A Density-Based Basis-Set Correction For Wave Function Theory
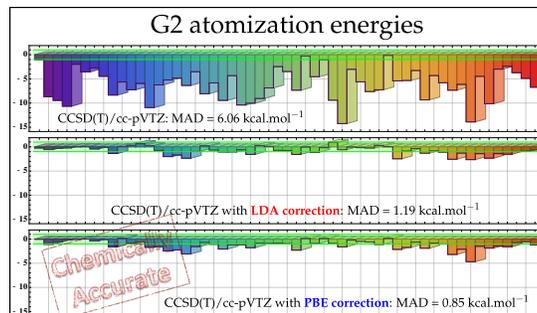
Pierre-François Loos,[1, a)] Bathélemy Pradines,[2, 3] Anthony Scemama,[1] Julien Toulouse,[2, b)] and Emmanuel Giner[2, c)]

[1)] *Laboratoire de Chimie et Physique Quantiques (UMR 5626), Université de Toulouse, CNRS, UPS, France*

[2)] *Laboratoire de Chimie Théorique, Sorbonne Université, CNRS, Paris, France*

[3)] *Institut des Sciences du Calcul et des Données, Sorbonne Université, Paris, France*

We report a universal density-based basis-set incompleteness correction that can be applied to any wave function method. The present correction, which appropriately vanishes in the complete basis set (CBS) limit, relies on short-range correlation density functionals (with multi-determinant reference) from range-separated density-functional theory (RS-DFT) to estimate the basis-set incompleteness error. Contrary to conventional RS-DFT schemes which require an *ad hoc* range-separation *parameter* μ, the key ingredient here is a range-separation *function* $\mu(\mathbf{r})$ that automatically adapts to the spatial non-homogeneity of the basis-set incompleteness error. As illustrative examples, we show how this density-based correction allows us to obtain CCSD(T) atomization and correlation energies near the CBS limit for the G2 set of molecules with compact Gaussian basis sets.

G2 atomization energies

CCSD(T)/cc-pVTZ: MAD = 6.06 kcal.mol^{-1}

CCSD(T)/cc-pVTZ with **LDA correction**: MAD = 1.19 kcal.mol^{-1}

CCSD(T)/cc-pVTZ with **PBE correction**: MAD = 0.85 kcal.mol^{-1}

Contemporary quantum chemistry has developed in two directions — wave function theory (WFT)[1] and density-functional theory (DFT).[2] Although both spring from the same Schrödinger equation, each of these philosophies has its own *pros* and *cons*.

WFT is attractive as it exists a well-defined path for systematic improvement as well as powerful tools, such as perturbation theory, to guide the development of new WFT *ansätze*. The coupled cluster (CC) family of methods is a typical example of the WFT philosophy and is well regarded as the gold standard of quantum chemistry for weakly correlated systems. By increasing the excitation degree of the CC expansion, one can systematically converge, for a given basis set, to the exact, full configuration interaction (FCI) limit, although the computational cost associated with such improvement is usually high. One of the most fundamental drawbacks of conventional WFT methods is the slow convergence of energies and properties with respect to the size of the one-electron basis set. This undesirable feature was put into light by Kutzelnigg more than thirty years ago.[3] To palliate this, following Hylleraas' footsteps,[4] Kutzelnigg proposed to introduce explicitly the interelectronic distance $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ to properly describe the electronic wave function around the coalescence of two electrons.[3,5,6] The resulting F12 methods yield a prominent improvement of the energy convergence, and achieve chemical accuracy for small organic molecules with relatively small Gaussian basis sets.[7–10] For example, at the CCSD(T) level, one can obtain quintuple-ζ quality correlation energies with a triple-ζ basis,[11] although computational overheads are introduced by the large auxiliary basis used to resolve three- and four-electron integrals.[12] To reduce further the computational cost and/or ease the transferability of the F12 correction, approximated and/or universal schemes have recently emerged.[13–18]

Present-day DFT calculations are almost exclusively done within the so-called Kohn-Sham (KS) formalism, which corresponds to an exact dressed one-electron theory.[19] The attractiveness of DFT originates from its very favorable accuracy/cost ratio as it often provides reasonably accurate energies and properties at a relatively low computational cost. Thanks to this, KS-DFT[19,20] has become the workhorse of electronic structure calculations for atoms, molecules and solids.[21] Although there is no clear way on how to systematically improve density-functional approximations,[22] climbing Perdew's ladder of DFT is potentially the most satisfactory way forward.[23,24] In the context of the present work, one of the interesting feature of density-based methods is their much faster convergence with respect to the size of the basis set.[25]

Progress toward unifying WFT and DFT are on-going. In particular, range-separated DFT (RS-DFT) (see Ref. [26] and references therein) rigorously combines these two approaches via a decomposition of the electron-electron (e-e) interaction into a non-divergent long-range part and a (complementary) short-range part treated with WFT and DFT, respectively. As the WFT method is relieved from describing the short-range part of the correlation hole around the e-e coalescence points, the convergence with respect to the one-electron basis set is greatly improved.[25] Therefore, a number of approximate RS-DFT schemes have been developed within single-reference[27–32] or multi-reference[33–38] WFT approaches. Very recently, a major step forward has been taken by some of the present authors thanks to the development of a density-based basis-set correction for WFT methods.[39] The present work proposes an extension of this new methodological development alongside the first numerical tests on molecular systems.

The present basis-set correction relies on the RS-DFT formalism to capture the missing part of the short-range cor-

relation effects, a consequence of the incompleteness of the one-electron basis set. Here, we only provide the main working equations. We refer the interested reader to Ref. 39 for a more formal derivation.

