Accurate energies of transition metal atoms, ions, and monoxides using selected configuration interaction and density-based basis-set corrections

Yuan Yao,[1,a] (iD) Emmanuel Giner,[2,b] (iD) Tyler A. Anderson,[1,c] (iD) Julien Toulouse,[2,3,d] (iD) and C. J. Umrigar[1,a] (iD)

AFFILIATIONS

[1] Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA
[2] Laboratoire de Chimie Théorique, Sorbonne Université and CNRS, F-75005 Paris, France
[3] Institut Universitaire de France, F-75005 Paris, France

[a] Authors to whom correspondence should be addressed: yy682@cornell.edu and cyrusumrigar@cornell.edu
[b] emmanuel.giner@lct.jussieu.fr
[c] taa65@cornell.edu
[d] toulouse@lct.jussieu.fr

ABSTRACT

The semistochastic heat-bath configuration interaction method is a selected configuration interaction plus perturbation theory method that has provided near-full configuration interaction (FCI) levels of accuracy for many systems with both single- and multi-reference character. However, obtaining accurate energies in the complete basis-set limit is hindered by the slow convergence of the FCI energy with respect to basis size. Here, we show that the recently developed basis-set correction method based on range-separated density functional theory can be used to significantly speed up basis-set convergence in SHCI calculations. In particular, we study two such schemes that differ in the functional used and apply them to transition metal atoms and monoxides to obtain total, ionization, and dissociation energies well converged to the complete-basis-set limit within chemical accuracy.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0072296

I. INTRODUCTION

Selected configuration interaction plus perturbation theory (SCI + PT) methods[1–8] are capable of providing excellent approximations to full configuration interaction (FCI) energies, i.e., the exact energies within a given one-electron basis, for systems and basis-sets that are considerably larger than those for which FCI is feasible. In this paper, we employ a particularly efficient SCI method developed by some of the current authors, namely, the semistochastic heat-bath configuration interaction (SHCI) method,[9–12] which has been used for a number of challenging problems, including the potential energy curve of the Cr_2 molecule,[13] the dissociation energies of the 55 molecules comprising the G2 set,[14] and the ionization and dissociation energies of seven transition metal elements (their atoms, ions, and monoxides).[15]

In particular, the SHCI results on transition metal systems have served as accurate benchmark energies for 19 other quantum chemistry methods for basis sets from DZ to 5Z as well as the extrapolated complete-basis-set (CBS) limit.[15] However, unlike for the extrapolation to the FCI limit, whose extrapolation error is on the order of a couple of mHa for the largest monoxide molecules and much less for most other systems in the study, the largest extrapolation error to the CBS values can be on the order of 10 mHa.[16] In addition, the computational cost goes up rapidly with basis size, so efficient techniques

for accelerating the convergence to the CBS limit become important in situations where the available computational resources limit the basis sizes one can use.

The slow convergence of SCI and other wave-function methods with respect to the basis-set size is due to the failure of wave functions expanded in finite one-electron basis sets to reproduce the short-range correlation hole around the electron–electron cusp present in exact wave functions. F12 methods[17–21] accelerate the basis-set convergence of wave-function methods by restoring the electron–electron cusp with a correlation factor at the cost of introducing an auxiliary basis. In particular, F12 methods have been used to accelerate the basis-set convergence of coupled-cluster calculations on transition metal systems.[22–25] In this paper, we instead use two variants of a method based on range-separated density-functional theory (RS-DFT), recently developed by some of the current authors.[26–30] The basic idea is that the electron–electron Coulomb interaction projected in a one-electron basis resembles the long-range potential, $w_{ee}^{lr}(r_{12}) = \text{erf}(\mu r_{12})/r_{12}$, used in RS-DFT. The on-top (OT) value of the projected Coulomb interaction provides the local value of the range-separation parameter μ and functionals derived in the context of RS-DFT can be used to provide a basis-set correction.

The outline of this paper is as follows: In Sec. II, we briefly review the SHCI method. In Sec. III, we discuss the two basis-set correction schemes we employ in this paper. Section IV contains the computational details. In Sec. V, we present the energies obtained with and without the two basis-set correction formulas. The conclusions are given in Sec. VI.

II. REVIEW OF THE SHCI METHOD

In this section, we give a brief review of the SHCI method.

The variational stage of SHCI starts from an initial determinant, such as the Hartree–Fock (HF) determinant, and generates the variational wave function, Ψ_V, through an iterative procedure. At each iteration, Ψ_V is written as a linear combination of the determinants in the variational set $\mathcal{V}$,

$$|\Psi_V\rangle = \sum_{D_i \in \mathcal{V}} c_i |D_i\rangle, \tag{1}$$

and new determinants, D_a, from the perturbative set $\mathcal{P}$ (i.e., the determinants not in $\mathcal{V}$ but connected through the Hamiltonian to the determinants in $\mathcal{V}$) that satisfy the criterion

$$\exists \, D_i \in \mathcal{V} \text{ such that } |H_{ai} c_i| \geq \epsilon_1 \tag{2}$$

are added to the set $\mathcal{V}$, where H_{ai} is the Hamiltonian matrix element between determinants D_a and D_i, and ϵ_1 is a user-defined parameter that controls the accuracy of the variational stage.[31] (When $\epsilon_1 = 0$, the method becomes equivalent to FCI.) After adding the new determinants to the set $\mathcal{V}$, the Hamiltonian matrix is constructed and diagonalized using the diagonally preconditioned Davidson method[32] to obtain an improved estimate of the lowest eigenvalue, E_V, and eigenvector, Ψ_V. This process is repeated until the change in the variational energy E_V falls below a certain threshold.

Although the selection criterion in Eq. (2) is only an approximation to selection criteria based on the largest contributions to

the first-order perturbative correction to the wave function or the second-order perturbative correction to the energy, the selected determinants are only slightly inferior to those selected by either of these perturbative criteria. This is greatly outweighed by the improved selection speed.

In the perturbative stage of SHCI, a second-order Epstein–Nesbet perturbative energy correction $\Delta E^{(2)}$ is calculated using the determinants connected to the final set $\mathcal{V}$ that satisfy Eq. (2) but with ϵ_1 replaced by ϵ_2, which is typically orders of magnitude smaller than ϵ_1,

$$\Delta E^{(2)} = \sum_{D_a \in \mathcal{P}} \frac{\left(\sum_{D_i \in \mathcal{V}}^{(\epsilon_2)} H_{ai} c_i \right)^2}{E_V - E_a}, \tag{3}$$

where $E_a = H_{aa}$ and $\sum^{(\epsilon_2)}$ indicates that only terms in the sum for which $|H_{ai} c_i| \geq \epsilon_2$ are included. The final SHCI energy is then $E_{SHCI} = E_V + \Delta E^{(2)}$.

