The Journal of Chemical Physics

RESEARCH ARTICLE | AUGUST 22 2024

Compactification of determinant expansions via transcorrelation

Abdallah Ammar ; Anthony Scemama ; Pierre-François Loos ; Emmanuel Giner

Check for updates

J. Chem. Phys. 161, 084104 (2024)
https://doi.org/10.1063/5.0217650

View Online
Export Citation

22 August 2024 16:47:33

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors

Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Compactification of determinant expansions via transcorrelation

Cite as: J. Chem. Phys. 161, 084104 (2024); doi: 10.1063/5.0217650
Submitted: 6 May 2024 • Accepted: 22 July 2024 •
Published Online: 22 August 2024

View Online Export Citation CrossMark

Abdallah Ammar,[1,a] Anthony Scemama,[1,b] Pierre-François Loos,[1,c] and Emmanuel Giner[2,d]

AFFILIATIONS

[1] Laboratoire de Chimie et Physique Quantiques (UMR 5626), Université de Toulouse, CNRS, UPS, Toulouse, France
[2] Laboratoire de Chimie Théorique, Sorbonne Université and CNRS, F-75005 Paris, France

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

ABSTRACT

Although selected configuration interaction (SCI) algorithms can tackle much larger Hilbert spaces than the conventional full CI method, the scaling of their computational cost with respect to the system size remains inherently exponential. In addition, inaccuracies in describing the correlation hole at small interelectronic distances lead to the slow convergence of the electronic energy relative to the size of the one-electron basis set. To alleviate these effects, we show that the non-Hermitian, transcorrelated (TC) version of SCI significantly compactifies the determinant space, allowing us to reach a given accuracy with a much smaller number of determinants. Furthermore, we note a significant acceleration in the convergence of the TC-SCI energy as the basis set size increases. The extent of this compression and the energy convergence rate are closely linked to the accuracy of the correlation factor used for the similarity transformation of the Coulombic Hamiltonian. Our systematic investigation of small molecular systems in increasingly large basis sets illustrates the magnitude of these effects.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0217650

22 August 2024 16:47:33

I. INTRODUCTION

Obtaining an accurate description of the electronic structure of molecular systems remains one of the main challenges of theoretical chemistry. The primary challenge stems from electronic correlation effects driven by Coulomb's repulsion among electrons. To address this issue, wave function theory appears as a promising tool as it allows for a systematically improvable solution of the Schrödinger equation following two complementary directions: (i) an improvement of the wave function method to get as close as possible to the full configuration interaction (FCI) solution and (ii) an increase in the size of the one-electron basis set toward the complete basis set (CBS) limit. These two aspects are usually considered separately and addressed by different approaches.

Regarding the basis set expansion, it was acknowledged long ago[1] that the slow convergence of properties with respect to the size of the one-particle basis set, computed within wave function methods, arises from correlation effects when two electrons are close to each other, that is, near the derivative discontinuity (the so-called electron–electron cusp) that originates from the divergence of the Coulomb repulsion as the interelectronic distance $r_{12} \to 0$.[2–6]

To mitigate this issue, it was proposed to complement the wave function with a two-electron function (or geminal) explicitly depending on r_{12}, leading to the so-called "R12" and the more modern "F12" methods.[7–13] In the latter method, a correlation factor is included to introduce short-range correlation effects, hence improving the shape of the correlation hole at small r_{12}. This leads to faster basis set convergence of ground-state properties.[14–17] Nevertheless, F12 methods do not use the full correlation factor *per se* but only its orthogonal complement with respect to the space spanned by the one-electron basis, leading to an operator that, by design, only describes correlation effects that are absent of the one-electron basis set. Therefore, F12 methods do not alter the expansion and potential complexity of the wave function within the basis set.

Considering the approximations of the FCI wave function and energy, there exist a wide variety of systematically improvable approaches starting from the mean-field Hartree–Fock (HF) solution: perturbation theory (PT), CI expansions, coupled-cluster (CC) theory,[18–23] or matrix product states (MPS).[24,25] Within this zoo of wave function formalisms, one can classify them into two categories: methods for which the wave function Ansatz is fixed by design (e.g., CC with single and double excitations), and adaptive methods where

Published under an exclusive license by AIP Publishing

the wave function automatically adapts until reaching a given accuracy. The former has a clear advantage as their computational cost scales polynomially with the system size, with the potential drawback that the Ansatz might not always be adapted to the specific problem under study.

Among the adaptive methods, one can mention the vast family of selected configuration interaction (SCI)[26–40] approaches whose early stages originated in the late 1960s, the MPS approaches,[41,42] which started in the late 1990s,[24] or the more recent FCI quantum Monte Carlo (FCIQMC),[43,44] many-body expanded FCI (MBE-FCI),[45,46] full CC reduction (FCCR),[47,48] and moment expansion CC [CC(P,Q)].[49–51] Although formulated in distinct ways, these adaptive methods are characterized by the idea of selecting the most important components of the wave function among a very large Hilbert space.

One of the particularities of FCIQMC and SCI is that they both rely on a linear parameterization of the electronic wave function. The advantage of such linear expansions is certainly the ability to rapidly and easily compute the Hamiltonian matrix elements between the numerous Slater determinants involved in the SCI or FCIQMC calculations. At the SCI level, this is further exploited for the computation of a second-order PT (PT2) correction, which can be then used to extrapolate the variational energy to the FCI limit, as originally proposed by Holmes *et al.*[52] and lately rationalized by Burton and Loos.[53] The resulting SCI + PT2 approach is now regarded as powerful machinery to generate FCI-quality energies for both ground and excited states.[37,38,52,54–61] However, due to the intrinsic exponential nature of the exact wave function, both SCI and FCIQMC rapidly run out of steam as the size of systems and/or basis sets increases. Therefore, being able to compactify the CI expansion within a given basis set can substantially expand the range of applicability, in terms of system size, of these state-of-the-art approaches.

A potential solution allowing for simultaneous compactification of the CI expansion and reduction of the basis set error is to supplement the wave function with a correlation factor without projection schemes.[15,62–66] The two main families of projection-free methods dealing with correlation factors are the transcorrelation (TC) method of Boys and Handy,[67] which is related to earlier work by Hirschfelder,[68] and the variational Monte Carlo (VMC) approach.[69] Importantly, while VMC treats the correlation factor in a Hermitian framework, TC involves a similarity transformation of the Hamiltonian by the same correlation factor, leading to a non-Hermitian effective Hamiltonian.

The main consequences are fourfold: (i) the VMC framework being Hermitian, it necessarily provides variational energies unlike the TC formalism; (ii) the Baker–Campbell–Hausdorff (BCH) expansion of the TC Hamiltonian naturally truncates at second order with at most three-body terms, while the BCH expansion of the effective VMC Hamiltonian does not truncate and, therefore, generates up to N-body terms; (iii) because TC involves at most three-body terms, the corresponding matrix elements between two Slater determinants can be computed in a deterministic way, in contrast to VMC where one must rely on stochastic integration techniques to compute the N-body integrals and avoid the curse of dimensionality; and (iv) the similarity transformation of the TC Hamiltonian maintains the original overlap metric of the Slater determinant basis while VMC necessarily introduces nonorthogonality.

Despite these differences, for a given correlation factor and in the limit of a complete basis, the right eigenvector of the TC Hamiltonian and the Slater part of the VMC trial wave function coincide, resulting in identical TC and VMC energies.[70–73] Because the TC Hamiltonian can be written in second-quantized form, it can then be employed in any post-HF methods, which nevertheless have to be adapted to the non-Hermitian TC formalism. Transcorrelated versions of most wave function Ansätze have been reported, starting from the single-determinant formalism[67,74–99] to PT,[100–102] CI and SCI,[65,103–108] CC,[73,109–113] FCIQMC,[64,114–118] and MPS.[119–121] In a similar fashion, Ten-No recently introduced the so-called projected TC[122] (pTC), which combines the similarity transformation of the TC framework with an F12-type correlation factor. In contrast to the correlation factors used in TC fulfilling only the singlet-pairs cusp condition, the pTC allows for the fulfillment of the triplet-pairs cusp condition at the price of using an orthogonal projector as in F12, which, therefore, reduces the role of the correlation factor to capturing correlation effects outside the basis set.

As mentioned in previous studies,[15,62–66] the inclusion of a correlation factor, either within the TC or VMC framework, allows for the compactification of the determinantal part of the wave function. This key feature could potentially expand the applicability of SCI while decreasing the finite basis set error. Moreover, the superior quality of TC energy differences has been previously reported in several studies.[106,108,112,113,116,121] The aim of the present work is to focus on the improved convergence properties of TC-SCI with respect to the standard SCI implementation for both absolute and relative energies. More precisely, we aim to study the impact of the correlation factor on three key aspects of the SCI algorithm: (i) the convergence of the TC-SCI energy, (ii) the selection of the Slater determinants, and (iii) the choice of the orbital set.

This paper is organized as follows: In Sec. II, we recall the theoretical background of the various methods employed here. The general theory of TC is summarized in Sec. II A. Then, the biorthogonal basis set representation of the TC Hamiltonian is presented in Sec. II B, and its application to orbital optimization is in Sec. II C. The working equations for the normal-ordering approximation of the three-body interaction are gathered in Sec. II D, the general TC-SCI algorithm is presented in Sec. II E, and the correlation factor is described in Sec. II F. In Sec. II G, we discuss the computation of the various additional integrals required within the TC formalism, while Sec. II H analyzes the main computational bottleneck associated with TC calculations. Then, Sec. III presents the main numerical results for atomic and molecular systems. Section III A gathers the computational details. Considering the water molecule as an example, we carry out a detailed investigation of (i) the role of the correlation factor and orbital optimization on the distribution of the determinant weights in the TC-SCI expansion (Sec. III B), (ii) the convergence of the non-variational TC-SCI energy and its extrapolation toward the FCI limit (Sec. III C), and (iii) the compactification of the determinant expansion provided by the TC formalism (Sec. III D). We also study, in Secs. III E and III F, the convergence of total energies and ionization potentials (IPs) obtained within the TC-SCI approach. Comparisons with the best estimates from the literature and CBS values are reported. Finally,

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

our conclusions are drawn in Sec. IV. Unless otherwise stated, atomic units are used throughout.

II. THEORETICAL FRAMEWORK

From hereon, the indices $p, q, r, s, \ldots$ denote arbitrary spin-orbitals, while $i, j, k, \ldots$ and $a, b, c, \ldots$ designate the occupied and virtual spin-orbitals, respectively. Moreover, the indices $\mu, \nu, \lambda, \sigma, \ldots$ represent the basis functions, i.e., atomic orbitals.

A. Transcorrelated approach

The central concept of the TC theory is to apply a similarity transformation

$$\hat{H}_{\mathrm{TC}} \equiv e^{-\hat{\tau}} \hat{H} e^{\hat{\tau}} \tag{1}$$

to the bare Coulombic Hamiltonian

$$\hat{H} = -\sum_i \frac{\nabla_i^2}{2} - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_{i<j} \frac{1}{r_{ij}}, \tag{2}$$

where $\boldsymbol{r}_i$ represents the position of electron i, the index A runs over the nuclei of charge Z_A, r_{iA} is the distance between the ith electron and the Ath nucleus, and $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$. The similarity transformation defined in Eq. (1) involves a correlation (or Jastrow) factor,

$$\hat{\tau} \equiv \sum_{i<j} u_{ij}, \tag{3}$$

where the two-electron function $u_{ij} \equiv u(\boldsymbol{r}_i, \boldsymbol{r}_j)$ is symmetric with respect to the exchange of two electrons, i.e., $u_{ji} = u_{ij}$.

The two Hamiltonians, $\hat{H}$ and $\hat{H}_{\mathrm{TC}}$, defined in Eqs. (2) and (1), respectively, share the same spectrum of eigenvalues on a complete basis. The similarity transformation allows the transfer of physical effects from the correlation factor to the TC Hamiltonian. Therefore, by incorporating correlation effects and exact conditions in $\hat{\tau}$, such as the electron–electron cusp, one would anticipate the eigenvalues and eigenvectors of $\hat{H}_{\mathrm{TC}}$ to converge faster toward the CBS limit when employing finite basis sets.

Using the BCH expansion, which here truncates naturally in second order, one can show that the TC Hamiltonian reads

$$\hat{H}_{\mathrm{TC}} = \hat{H} - \sum_{i<j} \hat{K}_{ij} - \sum_{i<j<k} \hat{L}_{ijk}, \tag{4}$$

where the two-electron operator $\hat{K}_{ij}$ and the three-electron operator $\hat{L}_{ijk}$ are explicitly given by

$$\begin{aligned}\hat{K}_{12} \equiv\ & (\nabla_1 u_{12}) \cdot \nabla_1 + (\nabla_2 u_{21}) \cdot \nabla_2 + \frac{1}{2}\nabla_1^2 u_{12} \\ & + \frac{1}{2}\nabla_2^2 u_{21} + \frac{1}{2}(\nabla_1 u_{12})^2 + \frac{1}{2}(\nabla_2 u_{21})^2\end{aligned} \tag{5}$$

and

$$\begin{aligned}\hat{L}_{123} \equiv\ & (\nabla_1 u_{12}) \cdot (\nabla_1 u_{13}) + (\nabla_2 u_{21}) \cdot (\nabla_2 u_{23}) \\ & + (\nabla_3 u_{32}) \cdot (\nabla_3 u_{31}).\end{aligned} \tag{6}$$

Clearly, $\hat{H}_{\mathrm{TC}}$ is the non-Hermitian and includes three-electron operators. These notable features will be thoroughly examined and addressed in the subsequent sections.

