Mol. Phys. **103**, 279 (2005).

[71] T. Tsuchimochi, G. E. Scuseria, and A. Savin, J. Chem. Phys. **132**, 024111 (2010).

[72] G. Li Manni, R. K. Carlson, S. Luo, D. Ma, J. Olsen, D. G. Truhlar, and L. Gagliardi, J. Chem. Theory Comput. **10**, 3669 (2014).

[73] A. J. Garza, I. W. Bulik, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **142**, 044109 (2015).

[74]A. J. Garza, I. W. Bulik, T. M. Henderson, and G. E. Scuseria, Phys. Chem. Chem. Phys. **17**, 22412 (2015).

[75]R. K. Carlson, D. G. Truhlar, and L. Gagliardi, J. Chem. Theory Comput. **11**, 4077 (2015).

[76]J. P. Perdew, A. Savin, and K. Burke, Phys. Rev. A **51**, 4531 (1995).

[77]V. N. Staroverov and E. R. Davidson, Chem. Phys. Lett. **340**, 142 (2001).

[78]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[79]A. A. Holmes, C. J. Umrigar, and S. Sharma, J. Chem. Phys. **147**, 164111 (2017).

[80]A. Scemama, Y. Garniron, M. Caffarel, and P.-F. Loos, J. Chem. Theory Comput. **14**, 1395 (2018).

[81]P.-F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, J. Chem. Theory Comput. **14**, 4360 (2018).

[82]A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P.-F. Loos, J. Chem. Phys. **149**, 034108 (2018).

[83]P. F. Loos, M. Boggio-Pasqua, A. Scemama, M. Caffarel, and D. Jacquemin, J. Chem. Theory Comput. **15**, 1939 (2019).

[84]Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Véril, M. Caffarel, P.-F. Loos, E. Giner, and A. Scemama, J. Chem. Theory Comput. **15**, 3591 (2019).

[85]L. Bytautas, T. Nagata, M. S. Gordon, and K. Ruedenberg, J. Chem. Phys. **127**, 164317 (2007).

[86]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[87]G. C. Lie and E. Clementi, J. Chem. Phys. **60**, 1275 (1974).

[88]M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, S. Zhang, and Simons Collaboration on the Many-Electron Problem, Phys. Rev. X **7**, 031059 (2017).

[89]M. S. Gordon and M. W. Schmidt, "Advances in electronic structure theory: GAMESS a decade later," in *Theory and Applications of Computational Chemistry: The First Forty Years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005), pp. 1167–1189.

[90]L. Bytautas and K. Ruedenberg, J. Chem. Phys. **122**, 154110 (2005).

[91]P. C. Hiberty, S. Humbel, C. P. Byrman, and J. H. van Lenthe, J. Chem. Phys. **101**, 5969 (1994).

[92]A. Savin, Chem. Phys. **356**, 91 (2009).