In both the variational and the perturbative stages, the fact that the number of distinct values of the double-excitation matrix elements scales only as N_{orb}^4, where N_{orb} is the number of orbitals, is used to avoid ever looking at the very large number of unimportant determinants that do not contribute to the final energy. Nevertheless, straightforward approaches for evaluating the perturbative correction entail either storing all determinants of the perturbative set, $\mathcal{P}$, which creates a memory bottleneck for large systems, or a significant increase in computer time.

SHCI circumvents this using a three-step semistochastic approach for the second-order perturbation calculation, where the energy correction is split into deterministic, "pseudo-stochastic," and stochastic contributions.[12] The deterministic step chooses both variational and perturbative determinants deterministically, the "pseudo-stochastic" step chooses variational determinants deterministically and perturbative determinants stochastically, and the stochastic step chooses both variational and perturbative determinants stochastically. A threshold value ϵ_2^{dtm} ($< \epsilon_1$), automatically determined to correspond to a determinant set of manageable size depending on available computer memory, is first used to obtain a deterministic energy correction. The remaining correlation is calculated as the sum of a "pseudo-stochastic" part obtained as the difference in the second-order corrections evaluated with ϵ_2^{dtm} and a smaller threshold ϵ_2^{psto}, and a stochastic part obtained as the difference in corrections between ϵ_2^{psto} and the final threshold ϵ_2 ($\epsilon_2 < \epsilon_2^{psto} < \epsilon_2^{dtm} < \epsilon_1$, separated by one order of magnitude from one another). Stochastic samples are taken until the statistical error falls below a specified threshold, which is typically on the order of a few microHartrees.

In a typical SHCI calculation, the variational energy and the corresponding perturbative correction are computed for several values of ϵ_1 (for the systems presented in this study, the lowest ϵ_1 values are in the $2–4 \times 10^{-5}$ range). To estimate the FCI energy, we perform a weighted quadratic fit of E_{SHCI} to $-\Delta E^{(2)}$ to obtain E_{SHCI} at $-\Delta E^{(2)} = 0$ using weights proportional to $(\Delta E^{(2)})^{-2}$. In order to reduce the extrapolation error, one can either go to larger variational wave functions by decreasing ϵ_1, incurring a large memory footprint, or optimize the orbitals[33] to minimize the variational energy for the same number of determinants. To limit the computational cost

of orbital optimization, the value of the threshold ϵ_1 used during optimization can be much larger than the value used to obtain the final energy.

III. BASIS-SET CORRECTION

One significant drawback of wave-function methods is the slow convergence of energies with respect to the size of the one-electron basis set. This is because wave functions expanded in finite one-electron basis sets usually poorly describe short-range correlation near the electron–electron cusp. To estimate the basis-set incompleteness error and speed up basis-set convergence, we employ the density-based basis-set correction recently developed in Refs. 26–30 by some of the current authors.

The starting idea of this basis-set correction scheme is that, in a given basis set $\mathcal{B}$, the exact ground-state energy E_0 of an N-electron system can be approximated by the energy $E_0^{\mathcal{B}}$ defined by the following minimization over $\mathcal{B}$-representable one-electron densities $n^{\mathcal{B}}(\mathbf{r})$, i.e., densities that can be obtained from a wave function $\Psi^{\mathcal{B}}$ belonging to the N-electron Hilbert space generated by the basis set $\mathcal{B}$:

$$E_0^{\mathcal{B}} = \min_{n^{\mathcal{B}}}\left\{ F[n^{\mathcal{B}}] + \int \mathrm{d}\mathbf{r}\, v_{\mathrm{ne}}(\mathbf{r}) n^{\mathcal{B}}(\mathbf{r}) \right\}, \qquad (4)$$

where $v_{\mathrm{ne}}(\mathbf{r})$ is the nuclei–electron potential and $F[n] = \min_{\Psi \to n}\langle\Psi|\hat{T} + \hat{W}_{\mathrm{ee}}|\Psi\rangle$ is the standard constrained-search Levy–Lieb universal density functional[34,35] with $\hat{T}$ and $\hat{W}_{\mathrm{ee}}$ being the kinetic and electron–electron Coulomb operators, respectively. Obviously, in the CBS limit, $E_0^{\mathcal{B}}$ tends to the exact ground-state energy, i.e., $\lim_{\mathcal{B}\to\mathrm{CBS}} E_0^{\mathcal{B}} = E_0$. Crucially, since the density usually has a fast convergence with the size of the basis set, the energy $E_0^{\mathcal{B}}$ also has a fast basis-set convergence.

As it stands, Eq. (4) is, of course, not practical due to the lack of a sufficiently accurate explicit approximation for the universal density functional $F[n]$. However, as shown in Ref. 26, by reintroducing a wave function, $\Psi^{\mathcal{B}}$, the energy $E_0^{\mathcal{B}}$ can be expressed as

$$E_0^{\mathcal{B}} = \min_{\Psi^{\mathcal{B}}}\left\{ \langle\Psi^{\mathcal{B}}|\hat{H}|\Psi^{\mathcal{B}}\rangle + \bar{E}^{\mathcal{B}}[n_{\Psi^{\mathcal{B}}}] \right\}, \qquad (5)$$

where the minimization is over normalized N-electron wave functions $\Psi^{\mathcal{B}}$ expanded in the basis set $\mathcal{B}$, $\hat{H}$ is the total Hamiltonian, and $\bar{E}^{\mathcal{B}}[n]$ is a complementary basis-set correction density functional,

$$\bar{E}^{\mathcal{B}}[n] = \min_{\Psi \to n}\langle\Psi|\hat{T} + \hat{W}_{\mathrm{ee}}|\Psi\rangle - \min_{\Psi^{\mathcal{B}} \to n}\langle\Psi^{\mathcal{B}}|\hat{T} + \hat{W}_{\mathrm{ee}}|\Psi^{\mathcal{B}}\rangle, \qquad (6)$$

