

RESEARCH ARTICLE | JUNE 17 2022

## Performance of a one-parameter correlation factor for transcorrelation: Study on a series of second row atomic and molecular systems

Werner Dobrutz  ; Aron J. Cohen  ; Ali Alavi  ; Emmanuel Giner  



*J. Chem. Phys.* 156, 234108 (2022)

<https://doi.org/10.1063/5.0088981>



View  
Online



Export  
Citation

CrossMark



The Journal of Chemical Physics  
Special Topic: Adhesion and Friction

Submit Today!

AIP  
Publishing

# Performance of a one-parameter correlation factor for transcorrelation: Study on a series of second row atomic and molecular systems

Cite as: J. Chem. Phys. 156, 234108 (2022); doi: 10.1063/5.0088981

Submitted: 21 February 2022 • Accepted: 26 May 2022 •

Published Online: 17 June 2022



View Online



Export Citation



CrossMark

Werner Dobrutz,<sup>1,2,a</sup> Aron J. Cohen,<sup>1,3</sup> Ali Alavi,<sup>1,4</sup> and Emmanuel Giner<sup>5,b</sup>

## AFFILIATIONS

<sup>1</sup> Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany

<sup>2</sup> Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Kemigården 4, 41258 Gothenburg, Sweden

<sup>3</sup> DeepMind, 6 Pancras Square, London N1C 4AG, United Kingdom

<sup>4</sup> Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

<sup>5</sup> Laboratoire de Chimie Théorique, Sorbonne Université and CNRS, F-75005 Paris, France

<sup>a</sup>E-mail: [dobrutz@chalmers.se](mailto:dobrutz@chalmers.se)

<sup>b</sup>Author to whom correspondence should be addressed: [emmanuel.giner@lct.jussieu.fr](mailto:emmanuel.giner@lct.jussieu.fr)

## ABSTRACT

In this work, we investigate the performance of a recently proposed transcorrelated (TC) approach based on a single-parameter correlation factor [E. Giner, *J. Chem. Phys.* **154**, 084119 (2021)] for systems involving more than two electrons. The benefit of such an approach relies on its simplicity as efficient numerical-analytical schemes can be set up to compute the two- and three-body integrals occurring in the effective TC Hamiltonian. To obtain accurate ground state energies within a given basis set, the present TC scheme is coupled to the recently proposed TC-full configuration interaction quantum Monte Carlo method [Cohen *et al.*, *J. Chem. Phys.* **151**, 061101 (2019)]. We report ground state total energies on the Li–Ne series, together with their first cations, computed with increasingly large basis sets and compare to more elaborate correlation factors involving electron–electron–nucleus coordinates. Numerical results on the Li–Ne ionization potentials show that the use of the single-parameter correlation factor brings on average only a slightly lower accuracy (1.2 mH) in a triple-zeta quality basis set with respect to a more sophisticated correlation factor. However, already using a quadruple-zeta quality basis set yields results within chemical accuracy to complete basis set limit results when using this novel single-parameter correlation factor. Calculations on the H<sub>2</sub>O, CH<sub>2</sub>, and FH molecules show that a similar precision can be obtained within a triple-zeta quality basis set for the atomization energies of molecular systems.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0088981>

## I. INTRODUCTION

At the heart of quantum chemistry lies the accurate description of the electronic structure of molecular systems, which is a very challenging task since the corresponding mathematical problem to be solved scales exponentially with the system size. Wave function theory (WFT) aims at solving the Schrödinger equation for a general molecular system and provides a systematic way of improving the accuracy of the computed properties following a twofold path: (i) improving the quality of the wave function in a given basis to

get as close as possible to the full-configuration interaction (FCI) and (ii) improving the quality of the one-electron basis set used to project the Schrödinger equation. The exact properties of the system would be obtained with the FCI wave function in a complete basis set (CBS). There exist many different flavors of wave function *Ansätze*, which approximate the FCI wave function and energy, and they all have an unfavorable computational scaling with the system size and, most importantly, with the size of the basis set in common. Therefore, a major drawback of WFT is the slow convergence of the results with respect to the basis set size, which mainly

originates from the poor description of the two-body density matrix near the electron–electron coalescence point (i.e.,  $r_{12} \approx 0$ ).

A central idea shared by the theories aiming to improve the convergence of WFT with respect to the basis set is related to the so-called electron–electron cusp-conditions derived by Kato:<sup>1</sup> one multiplies the cusp-less wave function developed in an incomplete basis set by a correlation factor explicitly depending on the  $r_{12}$  coordinate, which restores the cusp-conditions. Besides the cusp, the most important role of the correlation factor is to lower the probability of finding two electrons near one another, which is often referred to as “digging” the short-range part of the Coulomb hole.

There are mainly three classes of theories dealing with a correlation factor: (i) F12 theory,<sup>2–7</sup> where one projects out the effect of the correlation factor from the incomplete basis set used to compute the cusp-less wave function, (ii) variational Monte Carlo (VMC) methods,<sup>8</sup> where the full effect of the correlation factor is retained in the wave function and all parameters are variationally optimized, and (iii) the transcorrelated (TC) theory,<sup>9–11</sup> where the effect of the full correlation factor is incorporated through a non-Hermitian effective Hamiltonian. All these three theories have been shown to strongly reduce the basis set convergence problem of WFT.

The main advantage of the TC theory is that it combines favorable aspects of both VMC and F12: (i) since only up to effective three-electron terms are needed (compared to the N-body terms of VMC), usual post-Hartree–Fock methods can be designed to solve the TC Hamiltonian, (ii) no more than  $\mathbb{R}^6$  integrals are needed, and (iii) compact wave function can be obtained because the full correlation factor is taken into account. Despite these attractive features, the two main drawbacks of the TC theory are that (i) the non-Hermitian nature of the TC operator causes the loss of variationality, and (ii) the three-body terms generate an  $N^6$  tensor, where  $N$  is the number of basis set functions, which becomes rapidly prohibitive to store during calculations. Nevertheless, because it originates from a similarity transformation, the exact eigenvalues are obtained when reaching the CBS limit, which suggests that the loss of the variational property in the TC method could be a signature of an unbalanced treatment between the Slater determinant (SD) expansion and correlation factor in the wave function *Ansatz*. Regarding the functional form of the cusp-less wave function and of the correlation factors, the seminal work of Boys and Handy<sup>10,11</sup> proposed to optimize both the orbitals of a single SD and a sophisticated correlation factor. Then, Ten-No<sup>12</sup> proposed a significant change of paradigm since he used the combination of a rather simple universal correlation factor whose shape was optimized for the range of valence electrons, and a rather sophisticated *Ansatz* for the wave function (Møller–Plesset at the second order in Refs. 12 and 13 and linearized coupled-cluster in Ref. 14). The works of Ten-No have shown that TC theories, such as transcorrelated second order Møller–Plesset perturbation theory (TC-MP2), converge faster toward the exact energies with respect to the basis set compared to their non-TC counterpart, like regular MP2. Nevertheless, it should be mentioned that because the correlation factor was optimized for valence electrons, the use of basis sets explicitly optimized for core electrons (e.g., the cc-pCVXZ family) is mandatory in order to maintain a sensible value for the energy in all-electron calculations.