Let us assume we have both the energy $E_Y^{\mathcal{B}}$ and density $n_Z^{\mathcal{B}}$ of a N-electron system described by two methods Y and Z (potentially identical) in an incomplete basis set $\mathcal{B}$. According to Eq. (15) of Ref. 39, assuming that $E_Y^{\mathcal{B}}$ and $n_Z^{\mathcal{B}}$ are reasonable approximations of the FCI energy and density within $\mathcal{B}$, the exact ground state energy E may be approximated as

$$E \approx E_Y^{\mathcal{B}} + \bar{E}^{\mathcal{B}}[n_Z^{\mathcal{B}}], \tag{1}$$

where

$$\bar{E}^{\mathcal{B}}[n] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi \rangle - \min_{\Psi^{\mathcal{B}} \to n} \langle \Psi^{\mathcal{B}} | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi^{\mathcal{B}} \rangle \tag{2}$$

is the basis-dependent complementary density functional, $\hat{T}$ is the kinetic operator and $\hat{W}_{\mathrm{ee}} = \sum_{i<j} r_{ij}^{-1}$ is the interelectronic repulsion operator. In Eq. (2), $\Psi^{\mathcal{B}}$ and Ψ are two general N-electron normalized wave functions belonging to the Hilbert space spanned by $\mathcal{B}$ and the complete basis set (CBS), respectively. Both wave functions yield the same target density n (assumed to be representable in $\mathcal{B}$). Importantly, in the CBS limit (which we refer to as $\mathcal{B} \to \infty$), we have, for any density n, $\lim_{\mathcal{B}\to\infty} \bar{E}^{\mathcal{B}}[n] = 0$. This implies that

$$\lim_{\mathcal{B}\to\infty} \left(E_Y^{\mathcal{B}} + \bar{E}^{\mathcal{B}}[n_Z^{\mathcal{B}}] \right) = E_Y \approx E, \tag{3}$$

where E_Y is the energy associated with the method Y in the CBS limit. In the case where $Y = $ FCI in Eq. (3), we have a strict equality as $E_{\mathrm{FCI}} = E$. Provided that the functional $\bar{E}^{\mathcal{B}}[n]$ is known exactly, the only sources of error at this stage lie in the potential approximate nature of the methods Y and Z, and the lack of self-consistency in the present scheme.

The functional $\bar{E}^{\mathcal{B}}[n]$ is obviously *not* universal as it depends on $\mathcal{B}$. Moreover, as $\bar{E}^{\mathcal{B}}[n]$ aims at fixing the incompleteness of $\mathcal{B}$, its main role is to correct for the lack of cusp (i.e. discontinuous derivative) in $\Psi^{\mathcal{B}}$ at the e-e coalescence points, a universal condition of exact wave functions. Because the e-e cusp originates from the divergence of the Coulomb operator at $r_{12} = 0$, a cuspless wave function could equivalently originate from a Hamiltonian with a non-divergent two-electron interaction at coalescence. Therefore, as we shall do later on, it feels natural to approximate $\bar{E}^{\mathcal{B}}[n]$ by a short-range density functional which is complementary to a non-divergent long-range interaction. Contrary to the conventional RS-DFT scheme which requires a range-separation *parameter* μ, here we use a range-separation *function* $\mu^{\mathcal{B}}(\mathbf{r})$ that automatically adapts to quantify the incompleteness of $\mathcal{B}$ in $\mathbb{R}^3$.

The first step of the present basis-set correction consists in obtaining an effective two-electron interaction $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ "mimicking" the Coulomb operator in an incomplete basis $\mathcal{B}$. In a second step, we shall link $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ to $\mu^{\mathcal{B}}(\mathbf{r})$. As a final step, we employ short-range density functionals[40] with $\mu^{\mathcal{B}}(\mathbf{r})$ as range-separation function.

We define the effective operator as[39]

$$W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} f^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)/n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2), & \text{if } n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) \neq 0, \\ \infty, & \text{otherwise}, \end{cases} \tag{4}$$

where

$$n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pqrs\in\mathcal{B}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2)\Gamma_{pq}^{rs}\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2), \tag{5}$$

and $\Gamma_{pq}^{rs} = 2\langle \Psi^{\mathcal{B}} | \hat{a}_{r_\downarrow}^\dagger \hat{a}_{s_\uparrow}^\dagger \hat{a}_{p_\uparrow} \hat{a}_{q_\downarrow} | \Psi^{\mathcal{B}} \rangle$ are the opposite-spin pair density associated with $\Psi^{\mathcal{B}}$ and its corresponding tensor, respectively, $\phi_p(\mathbf{r})$ is a (real-valued) molecular orbital (MO),

$$f^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{pqrstu\in\mathcal{B}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2)V_{pq}^{rs}\Gamma_{rs}^{tu}\phi_t(\mathbf{r}_1)\phi_u(\mathbf{r}_2), \tag{6}$$

and $V_{pq}^{rs} = \langle pq|rs \rangle$ are the usual two-electron Coulomb integrals. With such a definition, $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ satisfies (see Appendix A of Ref. 39)

$$\iint \frac{n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 = \iint W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2, \tag{7}$$

which intuitively motivates $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ as a potential candidate for an effective interaction. Note that the divergence condition of $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ in Eq. (4) ensures that one-electron systems are free of correction as the present approach must only correct the basis-set incompleteness error originating from the e-e cusp. As already discussed in Ref. 39, $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ is symmetric, *a priori* non translational, nor rotational invariant if $\mathcal{B}$ does not have such symmetries. Thanks to its definition one can show that (see Appendix B of Ref. 39)

$$\lim_{\mathcal{B}\to\infty} W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) = r_{12}^{-1}, \tag{8}$$

for any $(\mathbf{r}_1, \mathbf{r}_2)$ such that $n_2^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2) \neq 0$.

A key quantity is the value of the effective interaction at coalescence of opposite-spin electrons, $W^{\mathcal{B}}(\mathbf{r}, \mathbf{r})$, which is necessarily *finite* for an incomplete basis set as long as the on-top pair density $n_2^{\mathcal{B}}(\mathbf{r}, \mathbf{r})$ is non vanishing. Because $W^{\mathcal{B}}(\mathbf{r}_1, \mathbf{r}_2)$ is a non-divergent two-electron interaction, it can be naturally linked to RS-DFT which employs a non-divergent long-range interaction operator. Although this choice is not unique, we choose here the range-separation function

$$\mu^{\mathcal{B}}(\mathbf{r}) = \frac{\sqrt{\pi}}{2} W^{\mathcal{B}}(\mathbf{r}, \mathbf{r}), \tag{9}$$

such that the long-range interaction of RS-DFT, $w^{\mathrm{lr},\mu}(r_{12}) = \mathrm{erf}(\mu r_{12})/r_{12}$, coincides with the effective interaction at coalescence, i.e. $w^{\mathrm{lr},\mu^{\mathcal{B}}(\mathbf{r})}(0) = W^{\mathcal{B}}(\mathbf{r}, \mathbf{r})$ at any $\mathbf{r}$.