B. Biorthogonal representation

Given that the eigenvectors of a non-Hermitian Hamiltonian are generally not orthogonal, the biorthogonal framework[123,124] emerges as the most natural approach to handle $\hat{H}_{\mathrm{TC}}$.[96–99,101–103,108] By introducing two sets of one-electron functions, $\{\tilde{\phi}_p\}$ and $\{\phi_p\}$, satisfying the biorthogonality condition

$$\langle \tilde{\phi}_p | \phi_q \rangle = \delta_{pq}, \tag{7}$$

one can define direct and dual creation and annihilation operators by their actions on the true vacuum $|0\rangle$, as follows:[123]

$$a_p^\dagger |0\rangle = |\phi_p\rangle, \qquad \tilde{a}_p^\dagger |0\rangle = |\tilde{\phi}_p\rangle. \tag{8}$$

The biorthogonal condition (7) ensures that the creation and annihilation operators satisfy the anticommutation relations within the context of biorthogonality, that is,

$$\{\tilde{a}_p, a_q^\dagger\} = \delta_{pq}, \qquad \{a_p^\dagger, a_q^\dagger\} = 0, \qquad \{\tilde{a}_p, \tilde{a}_q\} = 0. \tag{9}$$

Hence, in the second-quantization formalism, $\hat{H}_{\mathrm{TC}}$ can be decomposed as a sum of one-, two-, and three-body terms, with explicit forms

$$\begin{aligned}\hat{H}_{\mathrm{TC}} = & \sum_{pq} h_q^p a_p^\dagger \tilde{a}_q + \frac{1}{2!} \sum_{pqrs} V_{rs}^{pq} a_p^\dagger a_q^\dagger \tilde{a}_s \tilde{a}_r \\ & + \frac{1}{3!} \sum_{pqrstu} L_{stu}^{pqr}\ a_p^\dagger a_q^\dagger a_r^\dagger \tilde{a}_u \tilde{a}_t \tilde{a}_s,\end{aligned} \tag{10}$$

where

$$h_q^p = \langle \tilde{\phi}_p | \hat{h} | \phi_q \rangle, \tag{11a}$$

$$V_{rs}^{pq} = \langle \tilde{\phi}_p \tilde{\phi}_q | r_{12}^{-1} - \hat{K}_{12} | \phi_r \phi_s \rangle, \tag{11b}$$

$$L_{stu}^{pqr} = \langle \tilde{\phi}_p \tilde{\phi}_q \tilde{\phi}_r | - \hat{L}_{123} | \phi_s \phi_t \phi_u \rangle. \tag{11c}$$

C. Self-consistent field procedure

Unless one aims to perform a FCI calculation, the choice of the underlying orbitals considered to build the multideterminant expansion for a given approximated correlated treatment may have a significant impact on the overall accuracy and convergence properties. In the case of standard ground-state SCI calculations, one usually relies on mean-field HF orbitals as a starting point, which guarantees the fulfillment of Brillouin's theorem. Regarding TC-SCI calculations, as $\hat{H}_{\mathrm{TC}}$ differs from the bare Coulomb Hamiltonian, the usual HF orbitals are not optimal with respect to the TC effective potential. Relying on non-optimal orbitals may impact the computation of energy differences when enforcing the versatile frozen-core approximations[108] and can also deteriorate the convergence of TC-SCI calculations, as we will discuss in Sec. III C.

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

To address these challenges, the Brillouin theorem is generalized to the non-Hermitian case,[96–99,101–103] which implies two distinct left and right conditions. In the case of a closed-shell system, Brillouin's theorem translates into finding a pair of Slater determinants, $\tilde{\Phi}_0$ and Φ_0, built with the occupied orbitals of the sets $\{\tilde{\phi}_p\}$ and $\{\phi_p\}$, respectively, fulfilling the following conditions:

$$\langle\tilde{\Phi}_0|a_i^\dagger\tilde{a}_a\hat{H}_{\rm TC}|\Phi_0\rangle = 0, \tag{12a}$$

$$\langle\tilde{\Phi}_0|\hat{H}_{\rm TC}a_a^\dagger\tilde{a}_i|\Phi_0\rangle = 0. \tag{12b}$$

The sets of orbitals $\{\tilde{\phi}_p\}$ and $\{\phi_p\}$ are necessarily biorthogonal and can be obtained through the TC self-consistent field (SCF) procedure as detailed in Ref. 99.

This non-linear optimization process is carried out iteratively by diagonalizing the non-symmetric TC Fock matrix, which reads

$$F_q^p \equiv h_q^p + \sum_i \bar{V}_{iq}^{ip} + \frac{1}{2}\sum_{ij}\bar{L}_{ijq}^{ijp}, \tag{13}$$

where we have introduced the antisymmetrized set of integrals,

$$\bar{V}_{rs}^{pq} \equiv V_{rs}^{pq} - V_{sr}^{pq}, \tag{14a}$$

$$\bar{L}_{stu}^{pqr} \equiv L_{stu}^{pqr} - L_{sut}^{pqr} + L_{tus}^{pqr} - L_{tsu}^{pqr} + L_{ust}^{pqr} - L_{uts}^{pqr}. \tag{14b}$$

At the convergence of the TC-SCF procedure, the left and right TC-SCF orbitals are the left and right eigenvectors of the non-Hermitian TC Fock matrix defined in Eq. (13).

D. Normal-ordering with biorthogonal orbitals

Storing three-electron integrals poses a significant computational burden, demanding extensive memory resources in practice. To handle the three-electron interaction term in $\hat{H}_{\rm TC}$ (see Sec. II A), we apply the normal-ordering technique to the three-body operator,

$$\hat{L} = \frac{1}{3!}\sum_{pqrstu} L_{stu}^{pqr}a_p^\dagger a_q^\dagger a_r^\dagger\tilde{a}_u\tilde{a}_t\tilde{a}_s, \tag{15}$$

written in a biorthogonal basis, thus generalizing recent work performed in the context of an orthogonal basis.[125]

In our case, the Slater determinants $\tilde{\Phi}_0$ and Φ_0 obtained through the previous TC-SCF procedure are defined as the Fermi vacua. More specifically, we have

$$a_p^\dagger\tilde{a}_q = \mathcal{N}\left[a_p^\dagger\tilde{a}_q\right] + \overbrace{a_p^\dagger\tilde{a}_q}, \tag{16}$$

where

$$\overbrace{a_p^\dagger\tilde{a}_q} \equiv \langle\tilde{\Phi}_0|a_p^\dagger\tilde{a}_q|\Phi_0\rangle = \gamma_q^p, \tag{17}$$

and the normal-ordered term is defined such that

$$\langle\tilde{\Phi}_0|\mathcal{N}[a_p^\dagger a_q^\dagger\cdots\tilde{a}_r\tilde{a}_s\cdots]|\Phi_0\rangle = 0. \tag{18}$$

In Eq. (17), one can recognize the transition density matrix γ_q^p, which is the equivalent of the usual density matrix in Hermitian calculations.

By applying the following identity:

$$\begin{aligned} a_p^\dagger a_q^\dagger a_r^\dagger\tilde{a}_u\tilde{a}_t\tilde{a}_s = {} & \mathcal{N}\left[a_p^\dagger a_q^\dagger a_r^\dagger\tilde{a}_u\tilde{a}_t\tilde{a}_s\right] \\ & + \mathcal{N}\left[\overbrace{a_p^\dagger a_q^\dagger a_r^\dagger}\tilde{a}_u\tilde{a}_t\tilde{a}_s\right] + \cdots \\ & + \mathcal{N}\left[\overbrace{a_p^\dagger a_q^\dagger}a_r^\dagger\tilde{a}_u\tilde{a}_t\tilde{a}_s\right] + \cdots \\ & + \mathcal{N}\left[\overbrace{a_p^\dagger a_q^\dagger a_r^\dagger}\tilde{a}_u\tilde{a}_t\tilde{a}_s\right] + \cdots, \end{aligned} \tag{19}$$

considering all possible single, double, and triple contractions, along with the following relationships:

$$\mathcal{N}[a_p^\dagger\tilde{a}_q] = a_p^\dagger\tilde{a}_q - \gamma_q^p, \tag{20a}$$

$$\begin{aligned} \mathcal{N}[a_p^\dagger a_q^\dagger\tilde{a}_s\tilde{a}_r] = {} & a_p^\dagger a_q^\dagger\tilde{a}_s\tilde{a}_r - \gamma_r^p a_q^\dagger\tilde{a}_s + \gamma_r^q a_p^\dagger\tilde{a}_s \\ & - \gamma_s^q a_p^\dagger\tilde{a}_r + \gamma_s^p a_q^\dagger\tilde{a}_r - \gamma_r^q\gamma_s^p + \gamma_s^q\gamma_r^p, \end{aligned} \tag{20b}$$

one ends up with the following form for the three-body operator:

$$\hat{L} = \hat{L}^{(0)} + \hat{L}^{(1)} + \hat{L}^{(2)} + \hat{L}^{(3)}, \tag{21}$$

with

$$\hat{L}^{(0)} = \frac{1}{6}\sum_{ijk}\left(L_{ijk}^{ijk} + 2L_{kij}^{ijk} - 3L_{kji}^{ijk}\right), \tag{22a}$$

$$\hat{L}^{(1)} = \sum_{pq}\tilde{L}_q^p a_p^\dagger\tilde{a}_q, \tag{22b}$$

$$\hat{L}^{(2)} = \frac{1}{2}\sum_{pqrs}\tilde{L}_{rs}^{pq}a_p^\dagger a_q^\dagger\tilde{a}_s\tilde{a}_r, \tag{22c}$$

$$\hat{L}^{(3)} = \frac{1}{6}\sum_{pqrstu}L_{stu}^{pqr}\mathcal{N}[a_p^\dagger a_q^\dagger a_r^\dagger\tilde{a}_u\tilde{a}_t\tilde{a}_s], \tag{22d}$$

and

$$\tilde{L}_q^p = -\frac{1}{2}\sum_{ij}\left(L_{qij}^{pij} - L_{qji}^{pij} + 2L_{ijq}^{pij} - 2L_{jiq}^{pij}\right), \tag{23a}$$

$$\tilde{L}_{rs}^{pq} = \sum_i\left(L_{irs}^{ipq} - L_{ris}^{ipq} - L_{sri}^{ipq}\right), \tag{23b}$$

where we have used the fact that, in the case of single-determinant Fermi vacua, the only non-zero elements of the transition density matrix are $\gamma_i^i = 1$. Therefore, the three-body operator can be written as a sum of a scalar quantity, $\hat{L}^{(0)} = \langle\tilde{\Phi}_0|\hat{L}|\Phi_0\rangle$, one- and two-body operators, $\hat{L}^{(1)}$ and $\hat{L}^{(2)}$, together with a three-body operator involving normal-ordered annihilation and creation operators. Since, by definition, $\langle\tilde{\Phi}_0|\hat{L}^{(3)}|\Phi_0\rangle = 0$, a reasonable assumption is

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

that the effect of the pure three-body normal-ordered operator $\hat{L}^{(3)}$ is negligible. Therefore, the normal-ordered approximation,

$$\hat{L} \approx \hat{L}^{(0)} + \hat{L}^{(1)} + \hat{L}^{(2)}, \tag{24}$$

typically results in small errors and effectively reduces the TC Hamiltonian to a two-body operator,

$$\hat{H}_{\mathrm{TC}} = \hat{L}^{(0)} + \sum_{pq} \tilde{h}^p_q a^\dagger_p \tilde{a}_q + \frac{1}{2!}\sum_{pqrs} \tilde{V}^{pq}_{rs} a^\dagger_p a^\dagger_q \tilde{a}_s \tilde{a}_r, \tag{25}$$

with new effective one- and two-body integrals defined as

$$\tilde{h}^p_q = h^p_q + \tilde{L}^p_q, \qquad \tilde{V}^{pq}_{rs} = V^{pq}_{rs} + \tilde{L}^{pq}_{rs}, \tag{26}$$

where $\tilde{L}^p_q$ and $\tilde{L}^{pq}_{rs}$ are given by Eqs. (23a) and (23b), respectively.

E. Selected configuration interaction algorithm

In the TC framework, the optimal ground-state coefficients for a selected space of Slater determinants, referred to as the SCI wave function, are simply the ground-state left and right eigenvectors of the matrix representation of $\hat{H}_{\mathrm{TC}}$ within this space. This approach offers several advantages over other optimization schemes, such as those based on the expensive and noisy VMC method, making the TC theory highly appealing. Furthermore, an intriguing possibility is to go a step further and directly tailor the space of determinants to adapt to the correlation factor. This can be achieved through TC-SCI algorithms, which extend standard SCI versions.