which must be evaluated at the density of the wave function $\Psi^{\mathcal{B}}$. The basis-set correction density functional $\bar{E}^{\mathcal{B}}[n]$ compensates for the basis-set restriction on the wave function $\Psi^{\mathcal{B}}$ in Eq. (5) and vanishes in the CBS limit. This scheme can be trivially generalized to a basis-set correction functional depending on spin-resolved densities $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$, giving the same energy $E_0^{\mathcal{B}}$ as

$$E_0^{\mathcal{B}} = \min_{\Psi^{\mathcal{B}}}\left\{ \langle\Psi^{\mathcal{B}}|\hat{H}|\Psi^{\mathcal{B}}\rangle + \bar{E}^{\mathcal{B}}[n_{\uparrow,\Psi^{\mathcal{B}}}, n_{\downarrow,\Psi^{\mathcal{B}}}] \right\}, \qquad (7)$$

where $\bar{E}^{\mathcal{B}}[n_{\uparrow,\Psi^{\mathcal{B}}}, n_{\downarrow,\Psi^{\mathcal{B}}}]$ is a new basis-set correction functional evaluated at the spin-↑ and spin-↓ densities of the wave function $\Psi^{\mathcal{B}}$. Moreover, as shown in Ref. 30, the basis-set correction scheme can also be extended to a functional depending on both the density $n(\mathbf{r})$ and the on-top pair density $n_2(\mathbf{r})$, giving again the same energy $E_0^{\mathcal{B}}$ as

$$E_0^{\mathcal{B}} = \min_{\Psi^{\mathcal{B}}}\left\{ \langle\Psi^{\mathcal{B}}|\hat{H}|\Psi^{\mathcal{B}}\rangle + \bar{E}^{\mathcal{B}}[n_{\Psi^{\mathcal{B}}}, n_{2,\Psi^{\mathcal{B}}}] \right\}, \qquad (8)$$

where $\bar{E}^{\mathcal{B}}[n_{\Psi^{\mathcal{B}}}, n_{2,\Psi^{\mathcal{B}}}]$ is yet another basis-set correction functional evaluated at the density and on-top pair density of the wave function $\Psi^{\mathcal{B}}$.

In practice, we use two approximate basis-set correction functionals. For the scheme of Eq. (7), we use the Perdew–Burke–Ernzerhof-uniform electron gas (PBE-UEG) functional of the form[27]

$$\bar{E}_{\mathrm{PBE-UEG}}^{\mathcal{B}}[n_\uparrow, n_\downarrow] = \int \mathrm{d}\mathbf{r}\, f\big(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r}), \nabla n(\mathbf{r}), \mu^{\mathcal{B}}(\mathbf{r})\big), \qquad (9)$$

where f is a function of the spin-resolved densities $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$, the density gradient $\nabla n(\mathbf{r})$, and the local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$. The function f was constructed based on a short-range version[36] of the Perdew–Burke–Ernzerhof (PBE)[37] correlation functional. One key ingredient of this short-range functional is a parameterization (in terms of the spin-resolved densities) of the on-top pair density of the uniform electron gas (UEG),[38,39] hence its name. Its full explicit form was given in Refs. 27–30. For the scheme of Eq. (8), we use the spin-unpolarized PBE-OT functional of the form[29]

$$\bar{E}_{\mathrm{PBE-OT}}^{\mathcal{B}}[n, n_2] = \int \mathrm{d}\mathbf{r}\, g\big(n(\mathbf{r}), \nabla n(\mathbf{r}), n_2(\mathbf{r}), \mu^{\mathcal{B}}(\mathbf{r})\big), \qquad (10)$$

where g is a function of the density $n(\mathbf{r})$, the density gradient $\nabla n(\mathbf{r})$, the on-top pair density $n_2(\mathbf{r})$, and the local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$. (To avoid confusion, we point out that the PBE-OT functional of this paper was named the SU-PBE-OT functional in Ref. 29.) The PBE-OT functional has essentially the same physical content as the PBE-UEG functional with the exception that it uses the on-top (OT) pair density $n_2(\mathbf{r})$ of the wave function $\Psi^{\mathcal{B}}$ as an independent variable, instead of that of the UEG, which is more accurate for strongly correlated systems. In addition, it uses the total density instead of the spin-resolved densities since the additional information provided by the spin-resolved densities is largely already contained in the on-top pair density $n_2(\mathbf{r})$.[29] Its full explicit form was given in Refs. 28–30.

The common key ingredient in these basis-set correction functionals is the local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$, which provides a local measure of the incompleteness of the basis set, and is given by[26,30]

$$\mu^{\mathcal{B}}(\mathbf{r}) = \frac{\sqrt{\pi}}{2} W^{\mathcal{B}}(\mathbf{r}), \qquad (11)$$

where $W^{\mathcal{B}}(\mathbf{r})$ is the on-top value of an effective two-electron interaction representing the Coulomb two-electron interaction in the basis set $\mathcal{B}$. In an orthonormal orbital basis $\{\phi_p\}$ generated by the basis set $\mathcal{B}$, $W^{\mathcal{B}}(\mathbf{r})$ is defined by

$$W^{\mathcal{B}}(\mathbf{r}) = \begin{cases} f^{\mathcal{B}}_{\Psi^{\mathcal{B}}_{\text{loc}}}(\mathbf{r})/n_{2,\Psi^{\mathcal{B}}_{\text{loc}}}(\mathbf{r}) & \text{if } n_{2,\Psi^{\mathcal{B}}_{\text{loc}}}(\mathbf{r}) \neq 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{12}$$

with

$$f^{\mathcal{B}}_{\Psi^{\mathcal{B}}_{\text{loc}}}(\mathbf{r}) = \sum_{pqrstu}^{\mathcal{B}} \phi_p(\mathbf{r})\phi_q(\mathbf{r}) V^{rs}_{pq} \Gamma^{tu}_{rs} \phi_t(\mathbf{r})\phi_u(\mathbf{r}), \tag{13}$$

$$n_{2,\Psi^{\mathcal{B}}_{\text{loc}}}(\mathbf{r}) = \sum_{rstu}^{\mathcal{B}} \phi_r(\mathbf{r})\phi_s(\mathbf{r}) \Gamma^{tu}_{rs} \phi_t(\mathbf{r})\phi_u(\mathbf{r}), \tag{14}$$

