Development and parallel implementation of selected configuration interaction methods

Yann Garniron

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Yann GARNIRON

Le 3 décembre 2018

Développement et implémentation parallèle de méthodes d'interaction de configurations sélectionnées

---

Ecole doctorale : **SDM - SCIENCES DE LA MATIERE - Toulouse**

Spécialité : **Physico-Chimie Théorique**

Unité de recherche :
LCPQ-IRSAMC - Laboratoire de Chimie et Physique Quantiques

Thèse dirigée par
Anthony SCEMAMA

Jury

M. Jean-Philip Piquemal, Rapporteur
M. Carbonnière, Rapporteur
Mme Nathalie Guihéry, Examinateur
M. Nicolas Renon, Examinateur
M. Roland Assaraf, Examinateur
M. Anthony Scemama, Directeur de thèse

Acknowledgments

Trois années déjà, le temps s'est écoulé trop vite.

Avant tout, je remercie bien entendu mon directeur de thèse Anthony, son flot ininterrompu d'idées toutes intéressantes mais trop nombreuses pour espérer en exécuter ne serait-ce qu'une fraction, son humour toujours fin et sa bienveillance naturelle. Je remercie également Michel, Jean-Paul et Titou, véritables puits de science, pour les moult discussions enrichissantes.

Merci à l'ensemble du LCPQ pour son accueil, et un merci de plus à l'équipe GMO pour les réunions du jeudi matin, en particulier les 3 dont je suis reparti avec la bouteille de vin soigneusement sélectionnée par Trond ou Arjan. La dernière est conservée pour le pot de thèse.

Une petite excuse pour l'équipe administrative, pour ma rigueur très approximative en la matière. Un petit merci à Bryan, qui ne lira sûrement jamais ces lignes, pour les parties d'agar.io (en dehors des heures de stage bien sûr!) et sa ponctualité scandaleuse.

Un grand remerciement à ma femme Joany pour son soutien indéfectible pendant ces 3 années - et toutes celles qui ont précédées.

Et je laisse le mot de la fin à ma petite Aurore venue m'apporter son aide précieuse au beau milieu de ma thèse :

eszqqwxxcccvbvbkllkkjuuuèèèè-wxwxxde'"'s""za <

Contents

Notations

N_{orb} Number of molecular orbitals

N_{states} Number of considered eigenstates

N_{det} Number of determinants in the internal space

N_{gen} Number of generator determinants

N_{sel} Number of selector determinants

N_{elec} Number of electrons

$N_{\text{elec}}^{\uparrow}$ Number of α-spin electrons

$N_{\text{elec}}^{\downarrow}$ Number of β-spin electrons

N_{int} : Number of 64-bit integers required to store N_{orb} bits : $N_{\text{int}} = \lfloor \frac{N_{\text{orb}}-1}{64} \rfloor + 1$

N_{FCI} : Number of determinants in the FCI space

$|\mathcal{D}|$: Cardinality of the set $\mathcal{D}$

$|\alpha\rangle$: External determinant

$|D_I\rangle$: Internal determinant

I_σ : Bitstring representation of the $\sigma \in \{\uparrow, \downarrow\}$ spin part of $|I\rangle$

Chapter 1

Introduction

Quantum chemistry is a discipline which relies on very expensive computations. The scalings of wave function methods lie between $\mathcal{O}\left(N^5\right)$ and $\mathcal{O}\left(N^8\right)$, where N is proportional to the number of electrons in the system. Therefore, performing accurate calculations requires both approximations that can reduce the scaling, and an efficient implementation that can take advantage of modern architectures. The work presented in this thesis is more centered on this last aspect.

In 1965, Gordon Moore predicted that the number of transistors in an integrated circuit would double about every two years (the so-called Moore's law).[2] Rapidly, this "law" was interpreted as an expected $2\times$ gain in performance every 18 months, which became an industrial goal. The development of today's most popular codes of the community (Molpro[3], Molcas[4], or Gaussian[5]…) was initiated in the 1990's. At that time, the increase of computational power from one generation of supercomputers to the next one was mostly due to the increase of the frequency of the processors. the amount of random access memory was small, the time to access data from disk was slow, and the energy consumption of the most powerful computer was 236kW, which was not an economical concern.[6] At the beginning of the years 2000, having increased continuously both the number of processors and their frequency raised the power consumption of supercomputers by two orders of magnitude, raising accordingly the annual electricity bill. The only way to slow down this need for electricity while keeping alive Moore's law was to keep the frequency fixed (between 1 and 4 GHz), and increase the number of CPU cores. The consequence of such a choice was that "free lunch" was over, and the programmers now had to parallelize their programs to make them run faster.[7] At the same time, computer scientists realized that the increase of performance in memory access was slower than the increase in computational power,[8] and that the floating-point operation (or flop) count would soon not be the bottleneck, and that data movement would be the concern. This change was called the *memory wall*.

So today, the situation is completely different from the 1990's. Moore's law has

ended,[9] the CPU frequency tends to decrease, hundreds of thousands of cores need to be handled, data movement is the principal concern, and disk access times are prohibitively high. The work presented in this thesis is in the context of this change of paradigm that has been going on for the last decade. The traditional sequential algorithms of quantum chemistry are currently being redesigned so as to be replaced by parallel equivalents by multiple groups around the world, and this has also an influence on methodological development.

Initially, this work may been have expected to focus on methods that are by design adapted to massively parallel architectures, such as Monte-Carlo methods (stochastic methods), which are composed of a large number of independent tasks (*embarrassingly parallel* algorithms). In addition, they often are able to yield a satisfactory result for just a fraction of the cost of the equivalent deterministic, exact computation. An example of the move toward this type of method is the recently developed *Full Configuration Interaction Quantum Monte Carlo* (FCIQMC).[10] FCIQMC can be interpreted as a Monte-Carlo variant of older selected configuration interaction algorithms such as CIPSI,[11] that are iterative and thus a priori not well adapted to massively parallel architectures. But things turn out differently, and the focus of this thesis was to investigate how to make *configuration interaction* (CI) methods efficient on modern supercomputers.

The Quantum Package [12] developed at the LCPQ is a suite of wave function quantum chemistry methods, that strives to allow easy implementation and experimentation of new methods, and to make parallel computation as simple and efficient as possible. The main purpose of this package is to make experimentation on code design, algorithms and methods, more than to be used massively in production. Hence, the initial choice of the Quantum Package was to go in the direction of determinant-driven algorithms, as opposed to the more traditional integral-driven algorithms. A determinant-driven approach essentially implies that the wave function is expressed as a linear combination of determinants, and that the outermost loops of the algorithms loop over determinants. On the other hand, integral-driven algorithms have their outermost loop on the two-electron integrals which appear in the expression of the matrix elements in the determinant basis. In the context of configuration interaction or perturbative methods, the determinant-driven approach simplifies the development and allows the researchers to test new ideas very quickly. These algorithms allow more flexibility than their integral-driven counterparts,[13] but they have been known for years to be less efficient for solving problems that can be solved with an integral-driven variant. High-precision calculations are in a regime where the number of determinants is larger than the number of integrals, which justifies the integral-driven choice. Today, programming imposes parallelism, and if determinant-driven calculations prove to be better adapted to parallelism, such methods could regain in popularity. The work presented in this thesis focuses on determinant-driven approaches via the improvement

of the Quantum Package from the methodological, algorithmic and the implementational points of views.

Somewhat logically, the first focus was the acceleration and parallelization of the Davidson diagonalization, which is a pivotal point for CI methods. A naive determinant-driven algorithm implies a quadratic scaling with the number of determinants, while the integral-driven algorithm is expected to scale linearly. This fact gave us some insight that there was room for improvement in this step.

The second focus was the improvement of the determinant selection algorithm which is the main method used by the Quantum Package to build compact wave functions suitable for determinant-driven computations. In a nutshell, the principle is to incrementally build a variational wave function by scavenging its external space for determinants that interact with it. While the significant improvement that was brought to this implementation was in itself the most important part of this work, it also turned out to be the basis for the subsequent implementation of other algorithms. Indeed, efficiently implementing this method raised the fundamental question of connecting a variational wave function to its external space ; that is, gathering data to go beyond what is readily available in it. The next steps were partly guided by the aversion to waste data gathered during the selection.

Our selection algorithm, CIPSI, implies computing a perturbative contribution for external determinants, and including those with the largest contributions into the internal space in which the variational wave function is expressed. E_{PT2}, the sum of all the contributions of the external determinants, approximates how much energy the variational wave function is missing compared to the exact solution in the same basis set, namely the *Full Configuration Interaction* or Full CI. However to perform an acceptably accurate selection, not all external determinants need to be considered, nor does each contribution need to be known with a great accuracy.[14] This allows for approximations too severe for the sum of all computed contributions to yield an accurate estimation E_{PT2}, and incidentally the computation of E_{PT2} is much more time consuming than determinant selection. To make this step more affordable, we designed a hybrid deterministic-stochastic scheme which enabled us to get an accurate value for E_{PT2} by computing just a few percent of all the contributions.

The computation of E_{PT2} allows to correct the energy of the wave function by taking into account its external space. Unfortunately, it only improves the energy, but leaves the wave function unchanged. Based on the shifted-B_k algorithm, using our CIPSI implementation and the hybrid deterministic-stochastic scheme we were able to refine the wave function under the effect of a stochastically estimated external space using a Hamiltonian dressed by a matrix computed semi-stochastically.

In addition, we set up a general framework to enable the refining of the variational wave function under the effect of any external space, with a stochastic estimation. This was experimented by implementing a stochastic selected *multi-reference coupled*

cluster with singe and double substitutions (MR-CCSD).

The efficiency of the implemented algorithms is exposed, and the code was used in numerous applications, in particular to obtain reference excitation energies for difficult molecular systems. The high quality and compactness of the CIPSI wave function was also used for quantum Monte Carlo calculations to characterize the ground state of the Fe–S molecule.

Of course, the technical considerations were not the focus of the different articles that were produced. Because my work focused on the actual implementation of the methods at least as much as on the theory behind them, this thesis is an opportunity to discuss in depth the implementation. I hope this document will be one of the major pieces of documentation for developers willing to understand deeply the implementation of the Quantum Package , so I decided to write this thesis in English.

Chapter 2

Wave function methods

Contents

Quantum chemistry aims at describing the electronic structure of molecular systems. The velocity of the nuclei is considered negligible compared to that of the electrons (Born-Oppenheimer approximation), and for atoms of the first rows of the periodic table relativistic effects can be neglected. In this context, the model system is a cloud of N_{elec} electrons and a set of M nuclei considered punctual, immobile charges. It can be described by solving Shrödinger's equation for electrons :

$$\widehat{H}\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{elec}}}) = E\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{elec}}}) \tag{2.1}$$

where Ψ is the electronic wave function, E the associated energy, and $\mathbf{x}_i = (\mathbf{r}, m_s)$ contains the spatial coordinates of the electron $\mathbf{r}$, as well as a spin variable m_s. $\widehat{H}$ is the non-relativistic electronic Hamiltonian operator

$$\widehat{H} = \sum_{i=1}^{N_{\text{elec}}} \Big(-\frac{1}{2}\Delta_i - \sum_{j=1}^{M} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} \Big) + \sum_{i=1}^{N_{\text{elec}}} \sum_{k>i}^{N_{\text{elec}}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_k|} \tag{2.2}$$

where $\mathbf{R}_j$ and Z_j are respectively the spatial coordinate and charge of nucleus j.

2.1 Slater determinants

The simplest description of the wave function is the Hartree product. This consists in building the product of orthonormal one-electron functions, each function describing the state of one electron:

$$\Psi_{\text{Hartree}}(\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{elec}}}) = \prod_{i=1}^{N_{\text{elec}}} \phi_i(\mathbf{x}_i). \tag{2.3}$$

Because of the fermionic nature of electrons, Ψ must satisfy the condition of being antisymmetric with respect to the permutation of electrons coordinates, which is not verified by the Hartree product. Antisymmetrizing the Hartree product yields the so-called *Slater determinant*:

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{elec}}}) = \frac{1}{\sqrt{N_{\text{elec}}!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \ldots & \phi_1(\mathbf{x}_{N_{\text{elec}}}) \\ \vdots & \ddots & \vdots \\ \phi_{N_{\text{elec}}}(\mathbf{x}_1) & \ldots & \phi_N(\mathbf{x}_{N_{\text{elec}}}) \end{vmatrix} \tag{2.4}$$

which is the simplest possible antisymmetric wave function. The functions ϕ_i are called *spinorbitals*:

$$\phi_i(\mathbf{x}) = \varphi_i(\mathbf{r})\, \sigma_i(m_s) \tag{2.5}$$

where φ_i is a spatial function, or *molecular orbital (MO)*, and σ_i is a discrete spin function describing the spin state of the electron ($m_s = \pm\frac{1}{2}$). The spin function can be either $\alpha(m_s)$ or $\beta(m_s)$ defined as

$$\begin{aligned} \alpha\left(\frac{1}{2}\right) = 1 \; ; \; \alpha\left(-\frac{1}{2}\right) = 0 \\ \beta\left(\frac{1}{2}\right) = 0 \; ; \; \beta\left(-\frac{1}{2}\right) = 1, \end{aligned} \tag{2.6}$$

and for convenience, one will rewrite

$$\begin{aligned} \phi_i(\mathbf{x}) &= \varphi_i(\mathbf{r})\alpha(m_s) && \uparrow \text{ spinorbitals} \\ \overline{\phi}_i(\mathbf{x}) &= \varphi_i(\mathbf{r})\beta(m_s) && \downarrow \text{ spinorbitals} \end{aligned} \tag{2.7}$$

Packing together the $\uparrow$ spinorbitals, and then the $\downarrow$ spinorbitals in the representa-

tion of the determinant, one can express the Slater determinant as

$$\frac{1}{\sqrt{N_{\text{elec}}!}}\begin{vmatrix} \phi_1(\mathbf{x}_1) & \dots & \phi_1(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & \phi_1(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \phi_1(\mathbf{x}_{N_{\text{elec}}}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_1) & \dots & \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_{N_{\text{elec}}}) \\ \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_1) & \dots & \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_{N_{\text{elec}}}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_1) & \dots & \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_{N_{\text{elec}}}) \end{vmatrix} \tag{2.8}$$

where $N_{\text{elec}}^{\uparrow}$ is the number of $\uparrow$ electrons, i.e. the number of electrons with $m_s = 1/2$. By convention, we choose $N_{\text{elec}}^{\uparrow} \geq N_{\text{elec}}^{\downarrow}$. If one chooses the permutation in which the first $N_{\text{elec}}^{\uparrow}$ electrons have $m_s = 1/2$, and the other electrons have $m_s = -1/2$, one always has

$$\begin{aligned} \phi_i(\mathbf{x}_j) &= 0 \ \text{ for } 1 \leq i \leq N_{\text{elec}}^{\uparrow} \text{ and } N_{\text{elec}}^{\uparrow} < j \leq N_{\text{elec}} \\ \overline{\phi}_i(\mathbf{x}_j) &= 0 \ \text{ for } N_{\text{elec}}^{\uparrow} < i \leq N_{\text{elec}} \text{ and } 1 \leq j \leq N_{\text{elec}}^{\uparrow}, \end{aligned} \tag{2.9}$$

and the Slater determinant is the determinant of a block-diagonal matrix:

$$\frac{1}{\sqrt{N_{\text{elec}}!}}\begin{vmatrix} \phi_1(\mathbf{x}_1) & \dots & \phi_1(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & & & \\ \vdots & \ddots & \vdots & & 0 & \\ \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_1) & \dots & \phi_{N_{\text{elec}}^{\uparrow}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}}) & & & \\ & & & \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \overline{\phi}_{N_{\text{elec}}^{\uparrow}+1}(\mathbf{x}_{N_{\text{elec}}}) \\ & 0 & & \vdots & \ddots & \vdots \\ & & & \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \overline{\phi}_{N_{\text{elec}}}(\mathbf{x}_{N_{\text{elec}}}) \end{vmatrix}. \tag{2.10}$$

This allows us to rewrite the Slater determinant in a spin-free formalism as the *Waller-Hartree double determinant*,[15] namely the product of two determinants associated with $\uparrow$ and $\downarrow$ electrons respectively.

$$\begin{aligned} &\left[\Psi_{\text{HF}}(\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{elec}}^{\uparrow}}, \mathbf{r}_{N_{\text{elec}}^{\uparrow}+1}, \dots, \mathbf{r}_{N_{\text{elec}}} \,;\, \tfrac{1}{2}, \dots, \tfrac{1}{2}, -\tfrac{1}{2}, \dots, -\tfrac{1}{2})\right]^2 = \\ &\frac{1}{\sqrt{N_{\text{elec}}!}}\Psi_{\uparrow}(\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{elec}}^{\uparrow}}) \times \Psi_{\downarrow}(\mathbf{r}_{N_{\text{elec}}^{\uparrow}+1}, \dots, \mathbf{r}_{N_{\text{elec}}}) = \\ &\frac{1}{\sqrt{N_{\text{elec}}!}}\begin{vmatrix} \varphi_1(\mathbf{r}_1) & \dots & \varphi_1(\mathbf{r}_{N_{\text{elec}}^{\uparrow}}) \\ \vdots & \ddots & \vdots \\ \varphi_{N_{\text{elec}}^{\uparrow}}(\mathbf{r}_1) & \dots & \varphi_{N_{\text{elec}}^{\uparrow}}(\mathbf{r}_{N_{\text{elec}}^{\uparrow}}) \end{vmatrix} \begin{vmatrix} \varphi_1(\mathbf{r}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \varphi_1(\mathbf{r}_{N_{\text{elec}}}) \\ \vdots & \ddots & \vdots \\ \varphi_{N_\beta}(\mathbf{r}_{N_{\text{elec}}^{\uparrow}+1}) & \dots & \varphi_{N_\beta}(\mathbf{r}_{N_{\text{elec}}}) \end{vmatrix} \end{aligned} \tag{2.11}$$

Molecular orbitals are typically defined as linear combinations of *atomic orbitals*, or *AO*, here noted χ_k

$$\varphi_i(\mathbf{r}) = \sum_k C_{ik}\chi_k(\mathbf{r}). \tag{2.12}$$

These functions qualify the used one-electron *basis set*, and are usually themselves pre-defined linear combinations of Gaussian functions. This is a restriction put on the form of the wave function, therefore it is known as the *finite basis set approximation*. In the Hartree-Fock method, the wave function is a single Slater determinant, where the C_{ik} coefficients associated with molecular orbitals are optimized so as to minimize the energy. This method, however is missing some important physical effects. For instance, using Eq. (2.11) one can see that in this model opposite-spin electrons are statistically independent (or *uncorrelated*):

$$\begin{aligned}\left[\Psi_{\mathrm{HF}}(\mathbf{r}_1,\ldots,\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}},\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}+1},\ldots,\mathbf{r}_{N_{\mathrm{elec}}}\,;\,\tfrac{1}{2},\ldots,\tfrac{1}{2},-\tfrac{1}{2},\ldots,-\tfrac{1}{2})\right]^2 = \\ \left[\Psi_{\uparrow}(\mathbf{r}_1,\ldots,\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}})\times\Psi_{\downarrow}(\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}+1},\ldots,\mathbf{r}_{N_{\mathrm{elec}}})\right]^2 = \\ \left[\Psi_{\uparrow}(\mathbf{r}_1,\ldots,\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}})\right]^2\times\left[\Psi_{\downarrow}(\mathbf{r}_{N^{\uparrow}_{\mathrm{elec}}+1},\ldots,\mathbf{r}_{N_{\mathrm{elec}}})\right]^2.\end{aligned} \tag{2.13}$$

2.2 Electron correlation

Electron correlation is defined as[16]

$$E_{\mathrm{corr}} = E_{\mathrm{exact}} - E_{\mathrm{HF}} \tag{2.14}$$

where E_{HF} is the *Hartree-Fock limit*, i.e. the limit to which the Hartree-Fock energy converges when the size of the one-electron basis set increases.

To include electron correlation effects, Ψ may be expanded in $\{|D_I\rangle\}$, the set of all the possible Slater determinants that can be built by putting $N^{\uparrow}_{\mathrm{elec}}$ electrons in N_{orb} orbitals and $N^{\downarrow}_{\mathrm{elec}}$ electrons in N_{orb} orbitals. The eigenvectors of $\widehat{H}$ are consequently expressed as linear combinations of Slater determinants

$$|\Psi_n\rangle = \sum_I c_I^n\,|D_I\rangle\,. \tag{2.15}$$

Solving the eigenvalue equations in this basis is referred to as *Full Configurations Interaction (FCI)* and yields solutions for Schrödinger's equation that are exact for the given atomic basis set. But the FCI is usually computationally intractable because of its scaling with the size of the basis set. Indeed, the size of the FCI space is

$$N_{\mathrm{FCI}} = \frac{N_{\mathrm{orb}}!}{N^{\uparrow}_{\mathrm{elec}}!(N_{\mathrm{orb}}-N^{\uparrow}_{\mathrm{elec}})!}\times\frac{N_{\mathrm{orb}}!}{N^{\downarrow}_{\mathrm{elec}}!(N_{\mathrm{orb}}-N^{\downarrow}_{\mathrm{elec}})!}. \tag{2.16}$$

Post-Hartree-Fock methods are trying to circumvent this problem, and therefore are essentially approximations of FCI.

2.3 Matrix elements of $\widehat{H}$

In the N-electron basis of Slater determinants, one expects the matrix elements of $\widehat{H}$ to be integrals over $3N$ dimensions. However, given the two-electron nature of the Hamiltonian, and because the set of molecular orbitals is orthogonal, Slater determinants that differ by more than two electrons yield a zero matrix element, and the other elements can be expressed as sums of integrals over 3 or 6 spatial dimensions, which can be computed for a reasonable cost. These simplifications are known as *Slater-Condon's rules*:

$$\langle D|\hat{H}|D\rangle = \sum_{i\in|D\rangle} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i\in|D\rangle}\sum_{j\in|D\rangle}\Big[(ii|jj) - (ij|ij)\Big] \tag{2.17}$$

$$\left\langle D\middle|\hat{H}\middle|D_p^r\right\rangle = \langle p|\hat{h}|r\rangle + \sum_{i\in|D\rangle}\Big[(pr|ii) - (pi|ri)\Big] \tag{2.18}$$

$$\left\langle D\middle|\hat{H}\middle|D_{pq}^{rs}\right\rangle = (pr|qs) - (ps|qr) \tag{2.19}$$

where $\hat{h}$ is the one-electron part of the Hamiltonian (kinetic energy and electron-nucleus potential),

$$\langle p|\hat{h}|r\rangle = \int d\mathbf{x}\ \phi_p^*(\mathbf{x})\left(-\frac{1}{2}\nabla + V_1(\mathbf{x})\right)\phi_h(\mathbf{x}), \tag{2.20}$$

$i \in |D\rangle$ means that ϕ_i belongs to the Slater determinant $|D\rangle$, $\left|D_{pq}^{rs}\right\rangle$ is a determinant obtained from $|D\rangle$ by substituting orbitals ϕ_p and ϕ_q by ϕ_r and ϕ_s, and

$$(ij|kl) = \int d\mathbf{x}_1 \int d\mathbf{x}_2\ \phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\phi_k^*(\mathbf{x}_2)\phi_l(\mathbf{x}_2). \tag{2.21}$$

2.4 Two-electron integrals

In the Hartree-Fock method, Roothaan's equations allow to solve the problem in the basis of *atomic* orbitals.[17] In this context, one needs to compute the $\mathcal{O}\left({N_{\text{orb}}}^4\right)$ two-electron integrals $(pq|rs)$ over the atomic orbitals. Thanks to a large effort in algorithmic development and implementation,[18, 19, 20, 21, 22, 23, 24] these integrals can now be computed very fast on modern computers.