Once $\mu^{\mathcal{B}}(\mathbf{r})$ is defined, it can be used within RS-DFT functionals to approximate $\bar{E}^{\mathcal{B}}[n]$. As in Ref. 39, we consider here a specific class of short-range correlation functionals known as correlation energy with multi-determinantal reference (ECMD) whose general definition reads[40]

$$\bar{E}_{\mathrm{c,md}}^{\mathrm{sr}}[n,\mu] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi \rangle - \langle \Psi^\mu | \hat{T} + \hat{W}_{\mathrm{ee}} | \Psi^\mu \rangle, \tag{10}$$

where Ψ^μ is defined by the constrained minimization

$$\Psi^\mu = \arg \min_{\Psi \to n} \, \langle \Psi | \hat{T} + \hat{W}_{\text{ee}}^{\text{lr},\mu} | \Psi \rangle, \qquad (11)$$

with $\hat{W}_{\text{ee}}^{\text{lr},\mu} = \sum_{i<j} w^{\text{lr},\mu}(r_{ij})$. The ECMD functionals admit, for any n, the following two limiting forms

$$\lim_{\mu \to \infty} \bar{E}_{\text{c,md}}^{\text{sr}}[n,\mu] = 0, \qquad \lim_{\mu \to 0} \bar{E}_{\text{c,md}}^{\text{sr}}[n,\mu] = E_{\text{c}}[n], \quad (12)$$

where $E_{\text{c}}[n]$ is the usual universal correlation density functional defined in KS-DFT. The choice of ECMD in the present scheme is motivated by the analogy between the definition of $\bar{E}^{\mathcal{B}}[n]$ [Eq. (2)] and the ECMD functional [Eq. (10)]. Indeed, the two functionals coincide if $\Psi^{\mathcal{B}} = \Psi^\mu$. Therefore, we approximate $\bar{E}^{\mathcal{B}}[n]$ by ECMD functionals evaluated with the range-separation function $\mu^{\mathcal{B}}(\mathbf{r})$.

The local-density approximation (LDA) of the ECMD complementary functional is defined as

$$\bar{E}_{\text{LDA}}^{\mathcal{B}}[n,\mu^{\mathcal{B}}] = \int n(\mathbf{r}) \bar{\varepsilon}_{\text{c,md}}^{\text{sr,LDA}}\Big(n(\mathbf{r}),\zeta(\mathbf{r}),\mu^{\mathcal{B}}(\mathbf{r})\Big) d\mathbf{r}, \quad (13)$$

where $\zeta = (n_\uparrow - n_\downarrow)/n$ is the spin polarization and $\bar{\varepsilon}_{\text{c,md}}^{\text{sr,LDA}}(n,\zeta,\mu)$ is the ECMD short-range correlation energy per electron of the uniform electron gas (UEG)[41] parameterized in Ref. [42]. The short-range LDA correlation functional relies on the transferability of the physics of the UEG which is certainly valid for large μ but is known to over correlate for small μ. In order to correct such a defect, inspired by the recent functional proposed by some of the authors[38], we propose here a new Perdew-Burke-Ernzerhof (PBE)-based ECMD functional

$$\bar{E}_{\text{PBE}}^{\mathcal{B}}[n,\mu^{\mathcal{B}}] = \int n(\mathbf{r}) \bar{\varepsilon}_{\text{c,md}}^{\text{sr,PBE}}\Big(n(\mathbf{r}),s(\mathbf{r}),\zeta(\mathbf{r}),\mu^{\mathcal{B}}(\mathbf{r})\Big) d\mathbf{r}, \tag{14}$$

where $s = |\nabla n|/n^{4/3}$ is the reduced density gradient. $\bar{\varepsilon}_{\text{c,md}}^{\text{sr,PBE}}(n,s,\zeta,\mu)$ interpolates between the usual PBE correlation functional,[43] $\varepsilon_{\text{c}}^{\text{PBE}}(n,s,\zeta)$, at $\mu = 0$ and the exact large-μ behavior,[26,42,44] yielding

$$\bar{\varepsilon}_{\text{c,md}}^{\text{sr,PBE}}(n,s,\zeta,\mu) = \frac{\varepsilon_{\text{c}}^{\text{PBE}}(n,s,\zeta)}{1 + \beta(n,s,\zeta)\mu^3}, \qquad (15a)$$

$$\beta(n,s,\zeta) = \frac{3}{2\sqrt{\pi}(1-\sqrt{2})} \frac{\varepsilon_{\text{c}}^{\text{PBE}}(n,s,\zeta)}{n_2^{\text{UEG}}(n,\zeta)}. \qquad (15b)$$

The difference between the ECMD functional defined in Ref. [38] and the present expression (15a)-(15b) is that we approximate here the on-top pair density by its UEG version, i.e. $n_2^{\mathcal{B}}(\mathbf{r},\mathbf{r}) \approx n_2^{\text{UEG}}(n(\mathbf{r}),\zeta(\mathbf{r}))$, where $n_2^{\text{UEG}}(n,\zeta) \approx n^2(1-\zeta^2)g_0(n)$ with the parametrization of the UEG on-top pair-distribution function $g_0(n)$ given in Eq. (46) of Ref. [45]. This represents a major computational saving without loss of accuracy for weakly correlated systems as we eschew the computation of $n_2^{\mathcal{B}}(\mathbf{r},\mathbf{r})$.

Depending on the functional choice, the complementary functional $\bar{E}^{\mathcal{B}}[n_Z^{\mathcal{B}}]$ is approximated by $\bar{E}_{\text{LDA}}^{\mathcal{B}}[n_Z^{\mathcal{B}},\mu^{\mathcal{B}}]$ or $\bar{E}_{\text{PBE}}^{\mathcal{B}}[n_Z^{\mathcal{B}},\mu^{\mathcal{B}}]$ where $\mu^{\mathcal{B}}(\mathbf{r})$ is given by Eq. (9).