where the sums run over all the (occupied and virtual) orbitals generated by the basis set $\mathcal{B}$, $V^{rs}_{pq} = \langle pq|rs \rangle$ are the two-electron integrals, and Γ^{tu}_{rs} is the (opposite-spin) two-body reduced density matrix associated with a localizing wave function $\Psi^{\mathcal{B}}_{\text{loc}}$. The only purpose of the wave function $\Psi^{\mathcal{B}}_{\text{loc}}$ is to localize the otherwise nonlocal effective interaction obtained by projecting the Coulomb two-electron interaction in the basis set $\mathcal{B}$. The local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$ is very weakly dependent on this wave function $\Psi^{\mathcal{B}}_{\text{loc}}$ [the dependence on $\Psi^{\mathcal{B}}_{\text{loc}}$ of the numerator and denominator in Eq. (12) largely cancels], and it should be thought of as essentially depending only on the basis set $\mathcal{B}$. Importantly, in the CBS limit, the local range-separation parameter diverges, i.e., $\lim_{\mathcal{B}\to\text{CBS}} \mu^{\mathcal{B}}(\mathbf{r}) = +\infty$, independently of $\Psi^{\mathcal{B}}_{\text{loc}}$, and in this case, the form of the PBE-UEG and PBE-OT basis-set correction functionals ensures that they properly vanish, i.e., $\lim_{\mathcal{B}\to\text{CBS}} \bar{E}^{\mathcal{B}}_{\text{PBE-UEG}}[n_\uparrow, n_\downarrow] = 0$ and $\lim_{\mathcal{B}\to\text{CBS}} \bar{E}^{\mathcal{B}}_{\text{PBE-OT}}[n, n_2] = 0$.

Even though it is possible to perform the minimizations in Eqs. (7) and (8) using the PBE-UEG and PBE-OT basis-set correction functionals,[30] in practice, for energy calculations, a non-self-consistent approximation can safely be used, in which the basis-set correction functionals are added *a posteriori* to an approximation of the FCI energy calculated with the basis set $\mathcal{B}$. Here, as in our previous work,[14] we use the total SHCI energy $E^{\mathcal{B}}_{\text{SHCI}}$ in the basis set $\mathcal{B}$. We then have two basis-set-corrected SHCI energies. First, we have the SHCI + PBE-UEG energy

$$E^{\mathcal{B}}_{\text{SHCI+PBE-UEG}} = E^{\mathcal{B}}_{\text{SHCI}} + \bar{E}^{\mathcal{B}}_{\text{PBE-UEG}}\left[n_{\uparrow,\Phi^{\mathcal{B}}_{\text{HF}}}, n_{\downarrow,\Phi^{\mathcal{B}}_{\text{HF}}}\right], \tag{15}$$

where the PBE-UEG functional is evaluated at the spin-resolved densities obtained from the HF single-determinant wave function $\Phi^{\mathcal{B}}_{\text{HF}}$. We calculate the spin densities at the HF level only since the PBE-UEG functional is weakly dependent on the level at which the spin densities are calculated. Moreover, we choose also $\Psi^{\mathcal{B}}_{\text{loc}} = \Phi^{\mathcal{B}}_{\text{HF}}$ in Eq. (12) for calculating the local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$ that enters into the PBE-UEG functional. Second, we have the SHCI + PBE-OT energy

$$E^{\mathcal{B}}_{\text{SHCI+PBE-OT}} = E^{\mathcal{B}}_{\text{SHCI}} + \bar{E}^{\mathcal{B}}_{\text{PBE-OT}}\left[n_{\Psi^{\mathcal{B}}_{\text{SHCI}}}, n_{2,\Psi^{\mathcal{B}}_{\text{SHCI}}}\right], \tag{16}$$

where the PBE-OT functional is evaluated with the density and on-top pair density of the variational SHCI wave function $\Psi^{\mathcal{B}}_{\text{SHCI}}$ available at the end of the SHCI calculation. We use the SHCI wave function and not the HF single-determinant wave function because the PBE-OT functional is accurate only if it is evaluated at the on-top pair density obtained from an accurate multideterminant wave function. Moreover, we choose also $\Psi^{\mathcal{B}}_{\text{loc}} = \Psi^{\mathcal{B}}_{\text{SHCI}}$ for calculating the local range-separation parameter $\mu^{\mathcal{B}}(\mathbf{r})$ that enters into the PBE-OT functional, even though the use of an accurate multideterminant wave function for $\Psi^{\mathcal{B}}_{\text{loc}}$ is not really important here. Comparing the two basis-set corrections, the PBE-UEG basis-set correction is simpler to compute since it uses the HF single-determinant wave function, but the PBE-OT basis-set correction which uses the on-top pair density from the SHCI wave function should be more accurate.

IV. COMPUTATIONAL SETUP

Our computations employ the eCEPP pseudopotentials of Trail and Needs[40] and their associated aug-cc-pVnZ-eCEPP basis sets with $n = 2, 3, 4,$ and 5. These are abbreviated as DZ, TZ, QZ, and 5Z, or generically as nZ, in what follows. The molecular geometries and the experimental energies are the same as in the previous benchmark paper of Ref. 15.

We calculate the total energies in each of the four basis sets in the following three different ways:

1. SHCI only. We first perform HF calculations with the software PySCF[41] and then perform SHCI calculations with orbital optimization[33] using the ARROW package.[42] These results have previously appeared in Ref. 15.

2. SHCI with the PBE-UEG basis-set correction. The HF wave function is fed into QUANTUM PACKAGE[43] to obtain the basis-set correction. The amplitude of the dominant determinant in the SHCI wave function ranges from 0.92 to 0.96 for the metal atoms and from 0.83 to 0.91 for the oxide molecules. Therefore, we expect the HF spin densities to be a reasonable approximation to the SHCI spin densities. The HF on-top pair density is not accurate, but the on-top pair density obtained from the UEG that has the HF spin densities is reasonably accurate.

3. SHCI with the PBE-OT basis-set correction. We perform SHCI variational calculations for several different values of the threshold ϵ_1 and use the corresponding two-body reduced density matrices for the basis-set correction using the QUANTUM PACKAGE.[43] As the size of the variational wave function increases, the on-top pair density decreases, leading to decreasing magnitudes of the PBE-OT basis-set correction. We plot the basis-set correction vs E_V and use a quadratic function to extrapolate to the SHCI total energy limit, $E_V \to E_{\text{SHCI}}$, which is then taken as the final PBE-OT basis-set correction for the given system and basis set.