But for post-Hartree-Fock methods, the computation of the two-electron integrals is a potential bottleneck. Indeed, when computing matrix elements of the Hamiltonian in the basis of Slater determinants, integrals over *molecular* orbitals are desired. Using Eq. (2.12), the cost of computing a single integral scales as $\mathcal{O}\left({N_{\text{orb}}}^4\right)$:

$$(ij|kl) = \sum_{pqrs} C_{pi} C_{qj} C_{rk} C_{sl} (pq|rs) \tag{2.22}$$

A naive computation of all integrals in the MO basis would cost $\mathcal{O}({N_{\text{orb}}}^8)$. Fortunately, computing all of them can be scaled down to $\mathcal{O}(N^5)$ by transforming the indices one by one:[25]

$$\begin{aligned}
(iq|rs) &= \sum_p C_{pi} (pq|rs) \\
(iq|ks) &= \sum_r C_{rk} (iq|rs) \quad \text{semi-transformed integrals} \qquad (2.23) \\
(ij|ks) &= \sum_q C_{qj} (iq|ks) \\
(ij|kl) &= \sum_s C_{sl} (ij|ks) \quad \text{fully transformed integrals} \qquad (2.24)
\end{aligned}$$

This step is known as the *four-index integral transformation*. In addition to being very costly, this step is no easy to parallelize in distributed way, because it implies multiple collective communications.[26, 27, 28, 29]

2.5 Variational post-Hartree-Fock methods

In variational methods, one tries to minimize the *variational energy*

$$E_{\text{var}} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_{\text{FCI}} \tag{2.25}$$

by optimizing the parameters of the wave function. Generally speaking, solving Schrödinger's equation in a basis of Slater determinants is called *Configuration Interaction (CI)*. In these methods, the molecular orbitals are kept fixed and the variational parameters are the coefficients associated with the Slater determinants. The general idea of CI methods is to select *a priori* a relevant subset of Slater determinants in which the CI problem will be solved, the FCI being the particular case where the whole $\{|D_i\rangle\}$ set is used.

One usual approach is to perform a FCI by allowing excitations from a reference determinant only within a reduced set of molecular orbitals. This is referred to as *Complete Active Space Configuration Interaction (CAS-CI)*. Choosing the CAS orbitals

often requires some chemical expertise. The CAS-SCF method minimizes the energy by performing iteratively a CAS-CI and optimizing the molecular orbitals.

Another usual CI approach is to select determinants according to their excitation degree — by how many occupied orbitals they differ — with respect to some reference. If the reference is the Hartree-Fock determinant and if only single and double excitations are considered, the method is known as *Configuration Interaction with Single and Double excitations (CISD)*. Alternatively, the reference can be a CAS-CI, in which case it is known as *Multi-Reference Configuration Interaction (MR-CI)*.

Regardless of the method, integrals involving all orbitals implied in at least one Slater determinant need to be computed in order to diagonalize $\widehat{H}$. Therefore CI methods cost at least $\mathcal{O}\left({N_{\text{orb}}}^5\right)$, due to the four-index transformation. In addition, the cost for fully diagonalizing $\widehat{H}$ is $\mathcal{O}\left({N_{\text{det}}}^3\right)$ with N_{det} the number of determinants in the considered subspace, which can be up to a few billion. This is usually not feasible, but only the few eigenvectors associated with lowest eigenvalues are typically of interest, so iterative methods can be used. The standard choice in quantum chemistry is to use the *Davidson diagonalization* originally developed by Ernest R. Davidson[30] specifically for CI methods.

2.6 Perturbative methods

One defines a *zeroth-order* Hamiltonian $\hat{H}^{(0)}$ as an approximate Hamiltonian which carries the dominant information of the exact Hamiltonian $\hat{H}$, and for which all the eigenvalues and eigenvectors are known.

$$\hat{H}^{(0)} \left|\Psi^{(0)}\right\rangle = E^{(0)} \left|\Psi^{(0)}\right\rangle \tag{2.26}$$

The difference between $\hat{H}$ and $\hat{H}^{(0)}$ is small enough to be considered as a *perturbation* $\hat{V}$:

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{V} \tag{2.27}$$

with λ a scalar connecting smoothly the approximate Hamiltonian ($\lambda = 0$) and the exact Hamiltonian ($\lambda = 1$). We will try to solve

$$\left(\hat{H}^{(0)} + \lambda \hat{V}\right) |\Psi(\lambda)\rangle = E(\lambda) |\Psi(\lambda)\rangle \tag{2.28}$$

assuming that solutions for $H(\lambda)$ can be written as a power series:

$$E(\lambda) = \sum_{l=0}^{\infty} \lambda^l e^{(l)} \tag{2.29}$$

$$|\Psi(\lambda)\rangle = \sum_{l=0}^{\infty} \lambda^l \left|\psi^{(l)}\right\rangle \tag{2.30}$$

Eq. (2.28) becomes

$$\left(\hat{H}^{(0)} + \lambda V\right)\left(\sum_{l=0}^{\infty} \lambda^l \left|\psi^{(l)}\right\rangle\right) = \left(\sum_{l=0}^{\infty} \lambda^l e^{(l)}\right)\left(\sum_{l=0}^{\infty} \lambda^l \left|\psi^{(l)}\right\rangle\right) \tag{2.31}$$

and equation at order n is obtained by isolating all terms that are multiplied by λ^n. For $n = 0$, we find Eq. (2.26). For $n = 1$ we have

$$\lambda\left[\hat{H}^{(0)}\left|\psi^{(1)}\right\rangle + V\left|\psi^{(0)}\right\rangle\right] = \lambda\left[e^{(0)}\left|\psi^{(1)}\right\rangle + e^{(1)}\left|\psi^{(0)}\right\rangle\right] \tag{2.32}$$

and so on…

The equation of order n involves all $e^{(m\leq n)}$ and $\left|\psi^{(m\leq n)}\right\rangle$. It is possible to iteratively solve the equations up a given value of n, each iteration i yielding $e^{(i)}$ and $\left|\psi^{(i)}\right\rangle$. Then the wave function and energy corrected at order n can be written

$$E^{(n)} = \sum_{i=0}^{n} e^{(i)} \tag{2.33}$$

$$\left|\Psi^{(n)}\right\rangle = \sum_{i=0}^{n} \left|\psi^{(i)}\right\rangle \tag{2.34}$$

The perturbation theory that is used is characterized by the choice of $\hat{H}^{(0)}$. If the zeroth-order (symmetric) Hamiltonian is chosen as the exact Hamiltonian on a subset of determinants ($1 \leq I \leq J \leq N_{\text{det}}$), and diagonal on the rest ($|\alpha\rangle = D_{K>N_{\text{det}}}$),

$$\langle D_I|\hat{H}^{(0)}|D_J\rangle = \langle D_I|\hat{H}|D_J\rangle \tag{2.35}$$

$$\langle D_I|\hat{H}^{(0)}|\alpha\rangle = 0 \tag{2.36}$$

$$\langle \alpha|\hat{H}^{(0)}|\alpha\rangle = \langle \alpha|\hat{H}|\alpha\rangle\,, \tag{2.37}$$

$$\tag{2.38}$$

the zeroth-order Hamiltonian is the so-called *Epstein-Nesbet* Hamiltonian. In that case,

$$\langle D_I|\hat{V}|\alpha\rangle = \langle D_I|\hat{H}|\alpha\rangle \tag{2.39}$$

$$\langle D_I|\hat{V}|D_J\rangle = 0 \tag{2.40}$$

$$\tag{2.41}$$

In Epstein-Nesbet perturbation theory, $e^{(0)}$ is the variational energy of the zeroth-order wave function, $e^{(1)} = 0$, and one needs to go to the second order to get an improvement on the energy:

$$E_{\text{PT2}} = e^{(2)} = \sum_{\alpha} \frac{\left\langle \alpha\middle|\hat{H}\middle|\Psi^{(0)}\right\rangle^2}{E^{(0)} - \langle \alpha|\hat{H}|\alpha\rangle} \tag{2.42}$$

2.7 Selected CI methods

These methods rely on the same principle as the usual CI approaches, except that determinants aren't chosen *a priori* based on an occupation or excitation criterion, but selected *on the fly* among the entire set of determinants based on their estimated contribution to the FCI wave function. Conventional CI methods can be seen as an exact resolution of Schrödinger's equation for a complete, well-defined subset of determinants (and for a given atomic basis set), while selected CI methods are more of a truncation of the FCI. The main advantages of these methods compared to the more conventional *a priori* selected ones, are that since the most relevant determinants are considered, they will typically yield a more accurate description of physical phenomena, and a much lower energy for an equivalent number of determinants. It has been noticed long ago that, even inside a predefined subspace of determinants, only a small number significantly contributes to the wave function.[31, 32] Therefore, an *on the fly* selection of determinants is a rather natural idea that has been proposed in the late 60's by Bender and Davidson[33] as well as Whitten and Hackmeyer[34] and is still very much under investigation, we can cite its stochastic variant the MC3I method[35] or the very recent *Machine Learning Configuration Interaction (MLCI).*[36]

The approach we are using is based on the *Configuration Interaction using a Perturbative Selection (CIPSI)* developed by Huron, Rancurel and Malrieu,[11] that iteratively selects *external* determinants $|\alpha\rangle$ (determinants which are not present in the wave function $|\Psi\rangle$) using a perturbative criterion

$$e_\alpha = \frac{\langle\Psi|\hat{H}|\alpha\rangle^2}{E_{\text{var}} - \langle\alpha|\hat{H}|\alpha\rangle} \tag{2.43}$$

with $|\alpha\rangle$ the external determinant being considered, and e_α the estimated gain in correlation energy that would be brought by the inclusion of $|\alpha\rangle$ in the wave function. E_{PT2} is an estimation of the total missing correlation energy:

$$E_{\text{PT2}} = \sum_\alpha e_\alpha \tag{2.44}$$

$$E_{\text{FCI}} \approx E_{\text{var}} + E_{\text{PT2}} \tag{2.45}$$

There is however a computational downside. In *a priori* selected methods, the rule by which determinants are selected is known *a priori*, and therefore, one can map a particular determinant to some row or column index.[37] As a consequence, it can be systematically determined to which matrix element of $\hat{H}$ a two-electron integral contributes. This allows for the implementation of so-called *integral-driven* methods, that work essentially by iterating over integrals. On the contrary, in selected methods an explicit list has to be maintained, and there is no immediate way to know whether a determinant has been selected, or what its index is. Consequently, so-called *determinant-driven* approaches will be used, in which iteration is done over determinants rather

than integrals. This can be a lot more expensive, since the number of determinants is typically much larger than the number of integrals. The number of determinants scales as $\mathcal{O}(N_{\text{orb}}!)$ while the number of integrals scales as $\mathcal{O}\left({N_{\text{orb}}}^4\right)$ with the number of MOs. Furthermore, determinant-driven methods require an effective way to compare determinants in order to extract the corresponding excitation operators, and a way to rapidly fetch the associated integrals involved, as described in section 2.3.

Because of this high computational cost, approximations have been proposed.[14] And recently, the *Heat-Bath Configuration Interaction (HCI)* algorithm has taken farther the idea of a more approximate but extremely cheap selection.[38, 39] Compared to CIPSI, the selection criterion is simplified to

$$e_\alpha^{\text{HCI}} = \max\left(\left|c_I \left\langle D_I|\hat{H}|\alpha\right\rangle\right|\right) \tag{2.46}$$

This algorithmically allows for an extremely fast selection of doubly excited determinants by an integral-driven approach.

Full Configuration Interaction Quantum Monte Carlo (FCI-QMC) is an alternate approach to selection recently proposed in 2009 by Alavi *et al.*,[10, 40, 41] where signed walkers spawn from one determinant to connected ones, with a probability that is a function of the associated matrix element. The average proportion of walkers on a determinant converges to its coefficient in the FCI wave function.

A more "bruteforce" approach at stochastic selection is *Monte-Carlo CI (MCCI)*,[42, 43] where determinants are randomly added to the variational wave function. After diagonalization, the determinants of smaller coefficient are removed, and new random determinants are added.

Chapter 3

Determinant-driven computation of the matrix elements

Contents

Generally speaking, implementing a wave function method requires iterations over either two-electron integrals, or determinants. Those approaches are referred to as *integral-driven* and *determinant-driven*. Because the number of determinants grows much faster than the number of double excitations, an efficient implementation would likely favor an integral driven approach. However, in most cases, the determinant-driven approach is more intuitive, as it stays closer to the basic CI equations. The Quantum Package is intended for developers, and thus prioritizes the approach that is easier for designing new methods.

For performance, it is vital that some basic operations are done efficiently. Notably, the computation of matrix elements of the Hamiltonian. This raises some questions

about the data structures used to represent the two-electron integrals and determinants, as well as their consequences from an algorithmic point of view.

This chapter is going to address these questions, by going through the basic concepts of our approach to determinant-driven computation.

3.1 Storage of the two-electron integrals

In all the algorithms presented, all the needed two-electron integrals are kept in memory and require a fast random access. A hash table is the natural choice which allows the storage of only non-zero values with a retrieval of data in nearly constant time,[44] but standard hashing algorithms tend to shuffle the data to limit the probability of collisions. Here, we favor instead the locality of the data over the probability of collision using the hash function given in Algorithm 1. It returns the same value for all the keys which are related by the permutation symmetry of the indices, keeps some locality in the storage of the data, and can be evaluated in the order of 10 CPU cycles if the integer division by two is replaced by a right bit shift instruction.

1 **Function** HASH(i,j,k,l): `/* Hash function for two-electron integrals. */`
  Data: i,j,k,l are the orbital indices
  Result: The corresponding hash
2 $p \leftarrow \min(i,k)$;
3 $r \leftarrow \max(i,k)$;
4 $t \leftarrow p + \frac{1}{2}r(r-1)$;
5 $q \leftarrow \min(j,l)$;
6 $s \leftarrow \max(j,l)$;
7 $u \leftarrow q + \frac{1}{2}s(s-1)$;
8 $v \leftarrow \min(t,u)$;
9 $w \leftarrow \max(t,u)$;
10 **return** $v + \frac{1}{2}w(w-1)$;

Algorithm 1: Hash function that maps the all the quartets of orbital indices related by permutation symmetry to a unique integer.

The hash table is such that each bucket can potentially store 2^{15} consecutive key-value pairs. The 15 least significant bits of the hash value are removed to give the index of the bucket ($i_{\text{bucket}} = \lfloor \text{hash}(i,j,k,l)/2^{15} \rfloor$), and only those 15 bits need to be stored in the bucket for the storage of the key ($\text{hash}(i,j,k,l) \mod 2^{16}$). Hence, the storage of the keys only requires two bytes per key. The keys within a bucket are sorted in increasing order, enabling a binary search within the bucket. The search of the key is

Table 3.1: Time to access the integrals (in nanoseconds/integral) with different access patterns. The time to generate random numbers (measured as 67 ns/integral) was not counted in the random access results.

Access	Array	Hash table
i, j, k, l	9.72	125.79
i, j, l, k	9.72	120.64
i, k, j, l	10.29	144.65
l, k, j, i	88.62	125.79
l, k, i, j	88.62	120.64
Random	170.00	370.00

always fast since the binary search is bounded by 15 misses and the maximum size of the array of keys is 64 kiB, the typical size of the L1 cache.

The efficiency of the storage as a hash table was measured on a dual socket Intel Xeon E5-2680 v2 @ 2.80GHz processor, taking the water molecule with the cc-pVQZ basis set (115 molecular orbitals). The time to access all the integrals was measured by looping over all integrals using different loop orderings. The results are given in table 3.1, the reference being the storage as a plain four-dimensional array.

In the array storage, the value of 170 ns/integral in the random access case is typical of the latency to fetch a value in the RAM modules when the data is not in any level of cache. When the data is accessed with a stride of one (the i, j, l, k storage) the cache levels accelerate the access by a factor of 18×, down to 9.71 ns/integral, corresponding mostly to the overhead of the function call, the retrieval of the data being negligible.

With the hash table, the random access is only 2.18× slower than the random access in the array. Indeed, two random accesses are required: one for the first element of the bucket of keys one for the corresponding value. The rest of the extra time corresponds to the binary search. The locality of the data can be exploited: when the access is done with a regular access pattern, the data is fetched $\sim$ 3× faster than using a random access.

A CIPSI calculation was run with the array storage and with the hash table storage. With the hash storage, the total wall clock time was increased only by a factor of two.

So to accelerate the access to the most frequently used integrals, all the integrals involving the 128 MOs closest to the Fermi level are copied in a dense array of 128^4 elements (2 GiB).

3.2 Internal representation of determinants

Determinants can be written as a string of creation operators applied to the vacuum state $|\rangle$.

$$a_i^\dagger a_j^\dagger a_k^\dagger |\rangle = |I\rangle \tag{3.1}$$

Because of the fermionic nature of electrons, a permutation of two contiguous creation operators results in a sign change, which makes their ordering relevant.

$$a_j^\dagger a_i^\dagger = -a_i^\dagger a_j^\dagger \tag{3.2}$$

$$a_j^\dagger a_i^\dagger a_k^\dagger |\rangle = -|I\rangle \tag{3.3}$$

This effectively allows to make any N_{perm} permutations and always get $-1^{N_{\text{perm}}}|I\rangle$. A determinant can be broken down into two pieces of information:

- A set of creation operators, corresponding to the set of occupied spinorbitals in the determinant.
- An ordering of the creation operators, responsible for the sign of the determinant. Once an ordering operator $\hat{\mathcal{O}}$ is chosen and applied to all the determinants, it is sufficient to store the sign change that occurs when applying this operator to the string of creation operators. This sign will be referred to as the *phase factor*.

Determinants are always associated with a coefficient. So if the determinants are always built after applying them the same ordering operator, we don't need to make the phase factor part of the determinant's internal representation. The sign may simply be reported on the associated coefficient.

All the determinants will be built using the order where all the $\uparrow$ spinorbitals are placed before the $\downarrow$ spinorbitals, as in the Waller-Hartree determinant representation:

$$\hat{\mathcal{O}}|I\rangle = \hat{I}|\rangle = \hat{I}_\uparrow \hat{I}_\downarrow |\rangle \tag{3.4}$$

and within each operator $\hat{I}_\uparrow$ and $\hat{I}_\downarrow$, the creation operators are sorted with increasing indices. For instance, consider the determinant built from the set of spinorbitals $\{i, j, k, \bar{i}\}$ with $i < j < k$,

$$|J\rangle = a_j^\dagger a_k^\dagger a_{\bar{i}}^\dagger a_i^\dagger |\rangle. \tag{3.5}$$

If we happen to encounter such a determinant, our choice of representation imposes us to consider its re-ordered expression

$$\hat{\mathcal{O}}|J\rangle = -a_i^\dagger a_j^\dagger a_k^\dagger a_{\bar{i}}^\dagger |\rangle = -|J\rangle \tag{3.6}$$

and the sign change (or *phase factor*) will need to be handled.

The indices of the creation operators (or equivalently the occupations of the spinorbitals), are stored using the so-called *bitstring* encoding. A bitstring is an array of bits ; typically, the 64-bit binary representation of an integer is a bitstring of size 64. Quite simply, the idea is to map each spinorbital to a single bit, with a value is set to its occupation number. In other words 0 and 1 are associated with the *unoccupied* and *occupied* states. By this definition, bitstrings encode the indices of the occupied spinorbitals.

For simplicity and performance considerations, the occupations of the $\uparrow$ and $\downarrow$ spinorbitals are stored on different bitstrings, rather than interleaved or otherwise merged in the same one. This allows to straightforwardly map orbital index n to bit index $n-1$ (orbitals are usually indexed from 1, while bits are indexed from 0), and makes a bitstring a set of orbitals. This makes the representation of a determinant a tuple of two bitstrings, associated with respectively $\uparrow$ and $\downarrow$ spinorbitals. Such objects are referred to as $\uparrow\downarrow$-*bitstrings*, and generally define a set of spinorbitals. When used to define a determinant, they imply the previously defined ordering.

- I is the $\uparrow\downarrow$-bitstring representation of $|I\rangle$
- $I_\uparrow$ is the bitstring representation of the set of occupied $\uparrow$ spinorbitals of $|I\rangle$
- $I_\downarrow$ is the bitstring representation of the set of occupied $\downarrow$ spinorbitals of $|I\rangle$

The storage space required for a single determinant is, in principle, one bit per spinorbital, or $2 \times N_{\text{orb}}$ bits. However, because CPUs are designed to handle efficiently 64-bit integers, each spin part is stored as an array of 64-bit integers, the unused space being padded with zeros. The actual storage needed for a determinant is $2 \times 64 \times N_{\text{int}}$ bits, where N_{int} is the number of 64-bits integers needed to store one spin part:

$$N_{\text{int}} = \left\lfloor \frac{N_{\text{orb}} - 1}{64} \right\rfloor + 1. \tag{3.7}$$

The Fortran representation of a bitstring is an array of N_{int} `integer*8` (64-bit integers). The Fortran representation of an $\uparrow\downarrow$-bitstring is a two dimensional array of `integer*8`, the first dimension of size N_{int} and the second of size 2, corresponding to the $\uparrow$ and $\downarrow$ spin parts.

```
! I is an updown-bitstring
! I_up and I_down are bitstrings

integer*8 :: I(N_int, 2)
integer*8 :: I_up(N_int), I_down(N_int)

... ! load some determinant in I
I_up   (:) = I(:,1)
I_down (:) = I(:,2)
```

In formulas or algorithms, depending on the level of detail desired, a bitstring or $\uparrow\downarrow$-bitstring may also be treated as a single mathematical integer (in $\mathbb{Z}$), avoiding the cumbersome separation into 64-bit packs. However, in algorithms we will usually try to stay closer to the actual implementation. I being the $\uparrow\downarrow$-bitstring associated with $|I\rangle$, we can explicitly refer to a single element of the 64-bit integer array as

$$I_\sigma[i] \ ; \ \sigma \in \{\uparrow, \downarrow\} \ ; \ 0 \leq i < N_{\text{int}} \tag{3.8}$$

which is the bitstring representation of the σ spinorbitals of determinant $|I\rangle$ in the range $[1 + i \times 64, \min((i+1) \times 64, N_{\text{orb}})]$, indexed from 0 to 63.