As most WFT calculations are performed within the frozen-core (FC) approximation, it is important to define an effective interaction within a subset of MOs. We then naturally split the basis set as $\mathcal{B} = \mathcal{C} \bigcup \mathcal{A}$ (where $\mathcal{C}$ and $\mathcal{A}$ are the sets of core and active MOs, respectively) and define the FC version of the effective interaction as

$$\widetilde{W}^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2) = \begin{cases} \widetilde{f}^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2)/\widetilde{n}_2^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2), & \text{if } \widetilde{n}_2^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2) \neq 0, \\ \infty, & \text{otherwise,} \end{cases}$$
$$(16)$$

with

$$\widetilde{f}^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{pq\in\mathcal{B}} \sum_{rstu\in\mathcal{A}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2) V_{pq}^{rs} \Gamma_{rs}^{tu} \phi_t(\mathbf{r}_1)\phi_u(\mathbf{r}_2),$$
$$(17a)$$

$$\widetilde{n}_2^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{pqrs\in\mathcal{A}} \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_2)\Gamma_{pq}^{rs}\phi_r(\mathbf{r}_1)\phi_s(\mathbf{r}_2), \quad (17b)$$

and the corresponding FC range-separation function $\widetilde{\mu}^{\mathcal{B}}(\mathbf{r}) = (\sqrt{\pi}/2)\widetilde{W}^{\mathcal{B}}(\mathbf{r},\mathbf{r})$. It is noteworthy that, within the present definition, $\widetilde{W}^{\mathcal{B}}(\mathbf{r}_1,\mathbf{r}_2)$ still tends to the regular Coulomb interaction as $\mathcal{B} \to \infty$.

Defining $\widetilde{n}_Z^{\mathcal{B}}$ as the FC (i.e. valence-only) one-electron density obtained with a method Z in $\mathcal{B}$, the FC contribution of the complementary functional is then approximated by $\bar{E}_{\text{LDA}}^{\mathcal{B}}[\widetilde{n}_Z^{\mathcal{B}},\widetilde{\mu}^{\mathcal{B}}]$ or $\bar{E}_{\text{PBE}}^{\mathcal{B}}[\widetilde{n}_Z^{\mathcal{B}},\widetilde{\mu}^{\mathcal{B}}]$.

The most computationally intensive task of the present approach is the evaluation of $W^{\mathcal{B}}(\mathbf{r},\mathbf{r})$ at each quadrature grid point. In the general case (i.e. $\Psi^{\mathcal{B}}$ is a multi-determinant expansion), we compute this embarrassingly parallel step in $\mathcal{O}(N_{\text{grid}}N_{\mathcal{B}}^4)$ computational cost with a memory requirement of $\mathcal{O}(N_{\text{grid}}N_{\mathcal{B}}^2)$, where $N_{\mathcal{B}}$ is the number of basis functions in $\mathcal{B}$. The computational cost can be reduced to $\mathcal{O}(N_{\text{grid}}N^2N_{\mathcal{B}}^2)$ with no memory footprint when $\Psi^{\mathcal{B}}$ is a single Slater determinant. As shown in Ref. [39], this choice for $\Psi^{\mathcal{B}}$ already provides, for weakly correlated systems, a quantitative representation of the incompleteness of $\mathcal{B}$. Hence, we will stick to this choice throughout the present study. In our current implementation, the computational bottleneck is the four-index transformation to get the two-electron integrals in the MO basis which appear in Eqs. (5) and (6). Nevertheless, this step usually has to be performed for most correlated WFT calculations. Modern integral decomposition techniques (such as density fitting[46]) or atomic-orbital-based algorithms could be employed to significantly speed up this step.

To conclude this section, we point out that, thanks to the definitions (4) and (9) as well as the properties (8) and (12), independently of the DFT functional, the present basis-set correction i) can be applied to any WFT method that provides an energy and a density, ii) does not correct one-electron systems, and iii) vanishes in the CBS limit, hence guaranteeing an unaltered CBS limit for a given WFT method.

We begin our investigation of the performance of the basis-set correction by computing the atomization energies of C_2, N_2, O_2 and F_2 obtained with Dunning's cc-pVXZ basis (X = D, T, Q and 5). N_2, O_2 and F_2 are weakly correlated systems and belong to the G2 set[47] (see below), whereas C_2 already
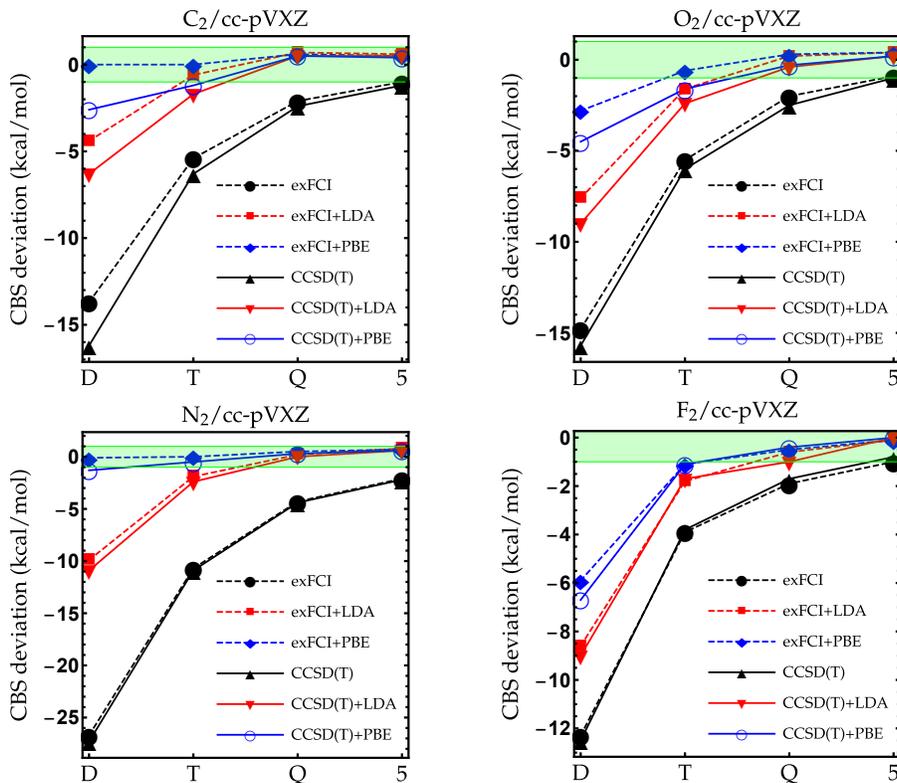
FIG. 1. Deviation (in kcal/mol) from CBS atomization energies of C_2 (top left), O_2 (top right), N_2 (bottom left) and F_2 (bottom right) obtained with various methods and basis sets. The green region corresponds to chemical accuracy (i.e. error below 1 kcal/mol). See supporting information for raw data.