The "*Configuration Interaction using a Perturbative Selection made Iteratively*" (CIPSI) algorithm[27] is highly efficient for constructing the space of determinants and providing accurate estimates of the FCI energy with compact wave functions. Recently, we extended the CIPSI method to accommodate the TC Hamiltonian.[106,108] In this section, we outline the TC-CIPSI algorithm, summarizing its key steps as follows:

1. Begin with a zeroth-order wave function $|\Psi^0\rangle = \sum_{I\in\mathcal{I}} c_I |D_I\rangle$ within a selected space of determinants $\mathcal{I}$, along with its dual $\langle\tilde{\Psi}^0| = \sum_{I\in\mathcal{I}} \tilde{c}_I \langle\tilde{D}_I|$, satisfying the following conditions:

$$\boldsymbol{H}_{\mathrm{TC}}|\Psi^0\rangle = E^{(0)}_{\mathrm{TC}}|\Psi^0\rangle, \tag{27a}$$

$$\boldsymbol{H}^\dagger_{\mathrm{TC}}|\tilde{\Psi}^0\rangle = E^{(0)}_{\mathrm{TC}}|\tilde{\Psi}^0\rangle, \tag{27b}$$

$$\langle\tilde{\Psi}^0|\Psi^0\rangle = 1, \tag{27c}$$

   where $\boldsymbol{H}_{\mathrm{TC}}$ is the matrix representation of $\hat{H}_{\mathrm{TC}}$ in the biorthogonal basis of selected determinants.
2. Compute the energetic contributions of all determinants outside $\mathcal{I}$,

$$e^{(2)}_\alpha = \frac{\langle\tilde{\Psi}^0|\hat{H}_{\mathrm{TC}}|D_\alpha\rangle\langle\tilde{D}_\alpha|\hat{H}_{\mathrm{TC}}|\Psi^0\rangle}{E^{(0)}_{\mathrm{TC}} - \langle\tilde{D}_\alpha|\hat{H}_{\mathrm{TC}}|D_\alpha\rangle}, \tag{28}$$

   which all together form the PT2 energy,

$$E^{(2)}_{\mathrm{TC}} = \sum_{\alpha\notin\mathcal{I}} e^{(2)}_\alpha. \tag{29}$$

   This second-order perturbative energy is computed using an efficient stochastic implementation, as proposed in Ref. 126. Because we rely on the normal-ordering approximation of Eq. (24), $\hat{H}_{\mathrm{TC}}$ remains a two-body operator, and therefore, in practice, the determinants $|D_\alpha\rangle$ and $|\tilde{D}_\alpha\rangle$ for which $e^{(2)}_\alpha \neq 0$ correspond to the singly and doubly excited determinants with respect to any selected determinant, as in usual SCI calculations.
3. Choose a new ensemble of determinants $\mathcal{A}$, identified by the largest energy contributions $|e^{(2)}_\alpha|$. Furthermore, when opting for a particular determinant, all other determinants within the corresponding configuration state function are automatically incorporated, ensuring pure spin states.[127]
4. Update the zeroth-order space $\mathcal{I} \leftarrow \mathcal{I}\cup\mathcal{A}$, and apply a non-symmetric Davidson algorithm[128] to update $\tilde{\Psi}^0$, Ψ^0, and $E^{(0)}_{\mathrm{TC}}$.
5. Repeat steps 2–4 until convergence is achieved.

We would like to highlight some of the subtleties of the TC-SCI algorithm that arise from the non-Hermitian character of $\hat{H}_{\mathrm{TC}}$. First, regarding the set of determinants $\{D_I\}$ and $\{\tilde{D}_I\}$, it is important to note that both are constructed from determinants with identical *orbital occupancies*. As a consequence, when working with a unique set of orthogonal orbitals, $\{D_I\}$ and $\{\tilde{D}_I\}$ are strictly identical, whereas when using a couple of biorthogonal orbital sets, they differ as each set of determinants is built with a different set of orbitals. Nevertheless, the biorthogonality relation between orbitals implies a biorthogonality relation between the determinants (see Ref. 108 for a more detailed discussion). Second, as shown in Ref. 106, the use of $e^{(2)}_\alpha$ [see Eq. (28)] corresponds to a balanced selection criterion for both the left and right eigenvectors as it involves both the left and right zeroth-order wave functions $\tilde{\Psi}^0$ and Ψ^0. Third, the use of the absolute value of $e^{(2)}_\alpha$ as a selection criterion is mandatory, as the expression of Eq. (28) is not necessarily negative, contrary to the Hermitian case.

To conclude this section, we briefly present the extrapolation scheme used in the context of TC-SCI, allowing us to estimate the TC-FCI energy.[108] The procedure is the direct application of the scheme initially proposed in Ref. 52 for the Hermitian case. For small enough $E^{(2)}_{\mathrm{TC}}$ values, one can approximate the TC-FCI energy by

$$E_{\mathrm{TC-FCI}} \approx E^{(0)}_{\mathrm{TC}} + E^{(2)}_{\mathrm{TC}}, \tag{30}$$

the TC-FCI limit being effectively and rigorously reached when $E^{(2)}_{\mathrm{TC}} = 0$. This implies that, as $E^{(2)}_{\mathrm{TC}} \to 0$, the zeroth-order energy $E^{(0)}_{\mathrm{TC}}$ becomes a linear function of the second-order energetic correction $E^{(2)}_{\mathrm{TC}}$. At each TC-SCI iteration, we compute both $E^{(0)}_{\mathrm{TC}}$ and $E^{(2)}_{\mathrm{TC}}$, and when the linear regime is reached, $E_{\mathrm{TC\text{-}FCI}}$ is estimated by linearly fitting the former as a function of the latter.

The main difference with the usual SCI extrapolation is the presence of both positive and negative contributions in $E^{(2)}_{\mathrm{TC}}$, which implies that the latter is not necessarily negative. Therefore, $E^{(0)}_{\mathrm{TC}}$ is not necessarily a monotonic function of $E^{(2)}_{\mathrm{TC}}$. In some cases, these two aspects can lead to situations where it is difficult to accurately estimate the TC-FCI energy, as shall be exemplified below.

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

F. Correlation factor

Various forms of correlation factors have been proposed in the literature, from the simplest universal two-body correlation factor based on Gaussian geminals[100,129,130] to more elaborate forms including electron–electron–nucleus terms.[62,67,99,131] In the present investigation, we consider the correlation factor proposed by Boys and Handy,[67] which we re-express as follows:

$$u_{ij} = \sum_{A} \sum_{p_A=1}^{P_A} C_{m_{p_A}}^{n_{p_A}} c_{p_A} [f_{\alpha_A}(r_{ij})]^{\ell_{p_A}} g_{\beta_A}^{m_{p_A} n_{p_A}}(r_{iA}, r_{jA}) \tag{31}$$

with

$$f_\alpha(x) = \frac{\alpha x}{1+\alpha x}, \tag{32a}$$

$$g_\alpha^{mn}(x,y) = [f_\alpha(x)]^m [f_\alpha(y)]^n + [f_\alpha(y)]^m [f_\alpha(x)]^n. \tag{32b}$$

For each nucleus, we employ a total of P_A terms, each term p_A being characterized by three positive integers $(\ell_{p_A}, m_{p_A}, n_{p_A})$ and a coefficient c_{p_A}.

Following the work of Schmidt and Moskowitz,[132] we set $\alpha_A = \beta_A = 1$ for all nuclei. In addition, we have

$$C_m^n = \begin{cases} 1/2, & \text{if } m = n, \\ 1, & \text{otherwise.} \end{cases} \tag{33}$$

Furthermore, we enforce the electron–electron cusp by systematically including in the parameterization the term characterized by $m_1 = n_1 = 0$ and $\ell_1 = 1$, with the corresponding coefficient $c_1 = 1/2$ (see Ref. 132 for more details).

It is important to highlight the three complementary effects introduced by the parameterization of Eq. (31). First, the terms characterized by $\ell > 0$ and $m = n = 0$ correspond to pure electron–electron terms, which explicitly depend on the inter-electronic distances. Their main purpose is to model the correlation hole, that is, the lowering of the probability of finding two electrons close together due to electron correlation effects. Second, for $\ell = m = 0$ and $n > 0$ (or $\ell = n = 0$ and $m > 0$), one introduces pure electron–nucleus terms, which provide flexibility to the one-electron density for its adjustment to accommodate the presence of the correlation hole. Third, components associated with $\ell > 0$, $m > 0$, and $n \geq 0$ (or $\ell > 0$, $m \geq 0$, and $n > 0$) include electron–electron–nucleus terms, effectively coupling the electron–electron and electron–nucleus terms previously mentioned. These three-body terms facilitate an even more accurate representation of the correlation hole.

In the present study, we rely on the parameterization of the correlation factor, optimized within a VMC framework at the single-determinant level, as reported in Refs. 132 and 133.

G. Integral evaluation

In this section, we focus on the computation of integrals necessary for TC calculations. In addition to the standard one- and two-electron integrals, one must also compute the following additional set of two-electron integrals:

$$K_{rs}^{pq} \equiv -\langle \tilde{\phi}_p \tilde{\phi}_q | \hat{K}_{12} | \phi_r \phi_s \rangle, \tag{34}$$

where $\hat{K}_{12}$ is defined in Eq. (5). In addition, three-electron integrals, as defined in Eq. (11c), must be computed.

The two-electron integrals are computed in the atomic orbital basis $\{\chi_\mu\}$, as follows:

$$K_{\eta\zeta}^{\mu\nu} \equiv -\langle \chi_\mu \, \chi_\nu | \hat{K}_{12} | \chi_\eta \, \chi_\zeta \rangle. \tag{35}$$

These integrals are subsequently transformed into orbital bases to obtain Eq. (34).

By exploiting the symmetry of the operator $\hat{K}_{12}$ with respect to electron exchange ($\hat{K}_{12} = \hat{K}_{21}$) and employing integration by parts, we obtain

$$K_{\lambda\sigma}^{\mu\nu} = \mathcal{K}_{\lambda\sigma}^{\mu\nu} + \mathcal{K}_{\sigma\lambda}^{\nu\mu}, \tag{36}$$

where we introduce

$$\begin{aligned} \mathcal{K}_{\lambda\sigma}^{\mu\nu} = \frac{1}{2} \int d\boldsymbol{r} [\chi_\mu(\boldsymbol{r}) \nabla \chi_\lambda(\boldsymbol{r}) - \nabla \chi_\mu(\boldsymbol{r}) \chi_\lambda(\boldsymbol{r})] \cdot \boldsymbol{G}_\sigma^\nu(\boldsymbol{r}) \\ + \frac{1}{2} \int d\boldsymbol{r} \chi_\mu(\boldsymbol{r}) \chi_\lambda(\boldsymbol{r}) J_\sigma^\nu(\boldsymbol{r}). \end{aligned} \tag{37}$$

The three-dimensional integrals $\boldsymbol{G}_\sigma^\nu$ and the scalar integrals J_σ^ν are defined by

$$\boldsymbol{G}_\sigma^\nu(\boldsymbol{r}_1) \equiv \int d\boldsymbol{r}_2 \chi_\nu(\boldsymbol{r}_2) \chi_\sigma(\boldsymbol{r}_2) \nabla_1 u_{12}, \tag{38a}$$

$$J_\sigma^\nu(\boldsymbol{r}_1) \equiv \int d\boldsymbol{r}_2 \chi_\nu(\boldsymbol{r}_2) \chi_\sigma(\boldsymbol{r}_2) [\nabla_1 u_{12}]^2. \tag{38b}$$

The three-electron integrals (11c) are computed directly in the orbital basis and are given by

$$\begin{aligned} L_{stu}^{pqr} = \int d\boldsymbol{r} \big[\tilde{\phi}_p(\boldsymbol{r}) \phi_s(\boldsymbol{r}) \boldsymbol{G}_t^q(\boldsymbol{r}) \cdot \boldsymbol{G}_u^r(\boldsymbol{r}) \\ + \tilde{\phi}_q(\boldsymbol{r}) \phi_t(\boldsymbol{r}) \boldsymbol{G}_s^p(\boldsymbol{r}) \cdot \boldsymbol{G}_u^r(\boldsymbol{r}) \\ + \tilde{\phi}_r(\boldsymbol{r}) \phi_u(\boldsymbol{r}) \boldsymbol{G}_s^p(\boldsymbol{r}) \cdot \boldsymbol{G}_t^q(\boldsymbol{r}) \big]. \end{aligned} \tag{39}$$

Note that $\boldsymbol{G}_q^p(\boldsymbol{r})$ corresponds to the potential (38a) in the orbital basis.

As the integrals previously defined cannot be obtained in closed form except for specific forms of correlation factors,[65,134] we compute these integrals numerically using Becke's quadrature grid.[135,136] Nevertheless, as the integrands in Eqs. (37) and (39) are much smoother than those appearing in the definition of the potentials in Eqs. (38a) and (38b), we employ two distinct grids: one for evaluating integrals over $\boldsymbol{r}_2$ [as in Eqs. (38a) and (38b)], which is denser than the one used for evaluating integrals over $\boldsymbol{r}$ [as in Eqs. (37) and (39)]. This approach allows for an efficient computation of two- and three-electron integrals via dense matrix–matrix multiplications using BLAS routines.

H. Computational cost

Here, we discuss some considerations regarding the computational overhead associated with a TC-SCI calculation compared to its Hermitian version.

Regarding strictly the SCI part, the dominant computational costs stem from the evaluation of the PT2 correction and the

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

Davidson diagonalization. These steps are limited by two operations: (i) applying the Slater–Condon rules between determinants (i.e., finding the particle–hole excitation operators together with the fermionic phase factor) and (ii) accessing and combining the integrals needed for the computation of matrix elements. Thanks to the normal-ordering approximation presented in Sec. II D, $\hat{H}_{\text{TC}}$ remains a two-body operator. Therefore, the Slater–Condon rules are unchanged and can be very efficiently computed.[137] In the present implementation, the left and right eigenvectors are computed separately, leading to an overall increase in computational time by a factor of 2.