In two of the above three schemes, we also extrapolate the total energies to the CBS limit. In earlier work by some of the present authors,[15] the SHCI-only energies were extrapolated to the CBS limit using separate extrapolations for the HF energy and the correlation energy,[44–46]

$$E_{\mathrm{HF}}^{\mathrm{CBS}} = E_{\mathrm{HF}}^{nZ} - a\,\exp(-bn), \qquad (17)$$

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = E_{\mathrm{corr}}^{nZ} - cn^{-3}. \qquad (18)$$

In this paper, the HF energy is still extrapolated in the same way using $n = 3, 4$, and 5, but the SHCI-only correlation energy is extrapolated using the expression

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = E_{\mathrm{corr}}^{nZ} - \frac{c}{n^3 + d} \qquad (19)$$

because we find that it gives a far better fit for the pseudopotentials and associated basis sets used in this study. Since Eq. (18) does not give a good fit, one would use only the larger basis sets $n = 4$ and 5 to get a reasonable extrapolation, whereas with Eq. (19) $n = 2, 3, 4$, and 5 can be used. [Even using just $n = 4$ and 5, Eq. (18) gives extrapolated energies that are too high, as shown in Fig. 4 of the supplementary material. For the heaviest system studied, CuO, the improved fit using Eq. (19) and $n = 2, 3, 4$, and 5 makes the CBS total energy 6 mHa lower than the fit using Eq. (18) and $n = 4$ and 5. A detailed comparison of the various fits can be found in the supplementary material.]

In the SHCI + PBE-OT scheme, we use a single exponential function

$$E_{\mathrm{SHCI+PBE-OT}}^{\mathrm{CBS}} = E_{\mathrm{SHCI+PBE-OT}}^{nZ} - a\,\exp(-bn) \qquad (20)$$

with $n = 3, 4$, and 5, since we expect an exponential convergence of the SHCI + PBE-OT energy with respect to n.[47] (In the supplementary material, we also show the SHCI + PBE-OT energies extrapolated to the CBS limit using $n = 2, 3$, and 4.)

Energies from the SHCI + PBE-UEG scheme are not extrapolated as the SHCI + PBE-UEG energies are nonmonotonic with basis size for many of the systems, as shown in Sec. V. Note, however, that our SHCI + PBE-UEG calculations on the G2 set[14] had a monotonic dependence on the basis size, and so we were able to extrapolate those energies.

V. RESULTS

In this section, we present energies for the seven monoxide molecules, the eight constituent atoms, and the corresponding ions using the three schemes described in Sec. IV. The results are shown for each of the four basis sets DZ, TZ, QZ, and 5Z as well as for the extrapolated CBS limit. The total energies, first ionization energies (for the eight atoms), and dissociation energies (for the seven monoxide molecules) are reported in Table I.

In what follows, we use as reference values the CBS results obtained from the SHCI + PBE-OT scheme. We think it likely

TABLE I. Total, ionization (for the atoms), and dissociation energies (for the monoxide molecules) obtained from the three schemes in the CBS limit. The SHCI + PBE-OT values (emphasized in bold) are used as reference values in Sec. V for comparison. Units: Ha.

| System	Total energy			Ionization/dissociation energy		
	SHCI	SHCI + PBE-UEG	SHCI + PBE-OT	SHCI	SHCI + PBE-UEG	SHCI + PBE-OT
O	−15.8477	−15.8490	**−15.8478**	0.5000	0.5009	**0.5004**
Sc	−46.4921	−46.4966	**−46.4939**	0.2409	0.2412	**0.2411**
Ti	−58.0068	−58.0117	**−58.0088**	0.2508	0.2510	**0.2510**
V	−71.2371	−71.2412	**−71.2381**	0.2473	0.2475	**0.2475**
Cr	−86.8078	−86.8111	**−86.8080**	0.2488	0.2489	**0.2489**
Mn	−104.1606	−104.1648	**−104.1617**	0.2726	0.2732	**0.2730**
Fe	−123.7718	−123.7784	**−123.7745**	0.2901	0.2905	**0.2903**
Cu	−197.6236	−197.6388	**−197.6328**	0.2832	0.2836	**0.2839**
O^+	−15.3477	−15.3481	**−15.3475**	⋯	⋯	⋯
Sc^+	−46.2511	−46.2554	**−46.2528**	⋯	⋯	⋯
Ti^+	−57.7560	−57.7607	**−57.7578**	⋯	⋯	⋯
V^+	−70.9898	−70.9938	**−70.9907**	⋯	⋯	⋯
Cr^+	−86.5591	−86.5622	**−86.5591**	⋯	⋯	⋯
Mn^+	−103.8879	−103.8917	**−103.8887**	⋯	⋯	⋯
Fe^+	−123.4817	−123.4879	**−123.4843**	⋯	⋯	⋯
Cu^+	−197.3404	−197.3552	**−197.3489**	⋯	⋯	⋯
ScO	−62.5983	−62.6040	**−62.6002**	0.2585	0.2584	**0.2585**
TiO	−74.1103	−74.1168	**−74.1128**	0.2557	0.2561	**0.2562**
VO	−87.3302	−87.3363	**−87.3318**	0.2453	0.2461	**0.2458**
CrO	−102.8299	−102.8358	**−102.8316**	0.1744	0.1758	**0.1758**
MnO	−120.1529	−120.1596	**−120.1550**	0.1446	0.1458	**0.1455**
FeO	−139.7769	−139.7861	**−139.7814**	0.1573	0.1587	**0.1591**
CuO	−213.5821	−213.5980	**−213.5903**	0.1108	0.1102	**0.1097**
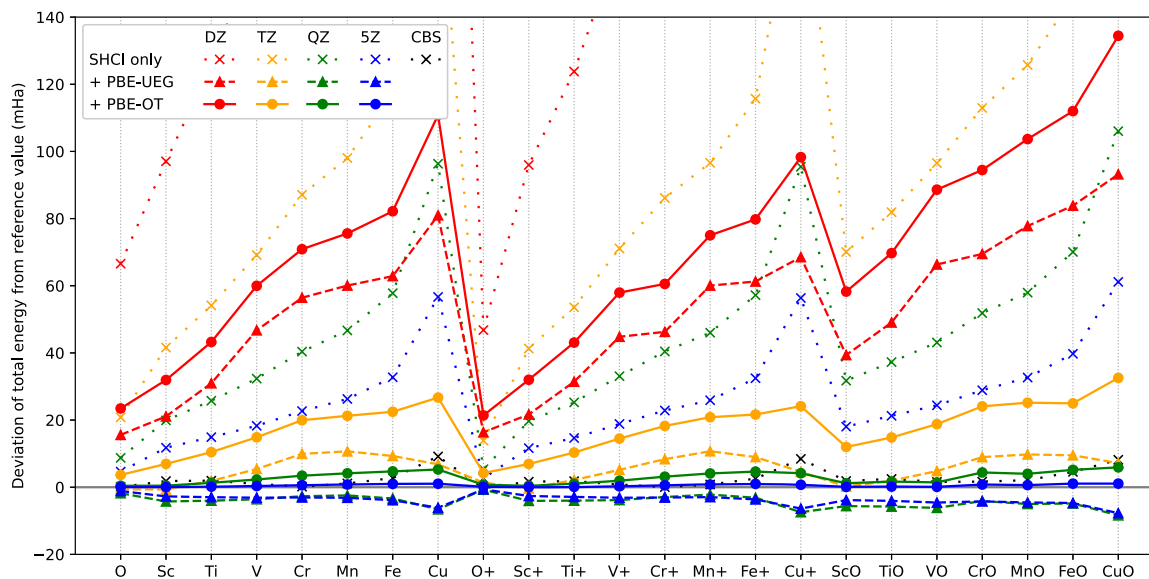
FIG. 1. Deviations of the total energies from reference (CBS SHCI + PBE-OT) values. Some of the SHCI-only values for the smaller basis sets are off the scale of the plot.