3.3 Bit manipulation

The bitstring encoding is a compact way of storing determinants, but it is more than just a data structure. It allows to perform a variety of operations on determinants by taking advantage of CPU's hardware aptitude to perform efficiently bitwise operations on integers.

In many of the presented algorithms, some Fortran intrinsics will be of use. Each of those maps to a CPU instruction that is available on modern CPUs.

- POPCNT(I) : Returns the number of non-zero bits for a given integer I.
  POPCNT$(00011000_2) = 2$.
- TRAILZ(I) : Returns the number of trailing zero bits for a given integer I.
  TRAILZ$(00000100_2) = 2$.
- IBCLR(I, n) : Returns the value of I with the bit at the n-th position set to zero (the rightmost bit is at position zero).
  IBCLR$(00001111_2, 2) = 00001011_2$.
- IOR(I, J) : Bitwise OR logical operation.
  IOR$(1100_2, 1010_2) = 1110_2$.
- IEOR(I, J) : Bitwise XOR (exclusive or) logical operation.
  IEOR$(1100_2, 1010_2) = 0110_2$.
- IAND(I, J) : Bitwise AND logical operation.
  IAND$(1100_2, 1010_2) = 1000_2$.
- NOT(I) : Bitwise NOT logical operation.
  NOT$(00001100_2) = 11110011_2$.

- $\mathtt{ISHFT}(I, n)$: Returns I with bits shifted $|n|$ places to the left if $n > 0$, otherwise to the right. Bits shifted out of the range are lost. Zeros are shifted from the opposite end.
  $\mathtt{ISHFT}(01001110_2, 2) = 00111000_2$,
  $\mathtt{ISHFT}(01001110_2, -2) = 00010011_2$.

- $\mathtt{BTEST}(I, n)$: Returns TRUE if the n-th bit of I is set, otherwise FALSE.
  $\mathtt{BTEST}(00001000_2, 3) = \mathtt{TRUE}$.

Those intrinsics apply to integers with at most 64-bits. This however is a purely implementational limitation, so depending on the level of detail desired, this constraint can be unambiguously lifted in formulas or algorithms. Different notations will be used for the 64-bit and the $\mathbb{Z}$ cases, as they are not always equivalent. For example, the $\mathtt{ISHFT}(\_, \_)$ Fortran intrinsic always returns zero for shifts larger than 64 bits, which is not the case for the shift_left$(_, _)$ function over mathematical integers. All binary operators are of same precedence and left-associative.

64-bit variant	mathematical variant
$\mathtt{ISHFT}(I, n)$	shift_left(I, n)
$\mathtt{TRAILZ}(I)$	trailing_zeros(I)
$\mathtt{IBCLR}(I, n)$	bit_clear(I, n)
$\mathtt{BTEST}(I, n)$	bit_test(I, n)
$\mathtt{NOT}(I)$	$\neg I$
$\mathtt{IAND}(I, J)$	$I \wedge J$
$\mathtt{IOR}(I, J)$	$I \vee J$
$\mathtt{IEOR}(I, J)$	$I \oplus J$
$\mathtt{POPCNT}(I)$	$\|\|I\|\|$

Some examples of how these instructions can be used are given below. They are key to understand how we can determine the holes and particles involved in the $\hat{T}_{I \to J}$ excitation operator defined by

$$|J\rangle = \hat{T}_{I \to J} |I\rangle . \tag{3.9}$$

Let I and J be the bitstring representations of $|I\rangle$ and $|J\rangle$, and P a bitstring with $N_{\text{int}} = 1$.

- $I_\uparrow$: bitstring representation of the set of $\uparrow$ spinorbitals of $|I\rangle$
- $\left\| I_\uparrow \right\|$: number of spinorbitals in $I_\uparrow$ (equal to the number of $\uparrow$ electrons).
- $I_\uparrow \oplus J_\uparrow$: bitstring representation of the set of $\uparrow$ spinorbitals that are present in either $I_\uparrow$ or $J_\uparrow$, but not in both (exclusive disjunction). This operator identifies all the $\uparrow$ spinorbitals involved in the excitation from $|I\rangle$ to $|J\rangle$.

- $I_\uparrow \wedge \left(I_\uparrow \oplus J_\uparrow\right)$: bitstring representation of the set of $\uparrow$ spinorbitals of $|I\rangle$ involved in the excitation from $|I\rangle$ to $|J\rangle$. This corresponds to the indices of the holes in the excitation $\hat{T}_{I\to J}$ or to the particles in $\hat{T}_{J\to I}$.
- $\left\|I_\uparrow \oplus J_\uparrow\right\|/2$: because the excitation of an electron involves 2 spinorbitals (one hole and one particle), this is the $\uparrow$ excitation degree between $|I\rangle$ and $|J\rangle$.
- $\mathrm{TRAILZ}(P)+1$: the index of the lowest orbital in P if $P \neq 0$. If $P = 0$, this function returns 65.
- $\mathrm{IBCLR}(P, \mathrm{TRAILZ}(P))$: P without its orbital of lowest index.

3.4 Identification of holes and particles

An algorithm used to compute the excitation degree is presented as algorithm 2, and one to compute the sets of created holes and particles as algorithm 3. Algorithm 3, however, returns the sets as bitstrings. Extracting the indices from a bitstring is another basic operation, presented as algorithm 4. Because computing excitations is a hotspot of the program, and because we are typically interested in double excitations at most, a more specialized algorithm can be used.[45]

```
1 Function EXC_DEGREE(I, J):
      Data: I, J: bitstring representations of determinants |I⟩ and |J⟩.
      Result: Returns the excitation degree between |I⟩ and |J⟩, namely
              ½‖I ⊕ J‖
2     X ← 0 ;
3     for σ ∈ {↑, ↓} do
4         for i ← 0, N_int − 1 do
5             X ← X + POPCNT(IEOR(I_σ[i], J_σ[i])) ;
6         end
7     end
8     return X/2;
```

Algorithm 2: Returns the degree of excitation between two determinants.

```
1  Function EXC(I,J):
       Data: I, J: the bitstring representations of determinants |I⟩ and
             |J⟩ = T̂_{I→J} |I⟩
       Result: Returns a tuple (P, H), where P and H are respectively the sets of
               particles and holes created by T̂_{I→J}, as ↑↓-bitstrings.
2      for σ ∈ {↑, ↓} do
3          for i ← 0, N_int − 1 do
4              C ← IEOR(I_σ[i], J_σ[i],);
5              P_σ[i] ← IAND(C, J_σ[i],);
6              H_σ[i] ← IAND(C, I_σ[i],);
7          end
8      end
9      return (P, H);
```

Algorithm 3: Returns the holes and particles created in an excitation as bitstrings.

```
1  Function LIST_FROM_BITSTRING(P):
       Data: P a bitstring. On output, P is destroyed.
       Result: L the list of orbital indices in P in increasing order.
2      k ← 0 ;
3      for i ← 0, N_int − 1 do
4          while P[i] ≠ 0 do
5              e ← TRAILZ(P[i]) + 1;
6              P[i] ← IBCLR(P[i], e);
7              L[k] ← e + i × 64;
8              k ← k + 1;
9          end
10         /* L contains k elements                                    */
11         return L
12     end
```

Algorithm 4: Transforms a bitstring into a list of orbital indices.

3.5 Phase factors

The computation of phase factors is slightly more complex. The following explanation is limited to one spin part. More detail will be given later about why spin parts can be treated independently. As we have seen in section 3.2, the ↑↓-bitstring representation of determinants implies an ordering of creation operators : first all the ↑ operators,

then all the $\downarrow$, both with increasing orbital indices.

Whenever we build a new determinant by applying an excitation operator, we obtain a determinant that is initially expressed not just with a different ordering, but with a mix of creation and annihilation operators.

First of all, we have to make this initial expression unambiguous by precisely defining excitation operators. We have defined an implicit ordering for the expression of determinants, we also need an implicit ordering for the expression of excitation operators. Like for determinants, we pack together $\uparrow$ and $\downarrow$ operators.

$$\hat{T} = \hat{T}_\uparrow \hat{T}_\downarrow. \tag{3.10}$$

Within $\hat{T}_\uparrow$ and $\hat{T}_\downarrow$, the creation and annihilation operators are separately sorted with increasing indices, then interleaved starting with a creation. In other words, $\hat{T}_\uparrow$ and $\hat{T}_\downarrow$ are written as products of single excitations, lowest particle with lowest hole, then second lowest particle with second lowest hole, *etc*, and we arbitrarily chose to put creation before annihilation operators. For example, the double excitation $\hat{T}_{ab}^{cd}$ with $a < b < c < d$ is expressed as

$$\hat{T}_{ab}^{cd} = a_c^\dagger a_a a_d^\dagger a_b = \hat{T}_a^c \hat{T}_b^d. \tag{3.11}$$

We can now express $\hat{T}$ as a series of operators. In most cases, permuting contiguous operators will still just result in a sign change.

$$a_j a_i = -a_i a_j \tag{3.12}$$

$$a_j^\dagger a_i = \begin{cases} -a_i a_j^\dagger & i \neq j \\ 1 - a_i a_i^\dagger & i = j. \end{cases} \tag{3.13}$$

A particular case is the permutation of a creation and an annihilation operator with the same index. Indeed, if spinorbital l is unoccupied in $|I\rangle$,

$$a_l a_l^\dagger |I\rangle = |I\rangle \tag{3.14}$$

$$a_l^\dagger a_l |I\rangle = 0. \tag{3.15}$$

In the first case, a particle is created then annihilated, resulting in the same determinant. In the second case, there is an attempt at annihilating a particle that does not exist, resulting in 0. It is of course the opposite if l is occupied in $|I\rangle$. These formulas will be used to remove annihilation operators from the expression of a determinant.

Let $|I\rangle$ and $|K\rangle$ be two determinants with spinorbitals ordered as in the $\uparrow\downarrow$-bitstring representation:

$$|I\rangle = a_i^\dagger a_j^\dagger a_k^\dagger |\rangle \tag{3.16}$$

$$|K\rangle = a_i^\dagger a_k^\dagger a_l^\dagger |\rangle \tag{3.17}$$

with $i < j < k < l$. When one applies the excitation operator $\hat{T}_j^l$ to $|I\rangle$,

$$\hat{T}_j^l |I\rangle = a_l^\dagger a_j a_i^\dagger a_j^\dagger a_k^\dagger |\rangle. \tag{3.18}$$

To build the corresponding $\uparrow\downarrow$-bitstring, one needs to reorder the operators by permuting contiguous operators. It takes $n = 1$ permutation to bring a_j behind $a_j^\dagger$:

$$\hat{T}_j^l |I\rangle = -a_l^\dagger a_i^\dagger a_j a_j^\dagger a_k^\dagger |\rangle. \tag{3.19}$$

Using equation 3.14,

$$\hat{T}_j^l |I\rangle = -a_l^\dagger a_i^\dagger a_k^\dagger |\rangle. \tag{3.20}$$

Then, it takes again n permutations to bring $a_l^\dagger$ to the position formerly occupied by $a_j^\dagger$, and $x = 1$ more permutation to bring it at its final position.

$$\hat{T}_j^l |I\rangle = -a_i^\dagger a_k^\dagger a_l^\dagger |\rangle = -|J\rangle\,. \tag{3.21}$$

The total number of permutations needed is $N_{\text{perm}} = 2n + x$. The parity of N_{perm} is the parity of x. As can be seen, x is the number of spinorbitals with indices in the $]j, l[$ range in $|I\rangle$ (regardless of whether $l > j$ or $l < j$). In our case, there was one occupied spinorbital k, so N_{perm} is odd and we ended with a negative phase factor, $-1^{N_{\text{perm}}} = -1$.

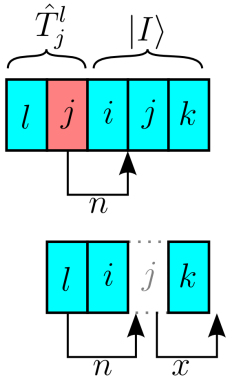


Figure 3.1: Computation of the phase factor.

3.5.1 Treating spin parts separately

It is not immediately obvious that $\uparrow$ and $\downarrow$ spin parts can be treated independently. It is possible because our ordering packs together operators of same spin in the expression

of both determinants (Eq. (3.4)) and excitations (Eq. (3.10)), and also because we never consider excitations where the spin is changed (spin-flips).

$$\hat{T}\,|I\rangle = \hat{T}_\uparrow \hat{T}_\downarrow\, \hat{I}_\uparrow \hat{I}_\downarrow |\rangle \tag{3.22}$$

$$= \left(\hat{T}_\uparrow \hat{I}_\uparrow\right)\left(\hat{T}_\downarrow \hat{I}_\downarrow\right)|\rangle = \hat{J}_\uparrow \hat{J}_\downarrow |\rangle = \pm\,|J\rangle\,. \tag{3.23}$$

The number of creation and annihilation operators in an excitation operator is always even. Hence, going from Eq. (3.22) to Eq. (3.23) by permuting $\hat{T}_\downarrow$ with $\hat{I}_\uparrow$ always requires an even number of permutations, keeping the phase factor unchanged. So the phase factors can be easily computed by reordering separately $\hat{T}_\uparrow \hat{I}_\uparrow$ and $\hat{T}_\downarrow \hat{I}_\downarrow$.

3.5.2 Single excitations

For a given determinant $|I\rangle$ and a singly excited determinant $|J\rangle = \hat{\mathcal{O}}\hat{T}^{j'}_{i'}\,|I\rangle$ where both use the ordering of spinorbitals defined previously, the phase factor can be determined by the parity of

$$N^I_{ij} = \sum_{k=i+1}^{j-1} \mathrm{occ}(k, I) \tag{3.24}$$

where $i = \min(i', j')$, $j = \max(i', j')$ and

$$\mathrm{occ}(k, I) = \delta\left(a^\dagger_k a_k\,|I\rangle - |I\rangle\right) = \begin{cases} 1 & \text{if } a^\dagger_k a_k\,|I\rangle = |I\rangle \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

is the occupation number of spinorbital k in determinant $|I\rangle$.[1] We will use the fact that the parity of an integer X can be obtained by extracting the least significant bit of its binary representation:

$$p(X) = X \wedge 1 = \begin{cases} 1 & \text{if } X \text{ is odd} \\ 0 & \text{if } X \text{ is even.} \end{cases}. \tag{3.26}$$

A bitstring R_{ij}, containing all orbitals in a given range $]i, j > i[$ can be built as

$$R_{ij} = \mathrm{shift\_left}(\neg 0, i) \oplus \mathrm{shift\_left}(\neg 0, j-1) \tag{3.27}$$

where $\neg 0$ denotes the bitstring where all the bits are set to one.

[1] Note that occupation numbers in $|I\rangle$ and $|J\rangle$ are by construction equal for any spinorbital $k \neq i \neq j$, so while we will use $|I\rangle$ in the following, we could as well use $|J\rangle$.

Let I be the bitstring representation of $|I\rangle$, the number of σ-spinorbitals in $|I\rangle$ in the given orbital range can be evaluated as

$$N_{ij}^{I} = \left\| I_\sigma \wedge R_{ij} \right\|. \tag{3.28}$$

So the phase factor to be applied when going from $|I\rangle$ to $|J\rangle$ is

$$\Phi(|I\rangle \to |J\rangle) = -1^{p(N_{ij}^{I})}. \tag{3.29}$$

Using the above formulas and taking into account the internal representation of $\uparrow\downarrow$-bitstring as arrays of N_{int} **integer*8**, we get algorithm 5.

Function PHASE_SINGLE(I_σ, a, b):
  Data: I_σ the bitstring representation of $\sigma \in \{\uparrow, \downarrow\}$ spinorbitals of $|I\rangle$
  Data: a, b indices of a hole and a particle created in σ spinorbitals of $|I\rangle$
  Result: The phase factor associated with $\hat{\mathcal{O}}\hat{T}_a^b |I\rangle$
  high $\leftarrow \max(a,b) - 1$;
  low $\leftarrow \min(a,b) - 1$;
  $il \leftarrow \frac{\text{low}}{64}$;
  $ih \leftarrow \frac{\text{high}}{64}$;
  $l \leftarrow$ low mod 64 ;
  $h \leftarrow$ high mod 64 ;
  for $i \leftarrow il, ih - 1$ **do**
    mask$[i]$ = NOT(0);
  end
  mask$[ih] \leftarrow$ ISHFT(NOT(0), $h+1$) ;
  mask$[il] \leftarrow$ IEOR(mask$[il]$, ISHFT(NOT(0), l)) ;
  $N_{\text{perm}} \leftarrow 0$;
  for $i \leftarrow il, ih$ **do**
    $N_{\text{perm}} \leftarrow N_{\text{perm}} +$ POPCNT(IAND(I_i, mask$[i]$)) ;
  end
  phase$[0:1] = [1.0\,;\ -1.0]$;
  return phase[IAND(N_{perm}, 1)] ;

Algorithm 5: Returns the phase factor of $\hat{\mathcal{O}}\hat{T}_a^b |I\rangle$.

3.5.3 Phase masks

Algorithm 5 is a general one, efficient for computing the phase factor for arbitrary determinants. However, a phase computation is typically needed every time two determinants are compared, resulting in a vast amount of computational power being

consumed. When only this method was used in the QUANTUM PACKAGE , it was not uncommon to find that a large fraction of the computational time was spent in phase computations. Advantage can be taken of the fact that, in most cases, the considered determinants aren't actually arbitrary. Usually, the phase computation will be performed repeatedly with the same determinant. For a fairly modest computational price, it is possible to compress the phase information from a particular determinant and make it cheaper to extract. The underlying principle is little more than a cumulative sum. If, for a determinant $|I\rangle$ that is going to be used repeatedly for phase computations, we pre-compute

$$E_i^I = \sum_{k=1}^{i} \mathrm{occ}(k, I) \tag{3.30}$$

we can access N_{ij}^I for any i and $j > i$ with no need to loop over k as

$$N_{ij}^I = E_{j-1}^I - E_i^I. \tag{3.31}$$

This requires to store E^I which is an integer array of size $2 \times (N_{\mathrm{orb}} + 1)$. This may be somewhat memory consuming if we want to pre-compute and store E for each determinant. However, because the actual information needed isn't N_{ij}^I, but merely its parity, we only need to store the so-called "phase mask" array P^I

$$P_i^I = p(E_i^I) \tag{3.32}$$

which is 1 bit of information per spinorbital, as opposed to an integer big enough to accommodate a number of electrons.

$$E_i^I = 2 \times \left\lfloor \frac{E_i^I}{2} \right\rfloor + P_i^I \tag{3.33}$$

$$N_{ij}^I = E_{j-1}^I - E_i^I \tag{3.34}$$

$$= 2 \times \left(\left\lfloor \frac{E_{j-1}^I}{2} \right\rfloor - \left\lfloor \frac{E_i^I}{2} \right\rfloor \right) + P_{j-1}^I - P_i^I \tag{3.35}$$

In the last equation, N_{ij}^I is expressed as a sum of three terms. The first one being even by construction, the parity of N_{ij}^I is the parity of the rest of the sum:

$$p(N_{ij}^I) = p(P_{j-1}^I - P_i^I) \tag{3.36}$$

This can be rewritten in a slightly more efficient way as

$$p(N_{ij}^I) = P_{j-1}^I \oplus P_i^I \tag{3.37}$$

We have used sorted indices $i < j$. In practice, we know which index refers to the particle and which refers to the hole. With an excitation operator $\hat{T}_h^p$ applied to $|I\rangle$, noticing that

- if $p > h$ we have

$$p(N_{hp}^I) = P_{p-1}^I \oplus P_h^I \tag{3.38}$$
$$= P_p^I \oplus P_h^I \oplus \mathrm{occ}(p, I) \tag{3.39}$$
$$= P_p^I \oplus P_h^I \tag{3.40}$$

- if $h > p$ we have

$$p(N_{ph}^I) = P_{h-1}^I \oplus P_p^I \tag{3.41}$$
$$= P_h^I \oplus P_p^I \oplus \mathrm{occ}(h, I) \tag{3.42}$$
$$= P_h^I \oplus P_p^I \oplus 1 \tag{3.43}$$

the phase factor can be computed as

$$\Phi\left(|I\rangle \to \hat{\mathcal{O}}\hat{T}_h^p |I\rangle\right) = \begin{cases} (-1)^{P_h^I \oplus P_p^I} & \text{if } p > h \\ -(-1)^{P_h^I \oplus P_p^I} & \text{if } h > p \end{cases} \tag{3.44}$$

Currently, the Quantum Package does not store a phase mask for each determinant of the wave function, but recomputes it whenever needed before a loop. The algorithm for computing P^I as a bitstring is shown as algorithm 6. It uses a trick to get the phase mask of a single integer with logarithmic complexity (loop at line 6 of the algorithm). With P^0 a single-integer bitstring for which we want to compute the phase mask, bits being indexed from 0:

1. $P^1 \leftarrow P^0 \oplus \mathrm{shift\_left}(P^0, 2^0)$.

$$P_i^1 = \begin{cases} p(P_i^0) & \text{if } i < 1 \\ p(P_{i-1}^0 + P_i^0) & \text{if } i \geq 1 \end{cases} \tag{3.45}$$

$$P_i^1 = p\left(\sum_{j=max(i-1,0)}^{i} P_j^0\right) \tag{3.46}$$

```
Function PHASEMASK(I):
    Data: I the bitstring representation of |I⟩
    Result: P is the phase mask associated with |I⟩, as described in Eq. (3.32),
            as a bitstring.
    for σ ∈ {↑, ↓} do
        r ← 0 ;
        for i ← 0, N_int − 1 do
            P_σ[i] ← IEOR(I_σ[i], ISHFT(I_σ[i], 1) ;
            for d ← 0, 5 do
                P_σ[i] ← IEOR(P_σ[i], ISHFT(P_σ[i], 2^d) ;
            end
            P_σ[i] ← IEOR(P_σ[i], r) ;
            if IAND(POPCNT(I_σ[i]), 1) == 1 then
                r ← NOT(r) ;
            end
        end
    end
    return P ;
```

Algorithm 6: Returns a phase mask as a bitstring.