More recently, Cohen *et al.*<sup>15</sup> applied the TC equations with an elaborate correlation factor and proposed to use the full configuration interaction quantum Monte Carlo (FCIQMC) method to obtain the exact ground state energy and the corresponding right eigenvector of the TC Hamiltonian in a given basis set. In their work,<sup>15</sup> the authors used the Jastrow factors of Schmidt and Moskowitz<sup>16</sup> optimized with VMC for the He–Ne neutral series, which explicitly take into account electron–electron–nucleus correlation effects. Their work also demonstrated the strong compression of the right eigenvector for both *ab initio* and model Hamiltonians.<sup>17</sup> Similar applications of the Gutzwiller *Ansatz*<sup>18,19</sup> were recently reported by Baiardi and Reiher<sup>20</sup> using the density matrix renormalization group, and various methods based on the TC approach have been used to reduce the resource requirements for accurate electronic structure calculations on state-of-the-art quantum computing hardware.<sup>21–25</sup>

Recently, one of the present authors introduced a single-parameter correlation factor<sup>26</sup> inspired by range-separated density functional theory (RS-DFT). The main idea developed in this work was to find a mapping between the leading order terms in  $1/r_{12}$  of the effective scalar potential obtained in the TC equations and the non-divergent long-range interaction  $\text{erf}(\mu r_{12})/r_{12}$  used in RS-DFT. The correlation factor obtained with such a procedure has an explicit analytical form, which depends on a single parameter  $\mu$ : the lower the  $\mu$ , the deeper the correlation hole dug by the correlation factor, and in the  $\mu \rightarrow \infty$  limit, the effect of the correlation factor vanishes. Preliminary tests on atomic and molecular two-electron systems have shown that this TC framework also improves the convergence of the energy, and a good value of the parameter  $\mu$  could be systematically obtained with nothing more than the knowledge of the Hartree–Fock (HF) density. The advantage of this simple correlation factor is that the corresponding TC Hamiltonian has a rather simple analytical form for which the two- and three-body integrals can be very efficiently obtained using a mixed numerical/analytical scheme.

The aim of this work is to study how this relatively simple correlation factor performs for systems with more than two electrons. In order to be able to eliminate any source of errors within a basis set, we use the FCIQMC approach to obtain the exact right eigenvector in a given basis set. We are then able to compare the results obtained with the more sophisticated correlation factor used in the recent work of Cohen *et al.*<sup>15</sup>

The remainder of this article is organized as follows: in Sec. II, we recap the main equations of the TC theory together with the explicit form of the TC Hamiltonian obtained in Ref. 26. Then, in Sec. III B 1, we investigate the sensitivity of the present approach with the quality of description of core electrons, in Sec. III B 2, the dependence of the results on the value of the parameter  $\mu$  is studied, and in Sec. III B 3, we investigate a possible approximation to the three-body term of the TC approach within the *Ansatz* considered here. In Sec. III C, we report total energies results of the neutral and first cation species of the Li–Ne series as a function of increasing basis set size. We also compare the quality of the ionization potentials (IPs) obtained with the present approach with the existing literature. We conclude this study by investigating the convergence of both total energies and atomization energies (AEs) on the CH<sub>2</sub>, FH, and H<sub>2</sub>O molecules.

## II. THEORY

### A. General equations and concepts of TC theory

The general form of the transcorrelated Hamiltonian for a symmetric correlation factor  $u(\mathbf{r}_1, \mathbf{r}_2)$  is given by

$$\begin{aligned}\tilde{H}[u] &\equiv e^{-\hat{\tau}_u} \hat{H} e^{\hat{\tau}_u} \\ &= H + [H, \hat{\tau}_u] + \frac{1}{2} [[H, \hat{\tau}_u], \hat{\tau}_u],\end{aligned}\quad (1)$$

where  $\hat{\tau}_u = \sum_{i < j} u(\mathbf{r}_i, \mathbf{r}_j)$  and  $\hat{H} = -\sum_i \frac{1}{2} \nabla_i^2 + v(\mathbf{r}_i) + \sum_{i < j} 1/r_{ij}$ . Equation (1) leads to the following transcorrelated Hamiltonian:

$$\tilde{H}[u] = H - \sum_{i < j} \hat{K}[u](\mathbf{r}_i, \mathbf{r}_j) - \sum_{i < j < k} \hat{L}[u](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k), \quad (2)$$

where the effective two- and three-body operators  $\hat{K}[u](\mathbf{r}_1, \mathbf{r}_2)$  and  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  are defined as

$$\begin{aligned}\hat{K}[u](\mathbf{r}_1, \mathbf{r}_2) &= \frac{1}{2} \left( \Delta_1 u(\mathbf{r}_1, \mathbf{r}_2) + \Delta_2 u(\mathbf{r}_1, \mathbf{r}_2) \right. \\ &\quad \left. + (\nabla_1 u(\mathbf{r}_1, \mathbf{r}_2))^2 + (\nabla_2 u(\mathbf{r}_1, \mathbf{r}_2))^2 \right) \\ &\quad + \nabla_1 u(\mathbf{r}_1, \mathbf{r}_2) \cdot \nabla_1 + \nabla_2 u(\mathbf{r}_1, \mathbf{r}_2) \cdot \nabla_2,\end{aligned}\quad (3)$$

and

$$\begin{aligned}\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) &= \nabla_1 u(\mathbf{r}_1, \mathbf{r}_2) \cdot \nabla_1 u(\mathbf{r}_1, \mathbf{r}_3) \\ &\quad + \nabla_2 u(\mathbf{r}_2, \mathbf{r}_1) \cdot \nabla_2 u(\mathbf{r}_2, \mathbf{r}_3) \\ &\quad + \nabla_3 u(\mathbf{r}_3, \mathbf{r}_1) \cdot \nabla_3 u(\mathbf{r}_3, \mathbf{r}_2).\end{aligned}\quad (4)$$

In practice, the TC Hamiltonian is projected into a basis set  $\mathcal{B}$ ,

$$\tilde{H}[u]^{\mathcal{B}} = P^{\mathcal{B}} \tilde{H}[u] P^{\mathcal{B}}, \quad (5)$$

where  $P^{\mathcal{B}}$  is the projector onto the  $N_e$ -electron Hilbert space spanned by a given basis set  $\mathcal{B}$ , where  $N_e$  is the number of electrons. Using real-valued orthonormal spatial molecular orbitals (MOs)  $\{\phi_i(\mathbf{r})\}$ ,  $\tilde{H}[u]^{\mathcal{B}}$  can be written in a second-quantized form as

$$\begin{aligned}\tilde{H}[u]^{\mathcal{B}} &= \sum_{i,j \in \mathcal{B}} \sum_{\sigma=\uparrow,\downarrow} h_{ij} a_{j,\sigma}^\dagger \hat{a}_{i,\sigma} + \frac{1}{2} \sum_{i,j,k,l \in \mathcal{B}} \sum_{\sigma,\lambda=\uparrow,\downarrow} \left( \tilde{W}_{ij}^{kl} - K_{ij}^{kl} \right) a_{k,\sigma}^\dagger a_{l,\lambda}^\dagger \hat{a}_{j,\lambda} \hat{a}_{i,\sigma} \\ &\quad - \frac{1}{6} \sum_{i,j,m,k,l,n \in \mathcal{B}} \sum_{\sigma,\lambda,\kappa=\uparrow,\downarrow} L_{ijm}^{kl} a_{k,\sigma}^\dagger a_{l,\lambda}^\dagger a_{n,\kappa}^\dagger \hat{a}_{m,\kappa} \hat{a}_{j,\lambda} \hat{a}_{i,\sigma},\end{aligned}\quad (6)$$

where  $h_{ij}$  are the usual one-electron integrals,  $\tilde{W}_{ij}^{kl}$  are the usual two-electron integrals,  $K_{ij}^{kl}$  are the two-electron integrals corresponding to the effective two-body operator  $\hat{K}[u](\mathbf{r}_1, \mathbf{r}_2)$  operator,

$$K_{ij}^{kl} = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_k(\mathbf{r}_1) \phi_l(\mathbf{r}_2) \hat{K}[u](\mathbf{r}_1, \mathbf{r}_2) \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2), \quad (7)$$

and  $L_{ijm}^{kl}$  are the three-electron integrals corresponding to the effective three-body operator  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ ,

$$\begin{aligned}L_{ijm}^{kl} &= \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 \phi_k(\mathbf{r}_1) \phi_l(\mathbf{r}_2) \phi_m(\mathbf{r}_3) \\ &\quad \times \hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \phi_i(\mathbf{r}_1) \phi_j(\mathbf{r}_2) \phi_n(\mathbf{r}_3).\end{aligned}\quad (8)$$