that this scheme gives the most accurate energies since the SHCI + PBE-OT scheme employs a more accurate on-top pair density than the SHCI + PBE-UEG scheme. In the supplementary material, the quantum Monte Carlo (QMC) energies for the oxygen atom also corroborate the selection of SHCI + PBE-OT as the reference. However, we also acknowledge that this choice of reference values is not certain. In our earlier work on the G2 set of 55 molecules (Ref. 14), we employed only the SHCI-only and the SHCI + PBE-UEG schemes since those molecules are sufficiently weakly correlated. In that study, we found that although for each of the finite basis sets the SHCI + PBE-UEG energies agreed considerably better with experiment than the SHCI-only energies, the CBS energies from SHCI-only gave slightly better agreement than those from SHCI + PBE-UEG.

A. Total energies

Figure 1 shows deviations of total energies from the reference values. Compared to the SHCI-only results, both basis-set correction schemes significantly reduce the basis incompleteness error. For each system in a particular basis, the correction from PBE-UEG is consistently larger in magnitude than from PBE-OT. This makes the PBE-UEG-corrected values closer to the CBS limit for smaller basis sets, but for larger basis sets, PBE-UEG tends to overcorrect. In fact, for many of the systems shown here, the PBE-UEG-corrected energies for 5Z are higher than for QZ, reversing the trend from DZ to QZ. Similar nonmonotonicity of the corrected values with basis size has been observed in previous studies as well.[30]

We have separately verified the total energy in the CBS limit by performing QMC calculations on the oxygen atom (see the supplementary material). The QMC total energy is in good agreement with the CBS SHCI-only and SHCI + PBE-OT energies and

differs considerably from the SHCI + PBE-UEG energy. The energy difference between the former two schemes may be viewed as an estimate of the uncertainty in the CBS energies. Table II reports the mean absolute deviations (MADs) from the reference values for the different basis sets and methods.

B. Ionization energies

Figure 2 shows the convergence of the first ionization energies of the eight atoms with respect to basis size. For the metal atoms, although the ionization energies obtained with the DZ basis sets for Cr and Cu show large errors, all three schemes are almost converged to within chemical accuracy at the level of TZ and converged to far better than chemical accuracy at the level of 5Z. The MADs are reported in Table III. Most of the MADs are more than an order of magnitude smaller than those for the total energies because of a large cancellation of errors between the atomic and ionic energies.

Accurate experimental ionization energies are also available for these systems. In Fig. 3, we compare the theoretical ionization energies obtained from the three schemes (i.e., extrapolated energies from the SHCI-only and SHCI + PBE-OT schemes and the 5Z energies from the SHCI + PBE-UEG scheme) to experiment. Much better agreement than chemical accuracy is obtained for all the

TABLE II. Mean absolute deviations of total energies from reference (CBS SHCI + PBE-OT) values for different basis sets. Units: mHa.

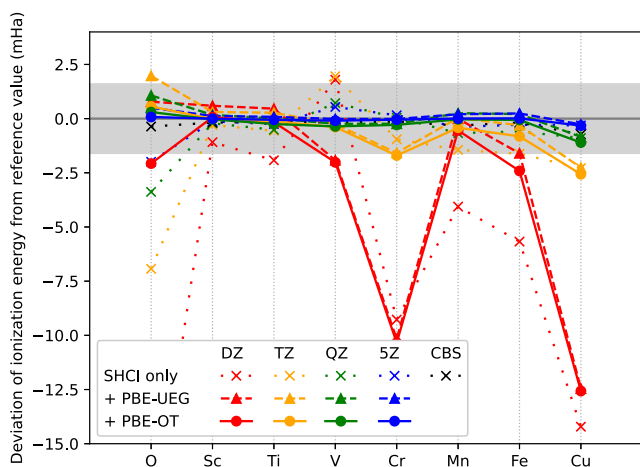|	DZ	TZ	QZ	5Z	CBS
SHCI	217.67	95.04	45.58	26.06	2.47
SHCI + PBE-UEG	52.35	5.63	4.21	3.70	···
SHCI + PBE-OT	70.75	17.37	2.87	0.51	0.00

FIG. 2. Deviations of the ionization energies from reference (CBS SHCI + PBE-OT) values. The shaded area indicates chemical accuracy, i.e., 1 kcal/mol.

TABLE III. Mean absolute deviations of ionization energies from reference (CBS SHCI + PBE-OT) values for different basis sets. Units: mHa.