2. $P^2 \leftarrow P^1 \oplus \text{shift\_left}(P^1, 2^1)$.

$$P_i^2 = \begin{cases} p(P_i^1) & \text{if } i < 2 \\ p(P_i^1 + P_{i-2}^1) & \text{if } i \geq 2 \end{cases} \tag{3.47}$$

$$P_i^2 = p\left(\sum_{j=max(i-3,0)}^{i} P_j^0\right) \tag{3.48}$$

3. $P^3 \leftarrow P^2 \oplus \text{shift\_left}(P^2, 2^2)$.

$$P_i^3 = \begin{cases} p(P_i^2) & \text{if } i < 4 \\ p(P_i^2 + P_{i-4}^2) & \text{if } i \geq 4 \end{cases} \tag{3.49}$$

$$P_i^3 = p\left(\sum_{j=max(i-7,0)}^{i} P_j^0\right) \tag{3.50}$$

4. etc...

Function PHASE_PHASEMASK (P^I, i, j)**:**

Data: P^I is the phase mask array associated with $|I\rangle$, as described in Eq. (3.32). i and j are spinorbitals of spin $\sigma \in \{\uparrow, \downarrow\}$ so that $\hat{T}_i^j |I\rangle \neq 0$.

Result: The phase factor associated with $\hat{\mathcal{O}}\hat{T}_i^j |I\rangle$.

if $j < i$ **then**
  $c \leftarrow 0$;
else
  $c \leftarrow 1$;
end
$i_n \leftarrow (i-1)/64$;
$j_n \leftarrow (j-1)/64$;
$i_b \leftarrow (i-1) \mod 64$;
$j_b \leftarrow (j-1) \mod 64$;
$B \leftarrow \texttt{ISHFT}(P^I_\sigma[i_n], -i_b) \oplus \texttt{ISHFT}(P^I_\sigma[j_n], -j_b)$;
if $B \wedge 1 = c$ **then**
  return -1;
else
  return 1 ;
end

Algorithm 7: Returns a phase factor associated with a single excitation using a phase mask. This routine may be optimized by replacing the integer divisions by bit shifts and the modulo by an AND instruction.

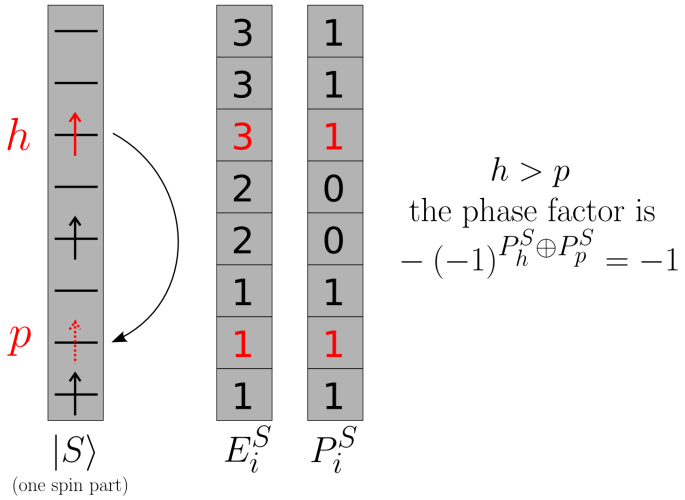


Figure 3.2: Illustrative exemple of phase mask use. h and p are the hole and particle involved in the excitation. E_i^S is the number of electrons in spinorbitals i and below (of same spin), P_i^S is its parity. Because $h > p$, $E_{h-1}^S - E_p^S$ gives the number of electrons that are "crossed" during the excitation. The phase factor is computed according to Eq.3.44

For a 64-bit integer we have to go to

$$P_{i<64}^6 = p\left(\sum_{j=0}^{i} P_j^0\right) \tag{3.51}$$

which is the definition of a phase mask (here indexed from 0).

The method to compute the phase factor associated with a single excitation from a phase mask is shown as algorithm 7. With this method, the cost of computing a phase factor — after paying the overhead of computing P — is accessing two bits and applying the **IEOR** function to them, in addition to the tests. Unlike the general method, its cost doesn't depend on N_{int}, and doesn't require to deal with the cumbersome boundaries of 64-bit integers.

3.5.4 Double excitations

A double excitation can be expressed as a product of two single excitations. In the case of an $\uparrow\downarrow$ double excitation, the two single excitations are independent, so the phase factor is merely the product of the phase factors computed for each spin part.

$$\Phi\Big(|I\rangle \to \hat{\mathcal{O}}\hat{T}_{p\bar{q}}^{r\bar{s}}|I\rangle\Big) = \Phi\Big(|I\rangle \to \hat{\mathcal{O}}\hat{T}_{p}^{r}|I\rangle\Big) \times \Phi\Big(|I\rangle \to \hat{\mathcal{O}}\hat{T}_{\bar{q}}^{\bar{s}}|I\rangle\Big) \tag{3.52}$$

There is a slight complication for $\uparrow\uparrow$ or $\downarrow\downarrow$ excitations. The ordering we defined for excitation operators is of importance. In order to write a double excitation as a product of two single excitations, it must be ensured that the ordering matches the one we defined : lowest particle with lowest hole, highest particle with highest hole.

As far as phase computation goes, it is irrelevant which index of an excitation operator is a creation and which is an annihilation. So, for convenience, we can define

$$\tilde{T}_{ab} = \hat{T}_a^b + \hat{T}_b^a \tag{3.53}$$

since at most one of $\hat{T}_a^b$ or $\hat{T}_b^a$ can be applied to a determinant.

Considering a double excitation $\tilde{T}^2$ involving 4 spinorbitals p, q, r, s of same spin, there are two possible situations, shown in figure 3.3.

- It can be expressed as two single excitations that do not cross, i.e.

$$\tilde{T}^2 = \tilde{T}_{pr}\tilde{T}_{qs} \; ; \; p < r < q < s \tag{3.54}$$

  In this case, the numbers of particles in the ranges $]p, r[$ and $]q, s[$ remain unchanged, so we can write

$$\Phi\Big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}^2 |I\rangle\Big) = \Phi\big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}_{pr} |I\rangle\big) \times \Phi\big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}_{qs} |I\rangle\big) \tag{3.55}$$

- It can be expressed as two single excitations that cross, i.e.

$$\tilde{T}^2 = \tilde{T}_{pr}\tilde{T}_{qs} \; ; \; p < q < r < s \tag{3.56}$$

  As we can see in figure 3.3, applying $\tilde{T}_{qs}$ results in a particle being created or annihilated in the range $]p, r[$, resulting in a change of parity for the number of particles in that range. Therefore,

$$\Phi\big(\hat{\mathcal{O}}\tilde{T}_{qs} |I\rangle \to \hat{\mathcal{O}}\tilde{T}_{pr}\tilde{T}_{qs} |I\rangle\big) = -\Phi\big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}_{pr} |I\rangle\big) \tag{3.57}$$

$$\Phi\Big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}^2 |I\rangle\Big) = -\Phi\big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}_{pr} |I\rangle\big) \times \Phi\big(|I\rangle \to \hat{\mathcal{O}}\tilde{T}_{qs} |I\rangle\big) \tag{3.58}$$

In practice, because LIST_FROM_BITSTRING returns indices with increasing order, if we determine an excitation operator $\hat{T}_{pq}^{rs}$ so that $\hat{T}_{pq}^{rs} |I\rangle = \pm |J\rangle$, we know $p < q$ and $r < s$. So for a $\uparrow\uparrow$ or $\downarrow\downarrow$ double excitation, we compute the phase factor from P^I the phase mask associated with I, as

$$\Phi\Big(|I\rangle \to \hat{\mathcal{O}}\hat{T}_{pq}^{rs} |I\rangle\Big) = \begin{cases} \kappa & \text{if } \neg(\max(p,r) > \min(q,s)) \\ -\kappa & \text{if } (\max(p,r) > \min(q,s)) \end{cases} \tag{3.59}$$

with

$$\kappa = \text{PHASE\_PHASEMASK}(P^I, p, r) \times \text{PHASE\_PHASEMASK}(P^I, q, s) \tag{3.60}$$

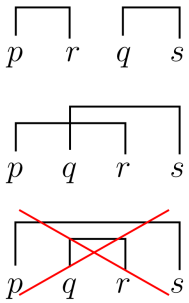


Figure 3.3: Crossing of two single excitations. The third situation doesn't fit the ordering we imposed on excitation operators. $p \leftrightarrow s$ is either lowest electron to highest orbital, or highest electron to lowest orbital.

3.6 Summary

Taking advantage of low-level hardware instructions, we are able, for a minimal cost, to find $\hat{T}$ so that $\hat{T}\,|I\rangle = |J\rangle$, which yields the i, j, k, l indices of the associated two-electron integral(s). Then, fetching its value can be done quickly using the proposed hash table. Because of the data structure we are using to store determinants, we also need to compute a phase factor, which can be done efficiently using the introduced phase masks.

Thanks to the algorithms presented in this chapter, we are able to efficiently compute the matrix elements of $\hat{H}$, which are the basic ingredients of determinant-driven methods. The instructions and functions we have defined here will be used in the following chapters for many different purposes.

Chapter 4

Diagonalization with Davidson's algorithm

Contents

Finding eigenvectors associated with the lowest eigenvalues of the Hamiltonian is a necessary step in configuration interaction. Standard diagonalization algorithms scale as $\mathcal{O}\left({N_{\text{det}}}^3\right)$ in terms of computation, and $\mathcal{O}\left({N_{\text{det}}}^2\right)$ in terms of storage, so the cost of the full diagonalization is prohibitive as N_{det} ranges usually between a few million and a few billion.

Fortunately, not all the spectrum of $\hat{H}$ is required: only the few lowest eigenstates are of interest. The Davidson algorithm[30, 46, 47, 48, 49] is an iterative algorithm which aims at extracting the first few N_{states} lowest eigenstates of a large matrix. This algorithm reduces the cost of the computation to $\mathcal{O}\left(N_{\text{states}}{N_{\text{det}}}^2\right)$, and of the storage to $\mathcal{O}(N_{\text{states}}N_{\text{det}})$. It is presented as algorithm 8.

Function DAVIDSON_DIAG(N_{states}, $\mathbf{U}$):
  Data: N_{states} : Number of requested states
  Data: N_{det} : Number of determinants
  Data: $\mathbf{U}$: Guess vectors, $N_{\text{det}} \times N_{\text{states}}$
  Result: N_{states} lowest eigenvalues eigenvectors of $\mathbf{H}$
  converged $\leftarrow$ FALSE ;
  while $\neg$*converged* **do**
    Gram-Schmidt orthonormalization of $\mathbf{U}$;
    $\mathbf{W} \leftarrow \mathbf{H}\,\mathbf{U}$;
    $\mathbf{h} \leftarrow \mathbf{U}^{\dagger}\,\mathbf{W}$;
    Diagonalize $\mathbf{h}$: eigenvalues E and eigenvectors $\mathbf{y}$;
    $\mathbf{U}' \leftarrow \mathbf{U}\,\mathbf{y}$;
    $\mathbf{W}' \leftarrow \mathbf{W}\,\mathbf{y}$;
    for $k \leftarrow 1, N_{states}$ **do**
      for $i \leftarrow 1, N_{det}$ **do**
        $\mathbf{R}_{ik} \leftarrow \frac{E_k \mathbf{U}'_{ik} - \mathbf{W}'_{ik}}{\mathbf{H}_{ii} - E_k}$;
      end
    end
    converged $\leftarrow \|\mathbf{R}\| < \epsilon$;
    $\mathbf{U} \leftarrow [\mathbf{U}, \mathbf{R}]$;
  end
  return $\mathbf{U}$;

Algorithm 8: Davidson's diagonalization algorithm

4.1 The computational bottleneck

Algorithmically, the expensive part of the Davidson diagonalization is the computation of the matrix product $\mathbf{H}\,\mathbf{U}$. Determinants are connected by $\mathbf{H}$ only if they differ by no more than two spinorbitals. Therefore, the number of non-zero elements per line of $\mathbf{H}$ is equal to the number of single and double excitation operators, namely $\mathcal{O}\left({N_{\text{elec}}^{\uparrow}}^2 \times (N_{\text{orb}} - N_{\text{elec}}^{\uparrow})^2\right)$. As $\mathbf{H}$ is symmetric, the number of non-zero elements per column is identical. This makes the $\mathbf{H}$ matrix very sparse, but for large basis sets the whole $\mathbf{H}$ matrix may still not fit in the memory of a single node, as the number of non-zero entries to store is $\mathcal{O}\left(N_{\text{det}} \times {N_{\text{elec}}^{\uparrow}}^2 \times (N_{\text{orb}} - N_{\text{elec}}^{\uparrow})^2\right)$. One possibility would be to distribute the storage of Hamiltonian among multiple compute nodes, and use a distributed library such as PBLAS[50] to perform the matrix-vector operations. Another approach is to use a so-called *direct* algorithm, where the matrix elements are computed *on the fly*, and this is the approach chosen in the QUANTUM PACKAGE .

This effectively means iterating over all pairs of determinants $|D_I\rangle$ an $|D_J\rangle$, checking whether $|D_I\rangle$ and $|D_J\rangle$ are connected by $\mathbf{H}$ and if they are, accessing the corresponding integral(s) and computing the phase factor. Even though we presented in section 3.4 a very efficient method to compute the excitation degree between two determinants, the number of such computations to be made scales as ${N_{\text{det}}}^2$, which can still be prohibitively high. To get an efficient determinant-driven implementation it is mandatory to filter out all pairs of determinants that are not connected by $\mathbf{H}$, and iterate only over connected pairs. To reach this goal, we have implemented an algorithm similar to the *Direct Selected Configuration Interaction Using Strings* (DISCIUS) algorithm.[13]

The finality is to build the matrix $\mathbf{W}$ as

$$W_{Ik} = \sum_J H_{IJ}\, U_{Jk}. \tag{4.1}$$

The diagonal of $\mathbf{H}$ is precomputed and stored in a one-dimensional array, as it is used both for the calculation of $\mathbf{W}$ and the residual $\mathbf{R}$. $\mathbf{W}$ is initialized with $W_{Ik} = H_{II}\, U_{Ik}$, and each time a connected pair of determinants $(I, J \neq I)$ is found, the I-th component of all states k stored in $\mathbf{W}$ is updated accordingly. To make this step efficient memory-wise, $\mathbf{U}$ and $\mathbf{W}$ are stored transposed, such that the state indices k are contiguous in memory.

We present our algorithm for iterating over only off-diagonal non-zero elements of $\mathbf{H}$ — in other words, pairs of connected determinants — as algorithms 9 and 10.

We define $N_{\text{det}}^{\uparrow}$ and $N_{\text{det}}^{\downarrow}$ as the number of different $\uparrow$ and $\downarrow$ spin parts present in the expression of the wave function. Computing the contributions the same-spin single

Data: $N_{\text{det}}^{\uparrow}$ the number of unique $\uparrow$ spin parts present in $|\Psi\rangle$
Data: D is the array of determinants present in $|\Psi\rangle$, sorted by $\uparrow$-major order (all determinants sharing the same $\uparrow$ part are next to each other)
Data: A the array so that $A[n]$ is the index of the first occurence of the n^{th} unique $\uparrow$ spin part in D. For algorithmic convenience we set $A[N_{\text{det}}^{\uparrow}+1] = N_{\text{det}}+1$

```
for a ← 1, N_det^↑ do
    /* All determinants sharing D[A(a)]_↑ ↑-spin part are in the
       range [A(a), A(a+1)−1]                                    */
    for b1 ← A(a), A(a+1) − 1 do
        for b2 ← b1 + 1, A(a+1) − 1 do
            if EXC_DEGREE(D[b1]_↓, D[b2]_↓) ≤ 2 then
                |D[b1]⟩ connected to |D[b2]⟩ by single or double ↓ excitation.
            end
        end
    end
end
/* Single and double ↑ excitations are found by the same
   algorithm after flipping spins                            */
```

Algorithm 9: Find internal determinants connected by purely $\uparrow$ or purely $\downarrow$ single or double excitations

```
   Data: see algorithm 9
1  for a1 ← 1, N_det^↑ do
2      for a2 ← a1 + 1, N_det^↑ do
3          if EXC_DEGREE(D[A(a1)]_↑, D[A(a2)]_↑) ≠ 1 then
4              cycle a2 loop;
5          end
6          for b1 ← A(a1), A(a1 + 1) − 1 do
7              for b2 ← A(a2), A(a2 + 1) − 1 do
8                  if EXC_DEGREE(D[b1]_↓, D[b2]_↓) = 1 then
9                      |D[b1]⟩ connected to |D[b2]⟩ by ↑↓ excitation.
10                 end
11             end
12         end
13     end
14 end
```

Algorithm 10: Find internal determinants connected by $\uparrow\downarrow$ double excitations (sequential, using the symmetry of $\mathbf{H}$).

and double excitations (algorithm 9) scales as $\mathcal{O}\left({N_{\text{det}}^{\uparrow}}^2\right)$, which is equal to $\mathcal{O}(N_{\text{det}})$. Indeed, when the FCI is reached $N_{\text{det}} = N_{\text{det}}^{\uparrow} \times N_{\text{det}}^{\downarrow}$.

The $\uparrow\downarrow$ double excitations are the most expensive (algorithm 10) as they scale as $\mathcal{O}\left(N_{\text{det}}^{\uparrow} {N_{\text{det}}^{\downarrow}}^2\right) = \mathcal{O}\left({N_{\text{det}}^{\uparrow}}^3\right)$. Indeed, the `cycle` instruction at line 4 makes the iterations over $b1$ and $b2$ do the computation only a number of times bounded by the number of possible single excitations ($N_{\text{elec}}^{\uparrow} \times \left(N_{\text{orb}} - N_{\text{elec}}^{\uparrow}\right)$). So at the FCI level, this step scales as $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$.

One can remark that during the computation of the contributions of the single excitations, one can store the lists of all singly-excited determinants for all $\uparrow$ and $\downarrow$ spin parts. These lists can be reused in the computation of the contributions of the $\uparrow\downarrow$ double excitations, so as to loop only over the single excitations on $D_{\uparrow}$, and on the single excitations on $D_{\downarrow}$. As the lengths of the lists of single excitations are bounded by $N_{\text{elec}}^{\uparrow} \times \left(N_{\text{orb}} - N_{\text{elec}}^{\uparrow}\right)$, the algorithm then scales linearly with N_{det}. However, in practice the CIPSI selection produces wave functions where $N_{\text{det}}^{\uparrow}$ and $N_{\text{det}}^{\downarrow}$ are much larger that ${N_{\text{det}}}^{1/2}$, and the storage of the single excitations can become a memory bottleneck that we want to avoid at the cost of more computation.

4.2 Sorting

The presented algorithm requires to sort the determinants by $\uparrow$-major order : all determinants sharing the same $\uparrow$ spin part next to each other. To perform this sort, we simply consider the bitstring representation of the determinants as tuples of integers and sort the list of tuples.

Surprisingly, the sort can be done in $\mathcal{O}(N_{\text{det}})$ instead of $\mathcal{O}(N_{\text{det}} \log(N_{\text{det}}))$ with the radix sort algorithm.[51] The principle of the radix sort is presented in algorithm 11. The key feature enabling the transition from $\mathcal{O}(N_{\text{det}})$ to $\mathcal{O}(N_{\text{det}} \log(N_{\text{det}}))$ is the fact that the set of sorted integers is bounded, and one can easily verify that the number of operations is proportional to $64 \times N_{\text{det}}$. So sorting the determinants of the wave function is not a bottleneck, and the flop-optimal algorithm still scales as $\mathcal{O}(N_{\text{det}})$.

4.3 Parallelization

To minimize the network communication, we separate the calculation in tasks such that the tasks build disjoint pieces of the result. A task corresponds a range of indices I in Eq (4.1). Therefore, the communication for the result is $\mathcal{O}(N_{\text{det}})$, and independent of the number of compute nodes. However, each task needs the complete $\mathbf{U}$ matrix, so its needs to be broadcast efficiently on every compute node at the beginning of the calculation. This broadcast if performed via an MPI library call for optimal performance,[52] and we use one MPI process per node such that the amount of communication scales with the number of nodes and not with the number of cores.

When idle, an MPI process requests a task to the server, and computes the corresponding result in parallel with OpenMP.[53] This allows the sharing of the $\mathbf{U}$ matrix, as well as the result array for $\mathbf{W}$, but also of all the large constant data needed for the calculation, such as the two-electron integrals. The OpenMP parallelization is made on the outermost loop, so each OpenMP thread loops over a smaller range of I (algorithm 12). The write access to the result is guaranteed to be safe, without requiring a lock. As the OpenMP tasks are not guaranteed to be balanced, we have used a dynamic scheduling, with a chunk size of 64 elements. The reason for this chunk size is to force that multiple threads accumulate their results in memory addresses far apart, avoiding the so-called *false sharing* performance degradation that occurs when multiple threads write simultaneously in the same cache line.[54] When the result is fully computed, it is sent back to the master process and a new task is requested, until the task queue is empty.

```
1  Function RADIX_SORT(D, N):
       Data: D: Array of integers to sort in input, sorted in output
       Data: N: Length of the array D
2      RADIX_SORT_rec(D, N, 64) ;
3  Function RADIX_SORT_rec(D, N, i):
       Data: D: Array of integers to sort in input, sorted in output
       Data: N: Length of the array D
       Data: r: index of the inspected bit
4      if r ≥ 0 then
5          Allocate temporary arrays D_0(1 : N) and D_1(1 : N);
6          l ← 1 ;
7          r ← 1 ;
8          for k ← 1, N do
9              if bit_test(D(k), i) then
10                 D_1[l] ← D[k] ;
11                 l ← l + 1 ;
12             else
13                 D_0[r] ← D[k] ;
14                 r ← r + 1 ;
15             end
16         end
17         RADIX_SORT_rec(D_0, r, i − 1) ;
18         RADIX_SORT_rec(D_1, l, i − 1) ;
19         for k ← 1, l do
20             D(k) ← D_0(k) ;
21         end
22         r ← 1 ;
23         for k ← l + 1, N do
24             D(k) ← D_1(r) ;
25             r ← r + 1 ;
26         end
27     end
```

Algorithm 11: Radix sort algorithm for non-negative integers

```
   Data: see algorithm 9
   Data: first, last : the boundaries of the range of determinants (in D) processed
         by the current OpenMP thread.
 1 for a1 ← 1, N_det^↑ do
 2 |   if A(a1 + 1) − 1 < first then
 3 |   |   cycle a1 loop;
 4 |   end
 5 |   if A(a1) > last) then
 6 |   |   return ;
 7 |   end
 8 |   f ← max(first, A(a1)) ;
 9 |   t ← min(last, A(a1 + 1) − 1) ;
10 |   for a2 ← 1, N_det^↑ do
11 |   |   if EXC_DEGREE(D[A(a1)]_↑, D[A(a2)]_↑) ≠ 1 then
12 |   |   |   cycle a2 loop;
13 |   |   end
14 |   |   for b1 ← f, t do
15 |   |   |   for b2 ← A(a2), A(a2 + 1) − 1 do
16 |   |   |   |   if EXC_DEGREE(D[b1]_↓, D[b2]_↓) = 1 then
17 |   |   |   |   |   |D[b1]⟩ connected to |D[b2]⟩ by ↑↓ excitation.
18 |   |   |   |   end
19 |   |   |   end
20 |   |   end
21 |   end
22 end
```

Algorithm 12: Find internal determinants connected by $\uparrow\downarrow$ double excitations, one in the range [first, last] the other in the range [first, N_{det}]

4.4 Symmetry of the Hamiltonian matrix

Taking into account the symmetry of $\widehat{H}$, each pair should be found only once, and the associated update would be

$$W_{Ik} \leftarrow W_{Ik} + U_{Jk} H_{IJ} \tag{4.2}$$
$$W_{Jk} \leftarrow W_{Jk} + U_{Ik} H_{IJ} \tag{4.3}$$

This reduces the computational effort by a factor of two, but the result of each task now has a size of N_{det} and no more the reduced size of $N_{\text{det}}/N_{\text{task}}$, since all the elements of $\mathbf{W}$ can potentially be modified. This increase of communication has the effect of killing the parallel efficiency.