The ground state eigenvalue and the associated right eigenvector fulfill the eigenvalue equation,

$$\tilde{H}[u]^{\mathcal{B}} |\Phi_0^{\mathcal{B}}[u]\rangle = \tilde{E}_0^{\mathcal{B}}[u] |\Phi_0^{\mathcal{B}}[u]\rangle, \quad (9)$$

and because of the properties of the similarity transformation, the exact ground state energy  $E_0$  is recovered in the CBS limit,

$$\lim_{\mathcal{B} \rightarrow \text{CBS}} \tilde{E}_0^{\mathcal{B}}[u] = E_0, \quad (10)$$

for all correlation factors  $u(\mathbf{r}_i, \mathbf{r}_j)$  chosen to obtain  $\tilde{H}[u]^{\mathcal{B}}$ . If  $u(\mathbf{r}_i, \mathbf{r}_j)$  is properly chosen, one expects a fast convergence of  $\tilde{E}_0^{\mathcal{B}}[u]$  toward  $E_0$ . Because of the loss of the variational principle of  $\tilde{E}_0^{\mathcal{B}}[u]$ , due to the non-Hermitian character of  $\tilde{H}[u]^{\mathcal{B}}$ , this convergence is not guaranteed to be monotonic as in the usual WFT calculations, and  $\tilde{E}_0^{\mathcal{B}}[u]$  can be below the exact ground state energy.

### B. One-parameter TC Hamiltonian: $\tilde{H}[\mu]$

Recently, one of the present authors<sup>26</sup> has introduced a one-parameter correlation factor  $u(r_{12}, \mu)$  based on a mapping between the  $r_{12} \approx 0$  limit of the TC Hamiltonian and the range-separated DFT effective Hamiltonian. The explicit form of  $u(r_{12}, \mu)$  derived in Ref. 26 reads as

$$u(r_{12}, \mu) = \frac{1}{2} r_{12} (1 - \text{erf}(\mu r_{12})) - \frac{1}{2\sqrt{\pi\mu}} e^{-(r_{12}\mu)^2}. \quad (11)$$

Because of the simple analytical expression of  $u(r_{12}, \mu)$ , the corresponding TC Hamiltonian  $\tilde{H}[\mu]$  defined as

$$\begin{aligned}\tilde{H}[\mu] &\equiv e^{-\hat{\tau}_{\mu}} \hat{H} e^{\hat{\tau}_{\mu}} \\ &= H - \sum_{i < j} \hat{K}[\mu](\mathbf{r}_i, \mathbf{r}_j) - \sum_{i < j < k} \hat{L}[\mu](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k),\end{aligned}\quad (12)$$

with  $\hat{\tau}_{\mu} = \sum_{i < j} u(r_{ij}, \mu)$ , has a relatively simple analytical form with the effective two- and three-body operators,

$$\begin{aligned}\hat{K}[\mu](\mathbf{r}_i, \mathbf{r}_j) &= \frac{1 - \text{erf}(\mu r_{12})}{r_{12}} - \frac{\mu}{\sqrt{\pi}} e^{-(\mu r_{12})^2} \\ &\quad + \frac{(1 - \text{erf}(\mu r_{12}))^2}{4} - (\text{erf}(\mu r_{12}) - 1) \frac{\partial}{\partial r_{12}},\end{aligned}\quad (13)$$

and

$$\begin{aligned}\hat{L}[\mu](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) &= \frac{1 - \text{erf}(\mu r_{12})}{2r_{12}} \mathbf{r}_{12} \cdot \frac{1 - \text{erf}(\mu r_{13})}{2r_{13}} \mathbf{r}_{13} \\ &\quad + \frac{1 - \text{erf}(\mu r_{12})}{2r_{12}} \mathbf{r}_{21} \cdot \frac{1 - \text{erf}(\mu r_{23})}{2r_{23}} \mathbf{r}_{23} \\ &\quad + \frac{1 - \text{erf}(\mu r_{13})}{2r_{13}} \mathbf{r}_{31} \cdot \frac{1 - \text{erf}(\mu r_{32})}{2r_{32}} \mathbf{r}_{32},\end{aligned}\quad (14)$$

respectively. The correlation factor  $u(r_{12}, \mu)$  exactly restores the cusp conditions and the scalar two- and three-body effective interactions in Eqs. (13) and (14) are non-divergent, yielding “cusp-less” eigenvectors as illustrated in Ref. 26. As apparent from the definitions of Eqs. (13) and (14), the global shape of  $\tilde{H}[\mu]$  depends on a unique parameter  $\mu$ , which can be seen either as the inverse of the typical range of the correlation effects or the typical value of the effective interaction at  $r_{12} = 0$ . In the  $\mu \rightarrow +\infty$  limit, one obtains the usual Hamiltonian, while in  $\mu \rightarrow 0$  limit, the correlation factor becomes singular, but one remains with a well-defined attractive non-Hermitian Hamiltonian.

Similar to Eq. (5), we define the projection onto a basis set  $\mathcal{B}$  of the TC Hamiltonian  $\tilde{H}[\mu]$ ,

$$\tilde{H}[\mu]^{\mathcal{B}} \equiv P^{\mathcal{B}} \tilde{H}[\mu] P^{\mathcal{B}}, \quad (15)$$

whose ground state eigenvalue and associated right-eigenvector satisfy

$$\tilde{H}[\mu]^{\mathcal{B}} |\Phi_0^{\mathcal{B}}[\mu]\rangle = \tilde{E}_0^{\mathcal{B}}[\mu] |\Phi_0^{\mathcal{B}}[\mu]\rangle. \quad (16)$$

Because of its relatively simple form, the correlation factor  $u(r_{12}, \mu)$  has the advantage that it leads to effective operators  $\hat{K}[\mu](\mathbf{r}_1, \mathbf{r}_2)$  and  $\hat{L}[\mu](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  with a simple-enough analytical form for which integrals can be computed efficiently using a mixed numerical-analytical scheme (see Ref. 26 for explicit formulas). This is in contrast to the  $\mathbb{R}^6$  numerical integrals needed when using more sophisticated correlation factors.