	DZ	TZ	QZ	5Z	CBS
SHCI	7.23	2.01	0.88	0.50	0.28
SHCI + PBE-UEG	3.50	0.87	0.38	0.18	⋯
SHCI + PBE-OT	3.77	0.82	0.30	0.08	0.00

atoms and all three schemes. Remarkably, the largest deviation of the energies obtained from either of the two correction schemes and experiment is only 0.01 eV. The MADs from experiment are 10.8, 5.9, and 4.3 meV for SHCI-only, SHCI + PBE-UEG, and
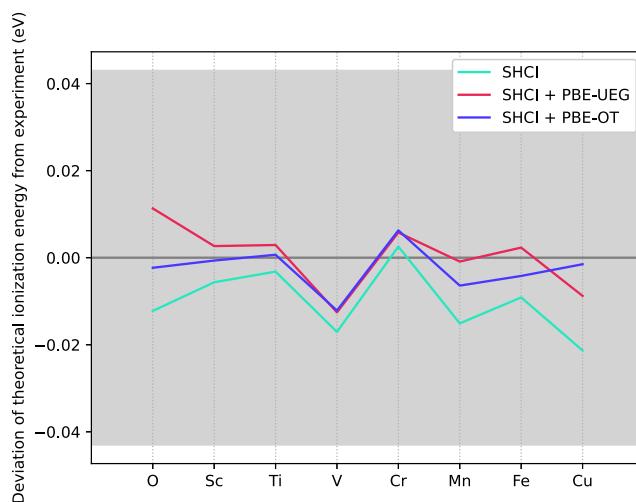

FIG. 3. Comparison of theoretical ionization energies with experiments. The theoretical values are obtained from the three schemes: SHCI only (CBS), SHCI + PBE-UEG (5Z), and SHCI + PBE-OT (CBS). The shaded area indicates chemical accuracy, i.e., 1 kcal/mol.

SHCI + PBE-OT, respectively. The high level of agreement of the SHCI + PBE-OT energies with experiment provides further support for using SHCI + PBE-OT energies as reference values and speaks to the accuracy of the experiments as well.

C. Dissociation energies

Figure 4 shows the dissociation energies in the different basis sets for the three schemes. As expected, the basis incompleteness
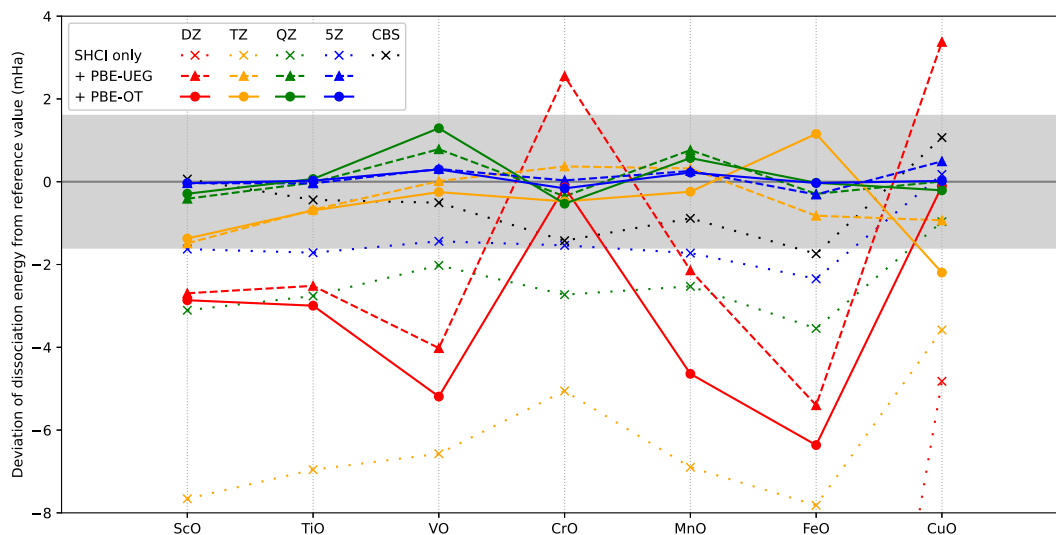

FIG. 4. Deviations of the dissociation energies from reference (CBS SHCI + PBE-OT) values. The shaded area indicates chemical accuracy, i.e., 1 kcal/mol. Some of the SHCI-only values for the smaller basis sets are off the scale of the plot.
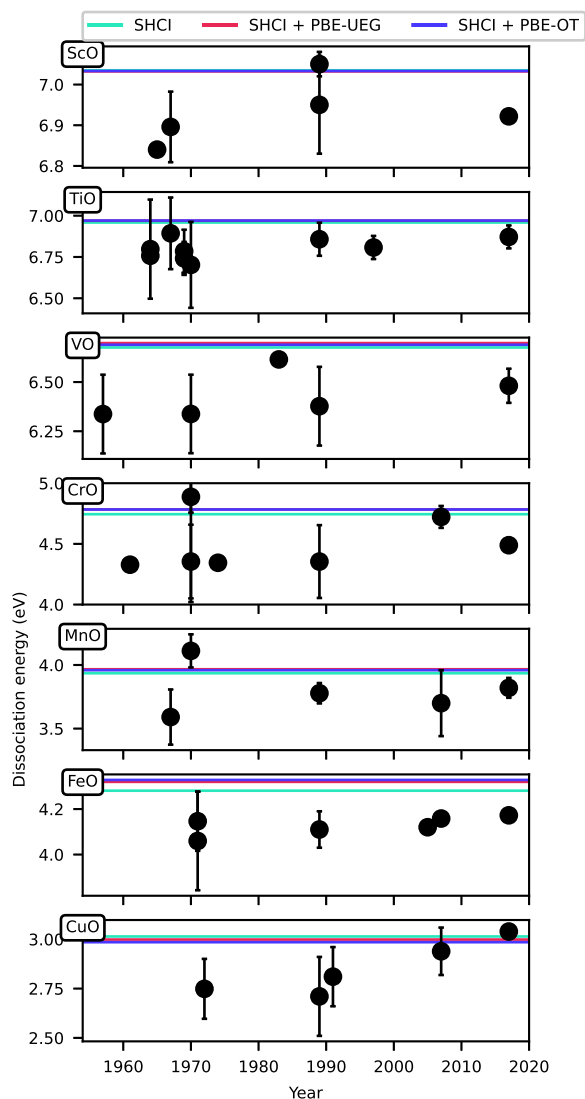
TABLE IV. Mean absolute deviations of dissociation energies from reference (CBS SHCI + PBE-OT) values for different basis sets. Units: mHa.