There are some additional drawbacks. First, in the non-symmetric case, a thread accumulates data in W_{Ik}, a memory location which is the same for multiple consecutive accesses, and in which no other thread can write. This pattern is memory-efficient. In the symmetric case, there is also an access to W_{Jk}. The access to W_{Jk} are non-contiguous and don't have a predictable pattern by the hardware. Such memory access patterns are terribly inefficient, especially when writing. In addition, a global memory lock should be acquired since there is no guarantee than another thread is not writing in that memory location at the same time. To avoid the lock, another solution is to use an output vector which is private to the thread, but it would make the memory grow as $N_{\text{CPU}} \times N_{\text{states}} \times N_{\text{det}}$, which is what we wanted to avoid using shared memory parallelism.

For a large number of nodes it is indisputably preferable not to use the symmetry of $\mathbf{H}$, even though it might seem surprising that increasing the number of operations can give a better time to solution.

4.5 Ensuring the solutions have the desired spin state

When working in a truncated space of determinants, there is no guarantee that the eigenstate of $\hat{H}$ will also be eigenstates of the spin operator $\hat{S}^2$. And when the proper conditions are fulfilled (see section 5.8), all the lowest eigenvectors may be of different spin states.

To help find solutions of the desired spin state, we use a penalty method in the diagonalization.[55] We modify the Hamiltonian as

$$\tilde{\mathbf{H}} = \mathbf{H} + \gamma \left(\mathbf{S}^2 - \mathbf{I} \langle S^2 \rangle_{\text{target}} \right)^2 \tag{4.4}$$

where γ is a fixed parameter. In the Davidson algorithm, this requires the additional computation of $\mathbf{S}\,\mathbf{U}$, for which the cost is expected to be the same as the cost of $\mathbf{H}\,\mathbf{U}$ as the expensive part is the search for the connections.

We have modified the function computing **H U** so that it also computes **S U** on the fly. Indeed, once a pair of connected determinants has been found, if they correspond to an $\uparrow\downarrow$ double excitation or to a diagonal term, the $\hat{S}^2$ contribution is added to an extra output vector, with almost no extra computational cost.

4.6 Summary

Davidson's diagonalization algorithm was implemented in its multi-state version. A direct algorithm was designed for arbitrary sets of determinants, and the formal scaling was reduced to $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$, and can be further reduced to $\mathcal{O}(N_{\text{det}})$ at the cost of additional storage. The implementation was parallelized using two levels of parallelism, MPI and OpenMP, keeping in mind the reduction of the communication. The empirical speedup measurements are presented in chapter 9.

Chapter 5

Selection with the CIPSI criterion

Contents

5.1 The basic algorithm

My initial and most important work has been the improvement of the implementation of the CIPSI algorithm present in the Quantum Package , that had been implemented by my predecessor.[56] As was briefly described in section 2.7, it is an *on the fly* iterative selection algorithm, where determinants are added to the variational wave function according to a perturbative criterion. Because it gathers a large amount of information, this CIPSI implementation has been the basis for other subsequent works presented in the next chapters.

The n^{th} iteration of CIPSI can be described like so:

1. The variational function $\left|\Psi^{(n)}\right\rangle$ is defined over a set of determinants $\{|D_I\rangle\}^{(n)}$ in which we diagonalize $\widehat{H}$

$$\left|\Psi^{(n)}\right\rangle = \sum_I c_I^{(n)} |D_I\rangle \tag{5.1}$$

   The determinants in $\{|D_I\rangle\}^{(n)}$ will be characterized as *internal*.

2. For all *external* determinants $|\alpha\rangle \notin \{|D_I\rangle\}^{(n)}$, we compute the perturbative contribution

$$e_\alpha = \frac{\left\langle \Psi^{(n)} \middle| \hat{H} \middle| \alpha \right\rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}. \tag{5.2}$$

   As we use Epstein-Nesbet perturbation theory, $E^{(n)} = E_{\text{var}}^{(n)}$ is the variational energy of the wave function at the current iteration (note that another perturbation theory could be used here).

3. Summing the contributions of all the external determinants gives the second order perturbative correction

$$E_{\text{PT2}}^{(n)} = \sum_\alpha e_\alpha \tag{5.3}$$

   and the FCI energy E_{FCI} can be estimated

$$E_{\text{FCI}} \approx E_{\text{var}}^{(n)} + E_{\text{PT2}}^{(n)} \tag{5.4}$$

4. We extract $\{|\alpha_\star\rangle\}^{(n)}$ the subset of determinants $|\alpha\rangle$ with the largest contributions e_α, and add them to the variational space

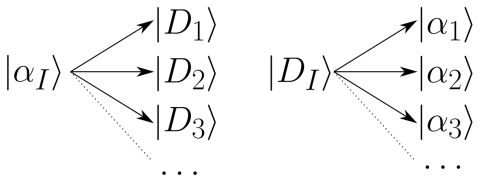
$$\{|D_I\rangle\}^{(n+1)} = \{|D_I\rangle\}^{(n)} \cup \{|\alpha_\star\rangle\}^{(n)} \tag{5.5}$$

5. Go to iteration $n+1$, or exit on some criterion (number of determinants in the wave function, low $E_{\text{PT2}}^{(n)}$, …).

As can be seen, CIPSI involves the creation of an external space and a precise knowledge of how it interacts with the internal space. Algorithmically speaking, we will need to enumerate all connections between all internal and all external determinants. There are, perhaps schematically, two ways to do this :

- "external to internal", looping over all possible $|\alpha\rangle$ and computing e_α.

- "internal to external", looping over all internal determinants $|D_I\rangle$ and all single or double excitations $\hat{T}$, creating $|\alpha\rangle = \delta(\langle\Psi|\alpha\rangle)\hat{\mathcal{O}}\hat{T}|D_I\rangle$, then incrementing $\tilde{e}_\alpha$ by $c_I \langle D_I|\hat{H}|\alpha\rangle$. Finally get $e_\alpha = \frac{\tilde{e}_\alpha^2}{E^{(n)} - H_{\alpha\alpha}}$.



The first approach is less tempting, as it means finding connections between the arbitrary set of internal determinants, and another arbitrary set of $|\alpha\rangle$ typically orders of magnitude larger.

While the second approach sounds more straightforward, it has the obvious issue of requiring all e_α to be stored in memory simultaneously. Unfortunately this is usually not feasible, since their number scales as $\mathcal{O}\left(N_{\text{det}} \times {N_{\text{elec}}^{\uparrow}}^2 \times \left(N_{\text{orb}} - N_{\text{elec}}^{\uparrow}\right)^2\right)$. The first approach therefore seems simpler when it comes to computing e_α, but it begs the question of how to generate all possible $|\alpha\rangle$ with no duplicates.

Both our former and newer implementations of CIPSI generate the external space in an "internal to external" way, that is, by applying single and double excitations to internal determinants ; a determinant used to generate $|\alpha\rangle$ is referred to as a *generator*. Ensuring each e_α is considered only once is done by checking that all the determinants $|\alpha\rangle$ generated by the generator $|D_I\rangle$ are not connected to any of the generators in $\{|D_{J<I}\rangle\}$. If a connection $\hat{T}$ is found, it means that $|\alpha\rangle$ is generated from $|D_J\rangle$ as $\hat{\mathcal{O}}\hat{T}|D_J\rangle$, and should not be considered by the current generator.

5.2 Approximations

Given the qualitative nature of this procedure — each $|\alpha\rangle$ is either selected or not — it is possible to save a vast amount of computational time with minimal approximations. These were present in the original implementation and retained in the new one.

From now on, we will consider that the determinants are sorted such that

$$c_I^2 \geq c_{I+1}^2 \tag{5.6}$$

Two approximations are made :

- The first approximation restricts the set $\{|\alpha\rangle\}$. It is very unlikely $|\alpha\rangle$ will be selected if it is not connected to any $|D_I\rangle$ with a large coefficient. Therefore, it is

possible to only consider the determinants of larger coefficient as generators. We choose a number of generators N_{gen} and only consider $\left|D_{I\leq N_{\text{gen}}}\right\rangle$ as generators. In practice we set N_{gen} according to a norm threshold n_g, picking N_{gen} as the highest value fulfilling

$$\sum_{I\leq N_{\text{gen}}} c_I^2 \leq n_g. \tag{5.7}$$

This approximation is a variant of the *three-class CIPSI*,[14] and typically $n_g = 0.99$ is used in the calculations.

- The second approximation reduces the cost of e_α. We do not need extremely accurate values for e_α as small differences are unlikely to substantially change the subset of the largest ones. So connections to $|D_I\rangle$ with small coefficients $|c_I|$ can be neglected in the expression of e_α. This approximation is achieved in a similar way by defining a threshold n_s on the norm of the wave function, and $N_{sel} \geq N_{\text{gen}}$ a number of so-called *selectors*. We approximate

$$\langle\Psi|\hat{H}|\alpha\rangle \approx \sum_{I\leq N_{sel}} c_I \langle D_I|\hat{H}|\alpha\rangle . \tag{5.8}$$

Typically, we use $n_s = 0.999$.

Note that generator determinants are a subset of selector determinants. See figure 5.1.

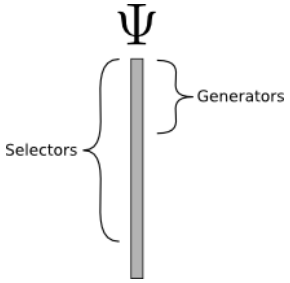


Figure 5.1: Determinants are sorted by decreasing c_I^2 ; generator and selector subsets are defined.

5.3 Initial implementation

Originally, the Quantum Package generated the external space in a "internal to external" way, by applying all excitations on all determinants ; but the computation of e_α itself was a straightforward "external to internal", computing a single e_α at a time, avoiding the problem of keeping track of all e_α simultaneously.

Data: $|\Psi\rangle$ with c_I^2 sorted in decreasing order.
Result: Guarantees all e_α are computed only once.

for $g \leftarrow 1, N_{gen}$ **do**
  /* apply all double excitations on $|D_g\rangle$ */
  forall $|\alpha\rangle$; $\langle D_g|H|\alpha\rangle \neq 0$ **do**
    for $p \leftarrow 1, g-1$ **do**
      if $|\alpha\rangle$ *connected to* $|D_p\rangle$ **then**
        /* $|\alpha\rangle$ has already been generated by $|D_p\rangle$ */
        discard this $|\alpha\rangle$;
      end
    end
    if $|\alpha\rangle \in \{D_{N_{sel}+1}, \ldots, D_{N_{det}}\}$ **then**
      /* $|\alpha\rangle \in \mathcal{D}$ */
      discard this $|\alpha\rangle$
    end
    $R \leftarrow 0$;
    for $s \leftarrow g, N_{sel}$ **do**
      $R \leftarrow R + c_s \langle D_s|\hat{H}|\alpha\rangle$;
      /* $|D_s\rangle = |\alpha\rangle$ is noticed when computing $\langle D_s|\hat{H}|\alpha\rangle$ */
      if $|D_s\rangle = |\alpha\rangle$ **then**
        /* $|\alpha\rangle \in \mathcal{D}$ */
        discard this $|\alpha\rangle$
      end
    end
    assert $R = \langle\Psi|\hat{H}|\alpha\rangle$;
    $e_\alpha = \frac{R^2}{E_{\text{var}} - \langle\alpha|\hat{H}|\alpha\rangle}$
  end
end

Algorithm 13: Simple CIPSI

While this relatively simple implementation has been abandoned, it is briefly presented for pedagogical reasons. A slightly more detailed algorithmic version is shown as algorithm 13.

1. Loop over generators $|G\rangle \in \left\{ \left| D_{I \leq N_{\text{gen}}} \right\rangle \right\}$.
2. Generate all singly and doubly excited determinants connected to $|G\rangle$.
3. From this set, discard those that appear in $\{|D_I\rangle\}$. This is now a set of $|\alpha\rangle$.
4. From this set, discard those that are connected $\{\left|D_{J \leq I}\right\rangle\}$. This is now a set of unique $|\alpha\rangle$.
5. Compute $e_\alpha = \frac{\langle \Psi | \hat{H} | \alpha \rangle^2}{E_{\text{var}} - \langle \alpha | \hat{H} | \alpha \rangle}$ for those new $|\alpha\rangle$.

5.4 Principle of the new algorithm

The current approach is intermediate between computing e_α one by one, and keeping track of all of them at the same time. It creates a subset, or *batch* of external determinants small enough to fit into memory, and importantly, that isn't arbitrary. A batch G_{pq} is defined by a doubly ionized generator

$$\left|G_{pq}\right\rangle = a_p a_q \left|G\right\rangle . \tag{5.9}$$

Determinants contained in the G_{pq} batch, some of which may be unique $|\alpha\rangle$, can be systematically defined by two indices r and s with

$$\hat{O} a_r^\dagger a_s^\dagger a_p a_q \left|G\right\rangle = \left|G_{pq}^{rs}\right\rangle . \tag{5.10}$$

Essentially, determinants in a batch are defined by their difference to $\left|G_{pq}\right\rangle$. Therefore, comparing $\left|G_{pq}\right\rangle$ to a selector determinant allows to systematically determine which $|\alpha\rangle$ of the batch it will connect to, and by what excitation. Additional filtering mechanisms are set up to avoid considering selectors that do not interact with the current batch. Those will be made explicit later on. Comparing figures 5.2 and 5.3 hints the differences between the former an newer algorithm. Note that because generators are a subset of selectors, a particular $|\alpha\rangle$ generated from $\left|D_g\right\rangle$ must be checked for connection to all selectors either as generators or as selectors.

- $\{|D_I\rangle \ ; \ I < g\}$ as generators to check if $|\alpha\rangle$ has been previously generated
- $\{|D_I\rangle \ ; \ g \leq I \leq N_{\text{sel}}\}$ as selectors to compute $\langle \Psi | \hat{H} | \alpha \rangle$.

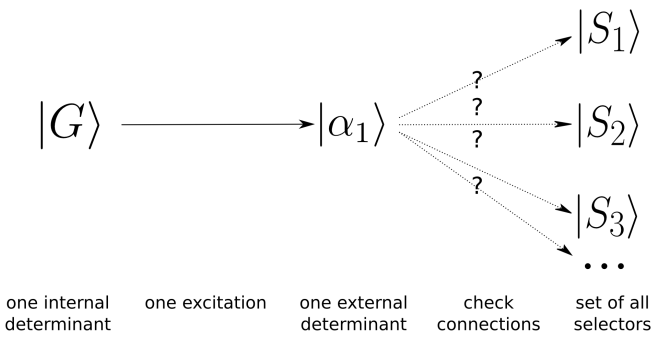


Figure 5.2: Original CIPSI schematic representation, some details omitted

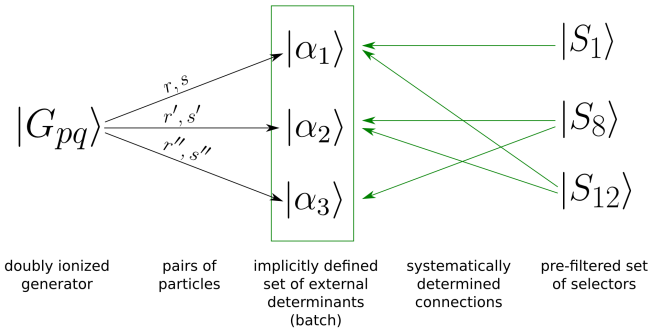


Figure 5.3: New CIPSI schematic representation, some details omitted.

5.4.1 Unfiltered algorithm

Filtering of selectors is a somewhat natural idea that was actually implemented before the batch approach. It however can easily be understood as something added "on top" of it, so it will be detailed in the next section and ignored in this one.

1. Iterate over $|G\rangle \in \left\{ \left| D_{I \leq N_{\text{gen}}} \right\rangle \right\}$.
2. Iterate over all possible $a_p a_q |G\rangle = |G_{pq}\rangle$.
3. Allocate a zero-initialized array for the matrix $P(G_{pq})$ indexed by r and s. Each cell is associated with $\hat{\mathcal{O}} a_r^\dagger a_s^\dagger a_p a_q |G\rangle = \left| G_{pq}^{rs} \right\rangle$. Some cells will be tagged as not being associated with a unique $|\alpha\rangle$, but either one of :
   - a determinant already present in the wave function
   - an *exclusion principle violating* determinant (EPV), i.e. $\left| G_{pq}^{rs} \right\rangle = 0$

- a non-unique $|\alpha\rangle$ (either a double excitation of a previous generator, or a single excitation of the current one)

4. Since two electrons cannot occupy the same spinorbital, tag cells where r or s is occupied in $\left|G_{pq}\right\rangle$ as well as those with $r = s$.
5. Apply single excitation tagging. This ensures single excitations of $|G\rangle$ are generated exactly once. It is described in section 5.4.3.
6. **selector loop**: Iterate over $|S\rangle \in \left\{\left|D_{J\leq N_{\text{sel}}}\right\rangle\right\}$
7. Determine whether there is an (r,s) pair so that $|S\rangle = \left|G_{pq}^{rs}\right\rangle$. In other words, look for $|S\rangle$ in the current batch. If it is found, tag the corresponding cell, $\left|G_{pq}^{rs}\right\rangle \in \{|D_I\rangle\}$.
8. Determine (r,s) pairs so that $\left|G_{pq}^{rs}\right\rangle$ is connected to $|S\rangle$
9. If $J < I$, tag the corresponding cells ; $\left|G_{pq}^{rs}\right\rangle$ is generated by $\left|D_J\right\rangle$.
10. If $J \geq I$, increment all untagged $P_{rs}(G_{pq})$ matrix elements by $\Delta P_{rs}(G_{pq}) = c_J \left\langle S\middle|\hat{H}\middle|G_{pq}^{rs}\right\rangle$. Note that the excitation operator $\hat{T}$ so that $|S\rangle = \pm\hat{T}\left|G_{pq}^{rs}\right\rangle$, useful for computing the associated matrix element, can be determined at the same time as the (r,s) pair.
11. End **selector loop**. All untagged cells are guaranteed to be associated with a unique $|\alpha\rangle$ and $P_{rs}(G_{pq}) = \left\langle\Psi\middle|\hat{H}\middle|G_{pq}^{rs}\right\rangle$. e_α's for the current batch can be computed, with $|\alpha\rangle = \left|G_{pq}^{rs}\right\rangle$, as

$$e_\alpha = \frac{P_{rs}(G_{pq})^2}{E_{\text{var}} - \langle\alpha|\hat{H}|\alpha\rangle} \tag{5.11}$$

12. End of other loops. All e_α have been computed a single time.

5.4.2 Tagging

Tagged cells are simply tracked using a boolean matrix $B(G_{pq})$ with $B_{rs}(G_{pq})$ keeping the tag status of $\left|G_{pq}^{rs}\right\rangle$, defaulting to FALSE. In some cases, full columns/rows are to be tagged. Keeping track of fully tagged rows or columns is useful for performance purpose, as it allows to bypass some loop iterations. A simple way to do it, is to add an

extra column and an extra row of index 0 to B ; $B_{0s}(G_{pq}) = \text{TRUE}$ means the whole s column is tagged, $B_{r0}(G_{pq}) = \text{TRUE}$ means the whole r line is tagged. The actual tag status of $\left|G_{pq}^{rs}\right\rangle$ becomes

$$B_{r0}(G_{pq}) \vee B_{0s}(G_{pq}) \vee B_{rs}(G_{pq}). \tag{5.12}$$

While significant, this optimization is fairly simple to set up and use, so for simplification purpose, it will be ignored in the text.

5.4.3 Single excitation tagging

The algorithm is designed to generate all $\left|G_{pq}^{rs}\right\rangle$, which are doubly excited from $|G\rangle$. The singly excited determinants are not explicitly generated, but are formally present as $\left|G_{pq}^{ps}\right\rangle$. The issue is that $\left|G_{pq}^{ps}\right\rangle$ refers to the same determinant $\hat{\mathcal{O}}a_s^\dagger a_q |G\rangle$ regardless of p, and the base algorithm only tags $\left|G_{pq}^{rs}\right\rangle$ with $|G\rangle = |D_I\rangle$ as duplicate if it can be generated by $|K\rangle = \left|D_{J<I}\right\rangle$, i.e. if

$$\left|G_{pq}^{rs}\right\rangle = \left|K_{p'q'}^{r's'}\right\rangle. \tag{5.13}$$

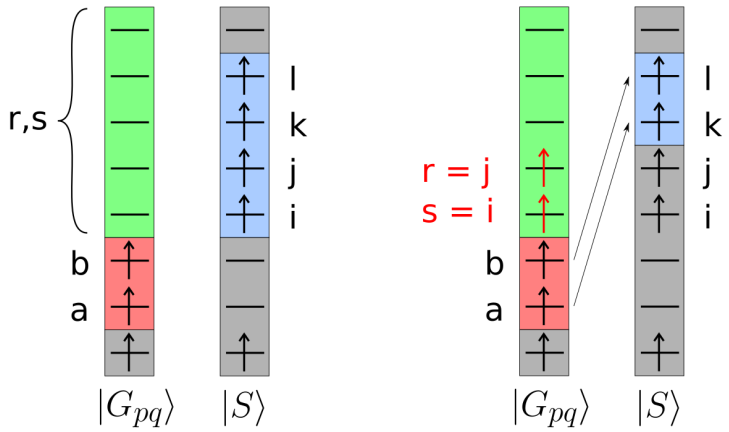
As can be seen this doesn't cover the case where $\left|G_{pq}^{ps}\right\rangle = \left|G_{p'q}^{p's}\right\rangle$.