### III. RESULTS

#### A. Computational details

To obtain the ground state eigenvalue  $\tilde{E}_0^{\mathcal{B}}[\mu]$  of a given TC Hamiltonian  $\tilde{H}[\mu]^{\mathcal{B}}$ , we use the recently developed transcorrelated full-configuration interaction quantum Monte Carlo (TC-FCIQMC) technique,<sup>15,27</sup> which extends the original stochastic projection technique of FCIQMC<sup>28–35</sup> to a non-Hermitian and three-body Hamiltonian. FCIQMC parameters were  $10^6$  walkers, an initiator threshold of  $n_{init} = 3$ , and a semi-stochastic space of  $N_D = 1000$ . After having investigated the dependence of the results with the parameter  $\mu$  in Sec. III B 2, we found that the so-called RSC + LDA system-dependent value defined in Eq. (57) of Ref. 26 is a good strategy to obtain a reasonable systematic value of the  $\mu$  parameter, and therefore, we use the latter throughout this work except for Sec. III B 2. Provided a  $\tilde{H}[\mu]^{\mathcal{B}}$  and a given basis set  $\mathcal{B}$ , the necessary one-, two-, and three-body integrals are computed using restricted Hartree-Fock (RHF) MOs. When the correlation factor is  $u(r_{12}, \mu)$ , we label the results by “ $\mu$ -TC,” whereas when using the correlation factor of Schmidt and Moskowitz,<sup>16</sup> we label the results by “SM-17,” as in Ref. 15. Regarding the integrals involved in  $\tilde{H}[\mu]^{\mathcal{B}}$ , the scalar two-body part is computed analytically and the non-Hermitian and three-body parts are computed using a mixed analytical-numerical scheme utilizing a Becke numerical grid<sup>36</sup> with 30 radial points and a Lebedev angular grid of 50 grid points. Numerical tests have shown that this relatively small number of grid points ensures a sub  $\mu\text{Ha}$  convergence of the total energies.

Estimates of the FCI in a given basis set  $\mathcal{B}$  and within a sub-mH precision were obtained with the configuration interaction perturbatively selected iteratively<sup>37</sup> (CIPSI) as implemented in the Quantum Package.<sup>38</sup> The CIPSI results are reported in Tables and Figures as “No-TC.” The estimated CBS all-electron results for atoms and cations are taken from Ref. 39. Except for the TC-FCIQMC, all calculations were performed using the Quantum Package.<sup>38</sup>

#### B. Preliminary investigation on B, B<sup>+</sup>, Ne, and Ne<sup>+</sup>

Before performing the study on the whole Li–Ne series together with their first cations, we perform a detailed study on the neutral and first cations of the boron and neon atoms. The main questions we address are (i) how to treat core electrons in all-electron calculations using  $\tilde{H}[\mu]$ , and (ii) to quantify the dependency of the results on the value of the parameter  $\mu$ , and (iii) to investigate a possible reduction of the computational cost involved in the three-body operator while maintaining high accuracy.

##### 1. Treatment of core electrons in all-electron calculations

We begin our preliminary investigation by studying the treatment of core electrons with  $\tilde{H}[\mu]$  in the case of the boron neutral atom. We performed all-electron  $\mu$ -TC calculations with the cc-pVXZ and cc-pCVXZ basis sets ( $X = D, T$ ) to study the impact of functions suited for core–valence correlation, and we report the results in Table I. The calculations with the cc-pVXZ and cc-pCVXZ basis sets are referred to as “w/o core–valence” (w/o standing for “without”) and “with core–valence,” respectively. From Table I, we can observe that the all-electron  $\mu$ -TC calculations without core–valence functions significantly underestimate the exact ground state energy of the boron atom by 36 and 26 mH in the cc-pVDZ and cc-pVTZ basis sets, respectively. This effect is strongly reduced when using core–valence functions as the underestimation of the ground state energy is 2.2 and 1.8 mH with the cc-pCVDZ and cc-pCVTZ basis sets, respectively. Regarding now the effect on the ionization potential (IP), it can be noted that while the energy difference computed using a cc-pCVDZ is already within a sub-mH precision with respect to the CBS value, the results obtained without the core–valence functions are far from such an accuracy as the error is 15 and 17 mH using the cc-pVDZ and cc-pVTZ, respectively. Therefore, as already shown by Ten-No *et al.*,<sup>13</sup> core–valence correlation functions are

**TABLE I.** Boron neutral and first cation total energies results (in a.u.) and ionization potentials (IP) from all-electron calculations with and without core–valence basis functions. cc-pVXZ basis sets are referred to as “w/o core valence,” while the cc-pCVXZ basis sets are referred to as “with core–valence.” The parameter  $\mu$  is obtained with the  $\mu_{\text{RSC+LDA}}$  approach and is about 1.02 and 1.15 for the neutral and first cation, respectively.

Atom		DZ	TZ	CBS <sup>39,40</sup>
B	With core–valence	−24.656 13	−24.655 68	
	W/o core–valence	−24.690 75	−24.680 63	−24.653 91
B <sup>+</sup>	With core–valence	−24.351 47	−24.349 60	
	W/o core–valence	−24.372 84	−24.364 95	−24.348 89
IP	With core–valence	0.304 66	0.306 08	
	W/o core–valence	0.317 91	0.315 68	0.305 02

mandatory when performing all-electron calculations in the context of TC methods, unless the correlation factor includes explicit electron–electron–nucleus correlation as, for instance, in the work of Cohen *et al.*<sup>15</sup>

## 2. Dependence of the results with respect to $\mu$

Having established that the use of basis sets able to treat core–core and core–valence correlation is mandatory to obtain sensible values for both total energies and energy differences in all-electron calculations, we investigate another aspect of the present method: the dependence of the results on  $\mu$ . The latter aspect has been quite intensively studied in Ref. 26 for two electron systems and this study has shown that the value of  $\mu$  must increase with the nuclear charge of the atomic system studied. Still in Ref. 26, one of the present authors has proposed three different system-dependent values of the  $\mu$  parameter: the so-called  $\mu_{\text{RSC}}$ ,  $\mu_{\text{LDA}}$ , and  $\mu_{\text{RSC+LDA}}$ . These values of  $\mu$  are efficiently obtainable and only depend on the density of the system under study, which essentially results in a parameter-free correlation factor. For more details, we refer the interested reader to Ref. 26. In this section, we perform a similar study on heavier atoms in order to quantify the dependence of the results on the parameter  $\mu$  in more realistic systems including core electrons. We report in Table II the total energies for the Ne and  $\text{Ne}^+$  atomic systems together with the ionization potential for the three different flavors of parameter  $\mu$  and the numerical values for the parameter  $\mu$  obtained with each approach in the cc-pCVDZ basis set. The  $\mu$ -values in the other basis sets used here are within less than 1%.

From Table II, we can observe that the lower the value of  $\mu$ , the lower the total energy, which can be related to the depth and

**TABLE II.** Total energies for the Ne atom and its first cation together with the corresponding ionization potential (IP) for different choices of the  $\mu$  parameter, in cc-pCVXZ ( $X = D, T, Q$ ) basis-sets, labeled here CVXZ. The numerical values of the  $\mu$  parameter are also reported in cc-pCVDZ only, with the values in the other basis sets being similar within less than 1%.

$\mu$ Type	Basis	Energy(Ne) (H)	Energy( $\text{Ne}^+$ ) (H)	IP (mH)
$\mu_{\text{LDA}}$	CVDZ	-128.852 76(1)	-128.062 38(1)	790.38(2)
$\mu_{\text{RSC}}$	CVDZ	-129.210 54(1)	-128.414 42(1)	796.12(2)
$\mu_{\text{RSC+LDA}}$	CVDZ	-128.964 35(1)	-128.163 45(1)	800.90(2)
$\mu_{\text{LDA}}$	CVTZ	-128.913 96(1)	-128.123 84(1)	790.12(2)
$\mu_{\text{RSC}}$	CVTZ	-128.950 26(1)	-128.160 04(1)	790.21(2)
$\mu_{\text{RSC+LDA}}$	CVTZ	-128.932 30(1)	-128.140 83(8)	791.47(9)
$\mu_{\text{LDA}}$	CVQZ	-128.931 80(1)	-128.139 51(1)	792.3(2)
$\mu_{\text{RSC}}$	CVQZ	-128.937 64(1)	-128.147 39(1)	790.24(2)
$\mu_{\text{RSC+LDA}}$	CVQZ	-128.935 69(3)	-128.142 95(8)	792.7(1)
<hr/> <b>Estimated exact</b> <hr/>				
		-128.937 60	-128.143 10	0.794 50
$\mu$ Type	Basis	Ne [a.u.]	$\text{Ne}^+$ [a.u.]	
$\mu_{\text{LDA}}$	CVDZ	1.93	2.043	
$\mu_{\text{RSC}}$	CVDZ	1.16	1.201	
$\mu_{\text{RSC+LDA}}$	CVDZ	1.55	1.622	