TABLE I. Statistical analysis (in kcal/mol) of the G2 atomization energies depicted in Fig. 2. Mean absolute deviation (MAD), root-mean-square deviation (RMSD), and maximum deviation (MAX) with respect to the CCSD(T)/CBS reference atomization energies. CA corresponds to the number of cases (out of 55) obtained with chemical accuracy. See supporting information for raw data.

Method	MAD	RMSD	MAX	CA
CCSD(T)/cc-pVDZ	14.29	16.21	36.95	2
CCSD(T)/cc-pVTZ	6.06	6.84	14.25	2
CCSD(T)/cc-pVQZ	2.50	2.86	6.75	9
CCSD(T)/cc-pV5Z	1.28	1.46	3.46	21
CCSD(T)+LDA/cc-pVDZ	3.24	3.67	8.13	7
CCSD(T)+LDA/cc-pVTZ	1.19	1.49	4.67	27
CCSD(T)+LDA/cc-pVQZ	0.33	0.44	1.32	53
CCSD(T)+PBE/cc-pVDZ	1.96	2.59	7.33	19
CCSD(T)+PBE/cc-pVTZ	0.85	1.11	2.64	36
CCSD(T)+PBE/cc-pVQZ	0.31	0.42	1.16	53

contains a non-negligible amount of strong correlation.[48] In a second time, we compute the atomization energies of the entire G2 set[47] composed by 55 molecules with the cc-pVXZ basis set family. This molecular set has been intensively studied in the last 20 years (see, for example, Refs. 49–57) and can be considered as a representative set of small organic

and inorganic molecules. As a method Y we employ either CCSD(T) or exFCI. Here, exFCI stands for extrapolated FCI energies computed with the CIPSI algorithm.[58–60] We refer the interested reader to Refs. 61–65 for more details. In the case of the CCSD(T) calculations, we have $Z = $ ROHF as we use the restricted open-shell HF (ROHF) one-electron density to compute the complementary basis-set correction energy. In the case of exFCI, the one-electron density is computed from a very large CIPSI expansion containing several million determinants. CCSD(T) energies are computed with Gaussian09 using standard threshold values,[66] while RS-DFT and exFCI calculations are performed with QUANTUM PACKAGE.[67] For the numerical quadratures, we employ the SG-2 grid.[68] Apart from the carbon dimer where we have taken the experimental equilibrium bond length (1.2425 Å), all geometries have been extracted from Ref. 69 and have been obtained at the B3LYP/6-31G(2df,p) level of theory. Frozen-core calculations are systematically performed and defined as such: a He core is frozen from Li to Ne, while a Ne core is frozen from Na to Ar. In the context of the basis-set correction, the set of active MOs, $\mathcal{A}$, involved in the definition of the effective interaction [see Eq. (16)] refers to the non-frozen MOs. The FC density-based correction is used consistently with the FC approximation in WFT methods. To estimate the CBS limit of each method, following Ref. 70, we perform a two-point X^{-3} extrapolation of the correlation energies using the quadruple- and quintuple-ζ data that we add up to the HF energies obtained in the largest
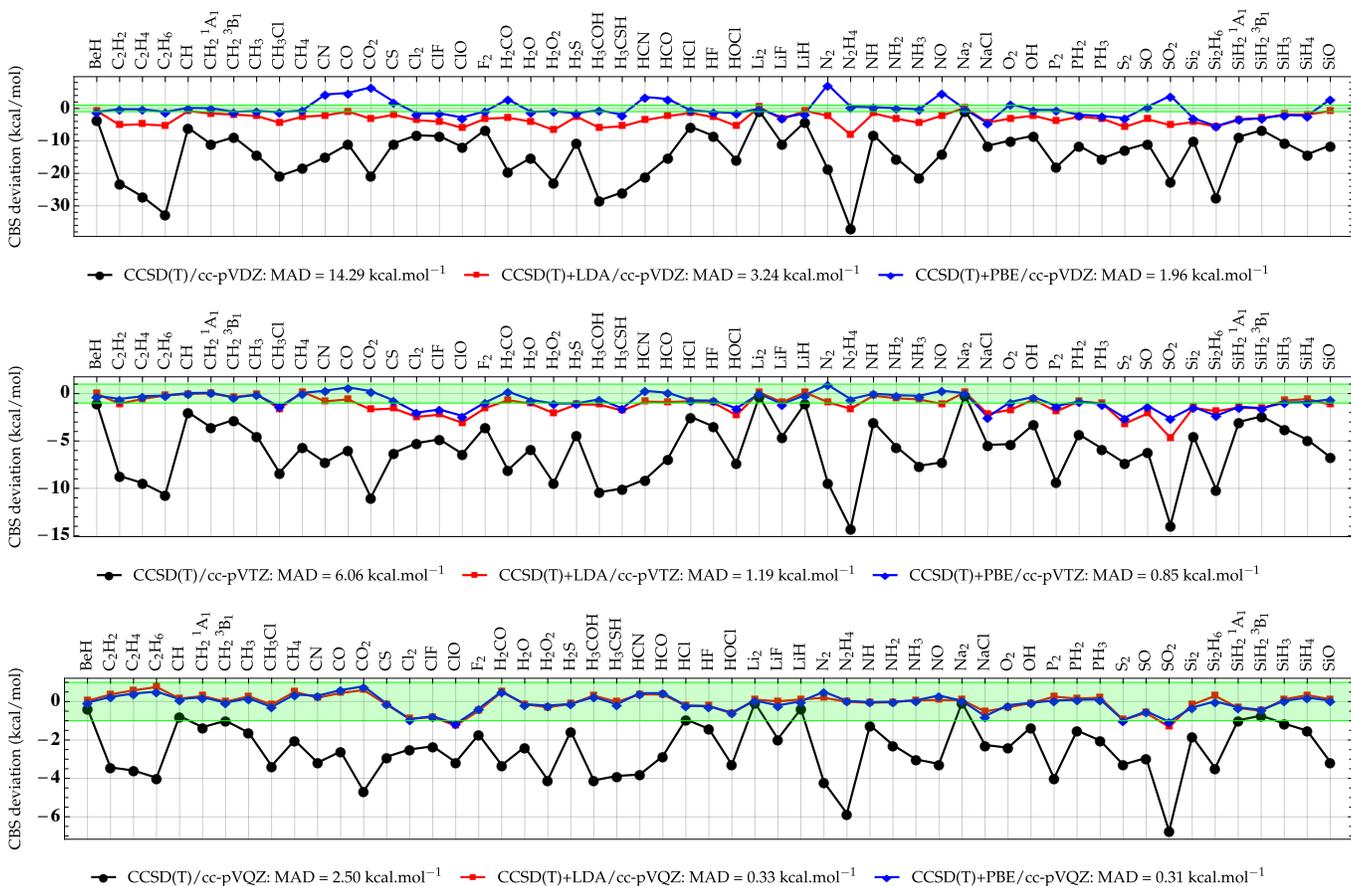
FIG. 2. Deviation (in kcal/mol) from the CCSD(T)/CBS atomization energy obtained with various methods with the cc-pVDZ (top), cc-pVTZ (center) and cc-pVQZ (bottom) basis sets. The green region corresponds to chemical accuracy (i.e. error below 1 kcal/mol). See supporting information for raw data.