To solve this issue, we default to tag $\left|G_{pq}^{ps}\right\rangle$, which prevents generating single excitations, and selectively untag in certain cases:

- **Untagging all ↑-spin single excitations of $|G\rangle$ exactly once**:
  Pick P any "non-frozen" ↓ spinorbital occupied in $|G\rangle$. We arbitrarily choose the lowest one. Untag $\left|G_{Pq}^{Ps}\right\rangle$ whenever q, s are of ↑ spin. Any ↑-spin single excitation $q \to s$ is untagged a single time.
  P cannot be chosen of ↑ spin, because single excitations $P \to s$ and $q \to P$ would be formally present as $\left|G_{PP}^{Ps}\right\rangle$ and $\left|G_{Pq}^{PP}\right\rangle$, which aren't ever generated, since for obvious reasons the base algorithm never considers the batch $\left|G_{qq}\right\rangle$ or the determinants $\left|G_{pq}^{rr}\right\rangle$.

- **Untagging all ↓-spin single excitations of $|G\rangle$ exactly once**:
  Pick Q any "non-frozen" ↑ spinorbital occupied in $|G\rangle$. Again we arbitrarily choose the lowest one. If p, q are of ↓ spin, untag $\left|G_{pQ}^{rQ}\right\rangle$. Any ↓-spin single excitation $p \to r$ is untagged a single time.

5.5 Systematic determination of connections



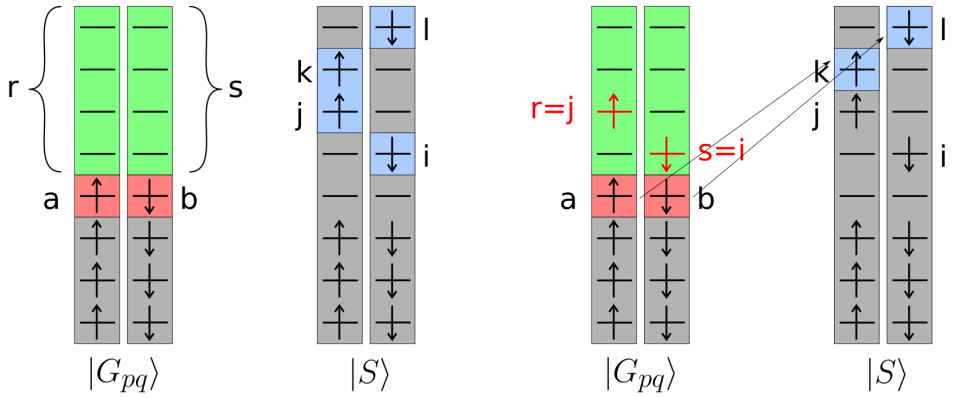
$$|S\rangle = \hat{T}^{kl}_{ab}|G^{ij}_{pq}\rangle$$
$$\Delta P_{ij}(G_{pq}) = c_S\langle G^{ij}_{pq}|\widehat{H}|S\rangle$$

Figure 5.4: Illustrative example of systematic determination of the connection between a selector $|S\rangle$ and determinants of the $|G_{pq}\rangle$ batch when p and q have the same spin. c_S is the coefficient of $|S\rangle$ in $|\Psi\rangle$.

The systematic determination of connections between $|S\rangle$ and determinants from the G_{pq} batch is done by comparing $|S\rangle$ to the doubly ionized determinant $|G_{pq}\rangle$. This yields a set of spinorbitals whose occupation status differ. Remembering $|S\rangle$ has two extra electrons compared to $|G_{pq}\rangle$, there are 4 cases of interest:

- i,j are occupied in $|S\rangle$ but not in $|G_{pq}\rangle$
- i,j,k are occupied in $|S\rangle$ but not in $|G_{pq}\rangle$; a is occupied in $|G_{pq}\rangle$, but not in $|S\rangle$
- i,j,k,l are occupied in $|S\rangle$ but not in $|G_{pq}\rangle$; a,b are occupied in $|G_{pq}\rangle$, but not in $|S\rangle$
- More differences : $|S\rangle$ isn't connected to any $|G^{rs}_{pq}\rangle$ and can be ignored.

Based on these indices, it is possible to immediately deduce any (r,s) pair so that $|G^{rs}_{pq}\rangle$ is at most a double excitation of $|S\rangle$, as well as the excitation operator $\hat{T}$ so that $|G^{rs}_{pq}\rangle = \hat{\mathcal{O}}\hat{T}\,|S\rangle$. Figures 5.4 and 5.5 show two possible cases as examples.



$$|S\rangle = \hat{T}^{kl}_{ab}|G^{ij}_{pq}\rangle$$
$$\Delta P_{ij}(G_{pq}) = c_S\langle G^{ij}_{pq}|\widehat{H}|S\rangle$$

Figure 5.5: Illustrative example of systematic determination of the connection between a selector $|S\rangle$ and determinants of the $|G_{pq}\rangle$ batch when p and q are of different spins. c_S is the coefficient of $|S\rangle$ in $|\Psi\rangle$.

While this could be done in a more compact way, we took a "case by case" approach, allowing more specialized code for each situation. Taking spin into account, the different cases are listed in table 5.1.

It is noticeable that, because of the "wildcard" indices X and Y :

- Cases of the form a, ijk cause full rows/columns of $P(G_{pq})$ to be tagged or incremented.
- Cases of the form ij cause the whole $P(G_{pq})$ matrix to be tagged or incremented. Obviously, tagging the whole matrix means stopping the computation for G_{pq}.

5.6 Filtering and loop breaking

A large amount of CPU time is wasted because every doubly ionized generator $|G_{pq}\rangle$ is compared to all internal determinants. In the vast majority of cases, it will show no connection can be made and the internal determinant will be ignored. Thus, it is interesting to filter internal determinants in the outermost loops (loop over generators, and loop over first ionization).

This can be done using the distance $f^B_A = f^A_B$, defined as the minimal number of operations — moving, annihilating or creating an electron — that must be done to go

```
/* For simplification purpose, a determinant |S⟩ is here
   represented by a single bitstring S of size 2N_orb where
   each bit is associated with a spinorbital                    */
Data: |Ψ⟩, i.e. {D_I} the set of internal determinants and their coefficients c_I
Data: N_gen, N_sel, N_det
Result: e_α ≠ 0 has been computed exactly once for any |α⟩ ∉ {D_I}
for g ← 1, N_gen do
    forall (p, q) ; a_p a_q |D_g⟩ ≠ 0 do
        |G_pq⟩ ← Ô a_p a_q |D_g⟩;
        /* B and P are indexed by spinrobitals                  */
        B a FALSE-initialized boolean matrix of size 2N_orb × 2N_orb ;
        P a zero-initialized real matrix size 2N_orb × 2N_orb ;
        Apply EPV and single excitations tagging (algorithm 15) ;
        for t ← 1, N_det do
            |S⟩ ← |D_t⟩ ;
            C ← S ∧ ¬G_pq ;
            if ||C|| = 2 then
                e ← LIST_FROM_BITSTRING(C);
                B_{e[0] e[1]} ← TRUE;
            end
            /* see table 5.1 for (r, s) pairs                   */
            if t < g then
                forall (r, s) ; ⟨S|Ĥ|G^rs_pq⟩ ≠ 0 do
                    B_rs ← TRUE ;
                end
            else if t ≤ N_sel then
                forall (r, s) ; ¬B_rs ∧ ⟨S|Ĥ|G^rs_pq⟩ ≠ 0 do
                    P_rs ← P_rs + c_t ⟨S|Ĥ|G^rs_pq⟩ ;
                end
            end
        end
        forall (r, s) ; ¬B_rs do
            |α⟩ = |G^rs_pq⟩ is a unique |α⟩ ;
            e_α = P_rs^2 / (E_var − ⟨α|Ĥ|α⟩) ;
        end
    end
end
```

Algorithm 14: Unfiltered CIPSI selection

Data: B, q, p and $\left|G_{pq}\right\rangle$ from outer scope.
Result: Updates B so as to tag EPVs, and determinants that are generated by a generator of lower index from a single excitation on $\left|G\right\rangle$

```
/* tag EPV                                                              */
forall r do
    B_rr ← TRUE ;
end
forall r ; (a_r |G_pq⟩ ≠ 0) ∨ (r = p) ∨ (r = q) do
    B_*r ← TRUE ;
    B_r* ← TRUE ;
end
/* tag duplicate single excitations                                     */
if (q is ↑) ∧ (p is the lowest "non-frozen" occupied ↓ spinorbital in |G⟩) then
    B_*p ← FALSE ;
    B_p* ← FALSE ;
end
if (p is ↓) ∧ (q is the lowest "non-frozen" occupied ↑ spinorbital in |G⟩) then
    B_*q ← FALSE ;
    B_q* ← FALSE ;
end
```

Algorithm 15: EPV and single excitations tagging

from a determinant $|A\rangle$ to a determinant $\pm|B\rangle$ (i.e. ignoring the phase factor) with respectively n_A and n_B electrons. Alternatively, it can be defined as the maximum between the number of annihilations and the number of creations required to go from $|A\rangle$ to $\pm|B\rangle$.

$$f_A^B = \frac{\left\|A_\uparrow \oplus B_\uparrow\right\| + \left\|A_\downarrow \oplus B_\downarrow\right\| + |n_A - n_B|}{2} \tag{5.14}$$

Considering $|S\rangle$ a selector determinant and $|X\rangle$ a generator determinant in a state of ionization from 0 to 2 (it essentially is a wildcard for $|G\rangle$, $a_p|G\rangle$ or $a_p a_q|G\rangle = \left|G_{pq}\right\rangle$).

- $f_X^\alpha + f_\alpha^S \geq f_X^S$
- $|\alpha\rangle$ can be generated from $|X\rangle$ iff $f_X^\alpha \leq 2$
- $|\alpha\rangle$ is connected to $|S\rangle$ iff $f_\alpha^S \leq 2$
- $0 \leq \left(f_Y^S - f_X^S\right) \leq 1$ with $|Y\rangle = a_p|X\rangle$

From the rules above, we can deduce that given any $|X\rangle$ and $|S\rangle$, there exists an $|\alpha\rangle$ generated from $|X\rangle$ so that $\langle\alpha|\hat{H}|S\rangle \neq 0$ only if $f_X^S \leq 4$. Based on this, a filtering mechanism can be set up, as shown on figure 5.6. The diagram is somewhat convoluted and deserves comments.

Internal determinants' path A triple loop is shown

1. over generators $|G\rangle$
2. over p a first ionization $\left|G_p\right\rangle$
3. over q a second ionization $\left|G_{pq}\right\rangle$, i.e. over batches.

In each one some filtering takes place. The internal determinants "flow" from the top $\{|D_I\rangle\}$ into intermediate lists, that are fully constructed before proceeding to the inner loop, as they will be the sources of determinants for that inner loop. A selector can only be duplicated at the node denoted by a black circle. Otherwise, it follows a single path, always going for the horizontal path if it satisfies the associated condition. If it doesn't satisfy the condition of a horizontal path, and there is no further vertical path, it is discarded.

"Drop" instructions *Drop* instructions are reached when, predictably, the current loop iteration will not yield any unique $|\alpha\rangle$. If a determinant reaches a *drop*, the current loop iteration ends immediately.

- *drop* G_{pq} is reached in the case where the whole $P(G_{pq})$ matrix is to be tagged, i.e. the possible values for (r, s) given by table 5.1 are two wildcards (X, Y and $X, \bar{Y}$). This corresponds to the case where $\left|G_{pq}\right\rangle$ has already been created from a previous generator $|K\rangle$, i.e. $\left|G_{pq}\right\rangle = \left|K_{p'q'}\right\rangle$, therefore for any pair (r, s) we have $\left|K^{rs}_{p'q'}\right\rangle = \left|G^{rs}_{pq}\right\rangle$.

- *drop* G_p, in the same fashion, is reached when $a_p\,|G\rangle = \left|G_p\right\rangle$ has already been created from a previous generator $|K\rangle$, i.e. $\left|G_p\right\rangle = \left|K_{p'}\right\rangle$. For any (q, r, s) triplet there will be $\left|K^{rs}_{p'q}\right\rangle = \left|G^{rs}_{pq}\right\rangle$, so no new $|\alpha\rangle$ will be created.

Paths and loops There are roughly a left and a right path. The reason for this, is that we want to reach *drop* instructions as fast as possible. Incidentally, in each loop, the implementation should prioritize operations that may cause a reach to *drop*.

1. The first loop discards some internal determinants and separates the others in two disjoint categories.

   - Right branch : determinants that may contribute to the $P(G_{pq})$ matrix or tag previously generated $|\alpha\rangle$. In other words, selectors that may connect to some $\left|G^{rs}_{pq}\right\rangle$.
   - Left branch : Determinants that aren't selectors, but are equal to some $\left|G^{rs}_{pq}\right\rangle$. Being non-selectors, those will not be checked for connection to any $\left|G^{rs}_{pq}\right\rangle$, but they still must be checked for equality in order to ensure $\left|G^{rs}_{pq}\right\rangle \notin \{|D_I\rangle\}$

   This step sets the complexity of the algorithm with respect to N_{det}. Naively, f_G^S must be computed for all pairs of internal determinants, setting the complexity to $\mathcal{O}({N_{\text{det}}}^2)$.

   Our current implementation quickly discards $f_G^S > 4$ by using a method similar to what we used in the Davidson diagonalization, adapted to seek excitation degrees ≤ 4 rather than ≤ 2. The key difference is that, for parallelism reasons, the research has to be done individually for each generator ; that is, we are not

Data: $|G\rangle$: a generator determinant.
Data: $N_{\text{det}}^{\uparrow}$: the number of unique $\uparrow$ spin parts present in $|\Psi\rangle$.
Data: $D^{\uparrow}$: the array of determinants present in $|\Psi\rangle$, sorted by $\uparrow$-major order (all determinants sharing the same $\uparrow$ part are next to each other).
Data: $A^{\uparrow}$: the arrays so that $A^{\uparrow}[n]$ is the index of the first occurence of the n^{th} unique $\uparrow$ spin part in $D^{\uparrow}$. For algorithmic convenience we set $A^{\uparrow}[N_{\text{det}}^{\uparrow}+1] = N_{\text{det}}+1$.
Data: $N_{\text{det}}^{\downarrow}$, $D^{\downarrow}$, $A^{\downarrow}$: the $\downarrow$ counterparts.

```
for a ← 1, N_det^↑ do
    e ← EXC_DEGREE(D^↑[A^↑(a)]_↑, G_↑) ;
    if e ≤ 2 then
        for b ← A^↑(a), A^↑(a+1) − 1 do
            if e + EXC_DEGREE(D^↑[b]_↓, G_↓) ≤ 4 then
                retain D^↑[b] ;
            end
        end
    end
end
for a ← 1, N_det^↓ do
    e ← EXC_DEGREE(D^↓[A^↓(a)]_↓, G_↓) ;
    if e ≤ 1 then
        for b ← A^↓(a), A^↓(a+1) − 1 do
            if e + EXC_DEGREE(D^↓[b]_↑, G_↑) ≤ 4 then
                retain D^↓[b] ;
            end
        end
    end
end
```

Algorithm 16: Filter internal determinants $|S\rangle$ so that $f_G^S \leq 4$

computing all f_G^S at the same time, but all f_G^S for a given $|G\rangle$ separately. The procedure is shown as algorithm 16. The complexity is reduced from $\mathcal{O}\left({N_{\text{det}}}^2\right)$ to $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$.

Note that the only point of separating those two categories rather than merging them in the same list, is to avoid additional *past* and *selector* tests in the second loop. This most likely is of little interest, depending on the implementation. But because it is the actual implementation and because it reduces the number of operations, it is still shown.

2. The second loop discards some internal determinants and separates the other in two categories, this time not disjoint.

   - Right branch : Selectors that may connect to some $\langle\Psi|\hat{H}|\alpha\rangle$.
   - Left branch : Determinants that may be equal to some $\left|G_{pq}^{rs}\right\rangle$. Those can be found in both lists built in the first loop.

   As previously discussed, if there is a previous generator $|K\rangle$ so that $a_{p'}|K\rangle = a_p|G\rangle$, it will result in $P(G_{pq})$ being fully tagged for any q, hence a need to reach *drop* G_p to avoid unnecessary computations. The reach for *drop* G_p can be put on the path between the right list of the first loop and the left list of the second loop.

   Indeed, $a_{p'}|K\rangle = a_p|G\rangle$ with $|K\rangle$ a previous generator translates to

$$\left(f_{G_p}^K = 1\right) \wedge \text{past} \tag{5.15}$$

   The right list of the first loop contains all internal determinants so that

$$\left(f_G^K \leq 4\right) \wedge \text{selector} \tag{5.16}$$

   However

$$f_{G_p}^K = 1 \implies f_G^K \leq 1 \implies f_G^K \leq 4 \tag{5.17}$$

$$\text{past} \implies \text{selector} \tag{5.18}$$

$$\left(f_{G_p}^K = 1\right) \wedge \text{past} \implies \left(f_G^K \leq 4\right) \wedge \text{selector} \tag{5.19}$$

   Therefore any internal determinant able to reach *drop* G_p will be present in that list. Trivially, from there it will always take the left path because $f_{G_p}^K = 1 \implies f_{G_p}^K \leq 2$.

3. Third loop :

   - Right branch : Final filtering to keep only selectors that do connect to some $\left|G_{pq}^{rs}\right\rangle$
   - Left branch : $f_{G_{pq}}^{S} = 2$ implies there exists (r, s) so that $|S\rangle = \left|G_{pq}^{rs}\right\rangle$. When one is found :
     - If past

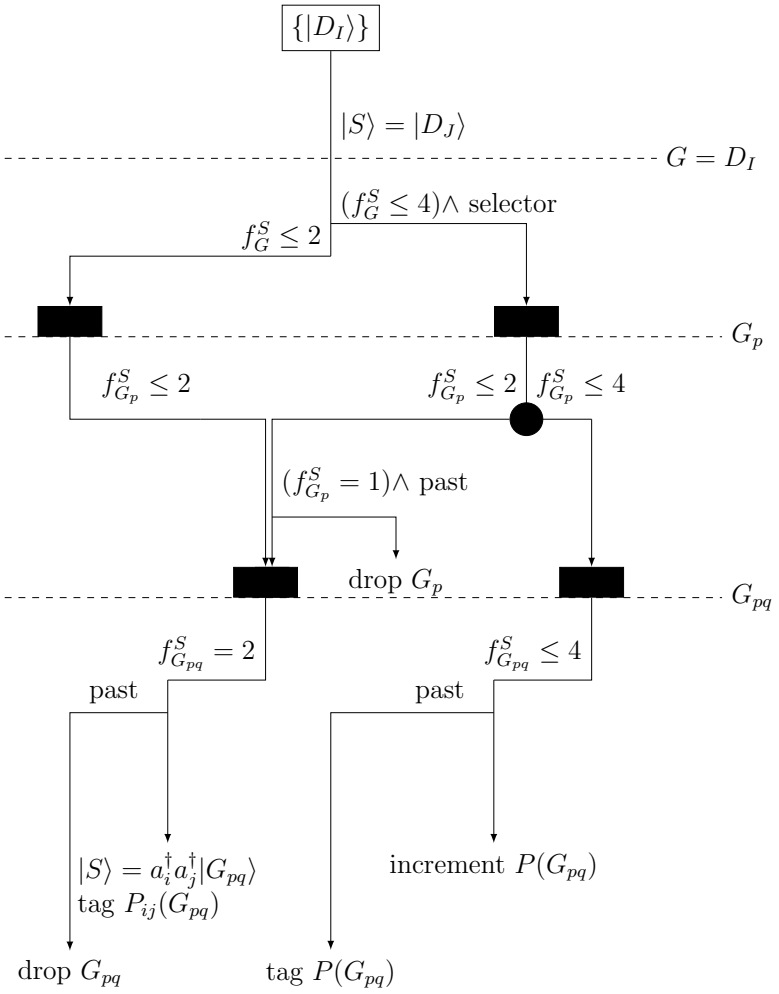
       $$|S\rangle = \left|G_{pq}^{rs}\right\rangle \implies |S_{rs}\rangle = \left|G_{pq}\right\rangle \tag{5.20}$$

       As explained above, it leads to $P(G_{pq})$ being fully tagged, and thus *drop* G_{pq} can be reached.
     - If ¬past, $\left|G_{pq}^{rs}\right\rangle$ must be tagged for referring to a determinant of the internal space.

Table 5.1: Systematic "case by case" determination of connections between a selector $|S\rangle$ and determinants of a batch G_{pq}

$\lvert S\rangle$	r, s	$\hat{T}$ such that $\hat{T}\lvert S\rangle = \pm\left\lvert G^{rs}_{pq}\right\rangle$
$a^\dagger_{ij}\lvert G_{pq}\rangle$	X, Y	$ij \to XY$
	X, i	$j \to X$
	i, j	$\hat{1}$
$a_a a^\dagger_{ijk}\lvert G_{pq}\rangle$	X, i	$aX \to jk$
	i, j	$a \to k$
$a_{\bar{a}} a^\dagger_{\bar{i}jk}\lvert G_{pq}\rangle$	X, j	$\bar{a}k \to \bar{i}X$
	j, k	$\bar{a} \to \bar{i}$
$a_{ab} a^\dagger_{ijkl}\lvert G_{pq}\rangle$	i, j	$ab \to kl$
$a_{a\bar{b}} a^\dagger_{ijk\bar{l}}\lvert G_{pq}\rangle$	i, j	$a\bar{b} \to k\bar{l}$
$a_{\bar{a}\bar{b}} a^\dagger_{ij\bar{k}\bar{l}}\lvert G_{pq}\rangle$	i, j	$\bar{a}\bar{b} \to \bar{k}\bar{l}$
$\lvert S\rangle$	$r, \bar{s}$	$\hat{T}; \hat{T}\lvert S\rangle = \left\lvert G^{r\bar{s}}_{p\bar{q}}\right\rangle$
$a^\dagger_{i\bar{j}}\lvert G_{p\bar{q}}\rangle$	$X, \bar{Y}$	$i\bar{j} \to X\bar{Y}$
	$i, \bar{X}$	$\bar{j} \to \bar{X}$
	$X, \bar{j}$	$i \to X$
	$i, \bar{j}$	$\hat{1}$
$a_a a^\dagger_{ij\bar{k}}\lvert G_{p\bar{q}}\rangle$	$X, \bar{k}$	$aX \to ij$
	$i, \bar{k}$	$a \to j$
	$i, \bar{X}$	$a\bar{k} \to j\bar{X}$
$a_{ab} a^\dagger_{ijk\bar{l}}\lvert G_{p\bar{q}}\rangle$	$i, \bar{l}$	$ab \to jk$
$a_{a\bar{b}} a^\dagger_{ijk\bar{l}}\lvert G_{p\bar{q}}\rangle$	i, j	$a\bar{b} \to k\bar{l}$
$a_{a\bar{b}} a^\dagger_{ij\bar{k}\bar{l}}\lvert G_{p\bar{q}}\rangle$	$i, \bar{k}$	$a\bar{b} \to j\bar{l}$

- the bar notation $\bar{a}$ is used to indicate relative spins
- $a_{ij\ldots}$ is a compact notation for $a_i a_j \ldots$
- X and Y are "wildcard" indices referring to any spinorbital unoccupied in both $|S\rangle$ and $\left|G_{pq}\right\rangle$



S : determinant in $\{|D_I\rangle\}$
G : generator
G_p : singly ionized generator
G_{pq} : doubly ionized generator
--- X : loop over all possible X
$\xrightarrow{c}$: path taken if c is true
drop X : cycle loop over X
past : $J < I$, so S has already been treated as a generator
selector : S is a selector
f_X^S : Excitation degree between X and S
■ : Filtered list of determinants to be used in the inner loop
● : Possible duplication of S

Figure 5.6: $|S\rangle$ is the internal determinant currently flowing down the chart. Tagging is fully computed, and *drop* instructions eventually reached before any update is done to $P(G_{pq})$.