	DZ	TZ	QZ	5Z	CBS
SHCI	15.52	6.36	2.52	1.51	0.88
SHCI + PBE-UEG	3.24	0.66	0.38	0.21	· · ·
SHCI + PBE-OT	3.18	0.91	0.43	0.11	0.00

errors in energy differences are around one order of magnitude smaller than in the individual total energies themselves. For both basis-set correction schemes, at the level of the TZ basis set, almost all systems are converged to within chemical accuracy (with the

single exception of CuO in the PBE-OT scheme). Table IV reports the MADs for different basis sets and methods.

In Fig. 5, we take the final dissociation energies obtained from the three schemes and compare with the experimental values reported over the years (see Ref. 15 and the references therein). These experimental values have been corrected for zero-point energy contributions but not for relativistic effects since the pseudopotentials we use incorporate scalar-relativistic effects. Unlike the ionization energies, the experimental dissociation energies have large uncertainties and differ significantly among themselves. For all systems studied, the theoretical values from our three schemes agree with each other to much higher precision than the differences between the various experimental values.

VI. CONCLUSIONS

In prior work,[15] we have shown that the SHCI method can be used to calculate near exact energies for transition metal atoms, ions, and oxides in basis sets up to 5Z. There, the SHCI energies were used as the reference for testing the accuracy of 19 other electronic-structure methods for each of the basis sets. The CBS energies were also estimated by extrapolation using the formulas in Eqs. (17) and (18). In this paper, we have shown that a more accurate extrapolation is possible using Eqs. (17) and (19). More importantly, we have shown that density-based basis-set corrections can be used to estimate the basis incompleteness error of a finite basis set and speed up convergence to the CBS limit. Two different schemes have been applied, namely, PBE-UEG and PBE-OT, differing in the functional used in the calculation. Both methods produce ionization and dissociation energies converged to within chemical accuracy of the CBS limit with only the TZ basis set, which avoids the high cost of going to larger basis sets using a wave-function method such as SHCI. We have also compared the ionization and dissociation energies from the two correction schemes as well as the uncorrected extrapolations to experimental values reported over the years. Remarkably, for ionization energies, the largest deviation of the experimental energies from the energies obtained by either of the two correction schemes is only 0.01 eV. Our computed dissociation energies are converged to much higher precision than the experimental uncertainties.

SUPPLEMENTARY MATERIAL

See the supplementary material for detailed comparisons of basis-set extrapolations for SHCI total energies and SHCI + PBE-OT energies as well as QMC results on the oxygen atom.

FIG. 5. Comparison of theoretical dissociation energies with experiments. The theoretical values obtained from the three schemes—SHCI only (CBS), SHCI + PBE-UEG (5Z), and SHCI + PBE-OT (CBS)—are converged to much higher precision than the experimental uncertainties.

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare no conflicts of interest.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1] B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).

[2] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).

[3] F. A. Evangelista, J. Chem. Phys. **140**, 124114 (2014).

[4] A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, J. Comput. Chem. **37**, 1866 (2016).

[5] Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, J. Chem. Phys. **147**, 034101 (2017).

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

[7] D. Hait, N. M. Tubman, D. S. Levine, K. B. Whaley, and M. Head-Gordon, J. Chem. Theory Comput. **15**, 5370 (2019).

[8] P.-F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama, and D. Jacquemin, J. Chem. Theory Comput. **16**, 1711 (2020).

[9] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).

[10] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).

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

[12] J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, J. Chem. Phys. **149**, 214110 (2018).

[13] J. Li, Y. Yao, A. Holmes, M. Otten, S. Sharma, and C. J. Umrigar, Phys. Rev. Res. **2**, 012015(R) (2020).

[14] Y. Yao, E. Giner, J. Li, J. Toulouse, and C. J. Umrigar, J. Chem. Phys. **153**, 124117 (2020).

[15] K. T. Williams, Y. Yao, J. Li, L. Chen, H. Shi, M. Motta, C. Niu, U. Ray, S. Guo, R. J. Anderson, J. Li, L. N. Tran, C.-N. Yeh, B. Mussard, S. Sharma, F. Bruneval, M. van Schilfgaarde, G. H. Booth, G. K.-L. Chan, S. Zhang, E. Gull, D. Zgid, A. Millis, C. J. Umrigar, and L. K. Wagner, Phys. Rev. X **10**, 011041 (2020).

[16] The largest differences between the CBS345 and CBS45 values in Table V of the supplementary material of Ref. 15 occur for the heaviest system, CuO, and are 8 mHa. However, the better extrapolation formula of Eq. (19) in this paper results in energies that are 6 mHa lower, and our best estimate using the PBE-OT basis set correction (described later) followed by extrapolation results in a further 8 mHa lower energy. The differences for the lighter transition metal systems are much smaller than this.

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

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

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

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

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

[22] J. G. Hill and K. A. Peterson, J. Chem. Theory Comput. **8**, 518 (2012).

[23] D. H. Bross, J. G. Hill, H.-J. Werner, and K. A. Peterson, J. Chem. Phys. **139**, 094302 (2013).

[24] M. K. Kesharwani and J. M. L. Martin, Theor. Chem. Acc. **133**, 1452 (2014).

[25] D. P. Tew, J. Chem. Phys. **145**, 074103 (2016).

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

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

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

[29] E. Giner, A. Scemama, P.-F. Loos, and J. Toulouse, J. Chem. Phys. **152**, 174104 (2020).

[30] E. Giner, D. Traore, B. Pradines, and J. Toulouse, J. Chem. Phys. **155**, 044109 (2021).

[31] Since the absolute values of c_i for the most important determinants tend to go down as more determinants are included in the wave function, a somewhat better selection of determinants is obtained by using a larger value of ϵ_1 in the initial iterations.

[32] E. R. Davidson, Comput. Phys. Commun. **53**, 49 (1989).

[33] Y. Yao and C. J. Umrigar, J. Chem. Theory Comput. **17**, 4183 (2021).

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

[35] E. H. Lieb, Int. J. Quantum Chem. **24**, 243 (1983).

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

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

[38] P. Gori-Giorgi and J. P. Perdew, Phys. Rev. B **64**, 155102 (2001).

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

[40] J. R. Trail and R. J. Needs, J. Chem. Phys. **146**, 204107 (2017).

[41] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).

[42] See https://github.com/QMC-Cornell/shci for the source code of the ARROW package.

[43] 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).

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

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

[46] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

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