range of the correlation hole induced by the correlation factor, which are typically inversely proportional to  $\mu$ . One can also note that the value of  $\mu_{\text{LDA}}$  is typically twice as large as  $\mu_{\text{RSC}}$ , and the differences in total energies are significant in the cc-pCVDZ basis set (about 350 mH). They decrease monotonically in the cc-pCVTZ (about 40 mH) and cc-pCVQZ (about 5 mH) basis sets. The dependence of the total energy on the value of  $\mu$  decreases when using larger basis sets. This is expected and coherent with the fact that when reaching the CBS limit, one recovers the exact ground-state energy irrespective of the value of  $\mu$  chosen for the correlation factor. Regarding the values of the IP, one can note that although there are significant variations of the total energies, the energy differences are not very sensitive to the value of  $\mu$ . The largest deviations of the IPs are about 10 mH in the cc-pCVDZ basis set between  $\mu_{\text{LDA}}$  and  $\mu_{\text{RSC+LDA}}$ . Nevertheless, the IPs computed with the three different flavors of the parameter  $\mu$  remain much closer to the exact IP than the CIPSI results without any correlation factor. To conclude this study, one can note that the IP computed with  $\mu_{\text{RSC}}$  tends to converge less rapidly than the two other choices and that the values  $\mu_{\text{RSC+LDA}}$  and  $\mu_{\text{LDA}}$  give sensibly the same quality of energy differences. Therefore, we choose to retain the  $\mu_{\text{RSC+LDA}}$  strategy for the rest of the present work.

## 3. The “5-idx” approximation on the three-body term

Another important computational aspect of the TC method are the numerous  $N^6$  integrals to be computed for the three-body effective operator  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ . The problems regarding these terms are twofold: (i) the computation of the intermediate quantities can be quite demanding and (ii) the computation and storage of all the  $N^6$  integrals. In the context of  $\tilde{H}[\mu]^B$ , point (i) is not really a problem since all intermediate quantities are computed analytically and not numerically in contrast to more complex correlation factors<sup>15,27</sup> (see Appendix of Ref. 26). Therefore, the main computational bottleneck is the computation and storage of the  $N^6$  integrals. One can note that when using methods such as MP2 or coupled cluster (CC) approaches, as proposed in Ref. 13, one can use a resolution of the identity approximation (RI) with only  $N^4$ -storage requirement for intermediate quantities. Nevertheless, in stochastic algorithms like TC-FCIQMC, the RI would be very costly as it

**TABLE III.** Effect of neglecting the full three-body terms on total energies (in H) for all-electron TC-FCIQMC calculations in cc-pCVXZ basis sets ( $X = D, T, Q$ ) labeled here CVXZ, for the neon and  $\text{Ne}^+$  atomic system and the corresponding ionization potential (in mH). The parameter  $\mu$  is obtained with the  $\mu_{\text{RSC+LDA}}$  approach.

Method	Basis	Ne (H)	$\text{Ne}^+$ (H)	IP (mH)
Full	CVDZ	-128.964 35(1)	-128.163 45(1)	800.90(2)
5idx	CVDZ	-128.964 37(2)	-128.163 44(2)	800.93(4)
Full	CVTZ	-128.932 30(1)	-128.140 83(8)	791.47(9)
5idx	CVTZ	-128.932 01(2)	-128.140 64(6)	791.38(8)
Full	CVQZ	-128.935 69(3)	-128.142 95(8)	792.7(1)
5idx	CVQZ	-128.935 45(2)	-128.142 64(1)	792.81(3)
<hr/> <b>Estimated exact</b> <hr/>				
		-128.937 60	-128.143 10	0.794 50

would imply to recompute the three-electron integrals whenever they are required in the matrix elements computations. Nevertheless, one can note that the most numerous terms in  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  are those corresponding to six different indices, which corresponds to pure triple excitation operators. We propose here the “5-idx” approximation of the three-body term that consists of neglecting all integrals with six different indices, which reduces the number of integrals to compute and store for the treatment of  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  to  $N^5$ . We performed numerical calculation with the full consideration of the  $\hat{L}[u](\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  operator and the “5-idx” approximation using the cc-pCVXZ (X = D, T, Q) basis sets for the neon

atom and its first cation, and report the results in **Table III**. From **Table III**, we can observe that the results are almost insensitive to the 5-idx approximation as the differences between the energies are of about  $10^{-5}$  H for both the neon and the first cation. This result, therefore, indicates that the 5-idx approximation drastically reduces both the memory and CPU bottleneck of the TC calculations while leaving the numerical results unchanged to a sub-mH precision. This is still more expensive than the normal-ordered approaches proposed by some of the present authors in Ref. 41, but has the advantage to not depend on the one- and two-body density of some reference wave function. One has to keep in mind that as the

**TABLE IV.** Errors (in mH) with respect to the estimated CBS ground state energies (in H) for the Li–Ne species, together with their first cations and the corresponding ionization potential (IP) (in mH) computed with cc-pCVXZ (X = D, T, Q) basis sets. SM-17 stands for TC-FCIQMC results with the correlation factor of Schmidt and Moskowitz.<sup>16</sup>  $\mu$ -TC calculations are performed with a parameter  $\mu$  obtained with the  $\mu_{RSC+LDA}$  approach. Estimated non-relativistic CBS results are obtained from Ref. 39.