Except for the nonsymmetric Davidson diagonalization for large determinant spaces, the computation of the two- and three-electron integrals is the main computational bottleneck of the TC-CIPSI approach. To avoid storing these integrals, which would require $\mathcal{O}N^6$ storage (where N is the number of basis functions), they are computed on the fly and contracted to form the operators defined in Eqs. (22a), (22b), and (22c). $\hat{L}^{(0)}$, $\hat{L}^{(1)}$, and $\hat{L}^{(2)}$ require O^3, O^2N^2, and ON^4 integrals, respectively, while the construction of the TC Fock operator involves O^2N^2 integrals [see Eq. (13)], where O is the number of occupied orbitals. To compute L^{pqr}_{stu}, one first evaluates the potentials $\boldsymbol{G}^q_p(\boldsymbol{r})$ on each grid point for each pair of orbitals, which necessitates KN^2 storage (where K is the number of grid points). Then, these values are employed in Eqs. (22a), (22b), and (22c) to compute the required integral batches.

III. RESULTS AND DISCUSSION

A. Computational details

The calculations reported in this paper were performed with the QUANTUM PACKAGE.[39] In the subsequent calculations, we employ either the 6-31G split-valence basis sets[138–140] or Dunning's cc-pVXZ and cc-pCVXZ basis sets, where X denotes D, T, Q, and 5.[141,142] The molecular geometries were extracted from the QUEST database[59,143,144] or experimental data[145] and are provided in the supplementary material. Unless otherwise stated, calculations are performed in the frozen-core approximation.

The Jastrow parameters were obtained from two sources: Refs. 132 and 133. In particular, for atoms, we utilized the parameters outlined in Table VI of Ref. 132, whereas for molecules, the Jastrow parameters were sourced from Table 2 of Ref. 133. Mean-field calculations on closed-shell systems are performed in the restricted formalism, while the restricted open-shell formalism is employed for open-shell systems.[99] For the latter case, an average of integrals corresponding to spin-up and spin-down densities is employed in the normal-ordering procedure. In the calculations presented herein, we use a grid with 30 radial and 50 angular points for each atom over $\boldsymbol{r}$ and 70 radial and 266 angular points over $\boldsymbol{r}_2$. These grid configurations ensure the stability of energy calculations to within at least 0.1 mEh. All calculations were carried out on a single dual-socket node equipped with 36 Intel Skylake 6140 cores, each running at 2.3 GHz.

B. Weights of the Slater determinants

We begin this study by comparing the weight c_I^2 of a given Slater determinant I (built with HF orbitals) in the ground-state FCI wave function and the pseudo-weight $|\tilde{c}_I c_I|$ of the same determinant in the ground-state TC-FCI wave function. As a first example, we consider the H_2O molecule in the 6-31G basis set, as this relatively small basis set enables the calculation of both the FCI and TC-FCI wave functions.

We report in the left panel of Fig. 1 the FCI weights c_I^2 associated with the 500 most important Slater determinants sorted by decreasing weights, as well as the corresponding TC-FCI pseudo-weights $|\tilde{c}_I c_I|$. In addition, the level of excitation with respect to the reference determinant is indicated by different colors. An important observation is that the pseudo-weights do not decrease monotonically, unlike the weights. This indicates that when using the TC Hamiltonian, the determinants are not selected in the same order as with the standard Hamiltonian when using the CIPSI algorithm. In addition, the pseudo-weights are globally smaller than the FCI weights,

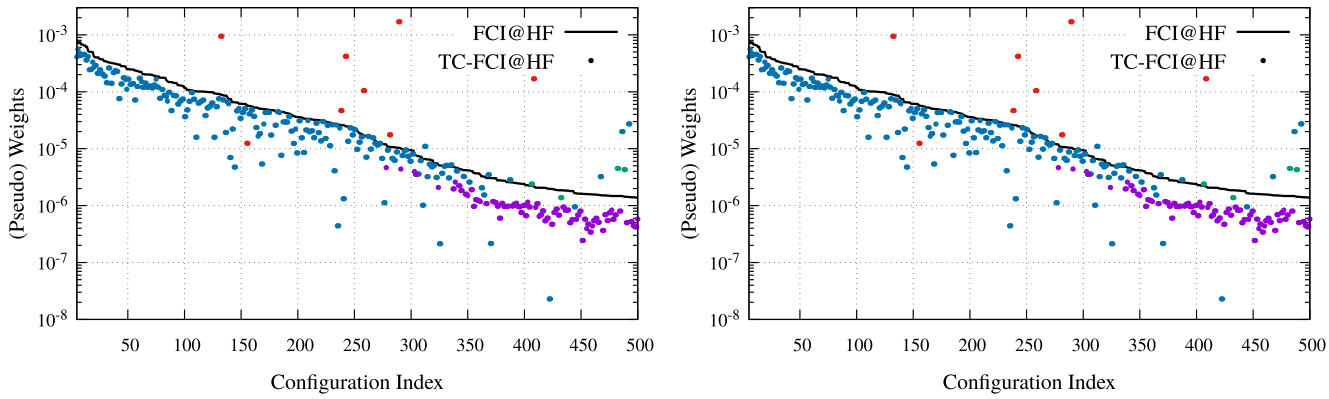

FIG. 1. Weights (c_I^2) of the FCI wave function (solid black line) and pseudo-weights ($|\tilde{c}_I c_I|$) of the TC-FCI wave function (colored markers) for H_2O in the 6-31G basis set. The excitation degree of each determinant with respect to the mean-field reference determinant is indicated by the following color code: red, blue, green, and purple for single, double, triple, and quadruple excitations, respectively. The left panel displays the weight distribution of the determinants sorted by c_I^2, using HF orbitals to construct the determinants. The right panel reports similar quantities, but TC-SCF orbitals are employed for constructing the TC-FCI wave function.

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

except for the single excitations, which are typically much larger in the TC wave function. The latter observation can be explained by the fact that while HF orbitals fulfill the Brillouin theorem for the standard Hamiltonian, it is no longer the case for the TC Hamiltonian as the effective interaction is no longer the bare Coulomb repulsion.

We report, in the right panel of Fig. 1, the weight distribution of the TC wave function using TC-SCF orbitals, which, therefore, fulfills both the left and right Brillouin conditions [see Eqs. (12a) and (12b)]. Consequently, the pseudo-weights of the single excitations are significantly reduced. For the sake of completeness, we should mention that in the latter case, as the orbitals are different, the determinants in the FCI and TC-FCI wave functions are no longer the same, but here we simply match a given determinant from the two wave functions with its orbital occupation, assuming that the physical meaning of the orbitals weakly changes with the TC-SCF optimization (which is clearly the case here).

More quantitatively, there are 2590 determinants with a weight larger than 10^{-7} in the FCI wave function, while there are only 1998 and 1704 determinants with a pseudo-weight larger than 10^{-7} in the TC wave function using the HF and TC-SCF orbitals, respectively.

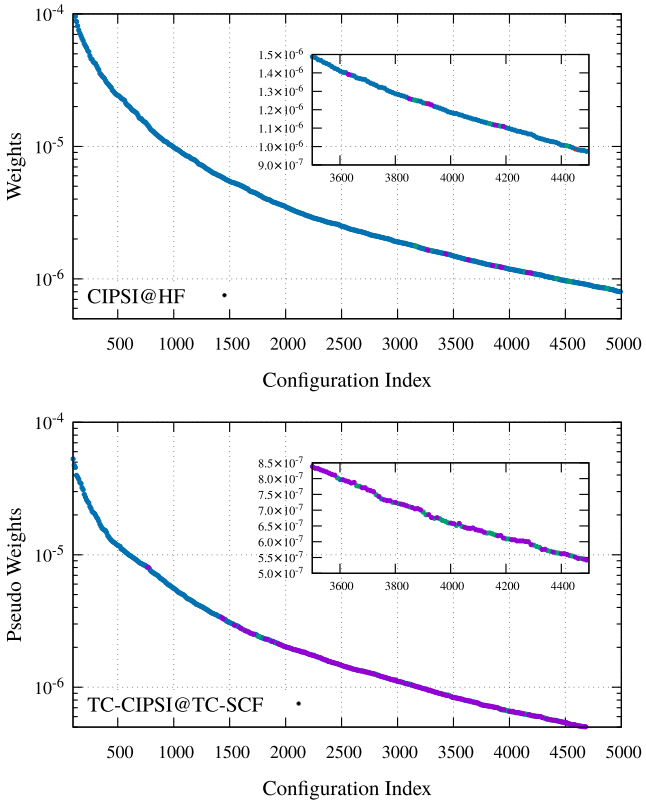

FIG. 2. Weights (c_I^2) or pseudo-weights ($|\tilde{c}_I c_I|$) of the first 5000 determinants composing the CIPSI (left) and TC-CIPSI (right) wave functions, sorted with respect to their corresponding (pseudo) weights, for H_2O in the cc-pVTZ basis set. For the sake of readability, one point out of 10 is represented. The excitation degree of each determinant with respect to the mean-field reference determinant is indicated by the following color code: red, blue, green, and purple for single, double, triple, and quadruple excitations, respectively.

This shows that the correlation factor has the effect of lowering the number of determinants that significantly contribute to energy. This is the first evidence of the compactification of the determinant expansion brought about by the TC approach.

Figure 2 shows similar quantities but in a larger basis set (cc-pVTZ) and computed at the CIPSI and TC-CIPSI levels employing HF and TC-SCF orbitals, respectively. In this case, the determinants are sorted with respect to their corresponding (pseudo) weights in their corresponding wave functions. The same color code is used to indicate the excitation degree of the determinants. In addition, to study the impact of the form of the correlation factor, we performed calculations using the simple correlation factor introduced previously by some of the authors.[107] The latter form consists of a universal two-body correlation factor designed for valence electrons multiplied by atom-centered Gaussian envelopes, suppressing the effect of the correlation factor in the core regions. The nuclear parameters for the Gaussian envelopes, used for these calculations were taken from Ref. 108. These complementary results are reported in the supplementary material.

Three key observations can be made: (i) as above, the pseudo weights are systematically lower than the weights; (ii) triple and higher excitations enter much earlier in the TC-CIPSI expansion, which demonstrates that the correlation factor allows for a significant reduction of the weights of doubly excited determinants; and (iii) the above effects are considerably less pronounced with the simpler correlation factor, highlighting the importance of flexible functional forms.

C. Convergence of the zeroth-order energy and its extrapolation

A critical aspect of the TC-SCI scheme is the convergence of the zeroth-order energy $E_{\mathrm{TC}}^{(0)}$ based on the criterion employed to select Slater determinants, a smooth convergence facilitating its extrapolation toward the TC-FCI limit. Here, we would like to investigate two distinct selection procedures that evidence the impact of the correlation factor: (i) selecting determinants via the second-order contribution computed with the TC Hamiltonian using both left and right eigenvectors [see Eq. (28)] (TC-CIPSI selection), or (ii) performing a standard selection based on the bare Hamiltonian (CIPSI selection).

To illustrate this point, we rely on the same representative example, the H_2O molecule in the cc-pVTZ basis set. We report the convergence of $E_{\mathrm{TC}}^{(0)}$ with respect to the number of Slater determinants in Fig. 3 with the CIPSI and TC-CIPSI selection procedures discussed above. We observe that starting from the HF reference determinant, the CIPSI selection leads rapidly (~3000 determinants) to an energy that is below the estimated TC-FCI energy, while, for expansions larger than 600 000 determinants, $E_{\mathrm{TC}}^{(0)}$ rises to eventually reach this limiting value. If one selects the determinants using the TC-CIPSI algorithm instead, we observe that, from about 20 determinants, the zeroth-order energy monotonically decays and remains above the TC-FCI energy. This clearly highlights the impact of the correlation factor on the selection of the determinants. In addition, Fig. 3 illustrates the convergence of $E_{\mathrm{TC}}^{(0)}$ using TC-SCF orbitals as opposed to HF orbitals, highlighting the impact of orbital optimization in the TC framework. One can see that although the

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

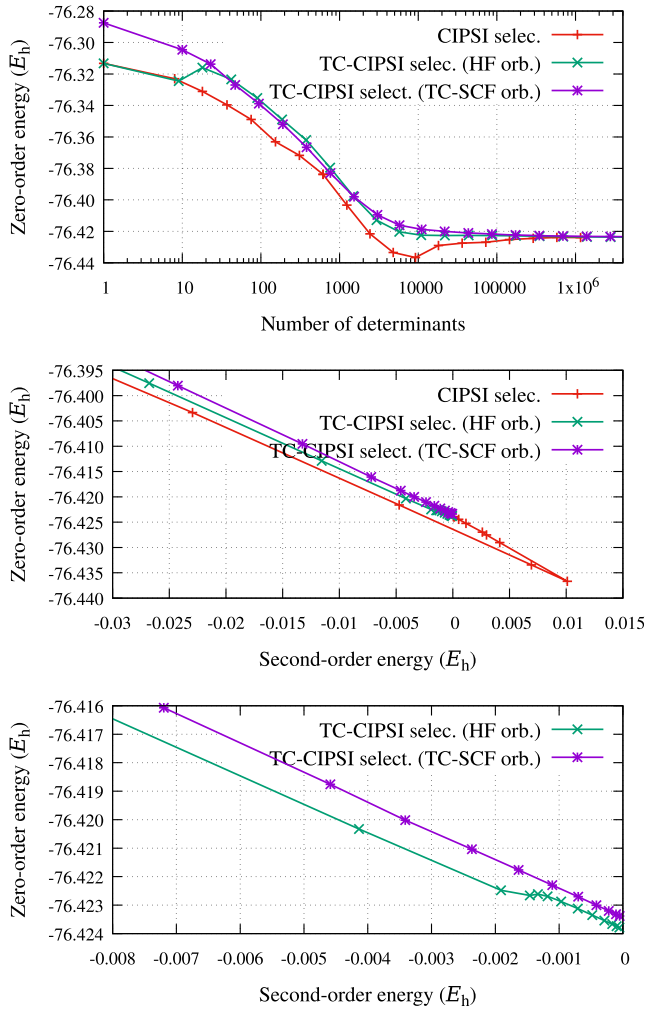

FIG. 3. Convergence of $E_{TC}^{(0)}$ as a function of the number of selected determinants (top) or $E_{TC}^{(2)}$ (center and bottom) for H_2O computed on the cc-pVTZ basis and obtained via the CIPSI and TC-CIPSI selection procedures. In the case of TC-CIPSI, HF or TC-SCF orbitals are employed. The bottom panel corresponds to a zoom of the central panel near the origin.