5.7 Parallel computation

Arguably the simplest way to make an algorithm parallel is, whenever possible, to create independent tasks corresponding to one iteration of the outermost loop. As figure 5.6 suggests, iterations for the outermost loop — over generators — are independent. This is due to our choice to perform the initial filtering on a "generator by generator" basis (algorithm 16). The cost for this initial filtering could be reduced with a "spin part by spin part" basis as in our Davidson algorithm, but since the CIPSI selection is more expensive than the Davidson diagonalization, the filtering steps only account for a few percent of the total CPU time, so for simplicity and load balancing we stuck to 1 task = 1 generator. Even so, the cost for different tasks is still very much unbalanced, the first few generators with large coefficients being very expensive, and the cost quickly decreasing.

For a better load balancing, we split the first, expensive tasks into smaller *fragments*, using a fairly simple approach. Essentially, some tasks will require computing just a subset of the G_{pq} batches associated with a generator $|G\rangle$, as opposed to all of them. This implies some overhead, since some filtering steps will be duplicated. Fortunately, only a relatively small number of expensive tasks need to be split.

Each generator $|D_i\rangle$ defines as a "logical" task i, and for each one we define F_i a fragmentation level, defining the number of independent "actual" tasks it should be decomposed in. We have empirically chosen the following expression for the fragmentation of the tasks:

$$F_{\max} = 1 + \min\left(N_{\text{elec}}(N_{\text{elec}} - 1)/2, \lfloor\sqrt{N_{\text{sel}}}\rfloor/10\right) \tag{5.21}$$

$$F_i = 1 + F_{\max}\left(\max_{k=1}^{N_{\text{states}}} |c_{ik}|^{1/2}\right) \tag{5.22}$$

where F_i is an estimated cost of task i. In practice $F_i = 1$ for the majority.

Task i is put in the task queue F_i times, each time associated with a 'fragment" index s ranging from 0 to $F_i - 1$, which together with F_i defines the subset of batches this task corresponds to.

This fragmentation scheme can be shown in a simpler and more general way when used to compute E_{PT2} in chapter 6. In a nutshell, we can see it as a toy problem where we want to print for each generator the sum of e_α over all unique $|\alpha\rangle$ it has generated. The algorithms for this on the master side and slave side are shown as algorithms 17 and 18 respectively.

For the determinant selection, we still use the mixed MPI/OpenMP paradigm, where one MPI process per node is used for efficient broadcasting of the replicated data. But, as opposed the Davidson implementation where each task was parallelized with OpenMP, here each OpenMP thread handles independently a task, computed with a single core. The first reason is that the number of tasks is larger than N_{det}, which

```
1  /* A logical task is the computation of e[i], the sum of e_α over
       all unique |α⟩ generated from generator |D_i⟩                    */
2  choose F_i for all i;
3  for i ← 1, N_gen do
4  |  for s ← 0, F_i − 1 do
5  |  |  add task (i, s) to the queue
6  |  end
7  end
8  f,e are arrays size N_gen initialized with 0 ;
9  while not all e[i] printed do
10 |  get (i, sum) from a slave ;
11 |  e[i] ← e[i] + sum ;
12 |  f[i] ← f[i] + 1 ;
13 |  if f[i] = F_i then
14 |  |  print e[i] ;
15 |  end
16 end
```

Algorithm 17: Task splitting, pseudocode for master.

```
1  while do
2  |  get task (i, s ∈ [0, F_i − 1]) from the queue ;
3  |  E ← 0 ;
4  |  c ← 0 ;
5  |  G ← D_i ;
6  |  /* duplicated computation if F_i > 1                             */
7  |  filtering for G (see figure 5.6) ;
8  |  foreach G_p do
9  |  |  /* duplicated computation if F_i > 1                         */
10 |  |  filtering for G_p (see figure 5.6) ;
11 |  |  foreach G_pq do
12 |  |  |  c ← c + 1 ;
13 |  |  |  if s = c  mod F_i then
14 |  |  |  |  increment E with all unique e_α in this batch ;
15 |  |  |  end
16 |  |  end
17 |  end
18 |  send (i, E) to master ;
19 end
```

Algorithm 18: Task splitting, pseudocode for slave.

is usually orders of magnitude larger than the number of CPU cores. Moreover, the computation of a task in parallel would require a synchronization barrier at the beginning and the end of the OpenMP section. Here, all the OpenMP threads are completely independent during the whole calculation of the selection, and this explains the very good scaling properties of the implementation, as shown in chapter 9.

5.8 Obtaining spin-pure states

The presented algorithm generates a wave function which is expressed on a truncated space of Slater determinants. Determinants are not necessarily eigenfunctions of the $\widehat{S}^2$ operator, so the eigenfunctions of the truncated Hamiltonian are not guaranteed to be also eigenfunctions of $\widehat{S}^2$.

A conventional solution to avoid this issue is to work in the basis of *Configuration State Functions* (CSFs). These are linear combinations of determinants which are eigenfunctions of $\widehat{S}^2$, and with the desired eigenvalue. Diagonalizing $\widehat{H}$ in this basis ensures that the solution is spin pure.

Working with CSFs instead of determinants has the additional advantage that the space of CSFs is smaller than the space of determinants. CSFs could in principle be used for CIPSI, but that makes the computation of the matrix elements of $\widehat{H}$ less straightforward.

We have chosen a more practical solution.[31] After the selection step, we add to the variational space all the determinants that are necessary to obtain a spin pure solution. These determinants correspond to all the possible spin flips in open shell determinants with the constraint that $N^{\uparrow}_{\text{elec}}$ and $N^{\downarrow}_{\text{elec}}$ are constant. The diagonalization of $\widehat{H}$ will automatically yield spin pure eigenfunctions at the cost of increasing the size of the wave function with many determinants with very small weights.

5.9 Conclusion

The novel implementation of the CIPSI algorithm runs orders of magnitude faster than the former one thanks to several key improvements.

- Excitations do not need to be computed explicitly using algorithms such as those presented in chapter 3. However the associated phase factors still have to be computed.
- Although not directly related to the CIPSI algorithm, the computation of the phase factors has been made much cheaper thanks to the use of phase masks (see section 3.5.3).

- Selector determinants are simultaneously compared to $\left(N_{\text{orb}} - N_{\text{elec}}^{\uparrow}\right)^2$ external determinants $|\alpha\rangle$ thanks to the batch approach.
- Selector determinants go through a filtering mechanism that narrows down the number of selector determinants to be considered against a particular $|\alpha\rangle$ (a particular batch of $|\alpha\rangle$).

This made way for applications that were not affordable before, such as those presented in chapter 10, as well as an additional application to copper complexes.[57] The algorithms presented in this chapter also lead to the methods presented in the next chapters.

Chapter 6

Computation of the second-order perturbative correction

Contents

6.1 Introduction

In the literature, the computation of E_{PT2} is part of the CIPSI method. This is understandable, as the selection and the computation of E_{PT2} involve gathering the same data.

$$E_{PT2} = \sum_{\alpha} \frac{\langle \Psi | \hat{H} | \alpha \rangle^2}{\Delta E_\alpha} \tag{6.1}$$

It is essentially the sum all e_α that are computed during a CIPSI selection.

$$E_{PT2} = \sum_{\alpha} e_\alpha \tag{6.2}$$

We have seen in section 5.2 that approximate calculations could be done to accelerate the selection. However, these approximations don't apply to the computation of E_{PT2}, so we designed a hybrid stochastic-deterministic scheme to get an accurate estimation of E_{PT2} for a much more reasonable cost.

The selection of determinants and the calculation of E_{PT2}, all approximations aside, both imply the computation of e_α for all $|\alpha\rangle$, so both can be computed at the same time. The selection is about identifying the set of the most important contributions, E_{PT2} is about computing the sum over all of them. There are two main consequences to this:

E_{PT2} is one iteration behind the selection At iteration n, identifying the most significant $|\alpha\rangle$ is about building $\left|\Psi^{(n+1)}\right\rangle$ while summing the contributions is about estimating the distance to full-CI for $\left|\Psi^{(n)}\right\rangle$. Computing E_{PT2} for a "final" wave function therefore requires an extra iteration, which applies to a larger number of determinants and thus is more expensive. Note that if only E_{PT2} is required, a more efficient algorithm can be used.[58]

The selection can take more approximations The computation of the sum has to be more costly than just identifying the largest terms. As was said in chapter 5, the CIPSI algorithm can take pretty drastic approximations for the selection.[14]

- N_{gen} allows to explore a reduced subset of $|\alpha\rangle$ in which we are almost sure to find those of interest.

- N_{sel} allows for a less accurate and less expensive computation of e_α, which is unlikely to significantly change the identified set.

These approximations do not apply when computing E_{PT2}. The very large number of smaller contributions makes them impossible to neglect without introducing a bias, and increasing the n_g and n_s thresholds dramatically increases the computational cost.

Unfortunately, a truncated computation of E_{PT2} always yields in a biased result. Since e_α is the contribution to the correlation energy brought by $|\alpha\rangle$, it is necessarily negative. Hence, E_{PT2} is a sum of same-sign contributions, and when the sum of e_α is truncated some correlation energy is missing.

Before the stochastic computation of E_{PT2} was implemented, our best choice was to set low values of n_g and n_s while performing the selection, and accept very approximate values for E_{PT2} for intermediate wave functions. Then, once the selection was completed, we would raise them just for a final, very expensive "E_{PT2} only" iteration. In fact, an exact computation with $N_{\text{gen}} = N_{\text{sel}} = N_{\text{det}}$ was often prohibitively long, so the final E_{PT2} was still biased, and the biases were not well controlled. The effect was particularly important when computing atomization energies, where the E_{PT2} values were much more approximate on the molecule than on the atoms. A practical way to circumvent this problem was already proposed 20 years ago, by giving an extrapolation of E_{PT2} when n_g goes to one.[59] However, the algorithm we propose here has the advantage of giving an unbiased result within a statistical confidence interval.

6.2 Stochastic estimation of E_{PT2}

We eventually solved the previously discussed problem by turning the bias into an error bar. The basic idea is that, instead of trying to get the largest possible chunk of contribution, we can randomly pick e_α contributions and make a Monte-Carlo estimate for the sum over all $|\alpha\rangle$. In this case, to avoid any bias we must set

$$N_{\text{gen}} = N_{\text{sel}} = {N_{\text{det}}}^\star \tag{6.3}$$

with ${N_{\text{det}}}^\star$ the number of internal determinants with non-zero coefficient. Not only the estimate will be unbiased and much closer to the actual E_{PT2}, but we will have an estimate for the error. Because E_{PT2} is itself used as an approximation

$$E_{\text{var}} + E_{\text{PT2}} \simeq E_{\text{FCI}}, \tag{6.4}$$

an error significantly smaller than the typical accuracy of $E_{\text{var}} + E_{\text{PT2}}$ *vs* E_{FCI} is certainly acceptable. Drawing randomly external determinants would probably not be efficient enough to improve significantly the computational time, so the algorithm we have designed is more convoluted.

6.2.1 Monte-Carlo sampling

We generally want to compute a quantity F which may be expressed as the expected value of a function $f(x)$ with respect to a probability distribution function $p(x)$:

$$F = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x \tag{6.5}$$

with

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1. \tag{6.6}$$

When X_i are samples randomly distributed according to p,

$$F = \langle f \rangle_p = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f(X_i) \tag{6.7}$$

So if one is able to draw M samples X_i with probability $p(X_i)$, F may be approximated as

$$\bar{F} = \frac{1}{M} \sum_{i=1}^{M} f(X_i) \tag{6.8}$$

The *Central Limit Theorem* states that when independent random variables are added together, their normalized sum tends to a normal distribution. The variance of this normal distribution,

$$\sigma^2(F) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} \left(f(X_i) - F\right)^2, \tag{6.9}$$

reflects the dispersion of the X_i. For a finite number of samples, the variance can be estimated as

$$\bar{\sigma}^2(F) = \frac{1}{M-1} \sum_{i=1}^{M} \left(f(X_i) - \bar{F}\right)^2, \tag{6.10}$$

and the 68.2% confidence interval (statistical error) is $\bar{F} \pm \sqrt{\frac{\bar{\sigma}^2(F)}{M}}$.

The simplest way to compute E_{PT2} with a Monte Carlo algorithm is to express E_{PT2} as

$$E_{\mathrm{PT2}} = \sum_{\alpha=1}^{N_\alpha} e_\alpha = \frac{1}{N_\alpha} \sum_{\alpha=1}^{N_\alpha} N_\alpha e_\alpha = \langle N_\alpha\, e_\alpha \rangle = \sum_{\alpha=1}^{N_\alpha} p(\alpha) N_\alpha e_\alpha \tag{6.11}$$

with $p(\alpha) = 1/N_\alpha$. This corresponds to a uniform sampling of the $|\alpha\rangle$ determinants.

The largest the variance of e_α, the slowest the convergence. Changing the sampling can reduce the variance, and one can make the sampling optimal by choosing the probability density

$$p(\alpha) = \frac{e_\alpha}{\mathcal{N}} \tag{6.12}$$

and computing

$$E_{\mathrm{PT2}} = \sum_{\alpha=1}^{N_\alpha} p(\alpha) \times \mathcal{N} \tag{6.13}$$

Unfortunately in this case, the normalization constant $\mathcal{N} = \sum_\alpha e_\alpha = E_{\mathrm{PT2}}$, and this would require to know already E_{PT2} before doing the calculation.

Choosing a probability density which can be computed very fast and which approximates $e_\alpha / E_{\mathrm{PT2}}$ is expected to improve the convergence. If one takes the expression of

$$E_{\mathrm{PT2}} = \sum_\alpha e_\alpha = \sum_I \sum_J c_I c_J \left(\sum_\alpha \frac{\langle D_I | \hat{H} | \alpha \rangle \ \langle \alpha | \hat{H} | D_J \rangle}{\Delta E_\alpha} \right), \tag{6.14}$$

one can remark that

1. As the determinants were selected with a CIPSI criterion, all the e_α are expected to be small. In this regime, all the ΔE_α are expected to be large, and large enough to consider that $1/\Delta E_\alpha$ is almost constant.

2. when $I = J$, each contribution to e_α has a negative sign due to the denominator. But when $I \neq J$, the sum is a sum of terms with alternating signs, almost cancelling each other. Hence, the dominant term of the sum is the diagonal term

$$\sum_I c_I^2 \left(\sum_\alpha \frac{\langle D_I | \hat{H} | \alpha \rangle^2}{\Delta E_\alpha} \right). \tag{6.15}$$

3. $\sum_\alpha \langle D_I | \hat{H} | \alpha \rangle^2 / \Delta E_\alpha$ can be seen as the correlation energy of determinant $|D_I\rangle$, and this quantity is expected to be of the same order of magnitude among all the determinants.

Therefore, we propose to pack together contributions of α such that the random variable becomes a quantity e_I indexed by I instead of α

$$e_I = \sum_{\alpha \in \mathcal{A}_I} e_\alpha \tag{6.16}$$

and use as a probability distribution

$$p(e_I) = c_I^2. \tag{6.17}$$

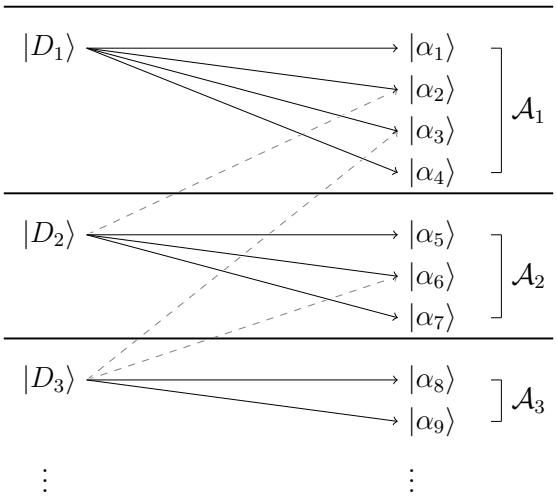


Figure 6.1: Construction of batches of $|\alpha\rangle$: disjoint sets related to the generator determinant.

6.2.2 Packing of e_α into elementary contributions e_I

Individual e_α are expensive to compute. In the CIPSI algorithm, each generator determinant creates a number of unique $|\alpha\rangle$, and computes e_α for each one of them. Essentially, the set of $|\alpha\rangle$ is split in N_{gen} disjoint sets, each associated with a generator determinant, as shown in figure 6.1.

$$\mathcal{A}_I = \left\{ e_\alpha \; ; \; \langle D_I|\hat{H}|\alpha\rangle \neq 0 \; ; \; \forall J < I, \langle D_J|\hat{H}|\alpha\rangle = 0 \right\}. \tag{6.18}$$

Because of the numerous tricks described in chapter 5, we are able to compute all the e_α of a set considerably faster than if we had to compute each contribution separately. Fortunately, this partition of $\{|\alpha\rangle\}$ fulfills the requirement of Eq. (6.16) and a large part of the implementation of the selection will be shared for the computation of the e_I to make the stochastic computation of E_{PT2} efficient.

To draw samples distributed with $p(e_I)$, we use the *inverse transform sampling method*.[60] The representation we are going to use is a collection of N_{det} boxes of width w_I, containing the determinant of index I. The determinant index I associated with drawing a random number $u \in [0,1)_{\mathbb{R}}$ is noted $w[u]$. It is determined using the cumulative probability distribution function W

$$W_I = \sum_{J \leq I} w_J \tag{6.19}$$

$$w(u) = I \; ; \; W_{I-1} \leq u < W_I. \tag{6.20}$$

To sample with $p(e_I)$, we set $w_I = p(e_I)$.

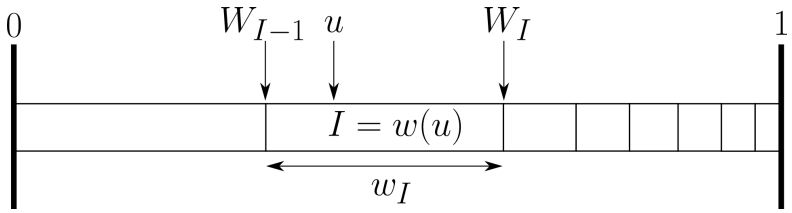


Figure 6.2: Schematic representation of Values inside the boxes are generator indices. Values outside are probabilities and drawn random numbers. Drawing the random number u using the probability density w yields generator index $I = w(u)$.

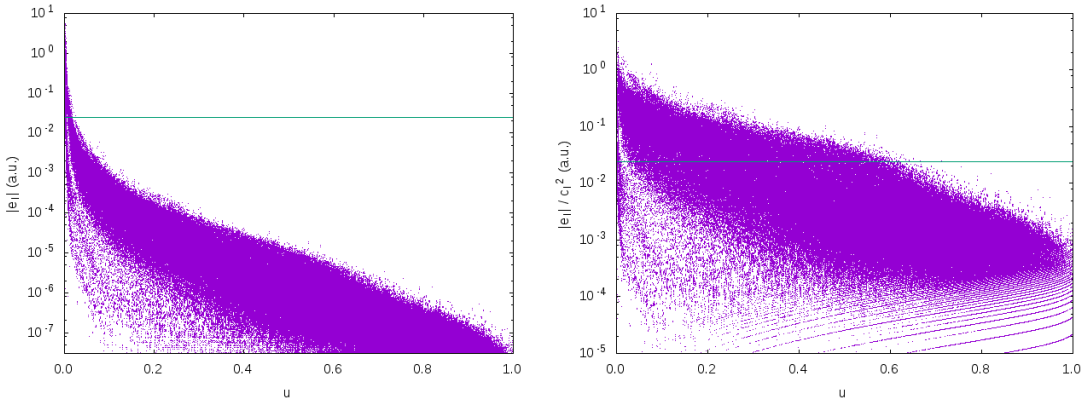


Figure 6.3: Contribution $|e_I|$ (left) and $|e_I|/c_I^2$ (right) as a function of the uniform random number u drawn. The green horizontal line is E_{PT2}.

Figure 6.3 confirms that using $p(e_I) = c_I^2$ as a probability density gives less fluctuations in the samples than when using a uniform sampling. The decreasing aspect of the curve comes from the elimination of the duplicate $|\alpha\rangle$, which makes the e_I smaller and smaller with respect to the correlation energy of $|D_I\rangle$.

To simplify the implementation, the actual distribution w_I we use isn't exactly c_I^2. This will be detailed in section 6.4.2.

6.2.3 Memoization of e_I's

Each contribution is now associated with a generator determinant. As a consequence, there are only N_{gen} elementary contributions to compute, but the computational cost of each contribution is increased. The number of contributions is small enough to make all the e_I's fit in memory, so when a e_I contribution is computed, its value is stored and simply reused if the same generator is drawn again. This optimization technique is known as *memoization*,[61] and can make exponentially-scaling algorithms become

polynomial.[62]

In our case, this optimization also leads to a drastic improvement in the computational time: it takes an infinite time for a Monte Carlo calculation to reach a zero statistical error, but using the memoization technique the exact result will eventually be known in finite time, once every contribution has been computed. The cost for the exact computation will essentially be the same as that of the purely deterministic full computation, with a negligible additional cost due to the Monte-Carlo related computations (drawing random numbers, finding the associated generators…).