Atom	Method	CVDZ	CVTZ	CVQZ	Est. CBS <sup>a</sup>	Atom	Method	CVDZ	CVTZ	CVQZ	Est. CBS <sup>a</sup>
Li	$\mu$ -TC	-1.03	-0.51	-0.26	-7.478 06	N	$\mu$ -TC	-9.76	-4.62	-0.99	-54.589 20
	SM-17	0.58	-0.18	...			SM-17	22.25	2.62	...	
	No-TC	12.04	3.82	1.70			No-TC	71.55	21.27	7.23	
$Li^+$	$\mu$ -TC	0.03	-0.12	-0.08	-7.279 91	$N^+$	$\mu$ -TC	-3.27	-2.61	-0.18	-54.054 60
	SM-17	0.40	-0.25	...			SM-17	15.85	3.01	...	
	No-TC	10.72	3.36	1.58			No-TC	58.48	17.88	6.10	
IP(Li)	$\mu$ -TC	1.06	0.38	0.18	198.15	IP(N)	$\mu$ -TC	6.49	2.01	0.85	534.60
	SM-17	-0.18	-0.07	...			SM-17	-6.40	0.39	...	
	No-TC	-1.32	-0.46	-0.12			No-TC	-13.07	-3.39	-1.13	
Be	$\mu$ -TC	-3.01	-1.19	-0.41	-14.667 36	O	$\mu$ -TC	-6.82	-0.44	0.01	-75.067 30
	SM-17	-2.33	-1.27	...			SM-17	40.54	6.48	...	
	No-TC	15.54	5.00	1.80			No-TC	116.79	36.08	12.83	
$Be^+$	$\mu$ -TC	-0.89	-0.62	-0.28	-14.324 76	$O^+$	$\mu$ -TC	-10.68	-3.75	-1.49	-74.566 80
	SM-17	-0.94	-0.64	...			SM-17	24.91	4.79	...	
	No-TC	13.74	4.28	1.59			No-TC	88.84	25.82	8.65	
IP(Be)	$\mu$ -TC	2.14	0.59	0.14	342.58	IP(O)	$\mu$ -TC	-3.86	-3.31	-1.50	500.50
	SM-17	1.41	0.65	...			SM-17	-15.63	-1.69	...	
	No-TC	-1.78	-0.70	-0.19			No-TC	-27.95	-10.26	-4.18	
B	$\mu$ -TC	-2.21	-1.77	-0.76	-24.653 91	F	$\mu$ -TC	-13.11	2.26	0.64	-99.733 90
	SM-17	2.22	-0.68	...			SM-17	63.89	11.06	...	
	No-TC	27.88	9.06	3.08			No-TC	164.25	52.05	18.81	
$B^+$	$\mu$ -TC	-2.55	-0.68	-0.61	-24.348 92	$F^+$	$\mu$ -TC	-11.18	-0.19	-0.78	-99.092 80
	SM-17	-1.36	-0.24	...			SM-17	51.64	10.52	...	
	No-TC	19.07	6.79	2.28			No-TC	137.54	41.23	14.30	
IP(B)	$\mu$ -TC	-0.33	1.09	0.14	304.99	IP(F)	$\mu$ -TC	1.93	-2.45	-1.42	641.10
	SM-17	-3.58	0.43	...			SM-17	-12.25	-0.54	...	
	No-TC	-8.81	-2.27	-0.80			No-TC	-26.71	-10.82	-4.51	
C	$\mu$ -TC	-3.88	-2.93	-1.17	-37.845 00	Ne	$\mu$ -TC	-26.75	5.39	1.91	-128.937 60
	SM-17	9.63	0.38	...			SM-17	89.86	18.15	...	
	No-TC	47.02	14.97	5.38			No-TC	215.06	69.37	25.25	
$C^+$	$\mu$ -TC	-0.59	-1.11	-0.67	-37.431 03	$Ne^+$	$\mu$ -TC	-20.35	2.28	0.12	-128.143 10
	SM-17	4.97	0.85	...			SM-17	76.19	17.57	...	
	No-TC	36.16	11.71	3.92			No-TC	188.73	58.12	20.51	
IP(C)	$\mu$ -TC	3.30	1.82	0.50	413.97	IP(Ne)	$\mu$ -TC	6.40	-3.11	-1.79	794.50
	SM-17	-4.66	0.47	...			SM-17	-13.67	-0.58	...	
	No-TC	-10.52	-2.74	-0.94			No-TC	-26.33	-11.25	-4.74	

three-body operator is truncated, the 5-idx approximation formally introduces a dependence on the choice of orbitals used in the calculations. We have investigated the dependency of the results in the cc-pCVDZ basis sets for both the Ne and CH<sub>2</sub> using different sets of orbitals (HF, localized, and natural orbitals from correlated calculations) and found very small differences in the 5-idx results (typically 10<sup>-5</sup> H).

### C. All-electrons calculations on the Li-Ne species and first cations

In Table IV, we report the performance of all-electron ground state calculations on the Li-Ne series in the cc-pCVXZ basis sets (X = D, T, Q), together with their first cations at the  $\mu$ -TC and CIPSI levels of theory. We also report in Table IV the estimated non-relativistic CBS results of Ref. 39, together with the so-called SM-17 results that are the TC-FCIQMC calculations in the cc-pCVXZ basis set family (X = D, T) using the same methodology of Ref. 15 where the correlation factor of Schmidt and Moskowitz<sup>16</sup> was used to obtain the TC Hamiltonian. The SM-17 correlation factor is very flexible as it contains explicit electron-nucleus, electron-electron and electron-electron-nucleus terms and has been optimized at the VMC level for each neutral species considered here. The mean absolute errors (MAEs) of the IPs are reported in Table V. A

**TABLE V.** Mean absolute errors (MAEs) in mH for the ionization potentials at the  $\mu$ -TC and No-TC levels of theory for the Li-Ne series in the cc-pCVXZ basis sets.  $\mu$ -TC calculations are performed with a parameter  $\mu$  obtained with the  $\mu_{RSC+LDA}$  approach. The results labeled by SM-17 are the TC-FCIQMC results of Ref. 15 using a flexible correlation factor.

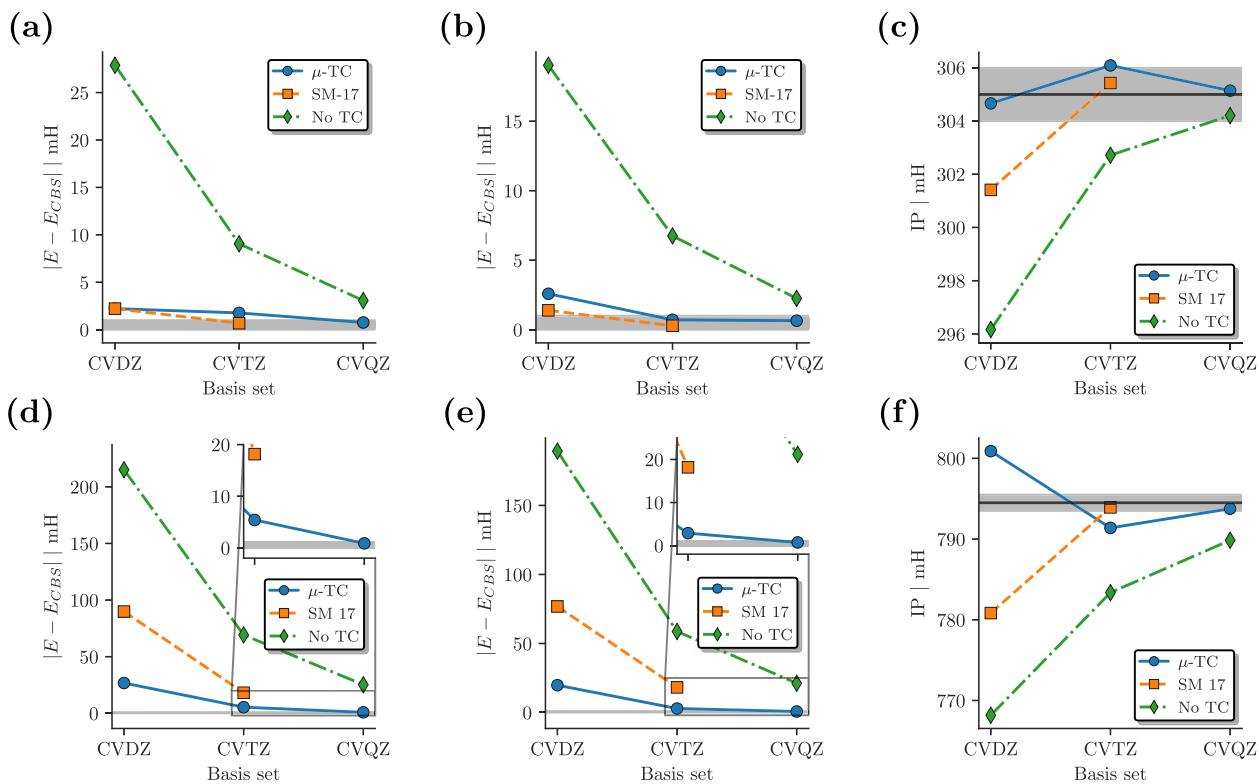
	CVDZ	CVTZ	CVQZ
No-TC	14.65	5.24	2.08
$\mu$ -TC	3.19	1.85	0.81
SM-17	7.22	0.60	...

graphical representation of the results B, B<sup>+</sup>, Ne, and Ne<sup>+</sup> is shown in Fig. 1.