TC-SCF reference determinant is higher in energy, $E_{TC}^{(0)}$ rapidly and monotonically decays toward the TC-FCI energy.

In Fig. 3, we also report the evolution of $E_{TC}^{(0)}$ as a function of the second-order energy correction $E_{TC}^{(2)}$ [see Eq. (29)] for the various cases investigated here. This helps us to appreciate the different behaviors of $E_{TC}^{(0)}$ from the perspective of its extrapolation. On the one hand, we observe that the erratic behavior of $E_{TC}^{(0)}$ obtained from the CIPSI selection makes it hard to extrapolate $E_{TC}^{(0)}$ to the TC-FCI energy, due to the presence of a "turning" point stemming from the interplay between the positive and negative contributions of $E_{TC}^{(2)}$, as mentioned in Sec. II E. On the other hand, using the TC-CIPSI selection leads to a globally linear curve. However, when one employs HF orbitals, we observe a sudden change in slope as $E_{TC}^{(2)} \to 0$, leading also to a possible untrustworthy extrapolation. However, using TC-SCF orbitals, $E_{TC}^{(0)}$ has a consistent linear behavior on a larger range of $E_{TC}^{(2)}$, allowing us to estimate the TC-FCI energy more safely. It is worth mentioning the small difference (roughly 0.4 mEh) in the TC-FCI limiting values obtained with HF and TC-SCF orbitals, which is due to the normal-ordering approximation that necessarily creates a weak dependence on the orbitals.

We can then conclude that the best strategy to perform a TC-SCI calculation is to rely on TC-SCF orbitals in combination with the TC-CIPSI algorithm for the determinant selection. In such a way, one obtains rapidly and monotonically convergent zeroth-order energies that permit a reliable extrapolation to the TC-FCI limit. From hereon, TC-CIPSI calculations are systematically performed with TC-SCF orbitals.

D. Compactification of the CI expansion

To demonstrate the disparity in convergence rate between energies computed via the CIPSI and TC-CIPSI methods, we conduct calculations on the water molecule using increasingly large Dunning basis sets (cc-pVDZ, cc-pVTZ, and cc-pVQZ). In order to quantify the acceleration brought by the TC approach, we primarily focus on two indicators of convergence. The first criterion corresponds to the number of determinants required to reach a specified value of the second-order energies, $E^{(2)}$ for CIPSI and $E_{TC}^{(2)}$ for TC-CIPSI. The second criterion involves achieving stable extrapolations toward the FCI and TC-FCI limits, respectively, through fitting functions represented as

$$E^{(0)}(E^{(2)}) = aE^{(2)} + b, \tag{40a}$$

$$E_{TC}^{(0)}(E_{TC}^{(2)}) = a_{TC}E_{TC}^{(2)} + b_{TC}. \tag{40b}$$

The data associated with the following analysis can be found in the supplementary material.

The convergence behavior of the zeroth-order energies, $E^{(0)}$ and $E_{TC}^{(0)}$, is depicted in Fig. 4. The corresponding extrapolated FCI and TC-FCI energies, E_{FCI} and $E_{TC\text{-}FCI}$, are presented by shaded lines. These plots illustrate that, as the number of selected determinants increases, $E_{TC}^{(0)}$ converges toward $E_{TC\text{-}FCI}$ at a notably faster rate compared to the convergence of $E^{(0)}$ toward E_{FCI}. A similar trend is observed for the second-order perturbative energies, as shown in Fig. 5, where we report the evolution of $E^{(2)}$ and $E_{TC}^{(2)}$ as functions of the number of determinants.

More quantitatively, achieving chemical accuracy, which corresponds to values of E_{PT2} below 1.5 mEh, requires ~96 883, 3 105 054, and 23 609 437 determinants in the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets, respectively. Conversely, in the TC case, reaching $E_{PT2}^{TC} \approx 1.5$ mEh occurs much earlier, with ~12 195, 148 821, and 781 281 determinants in these bases, indicating a compactification of the CI expansion by factors of ~8, 21, and 30, respectively.

In Fig. 6, we depict the evolution of $E^{(0)}$ as a function of $E^{(2)}$, as well as $E_{TC}^{(0)}$ as a function of $E_{TC}^{(2)}$. We perform a linear fit of these data using Eqs. (40a) and (40b) to target the FCI and the TC-FCI energies, respectively. Across all three basis sets, both CIPSI and

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

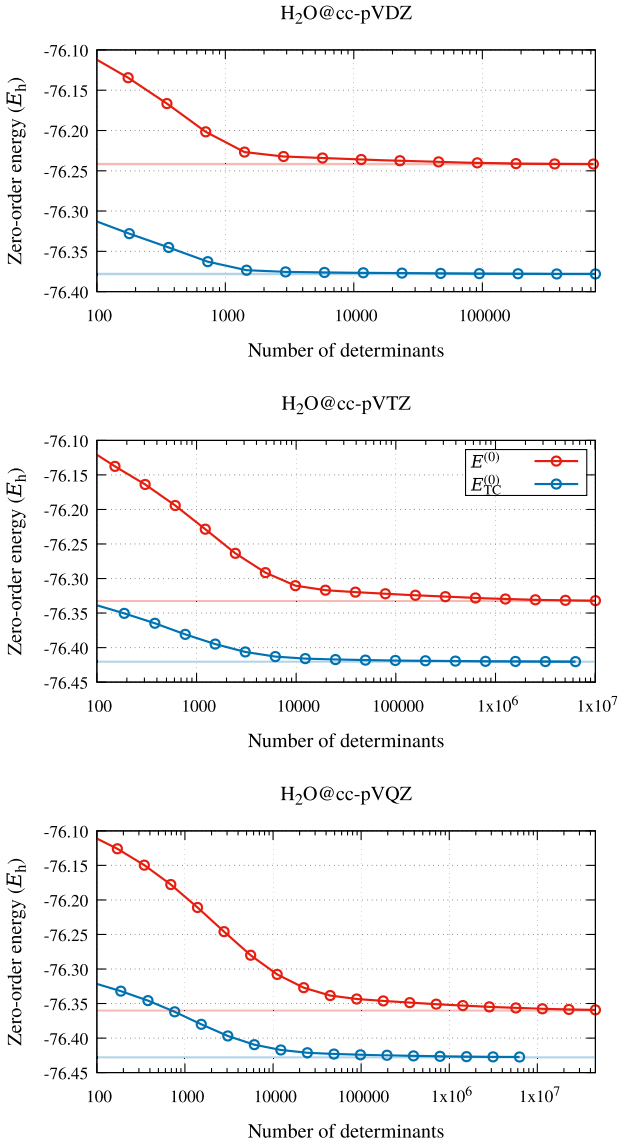

FIG. 4. Zeroth-order energies, $E^{(0)}$ (red) and $E^{(0)}_{TC}$ (blue), as functions of the number of selected determinants for H_2O on the cc-pVDZ (top), cc-pVTZ (center), and cc-pVQZ (bottom) basis. The shaded horizontal lines represent the extrapolated FCI (red) and TC-FCI (blue) energies.

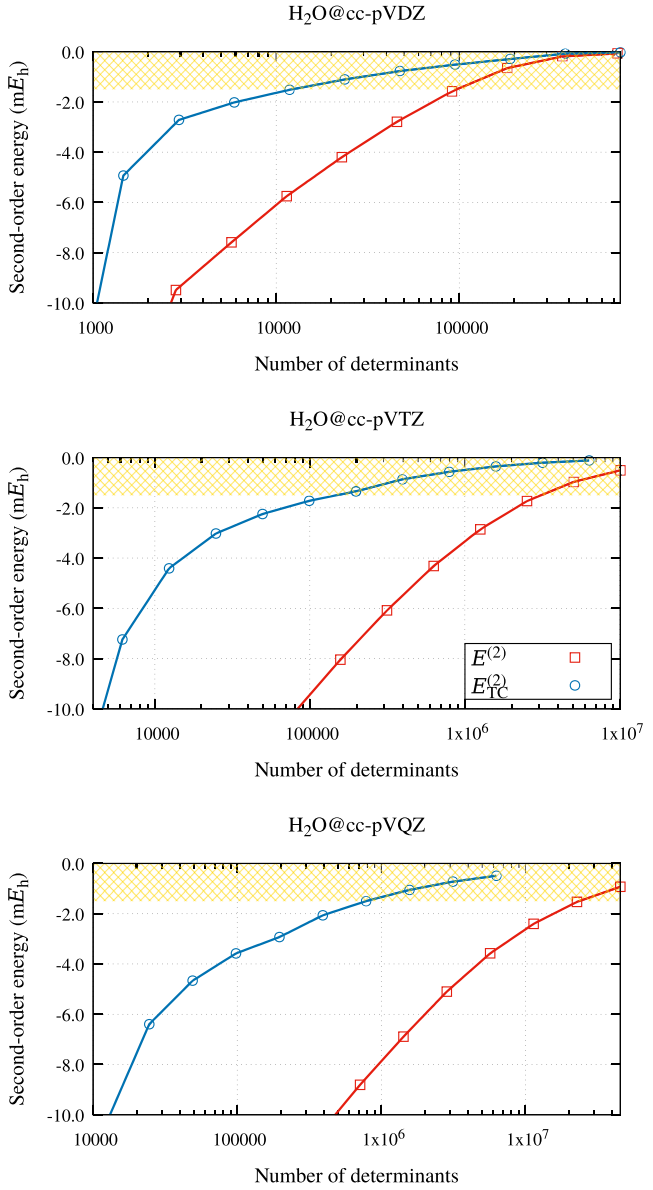

FIG. 5. Second-order perturbative energies, $E^{(2)}$ (red) and $E^{(2)}_{TC}$ (blue), as functions of the number of selected determinants for H_2O in the cc-pVDZ (top), cc-pVTZ (center), and cc-pVQZ (bottom) basis. The shaded yellow region corresponds to 1.5 mEh accuracy.

TC-CIPSI demonstrate stable and smooth linear extrapolations. However, upon closer examination, we observe that the TC-CIPSI extrapolation converges significantly faster.

In Fig. 7, we assess the error on the extrapolated energies as additional iterations are performed (or, equivalently, more determinants are added) with each extrapolation conducted using five consecutive iterations of the CIPSI or TC-CIPSI algorithm. The comparison is carried out for cc-pVDZ, cc-pVTZ, and cc-pVQZ. We establish convergence over extrapolation when performing an additional CIPSI (TC-CIPSI) iteration no longer alters the estimated FCI (TC-FCI) energy by more than 1.5 mEh. Under this criterion, the conventional CIPSI extrapolation necessitates ~2834, 9821, and 22 248 determinants to achieve convergence, while TC-CIPSI requires only 729, 3085, and 3063 in the cc-pVDZ, cc-pVTZ, and cc-pVQZ bases, respectively.

Next, we compare the computational costs of CIPSI and TC-CIPSI calculations. Figure 8 illustrates the convergence rates of the second-order energies as a function of the elapsed wall time (in seconds) for the same collection of basis sets (cc-pVDZ, cc-pVTZ,

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

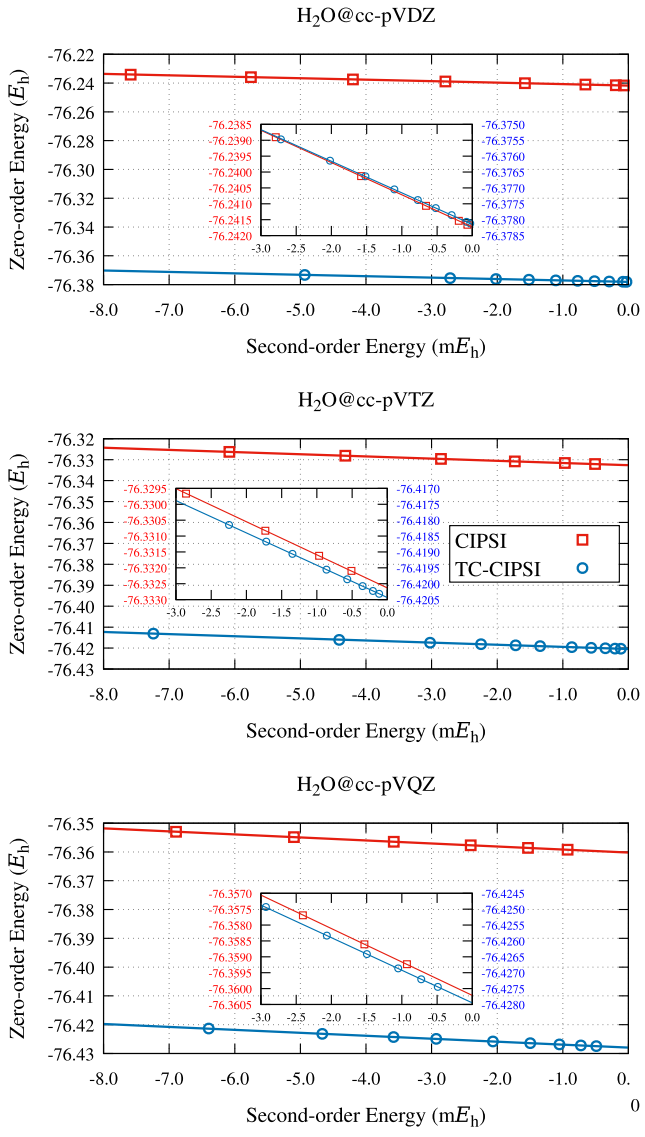

FIG. 6. Linear extrapolations of $E^{(0)}$ (red) and $E^{(0)}_{TC}$ (blue) toward the FCI and TC-FCI limiting energies for H_2O on the cc-pVDZ (top), cc-pVTZ (center), and cc-pVQZ (bottom) basis.