(i.e. quintuple-ζ) basis.

As the exFCI atomization energies are converged with a precision of about 0.1 kcal/mol, we can label these as near FCI. Hence, they will be our references for C_2, N_2, O_2 and F_2. The results for these diatomic molecules are reported in Fig. 1. The corresponding numerical data can be found in the supporting information. As one can see, the convergence of the exFCI atomization energies is, as expected, slow with respect to the basis set: chemical accuracy (error below 1 kcal/mol) is barely reached for C_2, O_2 and F_2 even with the cc-pV5Z basis set, and the atomization energies are consistently underestimated. A similar trend holds for CCSD(T). Regarding the effect of the basis-set correction, several general observations can be made for both exFCI and CCSD(T). First, in a given basis set, the basis-set correction systematically improves the atomization energies (both at the LDA and PBE levels). A small overestimation can occur compared to the CBS value by a few tenths of a kcal/mol (the largest deviation being 0.6 kcal/mol for N_2 at the CCSD(T)+PBE/cc-pV5Z level). Nevertheless, the deviation observed for the largest basis set is typically within the CBS extrapolation error, which is highly satisfactory knowing the marginal computational cost of the present correction. In most cases, the basis-set corrected triple-ζ atomization

energies are on par with the uncorrected quintuple-ζ ones. Importantly, the sensitivity with respect to the RS-DFT functional is quite large for the double- and triple-ζ basis sets, where clearly the PBE functional performs better. However, from the quadruple-ζ basis, the LDA and PBE functionals agree within a few tenths of a kcal/mol. Such weak sensitivity to the density-functional approximation when reaching large basis sets shows the robustness of the approach.

As a second set of numerical examples, we compute the error (with respect to the CBS values) of the atomization energies from the G2 test set with $Y = $ CCSD(T), $Z = $ ROHF and the cc-pVXZ basis sets. Here, all atomization energies have been computed with the same near-CBS HF/cc-pV5Z energies; only the correlation energy contribution varies from one method to the other. Investigating the convergence of correlation energies (or difference of such quantities) is commonly done to appreciate the performance of basis-set corrections aiming at correcting two-electron effects.[11,18,71] The "plain" CCSD(T) atomization energies as well as the corrected CCSD(T)+LDA and CCSD(T)+PBE values are depicted in Fig. 2. The raw data can be found in the supporting information. A statistical analysis of these data is also provided in Table I, where we report the mean absolute deviation (MAD),

root-mean-square deviation (RMSD), and maximum deviation (MAX) with respect to the CCSD(T)/CBS atomization energies. Note that the MAD of our CCSD(T)/CBS atomization energies is only 0.37 kcal/mol compared to the values extracted from Ref. 54 which corresponds to frozen-core non-relativistic atomization energies obtained at the CCSD(T)(F12)/cc-pVQZ-F12 level of theory corrected for higher-excitation contributions ($E_{\mathrm{CCSDT(Q)/cc\text{-}pV(D+d)Z}} - E_{\mathrm{CCSD(T)/cc\text{-}pV(D+d)Z}}$). From double-$\zeta$ to quintuple-ζ basis, the MAD associated with the CCSD(T) atomization energies goes down slowly from 14.29 to 1.28 kcal/mol. For a commonly used basis like cc-pVTZ, the MAD of CCSD(T) is still 6.06 kcal/mol. Applying the basis-set correction drastically reduces the basis-set incompleteness error. Already at the CCSD(T)+LDA/cc-pVDZ and CCSD(T)+PBE/cc-pVDZ levels, the MAD is reduced to 3.24 and 1.96 kcal/mol. With the triple-ζ basis, the MAD of CCSD(T)+PBE/cc-pVTZ is already below 1 kcal/mol with 36 cases (out of 55) where we achieve chemical accuracy. CCSD(T)+LDA/cc-pVQZ and CCSD(T)+PBE/cc-pVQZ return MAD of 0.33 and 0.31 kcal/mol (respectively) while CCSD(T)/cc-pVQZ still yields a fairly large MAD of 2.50 kcal/mol.

Therefore, similar to F12 methods,[11] we can safely claim that the present basis-set correction provides significant basis-set reduction and recovers quintuple-ζ quality atomization and correlation energies with triple-ζ basis sets for a much cheaper computational cost. Encouraged by these promising results, we are currently pursuing various avenues toward basis-set reduction for strongly correlated systems and electronically excited states.