From Tables IV and V, one can observe that the convergence of both the total energies and the ionization potential using  $\mu$ -TC and SM-17 is strongly improved with respect to usual WFT calculations, which is expected due to the presence of explicit correlation.

Several specific aspects have to be pointed out in Tables IV and V:

- For  $Z > 5$ , the total energies provided by  $\mu$ -TC are always closer to the exact ones than that of SM-17.



**FIG. 1.** Boron (a), B<sup>+</sup> (b), Ne (d), and Ne<sup>+</sup> (e) absolute energy difference with respect to the estimated exact total energies<sup>39</sup> and ionization potential for boron (c) and neon (f) of the range separated TC ( $\mu$ -TC), Cohen et al.<sup>15</sup> (SM-17) and non-transcorrelated results (no TC) for all-electron calculations with cc-pCVXZ core-valence basis sets.  $\mu$ -TC calculations are performed with a parameter  $\mu$  obtained with the  $\mu_{RSC+LDA}$  approach. The gray areas indicate sub-mH accuracy.

- (ii) With increasing nuclear charge, the discrepancy between the total energies at the triple-zeta basis set level using  $\mu$ -TC and SM-17 increases. This suggests that the electron-electron-nucleus term of the SM-17 takes into account only a part of the correlation effects arising from the core (i.e., core–core and core–valence correlation effects).
- (iii) Although the total energies obtained with  $\mu$ -TC in double-zeta basis sets can be way below the exact ground state energy (by about 26 and 20 mH in the case of the Ne and  $\text{Ne}^+$  in the cc-pCVDZ basis set, respectively), the energy differences are of good quality (the accuracy of the ionization potential of Ne using  $\mu$ -TC in the cc-pCVDZ basis set is comparable to that of regular WFT in a cc-pCVQZ basis set).
- (iv) While at the double-zeta level, the error with respect to the exact IP is significantly smaller using the  $\mu$ -TC than the SM-17 approach for all systems, the errors at the triple-zeta level are smaller with the SM-17 by approximately 1 kcal/mol on average. Nevertheless, the error with respect to the exact IP obtained with  $\mu$ -TC at the quadruple-zeta level decreased below 0.001 H, showing a systematic convergence pattern.

#### D. All-electron calculations on the $\text{CH}_2$ , FH, and $\text{H}_2\text{O}$ molecules

We conclude this study by performing all-electron calculations on the  $\text{CH}_2$ , FH, and  $\text{H}_2\text{O}$  molecules, whose equilibrium geometries have been taken from Refs. 42 and 43, respectively. In Table VI, we report the total energies of these three molecules computed at the

**TABLE VI.** Error (in mH) with respect to the estimated CBS ground state energies (in H) for the  $\text{H}_2\text{O}$ ,  $\text{CH}_2$ , and FH molecules, together with their respective atomization energies (AE) (in mH) computed with cc-pCVXZ ( $X = D, T, Q$ ) basis sets.  $\mu$ -TC calculations are performed with a parameter  $\mu$  obtained with the  $\mu_{\text{RSC+LDA}}$  approach. The molecular geometries are as follows:  $\text{CH}_2$ :CH distance = 1.1063 Å, HCH angle = 102.30°,  $\text{H}_2\text{O}$ :OH distance = 0.9572 Å, HOH angle = 104.52°, and FH:FH distance = 0.91680 Å.

System	Method	CVDZ	CVTZ	CVQZ	Est. CBS
$\text{H}_2\text{O}$	No-TC	156.07	4.90	1.74	-76.438 94(12) <sup>a</sup>
	$\mu$ -TC	-4.44	1.95	...	
AE( $\text{H}_2\text{O}$ )	No-TC	-37.84	-12.55	-4.45	371.64(12) <sup>b</sup>
	$\mu$ -TC	-0.94	-2.0	...	
$\text{CH}_2$	No-TC	72.95	21.94	7.36	-39.134 25 <sup>c</sup>
	$\mu$ -TC	0.35	-1.70	...	
AE( $\text{CH}_2$ )	No-TC	-24.45	-6.55	-1.84	289.22 <sup>c</sup>
	$\mu$ -TC	-5.90	-0.81	...	
FH	No-TC	189.14	59.94	21.89	-100.460 08 <sup>c</sup>
	$\mu$ -TC	-13.23	4.22	...	
AE(FH)	No-TC	-24.18	-7.70	-3.03	226.18 <sup>c</sup>
	$\mu$ -TC	0.84	-1.77	...	

<sup>a</sup>Results from Ref. 43.

<sup>b</sup>Results from Ref. 43 and from the estimated CBS total energies of the atomic components of the molecule.

<sup>c</sup>Results from Ref. 42.

$\mu$ -TC and CIPSI levels of theory, together with the corresponding atomization energies (AE). Estimated CBS quantities are obtained from Refs. 42 and 43. From Table VI, one can note that also for molecular systems, the convergence of both the total energies and AE is much faster with respect to the basis set with the  $\mu$ -TC than with the usual CIPSI calculations. For the  $\text{CH}_2$  molecule, we obtain an error below 1 mH for the AE in a cc-pCVTZ basis set, while the error is 2.00 and 1.77 mH for the  $\text{H}_2\text{O}$  and FH molecules, respectively. These results illustrate the benefit of using the TC approach with a simple correlation factor even for simple molecular systems.

#### IV. CONCLUSION

In this work, we further investigated a new strategy based on the TC method that was previously applied to two-electron systems only.<sup>26</sup> One of the focuses of this paper is to test its validity in more realistic systems where many-body effects arise, and therefore, the effective three-body terms of the TC Hamiltonian have to be included.

In order to avoid particle-hole truncation errors due to approximations of the right eigenvector in a given basis set, the ground state energy of the TC Hamiltonian has been obtained using the recently proposed non-Hermitian and three-body variant of FCIQMC.<sup>15</sup>

The main feature of the recently introduced<sup>26</sup> TC correlation factor is that, besides producing a strictly non-divergent TC Hamiltonian, it has a simple parameterization that depends only on a single parameter  $\mu$ . Such a parameter  $\mu$  determines the impact of the correlation factor through the depth and typical range of the correlation hole that it induces in the wave function. In addition, due to the simple analytical structure of the obtained TC Hamiltonian, all necessary integrals can be efficiently computed in a mixed numerical-analytical scheme. The parameter  $\mu$  is determined for each system, according to the method described in Ref. 26, and depends only on the density of the system under study, which essentially results in a parameter-free correlation factor.

The main focus of this work is the study of the convergence of the TC eigenvalues and energy differences with respect to the quality of the basis set and its ability to treat both core and valence electrons. We performed calculations on the Li–Ne series together with their first cations in the cc-pCVXZ ( $X = D, T, Q$ ) basis sets, in order to investigate the convergence toward the CBS limit of both total energies and IPs. We also computed both the total energies and AEs of the  $\text{CH}_2$ , FH, and  $\text{H}_2\text{O}$  molecules.

The main conclusion of this study is that, provided that the basis set contains core–valence functions, very accurate total energies can already be obtained from the triple-zeta quality basis sets.

Regarding the accuracy of the IPs computed here, while the MAE is significantly smaller in double-zeta basis sets using the single-parameter  $\mu$ -TC compared to the more elaborate SM-17 correlation factor, the results at the triple-zeta level of theory are outperformed by the latter by about 1.2 mH on average. Nevertheless, the MAEs of both these methods are within chemical accuracy (below 1 mH) with CBS limit results already at the quadruple zeta level. Regarding the AEs computed on the three small molecules considered here, an error below 1 mH is reached with a triple-zeta

quality for the  $\text{CH}_2$  molecule, while it is 2.00 and 1.77 mH for the  $\text{H}_2\text{O}$  and FH molecules, respectively.