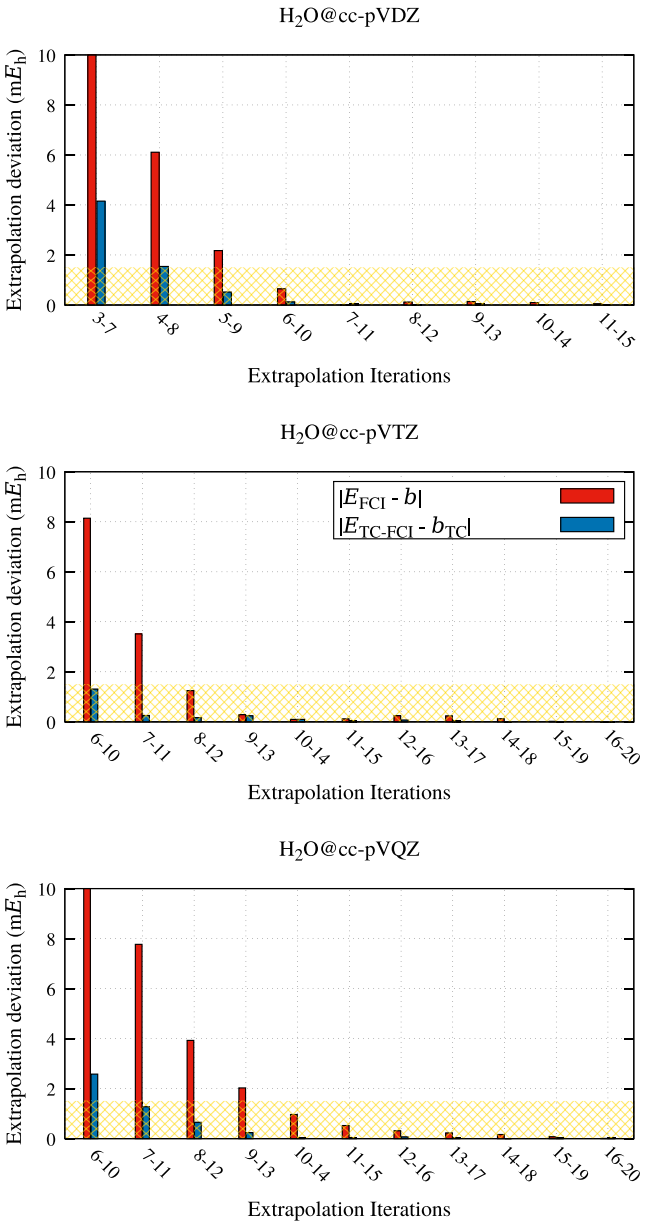

FIG. 7. Error in extrapolated energies as additional iterations (or selected determinants) are performed. The shaded yellow region corresponds to 1.5 m*E*h accuracy.

and cc-pVQZ). In the smallest cc-pVDZ basis set, TC-CIPSI initially appears slower than CIPSI due to the time required for preparing the normal-ordered intermediates, as defined in Eqs. (22a), (22b), and (22c). However, across all basis sets, TC-CISPI demonstrates a faster convergence to chemical accuracy compared to CIPSI. In particular, CIPSI takes 15 s, 22 min, and 14.6 h to reach a chemically accurate energy in cc-pVDZ, cc-pVTZ, and cc-pVQZ, respectively. These times are reduced in the case of TC-CIPSI, requiring only 11 s, 3 min, and 49 min, respectively. Consequently, the overall gain in wall time for chemically accurate energies using TC-CIPSI compared to CIPSI is 1.4, 8, and 18, respectively.

Similar calculations were conducted for Li_2, Be_2, NH_3, CH_4, and H_2CO molecules (the corresponding figures and tables are provided in the supplementary material), yielding comparable findings. The convergence behavior across different basis sets and the comparative analysis of the zeroth-order, second-order, and extrapolated energies between the CIPSI and TC-CIPSI methods consistently yielded similar results for these molecules.

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

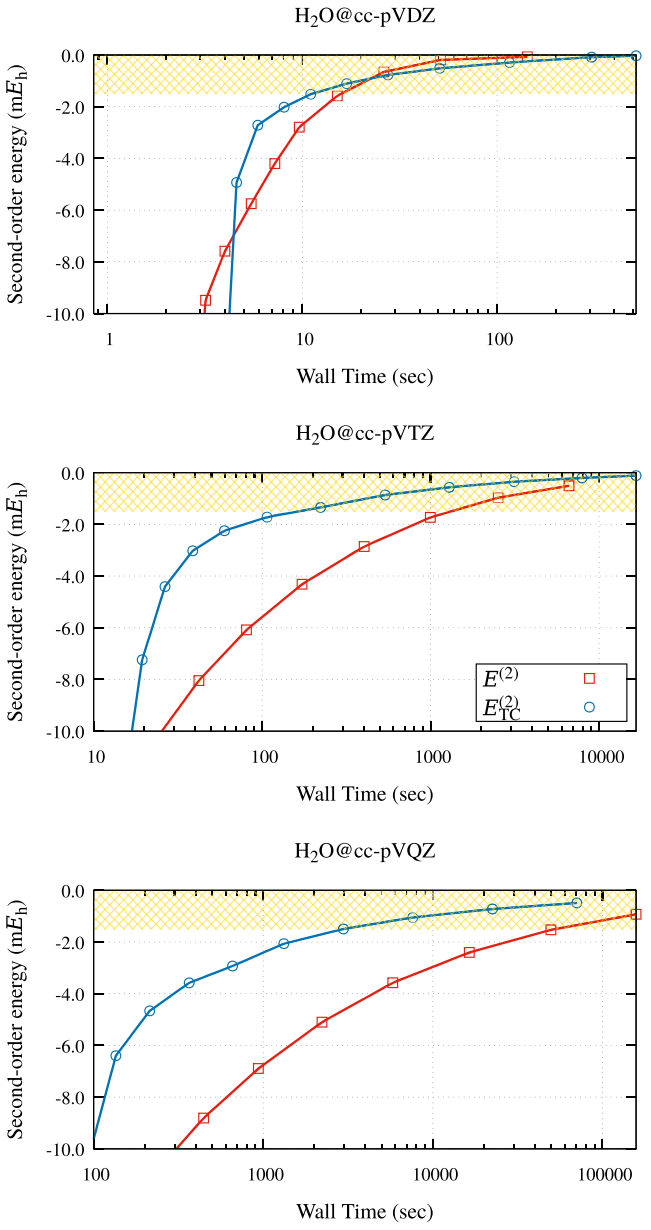

FIG. 8. Second-order perturbative energies, $E^{(2)}$ (red) and $E^{(2)}_{TC}$ (blue), as functions of the elapsed wall time (in seconds) for H_2O on the cc-pVDZ (top), cc-pVTZ (center), and cc-pVQZ (bottom) basis. The shaded yellow region corresponds to 1.5 mEh accuracy.

E. Total energies

Estimating the total nonrelativistic electronic energy of atomic and molecular systems presents significant challenges for standard CI techniques due to the need for very large basis sets to achieve chemical accuracy. However, at the TC-CI level, it is possible to considerably reduce the basis set error. We perform TC-CIPSI calculations without the frozen-core approximation using the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets for the neutral atoms from $Z = 2$ to $Z = 10$ (Table I) and for the following molecules reported in Table II: Li_2, Be_2, H_2O, NH_3, CH_4, and H_2CO. The TC-FCI energies are obtained through extrapolation of the TC-CIPSI energies, ensuring convergence of at least four digits (0.1 mEh). It is worth noting that we opted for calculations in cc-pVXZ rather than cc-pCVXZ basis sets, as we observed that with sufficiently large basis sets (typically when X > 3), the TC-FCI energies obtained in cc-pVXZ and cc-pCVXZ are very close thanks to the effect of the correlation factor. For instance, the energy difference between cc-pVQZ and cc-pCVQZ for systems such as Be, Ne, and H_2O is only 0.1 mEh, 0.6 mEh, and 0.2 mEh, respectively. Such differences are significantly smaller than those observed at the standard FCI level, often by two or three orders of magnitude.

The exact total energies for atomic systems, as estimated in Ref. 146, are also presented in Table I. A remarkable agreement between the TC-FCI and reference energies is observed, particularly with the cc-pVQZ basis set. The mean absolute errors are 32.5 mEh, 9.2 mEh, and 0.9 mEh in cc-pVDZ, cc-pVTZ, and cc-pVQZ, respectively. It is noteworthy that our results closely align with those obtained in Ref. 116 using the TC-FCIQMC algorithm, where three-electron integrals are treated exactly, that is, without normal-ordering. This suggests that the normal-ordering approximation outlined in Sec. II D introduces minimal bias, even in the case of open-shell systems.

Finding exact estimates of nonrelativistic total energies for molecular systems in the literature is significantly more challenging, with very few exceptions. All-electron fixed-node diffusion Monte Carlo (FN-DMC) calculations for the nonrelativistic ground-state energy of H_2O have been reported in Ref. 56 using the cc-pCVXZ basis set family. Consequently, we conducted additional calculations on the water molecule using these basis sets. The comparison between FCI,[147] TC-FCI, and FN-DMC[56] calculations is presented in Table III and visualized in Fig. 9, with the reference exact energy considered as −76.4389 Eh.[148] It is evident that while the standard FCI energy converges very slowly, the TC-FCI energy is found to be very close to the FN-DMC energy in cc-pVTZ and nearly identical in cc-pVQZ. We should nevertheless mention that, in the case of H_2O molecules, the results obtained with the CCSDT-R12 method[149] in the aug-cc-pVXZ basis sets using a correlation factor fulfilling both the singlet and triplet cusp[150] are significantly closer, for a given cardinal number X, to the estimated exact energy than the present TC-FCI results in the cc-pCVXZ basis sets. As the TC-FCI fulfills only the singlet cusp condition, the larger error found in our TC-FCI could be the signature of the lack of a triplet cusp condition, which, according to the work of Kutzelnigg and Morgan,[15] is directly related to the next order of the convergence rate of CI calculations.

F. Ionization potentials

In this section, we address the convergence of IPs (defined as the difference between the cation and neutral ground-state energies) as the one-electron basis set is enlarged. Concerning the computation of IPs within the TC formalism, the optimal approach would be to optimize the Jastrow parameters in a state-specific way. While not ideal, we have employed an alternative, cheaper strategy where the same Jastrow parameters are used for both the neutral and cationic

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

TABLE I. Total energies (in *E*h) obtained from TC-FCI calculations (without frozen-core approximation) in the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets for the neutral atoms from Z = 2 to Z = 10. For the sake of comparison, estimated exact energies are also provided.

Atom	He	Li	Be	B	C	N	O	F	Ne
cc-pVDZ	−2.8975	−7.4771	−14.6681	−24.6451	−37.8275	−54.5569	−75.0163	−99.6586	−128.8375
cc-pVTZ	−2.9033	−7.4781	−14.6681	−24.6499	−37.8390	−54.5795	−75.0527	−99.7139	−128.9099
cc-pVQZ	−2.9037	−7.4785	−14.6676	−24.6527	−37.8443	−54.5880	−75.0658	−99.7329	−128.9360
Exact[a]	−2.9037	−7.4781	−14.6674	−24.6539	−37.8450	−54.5892	−75.0673	−99.7339	−128.9376

[a]Values extracted from Ref. 146.

TABLE II. Total energies (in *E*h) obtained from TC-FCI calculations (without frozen-core approximation) in the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets for Li_2, Be_2, H_2O, NH_3, CH_4, and H_2CO.