SUPPORTING INFORMATION

See supporting information for raw data associated with the atomization energies of the four diatomic molecules and the G2 set.

ACKNOWLEDGMENTS

[1] Pople, J. A. Nobel Lecture: Quantum chemical models. *Rev. Mod. Phys.* **1999**, *71*, 1267.
[2] Kohn, W. Nobel Lecture: Electronic structure of matter - wave functions and density functionals. *Rev. Mod. Phys.* **1999**, *71*, 1253.
[3] Kutzelnigg, W. r12-Dependent terms in the wave function as closed sums of partial wave amplitudes for large l. *Theor. Chim. Acta* **1985**, *68*, 445.
[4] Hylleraas, E. A. Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium. *Z. Phys.* **1929**, *54*, 347.
[5] Kutzelnigg, W.; Klopper, W. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. *J. Chem. Phys.* **1991**, *94*, 1985.
[6] Noga, J.; Kutzelnigg, W. Coupled cluster theory that takes care of the correlation cusp by inclusion of linear terms in the interelectronic coordinates. *J. Chem. Phys.* **1994**, *101*, 7738.
[7] Ten-no, S. Explicitly correlated wave functions: summary and perspective. *Theor. Chem. Acc.* **2012**, *131*, 1070.
[8] Ten-no, S.; Noga, J. Explicitly correlated electronic structure theory from R12/F12 ansatze. *WIREs Comput. Mol. Sci.* **2012**, *2*, 114.
[9] Hattig, C.; Klopper, W.; Kohn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4.
[10] Kong, L.; Bischo, F. A.; Valeev, E. F. Explicitly Correlated R12/F12 Methods for Electronic Structure. *Chem. Rev.* **2012**, *112*, 75.
[11] Tew, D. P.; Klopper, W.; Neiss, C.; Hattig, C. Quintuple-ζ quality coupled-cluster correlation energies with triple-ζ basis sets. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1921.
[12] Barca, G. M.; Loos, P.-F. Three-and Four-Electron Integrals Involving Gaussian Geminals: Fundamental Integrals, Upper Bounds, and Recurrence Relations. *J. Chem. Phys.* **2017**, *147*, 024103.
[13] Torheyden, M.; Valeev, E. F. Universal perturbative explicitly correlated basis set incompleteness correction. *J. Chem. Phys.* **2009**, *131*, 171103.
[14] Kong, L.; Valeev, E. F. Perturbative correction for the basis set incompleteness error of complete-active- space self-consistent field. *J. Chem. Phys.* **2010**, *133*, 174126.
[15] Kong, L.; Valeev, E. F. SF-[2]$_{R12}$: A spin-adapted explicitly correlated method applicable to arbitrary electronic states. *J. Chem. Phys.* **2011**, *135*, 214105.
[16] Booth, G. H.; Cleland, D.; Alavi, A.; Tew, D. P. An explicitly correlated approach to basis set incompleteness in full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2012**, *137*, 164112.
[17] Irmler, A.; Hummel, F.; Gruneis, A. On the duality of ring and ladder diagrams and its importance for many-electron perturbation theories. *arXiv* **2019**, *1903.05458*.
[18] Irmler, A.; Gruneis, A. Particle-particle ladder based basis-set corrections applied to atoms and molecules using coupled-cluster theory. *arXiv* **2019**, *1903.05559*.
[19] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
[20] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B 864.
[21] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
[22] Becke, A. D. Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **2014**, *140*, 18A301.
[23] Perdew, J. P.; Schmidt, K. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.* **2001**, *577*, 1.
[24] Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits. *J. Chem. Phys.* **2005**, *123*, 062201.
[25] Franck, O.; Mussard, B.; Luppi, E.; Toulouse, J. Basis convergence of range-separated density-functional theory. *J. Chem. Phys.* **2015**, *142*, 074107.
[26] Toulouse, J.; Colonna, F.; Savin, A. Long-range–short-range separation of the electron-electron interaction in density-functional theory. *Phys. Rev. A* **2004**, *70*, 062505.
[27] Ángyán, J. G.; Gerber, I. C.; Savin, A.; Toulouse, J. van der Waals forces in density functional theory: perturbational long-range electron interaction corrections. *Phys. Rev. A* **2005**, *72*, 012510.
[28] Goll, E.; Werner, H.-J.; Stoll, H. A short-range gradient-corrected density functional in long-range coupled-cluster calculations for rare gas dimers. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3917.
[29] Toulouse, J.; Gerber, I. C.; Jansen, G.; Savin, A.; Ángyán, J. G. Adiabatic-connection fluctuation-dissipation density-functional theory based on range separation. *Phys. Rev. Lett.* **2009**, *102*, 096404.
[30] Janesko, B. G.; Henderson, T. M.; Scuseria, G. E. Long-range-corrected hybrids including random phase approximation correlation. *J. Chem. Phys.* **2009**, *130*, 081105.
[31] Toulouse, J.; Zhu, W.; Savin, A.; Jansen, G.; Ángyán, J. G. Closed-shell ring coupled cluster doubles theory with range separation applied on weak intermolecular interactions. *J. Chem. Phys.* **2011**, *135*, 084119.
[32] Mussard, B.; Reinhardt, P.; Ángyán, J. G.; Toulouse, J. Spin-unrestricted random-phase approximation with range separation: Benchmark on atomization energies and reaction barrier heights. *J. Chem. Phys.* **2015**, *142*,

154123, Erratum: J. Chem. Phys. **142**, 219901 (2015).

[33] Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **1997**, *275*, 151.

[34] Fromager, E.; Toulouse, J.; Jensen, H. J. A. On the universality of the long-/short-range separation in multiconfigurational density-functional theory. *J. Chem. Phys.* **2007**, *126*, 074111.

[35] Fromager, E.; Cimiraglia, R.; Jensen, H. J. A. Merging multireference perturbation and density-functional theories by means of range separation: Potential curves for Be$_2$, Mg$_2$, and Ca$_2$. *Phys. Rev. A* **2010**, *81*, 024502.