In the context of TC calculations, this study shows that the results obtained with a simple one-parameter correlation factor such as  $u(r_{12}, \mu)$  are comparable with those obtained with more sophisticated correlation factors including electron–electron–nucleus terms. On the other hand, it should be pointed out that in the TC work of Cohen *et al.*,<sup>15</sup> the employed SM17 correlation factors were taken from the literature and were not further optimized for use in the TC method. In subsequent work, some of the present authors will investigate a better suited method to optimize general Jastrow factors for the TC approach.

Among the perspectives of this work, the use of a correlation factor  $u(r_{12}, \mu)$  with a  $\mu$  varying in real space could be of interest as it could probably mimic the electron–electron–nucleus correlation effects.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors declare that they have no conflicts of interest.

### Author Contributions

**Werner Dobrautz:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Aron J. Cohen:** Resources (supporting). **Ali Alavi:** Formal analysis (supporting); Funding acquisition (supporting); Resources (supporting); Writing – original draft (supporting). **Emmanuel Giner:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Software (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal)

All authors read and approved the final manuscript.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

- <sup>1</sup>T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>2</sup>S. Ten-no, *Theor. Chem. Acc.* **131**, 1070 (2012).
- <sup>3</sup>S. Ten-no and J. Noga, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 114 (2012).
- <sup>4</sup>C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, *Chem. Rev.* **112**, 4 (2012).
- <sup>5</sup>L. Kong, F. A. Bischoff, and E. F. Valeev, *Chem. Rev.* **112**, 75 (2012).
- <sup>6</sup>A. Grüneis, S. Hirata, Y.-Y. Ohnishi, and S. Ten-no, *J. Chem. Phys.* **146**, 080901 (2017).
- <sup>7</sup>Q. Ma and H.-J. Werner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1371 (2018).
- <sup>8</sup>J. Toulouse, R. Assaraf, and C. J. Umrigar, in *Advances in Quantum Chemistry* (Academic Press, Cambridge, MA, 2016), Vol. 73, pp. 285–314.
- <sup>9</sup>J. O. Hirschfelder, *J. Chem. Phys.* **39**, 3145 (1963).
- <sup>10</sup>F. S. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 43 (1969).
- <sup>11</sup>F. S. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 63 (1969).
- <sup>12</sup>S. Ten-no, *Chem. Phys. Lett.* **330**, 169 (2000).
- <sup>13</sup>O. Hino, Y. Tanimura, and S. Ten-no, *J. Chem. Phys.* **115**, 7865 (2001).
- <sup>14</sup>O. Hino, Y. Tanimura, and S. Ten-no, *Chem. Phys. Lett.* **353**, 317 (2002).
- <sup>15</sup>A. J. Cohen, H. Luo, K. Guthier, W. Dobrautz, D. P. Tew, and A. Alavi, *J. Chem. Phys.* **151**, 061101 (2019).
- <sup>16</sup>K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1990).
- <sup>17</sup>W. Dobrautz, H. Luo, and A. Alavi, *Phys. Rev. B* **99**, 075119 (2019).
- <sup>18</sup>M. C. Gutzwiller, *Phys. Rev. Lett.* **10**, 159 (1963).
- <sup>19</sup>W. F. Brinkman and T. M. Rice, *Phys. Rev. B* **2**, 4302 (1970).
- <sup>20</sup>A. Baiardi and M. Reiher, *J. Chem. Phys.* **152**, 040903 (2020).
- <sup>21</sup>M. Motta, T. P. Gujarati, J. E. Rice, A. Kumar, C. Masteran, J. A. Latone, E. Lee, E. F. Valeev, and T. Y. Takeshita, *Phys. Chem. Chem. Phys.* **22**, 24270 (2020).
- <sup>22</sup>S. McArdle and D. P. Tew, “Improving the accuracy of quantum computational chemistry using the transcorrelated method,” [arXiv:2006.11181](https://arxiv.org/abs/2006.11181) (2020).
- <sup>23</sup>P. Schleich, J. S. Kottmann, and A. Aspuru-Guzik, “Improving the accuracy of the variational quantum eigensolver for molecular systems by the explicitly-correlated perturbative  $[2]_{R12}$ -correction,” [arXiv:2110.06812](https://arxiv.org/abs/2110.06812) (2021).
- <sup>24</sup>A. Kumar, A. Asthana, C. Masteran, E. F. Valeev, Y. Zhang, L. Cincio, S. Tretyak, and P. A. Dub, “Accurate quantum simulation of molecular ground and excited states with a transcorrelated Hamiltonian,” *Phys. Chem. Chem. Phys.* **24**, 13550–13564 (2022).
- <sup>25</sup>I. O. Sokolov, W. Dobrautz, H. Luo, A. Alavi, and I. Tavernelli, “Orders of magnitude reduction in the computational overhead for quantum many-body problems on quantum computers via an exact transcorrelated method,” [arXiv:2201.03049](https://arxiv.org/abs/2201.03049) (2022).
- <sup>26</sup>E. Giner, *J. Chem. Phys.* **154**, 084119 (2021).
- <sup>27</sup>K. Guthier, A. J. Cohen, H. Luo, and A. Alavi, *J. Chem. Phys.* **155**, 011102 (2021).
- <sup>28</sup>G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- <sup>29</sup>G. H. Booth and A. Alavi, *J. Chem. Phys.* **132**, 174104 (2010).
- <sup>30</sup>G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **135**, 084104 (2011).
- <sup>31</sup>K. Ghanem, A. Y. Lozovoi, and A. Alavi, *J. Chem. Phys.* **151**, 224108 (2019).
- <sup>32</sup>E. Vitale, A. Alavi, and D. Kats, *J. Chem. Theory Comput.* **16**, 5621 (2020).
- <sup>33</sup>K. Guthier, R. J. Anderson, N. S. Blunt, N. A. Bogdanov, D. Cleland, N. Dattani, W. Dobrautz, K. Ghanem, P. Jeszenszki, N. Liebermann, G. L. Manni, A. Y. Lozovoi, H. Luo, D. Ma, F. Merz, C. Overy, M. Rampp, P. K. Samanta, L. R. Schwarz, J. J. Shepherd, S. D. Smart, E. Vitale, O. Weser, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **153**, 034107 (2020).
- <sup>34</sup>W. Dobrautz, S. D. Smart, and A. Alavi, *J. Chem. Phys.* **151**, 094104 (2019).
- <sup>35</sup>W. Dobrautz, O. Weser, N. A. Bogdanov, A. Alavi, and G. Li Manni, *J. Chem. Theory Comput.* **17**, 5684 (2021).
- <sup>36</sup>A. D. Becke, *J. Chem. Phys.* **88**, 2547 (1988).
- <sup>37</sup>B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- <sup>38</sup>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).
- <sup>39</sup>S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. p Fischer, *Phys. Rev. A* **47**, 3649 (1993).
- <sup>40</sup>E. R. Davidson, S. A. Hagstrom, S. J. Chakravorty, V. M. Umar, and C. F. Fischer, *Phys. Rev. A* **44**, 7071 (1991).
- <sup>41</sup>T. Schraivogel, A. J. Cohen, A. Alavi, and D. Kats, *J. Chem. Phys.* **155**, 191101 (2021).
- <sup>42</sup>Y. Yao, E. Giner, J. Li, J. Toulouse, and C. J. Umrigar, *J. Chem. Phys.* **153**, 124117 (2020).
- <sup>43</sup>M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, *J. Chem. Phys.* **144**, 151103 (2016).