Molecule	Li_2	Be_2	H_2O	NH_3	CH_4	H_2CO
cc-pVDZ	−14.9776	−29.3245	−76.3775	−56.5162	−40.5031	−114.4012
cc-pVTZ	−14.9888	−29.3268	−76.4234	−56.5566	−40.5211	−114.4835
cc-pVQZ	−14.9918	−29.3395	−76.4367	−56.5655	−40.5207	−114.5097

TABLE III. Total energies (in *E*h) computed at the FCI, TC-FCI, and FN-DMC levels of theory for the water molecule in the cc-pCVXZ family of basis sets. The estimated exact energy is −76.4389 *E*h.[148]

	FCI[a]	TC-FCI[b]	FN-DMC[c]
cc-pCVDZ	−76.2829	−76.3884	−76.41571(20)
cc-pCVTZ	−76.3902	−76.4286	−76.43182(19)
cc-pCVQZ	−76.4212	−76.4365	−76.43622(14)
cc-pCV5Z	−76.4311	−76.4371	−76.43744(18)

[a]Values extracted from Ref. 147.
[b]This works.
[c]Values extracted from Ref. 56.

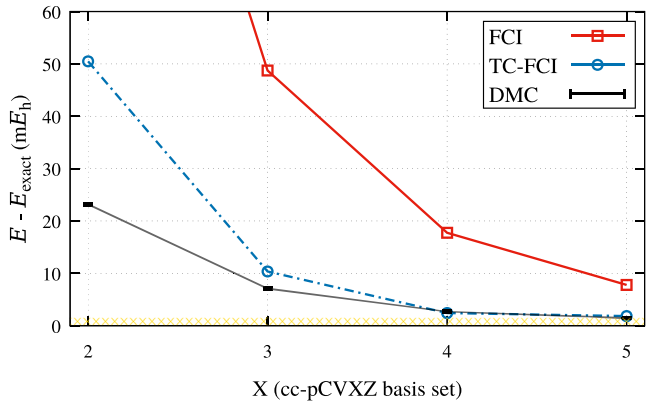

FIG. 9. Error (in m*E*h) in the total energy of H_2O computed at the FCI, TC-FCI, and FN-DMC levels with the cc-pCVXZ family of basis sets with respect to the estimated exact energy (−76.4389 Eh). The shaded yellow region corresponds to 1.5 m*E*h accuracy.

systems. (Here, we rely on the optimized parameters of the neutral system.)

We have computed the TC-FCI energies of the same set of atoms as in Sec. III E, both with and without the frozen-core approximation, using the cc-pVXZ basis sets (where X = D, T, and Q). The total energies are provided in the supplementary material, while the resulting IPs, expressed in eV, are depicted in Table IV. The absolute deviation caused by the frozen-core approximation is ∼1 meV on average. This suggests that the frozen-core approximation can be effectively employed in TC-CI calculations to reduce the size of the Hilbert space without deteriorating the precision. Furthermore, upon comparison with the exact IP estimates reported in Ref. 146, we observe mean absolute errors of 0.22, 0.02, and 0.03 eV for the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets, respectively. A slight increase in error is observed between cc-pVTZ and cc-pVQZ, which we attribute to the imbalance in accuracy between the neutral and cation species. Indeed, on the cc-pVQZ basis, the total energy of the neutral systems is nearly exact, while the energy of the cations is comparatively less precise. In addition, we report in Table IV the IPs computed at the FCI level for the same family of basis sets. The mean absolute errors are 0.37, 0.14, 0.06, and 0.03 eV for the cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z basis sets, respectively. In conclusion, while cc-pVTZ is sufficient to achieve chemical accuracy at the TC-FCI level, one must employ cc-pV5Z at the FCI level.

In Table V, we have conducted FCI and TC-FCI calculations, with or without the frozen-core approximation, to estimate the IP of the following molecules: Li_2, Be_2, H_2O, NH_3, CH_4, and H_2CO. Similar to the case of atoms, the mean absolute deviation induced by the frozen-core approximation is negligible (3 meV). To produce CBS estimates, we have extrapolated the IPs obtained at the FCI level using an inverse cubic parameterization across the cc-pVXZ basis sets (where X = T, Q, and 5). Considering the CBS estimates as references, the mean absolute errors are 0.40, 0.15, 0.06, and 0.03 eV at the FCI level for the cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z basis sets, respectively. These errors are reduced to 0.17, 0.05, and 0.02 eV at the TC-FCI level for cc-pVDZ, cc-pVTZ, and cc-pVQZ, respectively. Therefore, similar conclusions hold for the case of molecules. To further illustrate this, we present, in Fig. 10,

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

TABLE IV. IPs (in eV) computed at the FCI and TC-FCI levels with or without frozen-core (FC) approximation for the neutral atoms from $Z = 2$ to $Z = 10$. For the sake of comparison, estimated exact IPs are also provided.

	He		Li		Be	
	FCI	TC-FCI	FCI	TC-FCI	FCI	TC-FCI
cc-pVDZ	24.33	24.60	5.35	5.13	9.29	9.29
cc-pVTZ	24.53	24.61	5.35	5.40	9.29	9.32
cc-pVQZ	24.56	24.59	5.37	5.40	9.30	9.33
cc-pV5Z	24.58		5.38		9.31	
Exact[a]	24.59		5.39		9.32	

	B			C			N		
	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI
cc-pVDZ	8.07	8.15	8.15	10.98	11.08	11.08	14.19	14.33	14.33
cc-pVTZ	8.22	8.31	8.31	11.17	11.27	11.28	14.43	14.56	14.56
cc-pVQZ	8.24	8.30	8.30	11.21	11.27	11.27	14.49	14.56	14.56
cc-pV5Z	8.25			11.22			14.51		
Exact[a]	8.30			11.26			14.53		

	O			F			Ne		
	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI
cc-pVDZ	12.85	13.13	13.13	16.71	17.09	17.08	20.89	21.22	21.22
cc-pVTZ	13.32	13.58	13.58	17.13	17.45	17.45	21.30	21.62	21.62
cc-pVQZ	13.48	13.63	13.62	17.30	17.49	17.48	21.47	21.67	21.67
cc-pV5Z	13.55			17.37			21.54		
Exact[a]	13.62			17.42			21.56		

[a]Values extracted from Ref. 146.

TABLE V. IPs (in eV) computed at the FCI and TC-FCI levels with or without frozen-core (FC) approximation for Li_2, Be_2, H_2O, NH_3, CH_4, and H_2CO. The CBS estimates are obtained by extrapolation based on the FCI data.

	Li_2		Be_2		H_2O		
	FCI	TC-FCI	FCI	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI
cc-pVDZ	5.19	5.18	7.44	7.47	11.96	12.44	12.43
cc-pVTZ	5.22	5.26	7.45	7.50	12.49	12.72	12.71
cc-pVQZ	5.24	5.27	7.47	7.50	12.66	12.76	12.76
cc-pV5Z	5.25		7.48		12.73		
CBS	5.25		7.49		12.80		

	NH_3			CH_4			H_2CO		
	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI	FCI (FC)	TC-FCI (FC)	TC-FCI
cc-pVDZ	10.32	10.68	10.68	14.20	14.42	14.40	10.44	10.76	10.76
cc-pVTZ	10.76	10.92	10.91	14.36	14.47	14.47	10.80	10.95	10.95
cc-pVQZ	10.89	10.96	10.96	14.40	14.44	14.44	10.92	10.98	10.98
cc-pV5Z	10.95			14.41			10.95		
CBS	11.00			14.43			11.00		

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

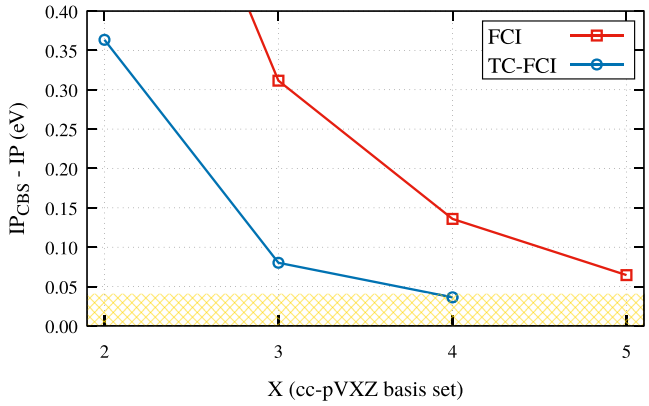

FIG. 10. Error (in eV) in the IP of H_2O computed at the FCI and TC-FCI levels with the cc-pVXZ family of basis sets with respect to the CBS estimate (12.80 eV). The shaded yellow region corresponds to 1.5 m*E*h accuracy.

the difference between the CBS estimate (12.80 eV) and the IP values obtained at the FCI and TC-FCI levels across the cc-pVXZ basis sets in the case of the water molecule. While chemical accuracy is nearly attained at the TC-FCI/cc-pVTZ level, the error is three times larger at the FCI/cc-pVTZ level.

IV. CONCLUSION

In the present study, we investigated how incorporating a correlation factor in the Hamiltonian within the TC framework markedly accelerates the convergence of SCI methods. To illustrate these results, we systematically studied the total energies and IPs in increasing large Dunning basis sets using optimized Jastrow factors available from the literature in the case of neutral atoms with $2 \leq Z \leq 10$ together with the Li_2, Be_2, H_2O, NH_3, CH_4, and H_2CO molecules. This acceleration has been demonstrated quantitatively through both the reduction in the size of the one-electron basis functions and the decrease in the number of important determinants in the Hilbert space.

Although the TC Hamiltonian exhibits certain difficulties related to three-electron terms and non-Hermiticity, the working equations derived here for TC-SCI demonstrate a scaling similar to that of SCI methods. Nonetheless, as our numerical analysis reveals, TC-SCI effectively increases the sparsity of the Slater determinant space when a flexible enough Jastrow factor is employed. This enhancement enables faster and more stable convergence toward the FCI limit within a given basis set.

By comparing with the best estimates (when available for total energies) or CBS values reported here for IPs, we have shown that achieving near-exact results, i.e., within chemical accuracy, is possible in TC-SCI using basis sets significantly smaller than those required in standard SCI calculations.

Expanding the applicability of TC-SCI methods to arbitrary systems necessitates the ability to optimize Jastrow factors at a reasonable cost. While this challenge has been considered in prior studies, it is widely acknowledged to be both difficult and expensive. In our upcoming efforts, we will explore how to systematically optimize Jastrow factors at a reasonable cost.

SUPPLEMENTARY MATERIAL

The supplementary material of the present paper contains (i) the molecular geometries used for all molecules studied in the present work, (ii) a detailed analysis of the decay of the weights of Slater determinants using two different types of correlation factors, (iii) the various curves studying the compaction of the CI expansion (similar to Sec. III D) for Li_2, Be_2, NH_3, CH_4, and H_2CO, (iv) a detailed study of the convergence of the IP's with respect to both the basis set and the size of the CI expansion, and (v) the total energies for all atomic and molecular systems studied here.

ACKNOWLEDGMENTS

This work was performed using HPC resources from GENCI-TGCC (Nos. gen1738, gen12363) and from CALMIP (Toulouse) under Allocation No. 2024-18005, and was also supported by the European Centre of Excellence in Exascale Computing TREX—Targeting Real Chemical Accuracy at the Exascale. This project received funding from the European Union's Horizon 2020—Research and Innovation Program—under Grant Agreement No. 952165. A.A., A.S., and P.-F.L. also acknowledge the funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 863481).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Abdallah Ammar: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Anthony Scemama**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Pierre-François Loos**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Emmanuel Giner**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