[36] Hedegård, E. D.; Knecht, S.; Kielberg, J. S.; Jensen, H. J. A.; Reiher, M. Density matrix renormalization group with efficient dynamical electron correlation through range separation. *J. Chem. Phys.* **2015**, *142*, 224108.

[37] Hedegård, E. D.; Toulouse, J.; Jensen, H. J. A. Multiconfigurational short-range density-functional theory for open-shell systems. *J. Chem. Phys.* **2018**, *148*, 214103.

[38] Ferté, A.; Giner, E.; Toulouse, J. Range-separated multideterminant density-functional theory with a short-range correlation functional of the on-top pair density. *J. Chem. Phys.* **2019**, *150*, 084103.

[39] Giner, E.; Pradines, B.; Ferté, A.; Assaraf, R.; Savin, A.; Toulouse, J. Curing basis-set convergence of wave-function theory using density-functional theory: A systematically improvable approach. *J. Chem. Phys.* **2018**, *149*, 194301.

[40] Toulouse, J.; Gori-Giorgi, P.; Savin, A. A short-range correlation energy density functional with multi-determinantal reference. *Theor. Chem. Acc.* **2005**, *114*, 305.

[41] Loos, P.-F.; Gill, P. M. W. The uniform electron gas. *WIREs Comput. Mol. Sci.* **2016**, *6*, 410.

[42] Paziani, S.; Moroni, S.; Gori-Giorgi, P.; Bachelet, G. B. Local-spin-density functional for multideterminant density functional theory. *Phys. Rev. B* **2006**, *73*, 155111.

[43] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

[44] Gori-Giorgi, P.; Savin, A. Properties of short-range and long-range correlation energy density functionals from electron-electron coalescence. *Phys. Rev. A* **2006**, *73*, 032506.

[45] Gori-Giorgi, P.; Savin, A. Properties of short-range and long-range correlation energy density functionals from electron-electron coalescence. *Phys. Rev. A* **2006**, *73*, 032506.

[46] Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496.

[47] Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian2 theory for molecular energies of first and secondrow compounds. *J. Chem. Phys.* **1991**, *94*, 7221.

[48] Booth, G. H.; Cleland, D.; Thom, A. J. W.; Alavi, A. Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods. *J. Chem. Phys.* **2011**, *135*, 084104.

[49] Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J. Chem. Phys.* **2008**, *129*, 204105.

[50] Grossman, J. C. Benchmark quantum Monte Carlo calculations. *J. Chem. Phys.* **2002**, *117*, 1434.

[51] Feller, D.; Peterson, K. A. Re-examination of atomization energies for the Gaussian-2 set of molecules. *J. Chem. Phys.* **1999**, *110*, 8384.

[52] Nemec, N.; Towler, M. D.; Needs, R. J. Benchmark all-electron ab initio quantum Monte Carlo calculations for small molecules. *J. Chem. Phys.*

[52] **2010**, *132*, 034111.

[53] Feller, D.; Peterson, K. A.; Hill, J. G. On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies. *J. Chem. Phys.* **2011**, *135*, 044102.

[54] Haunschild, R.; Klopper, W. New accurate reference energies for the G2/97 test set. *J. Chem. Phys.* **2012**, *136*, 164102.

[55] Petruzielo, F. R.; Toulouse, J.; Umrigar, C. J. Approaching chemical accuracy with quantum Monte Carlo. *J. Chem. Phys.* **2012**, *136*, 124116.

[56] Feller, D.; Peterson, K. A. An expanded calibration study of the explicitly correlated CCSD(T)-F12b method using large basis set standard CCSD(T) atomization energies. *J. Chem. Phys.* **2013**, *139*, 084110.

[57] Kesharwani, M. K.; Sylvetsky, N.; Kohn, A.; Tew, D. P.; Martin, J. M. L. Do CCSD and approximate CCSD-F12 variants converge to the same basis set limits? The case of atomization energies. *J. Chem. Phys.* **2018**, *149*, 154109.

[58] Huron, B.; Malrieu, J.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigura- tional zeroth-order wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745.

[59] Giner, E.; Scemama, A.; Caffarel, M. Using perturbatively selected configuration interaction in quantum Monte Carlo calculations. *Can. J. Chem.* **2013**, *91*, 879.

[60] Giner, E.; Scemama, A.; Caffarel, M. Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F2 using selected configuration interaction trial wavefunctions. *J. Chem. Phys.* **2015**, *142*.

[61] Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.

[62] Scemama, A.; Garniron, Y.; Caffarel, M.; Loos, P. F. Deterministic construction of nodal surfaces within quantum Monte Carlo: the case of FeS. *J. Chem. Theory Comput.* **2018**, *14*, 1395.

[63] Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.

[64] Scemama, A.; Benali, A.; Jacquemin, D.; Caffarel, M.; Loos, P.-F. Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes. *J. Chem. Phys.* **2018**, *149*, 034108.

[65] Loos, P.-F.; Boggio-Pasqua, M.; Scemama, A.; Caffarel, M.; Jacquemin, D. Reference Energies for Double Excitations. *J. Chem. Theory Comput.* **2019**, *15*, 1939–1956.

[66] Frisch, M. J. et al. Gaussian 09 Revision D.01. 2009; Gaussian Inc. Wallingford CT.

[67] Garniron, Y.; Gasperich, K.; Applencourt, T.; Benali, A.; Ferté, A.; Paquier, J.; Pradines, B.; Assaraf, R.; Reinhardt, P.; Toulouse, J.; Barbaresco, P.; Renon, N.; David, G.; Malrieu, J. P.; Véril, M.; Caffarel, M.; Loos, P. F.; Giner, E.; Scemama, A. Quantum Package 2.0: a open-source determinant-driven suite of programs. *J. Chem. Theory Comput.* **2019**, *submitted*.

[68] Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38*, 869–882.

[69] Haunschild, R.; Janesko, B. G.; Scuseria, G. E. Local hybrids as a perturbation to global hybrid functionals. *J. Chem. Phys.* **2009**, *131*, 154112.

[70] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243 – 252.

[71] Ten-no, S. Initiation of explicitly correlated Slater-type geminal theory. *Chem. Phys. Lett.* **2004**, *398*, 56 – 61.