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

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

REFERENCES

[1] E. A. Hylleraas, Z. Phys. **54**, 347 (1929).
[2] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[3] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
[4] R. T. Pack and W. B. Brown, J. Chem. Phys. **45**, 556 (1966).
[5] C. R. Myers, C. J. Umrigar, J. P. Sethna, and J. D. Morgan III, Phys. Rev. A **44**, 5537 (1991).
[6] J. D. I. Morgan and W. Kutzelnigg, J. Phys. Chem. **97**, 2425 (1993).
[7] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).
[8] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[9] J. Noga and W. Kutzelnigg, J. Chem. Phys. **101**, 7738 (1994).
[10] S. Ten-no and J. Noga, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 114 (2012).
[11] S. Ten-no, Theor. Chem. Acc. **131**, 1070 (2012).
[12] C. Hattig, W. Klopper, A. Kohn, and D. P. Tew, Chem. Rev. **112**, 4 (2012).
[13] A. Grüneis, S. Hirata, Y.-Y. Ohnishi, and S. Ten-no, J. Chem. Phys. **146**, 080901 (2017).
[14] C. Schwartz, Phys. Rev. **126**, 1015 (1962).
[15] W. Kutzelnigg and J. D. Morgan III, J. Chem. Phys. **96**, 4484 (1992).
[16] R. N. Hill, J. Chem. Phys. **83**, 1173 (1985).
[17] D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).
[18] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).
[19] J. Čížek, *Advances in Chemical Physics* (John Wiley & Sons, Ltd, Chichester, UK, 1969), pp. 35–89.
[20] J. Paldus, *Methods in Computational Molecular Physics* (Springer, Boston, MA, 1992), pp. 99–194.
[21] T. D. Crawford and H. F. Schaefer, *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2000), pp. 33–136.
[22] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
[23] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory, Cambridge Molecular Science* (Cambridge University Press, Cambridge, 2009).
[24] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[25] S. R. White, Phys. Rev. B **48**, 10345 (1993).
[26] C. F. Bender and E. R. Davidson, Phys. Rev. **183**, 23 (1969).
[27] B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).
[28] R. J. Buenker and S. D. Peyerimholf, Theor. Chim. Acta **35**, 33 (1974).
[29] R. J. Buenker, S. D. Peyerimhoff, and P. J. Bruna, *Computational Theoretical Organic Chemistry* (Springer, Dordrecht, The Netherlands, 1981), pp. 55–76.
[30] S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, Chem. Phys. **75**, 91 (1983).
[31] R. J. Harrison, J. Chem. Phys. **94**, 5021 (1991).
[32] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[33] E. Giner, A. Scemama, and M. Caffarel, J. Chem. Phys. **142**, 044115 (2015).
[34] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).
[35] J. B. Schriber and F. A. Evangelista, J. Chem. Theory Comput. **13**, 5354 (2017).
[36] E. Giner, D. P. Tew, Y. Garniron, and A. Alavi, J. Chem. Theory Comput. **14**, 6240 (2018).
[37] P.-F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, J. Chem. Theory Comput. **14**, 4360 (2018).
[38] P. F. Loos, M. Boggio-Pasqua, A. Scemama, M. Caffarel, and D. Jacquemin, J. Chem. Theory Comput. **15**, 1939 (2019).
[39] 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).
[40] N. Zhang, W. Liu, and M. R. Hoffmann, J. Chem. Theory Comput. **17**, 949 (2021).
[41] G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465 (2011).
[42] A. Baiardi and M. Reiher, J. Chem. Phys. **152**, 040903 (2020).
[43] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[44] G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **135**, 084104 (2011).
[45] J. J. Eriksen and J. Gauss, J. Chem. Theory Comput. **14**, 5180 (2018).
[46] J. J. Eriksen and J. Gauss, J. Chem. Theory Comput. **15**, 4873 (2019).
[47] E. Xu, M. Uejima, and S. L. Ten-no, Phys. Rev. Lett. **121**, 113001 (2018).
[48] E. Xu, M. Uejima, and S. L. Ten-no, J. Phys. Chem. Lett. **11**, 9775 (2020).
[49] J. E. Deustua, J. Shen, and P. Piecuch, J. Chem. Phys. **154**, 124103 (2021).
[50] K. Gururangan, J. E. Deustua, J. Shen, and P. Piecuch, J. Chem. Phys. **155**, 174114 (2021).
[51] K. Gururangan and P. Piecuch, J. Chem. Phys. **159**, 084108 (2023).
[52] A. A. Holmes, C. J. Umrigar, and S. Sharma, J. Chem. Phys. **147**, 164111 (2017).
[53] H. G. A. Burton and P.-F. Loos, J. Chem. Phys. **160**, 104102 (2024).
[54] J. J. Eriksen, T. A. Anderson, J. E. Deustua, K. Ghanem, D. Hait, M. R. Hoffmann, S. Lee, D. S. Levine, I. Magoulas, J. Shen, N. M. Tubman, K. B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G. K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S. L. Ten-no, C. J. Umrigar, and J. Gauss, J. Phys. Chem. Lett. **11**, 8922 (2020).
[55] P.-F. Loos, Y. Damour, and A. Scemama, J. Chem. Phys. **153**, 176101 (2020).
[56] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, J. Chem. Phys. **144**, 151103 (2016).
[57] A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, J. Phys. Chem. A **122**, 2714 (2018).
[58] P.-F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama, and D. Jacquemin, J. Chem. Theory Comput. **16**, 1711 (2020).
[59] M. Véril, A. Scemama, M. Caffarel, F. Lipparini, M. Boggio-Pasqua, D. Jacquemin, and P.-F. Loos, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **11**, e1517 (2021).
[60] Y. Damour, M. Véril, F. Kossoski, M. Caffarel, D. Jacquemin, A. Scemama, and P.-F. Loos, J. Chem. Phys. **155**, 134104 (2021).
[61] P.-F. Loos, Y. Damour, A. Ammar, M. Caffarel, F. Kossoski, and A. Scemama, "Go green: Selected configuration interaction as a more sustainable alternative for high accuracy," arXiv:2402.13111 [physics.chem-ph] (2024).
[62] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996).
[63] E. Giner, R. Assaraf, and J. Toulouse, Mol. Phys. **114**, 910 (2016).
[64] W. Dobrautz, H. Luo, and A. Alavi, Phys. Rev. B **99**, 075119 (2019).
[65] E. Giner, J. Chem. Phys. **154**, 084119 (2021).
[66] J. Ludovicy, R. Dahl, and A. Lüchow, J. Chem. Theory Comput. **19**, 2792 (2023).
[67] S. F. Boys and N. C. Handy, Proc. R. Soc. London, Ser. A **310**, 63 (1969).
[68] J. O. Hirschfelder, J. Chem. Phys. **39**, 3145 (1963).
[69] J. Toulouse, R. Assaraf, and C. J. Umrigar, *Advances in Quantum Chemistry* (Academic Press, Cambridge, MA, 2016), Vol. 73, pp. 285–314.
[70] H. Luo, W. Hackbusch, and H.-J. Flad, Mol. Phys. **108**, 425 (2010).
[71] H. Luo, J. Chem. Phys. **133**, 154109 (2010).
[72] H. Luo, J. Chem. Phys. **135**, 024109 (2011).
[73] H. Luo, J. Chem. Phys. **136**, 224111 (2012).
[74] S. F. Boys and N. C. Handy, Proc. R. Soc. London, Ser. A **310**, 43 (1969).
[75] S. F. Boys and N. C. Handy, Proc. R. Soc. London, Ser. A **309**, 209 (1969).
[76] S. F. Boys and N. C. Handy, Proc. R. Soc. London, Ser. A **311**, 309 (1969).
[77] N. C. Handy, J. Chem. Phys. **51**, 3205 (1969).
[78] S. F. Boys, Proc. R. Soc. London, Ser. A **309**, 195 (1969).
[79] N. C. Handy, Mol. Phys. **21**, 817 (1971).
[80] N. C. Handy, Mol. Phys. **23**, 1 (1972).
[81] E. A. G. Armour, Mol. Phys. **24**, 181 (1972).
[82] F. Bernardi and S. F. Boys, Mol. Phys. **25**, 35 (1973).
[83] F. Bernardi, J. Phys. **34**, 373 (1973).
[84] E. A. G. Armour, Chem. Phys. Lett. **25**, 614 (1974).
[85] R. Sakuma and S. Tsuneyuki, J. Phys. Soc. Jpn. **75**, 103705 (2006).
[86] N. Handy, *Computational Techniques in Quantum Chemistry and Molecular Physics* (Springer, 1975), pp. 425–433.
[87] P. N. Careless, D. Hyatt, and L. Stanton, Int. J. Quantum Chem. **12**, 569 (1977).
[88] G. G. Hall and C. J. Miller, Phys. Rev. A **18**, 889 (1978).
[89] E. A. G. Armour, J. Phys. C: Solid State Phys. **13**, 343 (1980).
[90] S. Tsuneyuki, Prog. Theor. Phys. Suppl. **176**, 134 (2008).
[91] M. W. C. Dharma-wardana and F. Grimaldi, Phys. Rev. A **13**, 1702 (1976).

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing

[92]M. Ochi, K. Sodeyama, R. Sakuma, and S. Tsuneyuki, J. Chem. Phys. **136**, 094108 (2012).
[93]M. Ochi and S. Tsuneyuki, J. Phys.: Conf. Ser. **454**, 012020 (2013).
[94]J. M. Wahlen-Strothman, C. A. Jiménez-Hoyos, T. M. Henderson, and G. E. Scuseria, Phys. Rev. B **91**, 041114 (2015).
[95]M. Ochi, R. Arita, and S. Tsuneyuki, Phys. Rev. Lett. **118**, 026402 (2017).
[96]M. Ochi, Comput. Phys. Commun. **287**, 108687 (2023).
[97]M. Ochi, Phys. Rev. A **108**, 032806 (2023).
[98]N. Lee and A. J. W. Thom, J. Chem. Theory Comput. **19**, 5743 (2023).
[99]A. Ammar, A. Scemama, and E. Giner, J. Chem. Theory Comput. **19**, 4883 (2023).
[100]S. Ten-no, Chem. Phys. Lett. **330**, 169 (2000).
[101]S. Ten-no and O. Hino, Int. J. Mol. Sci. **3**, 459 (2002).
[102]M. Ochi and S. Tsuneyuki, Chem. Phys. Lett. **621**, 177 (2015).
[103]W. R. Fimple and M. J. Unwin, Int. J. Quantum Chem. **10**, 643 (1976).
[104]M. Ochi and S. Tsuneyuki, J. Chem. Theory Comput. **10**, 4098 (2014).
[105]M. Ochi, Y. Yamamoto, R. Arita, and S. Tsuneyuki, J. Chem. Phys. **144**, 104109 (2016).
[106]A. Ammar, A. Scemama, and E. Giner, J. Chem. Phys. **157**, 134107 (2022).
[107]A. Ammar, E. Giner, and A. Scemama, J. Chem. Theory Comput. **18**, 5325 (2022).
[108]A. Ammar, A. Scemama, and E. Giner, J. Chem. Phys. **159**, 114121 (2023).
[109]O. Hino, Y. Tanimura, and S. Ten-no, J. Chem. Phys. **115**, 7865 (2001).
[110]O. Hino, Y. Tanimura, and S. Ten-no, Chem. Phys. Lett. **353**, 317 (2002).
[111]M. Ochi, K. Sodeyama, and S. Tsuneyuki, J. Chem. Phys. **140**, 074112 (2014).
[112]K. Liao, T. Schraivogel, H. Luo, D. Kats, and A. Alavi, Phys. Rev. Res. **3**, 033072 (2021).
[113]T. Schraivogel, E. M. Christlmaier, P. López Ríos, A. Alavi, and D. Kats, J. Chem. Phys. **158**, 214106 (2023).
[114]P. Jeszenszki, H. Luo, A. Alavi, and J. Brand, Phys. Rev. A **98**, 053627 (2018).
[115]H. Luo and A. Alavi, J. Chem. Theory Comput. **14**, 1403 (2018).
[116]A. J. Cohen, H. Luo, K. Guther, W. Dobrautz, D. P. Tew, and A. Alavi, J. Chem. Phys. **151**, 061101 (2019).
[117]P. Jeszenszki, U. Ebling, H. Luo, A. Alavi, and J. Brand, Phys. Rev. Res. **2**, 043270 (2020).
[118]W. Dobrautz, A. J. Cohen, A. Alavi, and E. Giner, J. Chem. Phys. **156**, 234108 (2022).
[119]A. Baiardi and M. Reiher, J. Chem. Phys. **153**, 164115 (2020).
[120]A. Baiardi, M. Lesiuk, and M. Reiher, J. Chem. Theory Comput. **18**, 4203 (2022).
[121]K. Liao, H. Zhai, E. M. Christlmaier, T. Schraivogel, P. L. Ríos, D. Kats, and A. Alavi, J. Chem. Theory Comput. **19**, 1734 (2023).
[122]S. L. Ten-no, J. Chem. Phys. **159**, 171103 (2023).
[123]M. Moshinsky and T. Seligman, Ann. Phys. **66**, 311 (1971).
[124]D. C. Brody, J. Phys. A: Math. Theor. **47**, 035305 (2013).
[125]E. M. Christlmaier, T. Schraivogel, P. López Ríos, A. Alavi, and D. Kats, J. Chem. Phys. **159**, 014113 (2023).
[126]Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, J. Chem. Phys. **147**, 034101 (2017).
[127]V. G. Chilkuri, T. Applencourt, K. Gasperich, P.-F. Loos, and A. Scemama, Adv. Quantum Chem. **83**, 65 (2021).
[128]K. Hirao and H. Nakatsuji, J. Comput. Phys. **45**, 246 (1982).
[129]B. S. Francis, Proc. R. Soc. London, Ser. A **258**, 402 (1960).
[130]J. V. L. Longstaff and K. Singer, Theor. Chim. Acta **2**, 265 (1964).
[131]C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).
[132]K. E. Schmidt and J. W. Moskowitz, J. Chem. Phys. **93**, 4172 (1990).
[133]N. C. H. Peter, T. A. Galek, and W. A. Lester, Jr., Mol. Phys. **104**, 3069 (2006).
[134]S. Ten-no, Chem. Phys. Lett. **330**, 175 (2000).
[135]A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).
[136]M. E. Mura and P. J. Knowles, J. Chem. Phys. **104**, 9848 (1996).
[137]A. Scemama and E. Giner, "An efficient implementation of slater-condon rules," arXiv:1311.6244 [physics.comp-ph] (2013).
[138]R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).
[139]W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[140]J. S. Binkley, J. A. Pople, and W. J. Hehre, J. Am. Chem. Soc. **102**, 939 (1980).
[141]J. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[142]D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).
[143]P.-F. Loos, A. Scemama, and D. Jacquemin, J. Phys. Chem. Lett. **11**, 2374 (2020).
[144]A. Marie and P.-F. Loos, "Reference energies for valence ionizations and satellite transitions," J. Chem. Theory Comput. **20**, 4751 (2024).
[145]K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure—IV. Constants of Diatomic Molecules* (van Nostrand Reinhold Company, 1979).
[146]S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. P Fischer, Phys. Rev. A **47**, 3649 (1993).
[147]C. X. Almora-Díaz, J. Chem. Phys. **140**, 184302 (2014).
[148]W. Klopper, Mol. Phys. **99**, 481 (2001).
[149]T. Shiozaki, M. Kamiya, S. Hirata, and E. F. Valeev, J. Chem. Phys. **130**, 054101 (2009).
[150]S. Ten-no, J. Chem. Phys. **121**, 117 (2004).

22 August 2024 16:47:33

Published under an exclusive license by AIP Publishing