6.3 Deterministic and stochastic ranges

Because e_I decreases rapidly, most of the contribution is contained in the first few e_I. We can compute the exact energy contribution for the first e_I, and only make a stochastic estimation for the sum over the smaller ones. This effectively splits the space of generators in two ranges, a deterministic one $\mathcal{D}_D$, then a stochastic one $\mathcal{D}_S$ (hence the hybrid characteristic of this method). Being ranges, they are always made of contiguous generators. The estimated energy can be written as

$$E_{\text{PT2}} = E_D + E_S \tag{6.21}$$

with E_D the exact energy for $\mathcal{D}_D$, and E_S the estimated energy for $\mathcal{D}_S$. The error bar only applies to E_S, which is typically much smaller than E_D. The number of generators in $\mathcal{D}_D$ increases during the computation. Initially, $\mathcal{D}_D = \mathcal{T}_0$ the initial deterministic range, for which the total weight is u_0.

$$\sum_{I \in \mathcal{T}_0} w_I = u_0 \tag{6.22}$$

$\mathcal{T}_0$ is always computed before any stochastic estimation can take place.

6.3.1 Partition of the stochastic range in N_{teeth} *teeth*

Generator determinants are sorted with decreasing values of c_I^2. As can be seen in figure 6.3, the values of e_I span many orders of magnitude and decrease rapidly with I, in an exponential-like way. Smoothed values for e_I are shown in figure 6.4. There are a few reasons for that.

- The values for the denominator ΔE_α used in the computation of e_α tend to increase, as internal determinants tend to be more and more excited and to populate higher orbitals

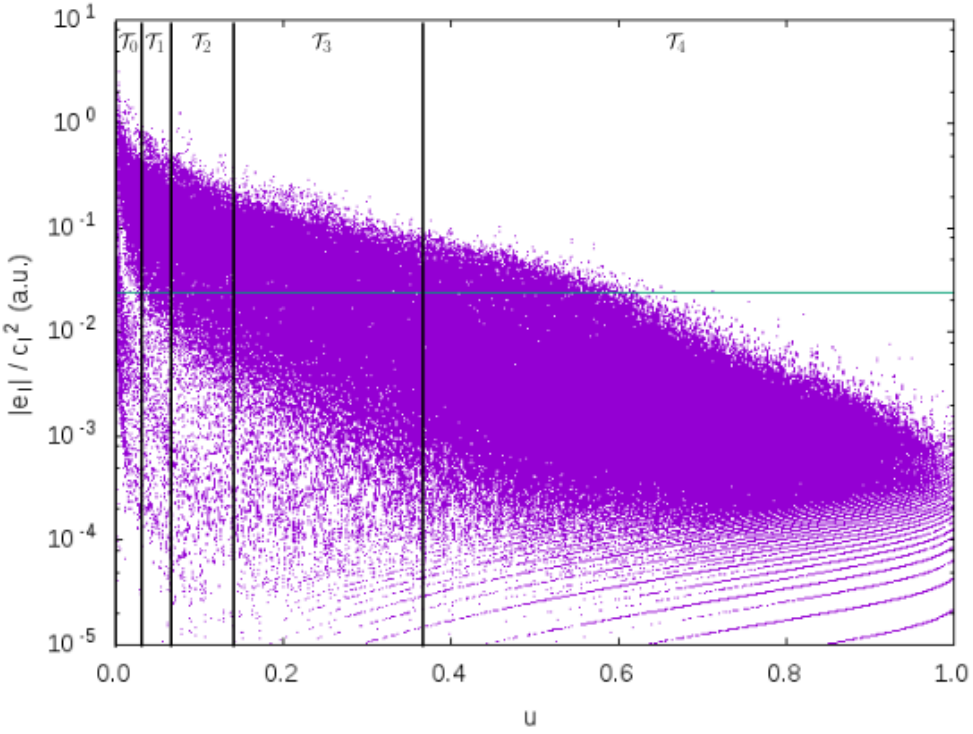


Figure 6.4: Contribution e_I associated with each generator, where the generators are sorted in decreasing order of c_I^2. The determinant space is divided into equal probability ranges, according do c_I^2.

- The number of unique $|\alpha\rangle$ per generator decreases. Indeed, the higher I, the likelier it is that the external determinants were were generated by other generators in $\{|D_{J<I}\rangle\}$.
- Unique $|\alpha\rangle$ are, by construction, disconnected from all previous generators, which mean they connect to a set of selectors with a decreasing norm.

Because of its original nature, this algorithm casts some ambiguity on what should be referred to as a *sample*. We are going to estimate a sum of elementary contributions e_I, compute and store them individually, and draw them based on a probability distribution function ; therefore they will be referred to as the *samples* and shown as such in the previously introduced representation. But the actual sample values are sums over several e_I, referred to as *combs*.

In a comb, the space of generators, minus the initial deterministic range $\mathcal{T}_0$, is split into ranges of equal probability (see ranges $\mathcal{T}_1, \dots \mathcal{T}_4$ in figure 6.4), and one sample is drawn in each range. Then, all those e_I are added together to make the contribution of the comb. Because the function e_I decreases smoothly, the contributions of the

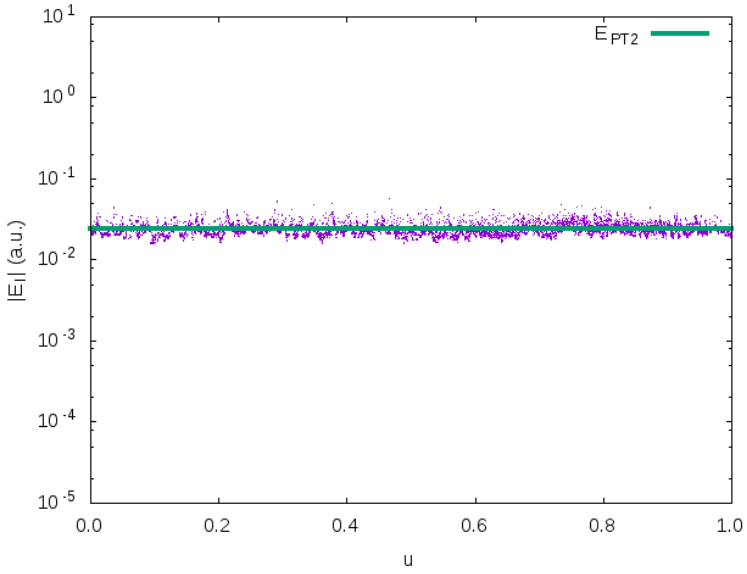


Figure 6.5: Comb contributions as a function of the uniform random number u drawn. The green horizontal line is E_{PT2}.

combs fluctuate less than the individual contributions, and the variance can be further reduced, as shown in figure 6.5.

As seen in section 6.3.1, because $\frac{e_I}{w_I}$ overall decreases, a range has a lower variance than the whole domain. The stochastic range is split in N_{teeth} ranges referred to as *teeth*, noted $\mathcal{T}_1, \ldots, \mathcal{T}_{N_{\text{teeth}}}$ (called $\mathcal{D}_t$ in the presented article) and sharing the same total weight W_T

$$\sum_{I \in \mathcal{T}_{t \geq 1}} w_I = W_T. \tag{6.23}$$

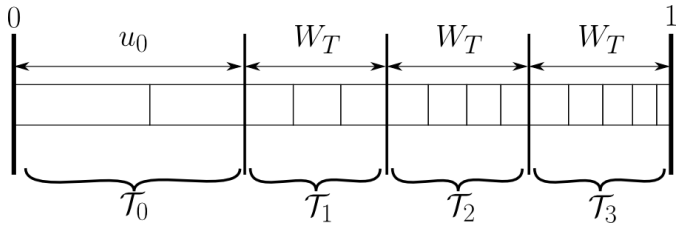


Figure 6.6: Partitioning of the generator space. The generator space is partitioned in so-called *teeth* from $\mathcal{T}_0$ to $\mathcal{T}_{N_{\text{teeth}}}$, $\mathcal{T}_0$ of total weight u_0, the others of total weight W_T

The partition of the generator space is as shown in figure 6.6. Intuitively, the sum of "one large, one medium and one small" has a lower variance than the sum of "three

at random". Instead of drawing individual indices, we are going to draw "combs" of indices, which are correlated sets of 1 index from each tooth. The associated sample value is the sum of $\frac{e_I}{w_I}$ over those indices that are in $\mathcal{D}_S$. The expression for this sample value is given below in Eq. (6.24).

6.3.2 Fully-computed teeth are moved to the deterministic range

Remembering we store e_I, given the first tooth $\mathcal{T}_t$ that contains an unknown e_I, the deterministic range extends to all $\mathcal{T}_{p<t}$. This makes $\mathcal{T}_t$ the first non-deterministic tooth. The number or generators in the deterministic range, therefore, is function of a tooth index t, and noted $n_0(t)$, as seen in figure 6.7

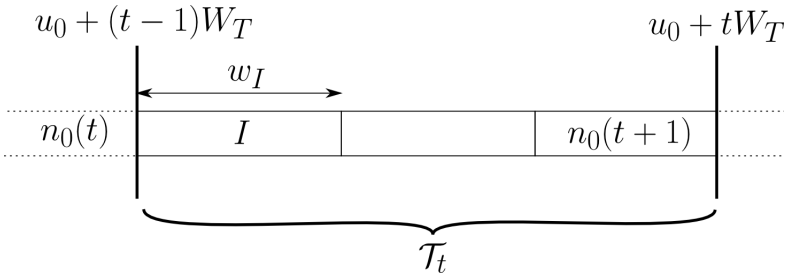


Figure 6.7: Boundaries of a tooth index-wise and probability-wise

Comb values We can write the expression of $B_t(u)$ the sample value associated with a comb. It is function of a random number $u \in [0,1)_{\mathbb{R}}$ and t the index of the first non-deterministic tooth.

$$B_t(u) = W_T \sum_{i=t-1}^{N_{\text{teeth}}-1} \frac{e_{y(u+i)}}{w_{y(u+i)}} \tag{6.24}$$

$$y(x) = w[u_0 + W_T \times x, W] \tag{6.25}$$

An illustrative example is given as figure 6.8.

Estimation of E_{PT2} After drawing n random numbers forming the set $\{\mathcal{U}\}$, and given $\mathcal{T}_t$ the first non-deterministic tooth, E_{PT2} can be estimated as

$$E_{\text{PT2}} = \sum_{I=1}^{n_0(t)} e_I + \frac{S_t}{n} \pm \text{err} \tag{6.26}$$

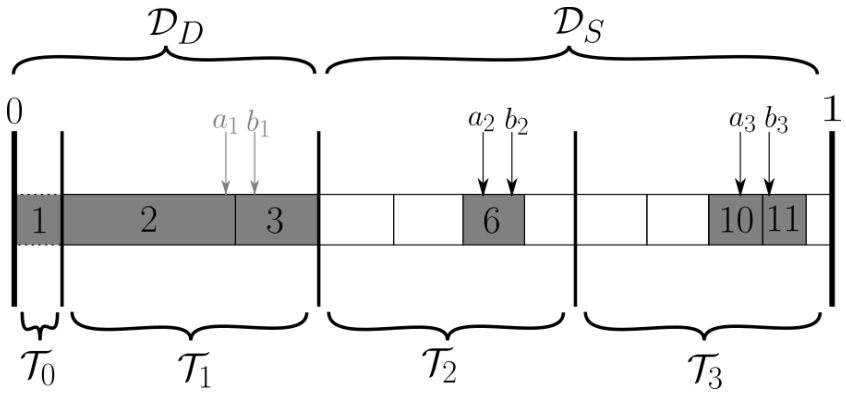


Figure 6.8: Illustrative example of drawing $n = 2$ combs a and b. Contributions that have been computed are greyed. $\mathcal{T}_1$ has been fully computed and is thus moved to $\mathcal{D}_D$. The first non-deterministic/not fully-computed tooth is $\mathcal{T}_{t=2}$. $E_D = e_1 + e_2 + e_3$, $B_t(a) = W_T\left(\frac{e_6}{w_6} + \frac{e_{10}}{w_{10}}\right)$; $B_t(b) = W_T\left(\frac{e_6}{w_6} + \frac{e_{11}}{w_{11}}\right)$, $E_S = \frac{B_t(a)+B_t(b)}{n}$

where the error is given by

$$\text{err} = \sqrt{\frac{S_t^{(2)} - {S_t}^2}{n-1}} \tag{6.27}$$

$$S_t = \sum_{u \in \{\mathcal{U}\}} B_t(u) \qquad S_t^{(2)} = \sum_{u \in \{\mathcal{U}\}} B_t(u)^2 \tag{6.28}$$

6.4 Technical considerations

6.4.1 Point of the initial deterministic part

The initial deterministic range $\mathcal{T}_0$ is a technical constraint. Because generators are sorted with decreasing $p(e_I)$, it is guaranteed that $|\mathcal{T}_{t>t'}| \geq |\mathcal{T}_{t'\neq 0}| - 1$ with $|\mathcal{T}_t|$ the cardinality of $\mathcal{T}_t$. For each comb, we draw a sample in each tooth, and when a tooth is entirely computed, it is moved to $\mathcal{D}_D$. Thus, it is immediately obvious that a tooth containing a single generator makes no sense, as it will be instantly moved to $\mathcal{D}_D$; it can as well be considered part of it. Because it is common practice to require at least 30 samples for the distribution of the average to become close enough to a Gaussian distribution,[63] we can go further and consider that a tooth with fewer than 5-10 generators will be moved to $\mathcal{D}_D$ too fast to be of real interest. Because the first c_I^2 are usually disproportionately large, it is not possible to fit that many in a tooth. Therefore they are immediately considered part of $\mathcal{D}_D$.

6.4.2 Desired *vs* effective distribution function

We have defined all teeth (except the special $\mathcal{T}_0$) as sharing the same total weight W_T. This is a constraint on our desired distribution function $p(e_I) = c_I^2$ that will lead to the effective distribution function w_I.

We are sampling comb values, but we have defined $p(e_I)$ a distribution function for e_I. Since we impose that the same number of e_I are drawn in each tooth — one per comb — the relative weight of teeth becomes irrelevant, and we effectively give all teeth the same weight

$$W_T = \frac{1 - u_0}{N_{\text{teeth}}} \tag{6.29}$$

in the Monte-Carlo scheme. Therefore, the effective weight given to $I \in \mathcal{T}_t$ is

$$w_{I \in \mathcal{T}_t} = W_T \times \frac{p(e_I)}{\sum_{J \in \mathcal{T}_t} p(e_J)} \tag{6.30}$$

To leave the distribution function unaltered, we need all teeth to actually weigh W_T

$$\sum_{J \in \mathcal{T}_t} p(e_J) = W_T. \tag{6.31}$$

Clearly this is not going to be the case for any distribution function $p(e_I)$. Schematically, as seen in figure 6.9, it would require the boundary between two teeth to exactly match the boundary between two generators. It is possible to artificially split in two a generator to get a matching boundary, but this adds some complexity in the implementation.

We have enforced that all teeth contain at least $5 - 10$ generators and they usually contain a lot more, up to hundreds of thousands. Therefore, a simpler solution is to "round" the teeth boundaries to the e_I thresholds directly above, which will result in teeth with weights close to W_T, and thus the effective distribution function will be little different from the desired one. Since our distribution function is an extremely rough estimation of e_I, this is unlikely to cause any significant change in the convergence rate.

Essentially, we will use $p(e_I)$ only to define the $\mathcal{T}_t$ sets, then use w as the actual distribution function. This gives us, by definition, teeth weighing exactly W_T. This is illustrated in figure 6.9.

6.4.3 Tooth filling

This is an empirical mechanism to balance the stochastic and deterministic aspect of this method. For a tooth containing n generators of equal probability, full computation

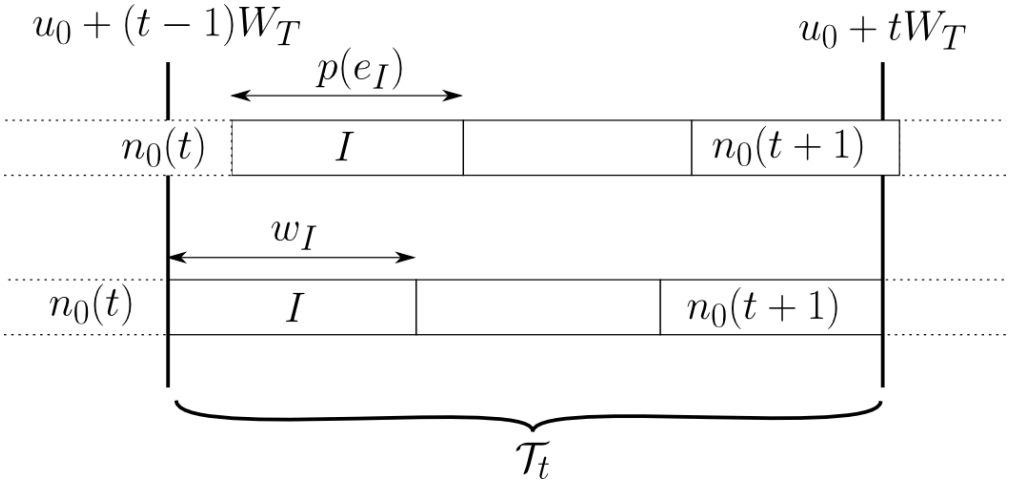


Figure 6.9: Modifying $p(e_I)$ such that the boundaries of the teeth match with the boundaries of the generators, and so as to satisfy $\sum_{I \in \mathcal{T}_{t \neq 0}} w_I = W_T$.

is achieved after on average

$$\sum_{i=0}^{n-1} \frac{n}{n-i} \tag{6.32}$$

combs are drawn. Thus, teeth containing thousands of determinants are very hard to move to $\mathcal{D}_D$. A tooth containing 10 000 generators with a single non-computed contribution only needs that particular generator to be drawn in order to be moved to $\mathcal{D}_D$, but it will take on average 10 000 more combs to be drawn until this happens by chance. With the non-uniform sampling, the situation is even worse. A convenient way to avoid this frustrating situation is, every time a comb is drawn, to additionally compute the first non-computed contribution of the whole space. This ensures smooth filling of teeth, and that that the full deterministic computation will be achieved before N_{gen} combs are drawn.

6.4.4 Comb drawing order

It must be noted that combs can only be taken into account in an initially defined random order (that is, the order of $u[i]$). It can happen that a comb becomes computable as its e_I are already computed, even though it appears first later down the list. Taking it into account would introduce a bias, as the subset of "computable combs" is biased toward the presence certain generators — those for which the contributions were computed — and thus isn't a random set of combs.

6.5 Implementation

6.5.1 Inverse Transform Sampling

The algorithm to find a generator index associated with drawing u in with a distribution function w is shown as algorithm 19.

```
1  Function FIND_SAMPLE(u,W):
       Data: 0 ≤ u < 1
       Data: W float array of size N with W[0] = 0, W[N] = 1, W[n+1] > W[n]
       Result: Returns i so that W[i−1] ≤ u < W[i]
2      /* The result must be in the range [l,r]_ℤ. We set them for
          the most general case.                                    */
3      l ← 0 ;
4      r ← N ;
5      while r − l > 1 do
6          i ← ⌊(r + l)/2⌋ ;
7          if W[i] < u then
8              l ← i ;
9          else
10             r ← i ;
11         end
12     end
13     return r ;
```

Algorithm 19: Finds generator index associated with drawing random value u in a cumulative probability distribution W

6.5.2 Building teeth

Input

- $p(e_I)$: The desired distribution function, which we chose to be c_I^2.
- N_{teeth}: A desired number of teeth
- minDetInT1 $>$ 1: A desired minimal number of generators in the first tooth. All subsequent teeth are guaranteed to contain minDetInT1 $-$ 1 generators.

Output

- w_I: The effective distribution function
- $n_0(t)$: An array of size $N_{\text{teeth}} + 1$ so that $n_0(t)$ is the number of samples in $\mathcal{D}_D$ with $\mathcal{T}_t$ the first non-deterministic tooth. For algorithmic convenience, $n_0(N_{\text{teeth}} + 1) = N_{\text{gen}}$.

There is no trivial way to ensure teeth building will succeed with a given set of parameters. With $n_0(1)$ the size of the initial deterministic set $\mathcal{T}_0$, teeth building is sure to fail if

$$N_{\text{gen}} - n_0(1) < (\text{minDetInT1} - 1) \times N_{\text{teeth}} + 1 \tag{6.33}$$

However, the opposite doesn't guarantee success. Because samples are sorted with decreasing $p(e_I)$, each tooth is guaranteed to contain at least minDetInT1 $-$ 1 samples, so building will success if $|\mathcal{T}_1| \geq$ minDetInT1. Relying on the fact that relative values of $p(e_I)$ get closer and closer, we increment $n_0(1)$ until either the first tooth contains minDetInT1 samples, or the impossibility condition is reached. If teeth building fails, we retry will $N_{\text{teeth}} - 1$ teeth (build always succeeds with $N_{\text{teeth}} = 1$). Then, $n_0(t)$ and the effective distribution function w_I can be fully computed. Teeth building is show as algorithms 20 and 21.

6.5.3 Building the task queue

Using memoization adds difficulties for the parallel implementation. In a standard Monte Carlo parallel implementation, all the CPU cores would repeatedly do the following independently: draw a random number, compute the associated e_I and send the result to the master process. The naive implementation of memoization would allocate a local memo array on each compute node, with shared memory. This would require to have locks in the write access to the memo array, and multiple contributions would be computed independently on different nodes, as the memo array would be local.

To avoid these issues while still conserving the benefits of memoization, we have chosen that only the master process performs the Monte Carlo sampling. The slave processes help by computing the contributions e_I. This is realized as follows. The master process draws combs, and each time a generator is drawn for the first time, it is appended to the task queue as a new task to be computed. When the results of the tasks are transmitted back, the master process is able to compute the running average and the error bar by identifying which combs were computed. With this scheme, each contribution e_I is computed at most once and all the CPU cores can work independently.

Because of the tooth filling mechanism, we know we will need at most N_{gen} combs.

- Q: is the task queue, $Q[I]$ the index of the I^{th} sample e_I that must be computed.
- d: keeps track of the computed samples. $d[I] =$ TRUE iff e_I has already been computed.

Data: $p(e_I)$, N_{teeth}, minDetInT1 as previously described
Result: $n_0(1), u_0, W_T$

$P_I \leftarrow \sum_{J \leq I} p(e_J)$;
$n_0(1) \leftarrow 0$;
while do
    $u_0 \leftarrow P[n_0(1)]$;
    $r \leftarrow P[n_0(1) + \text{minDetInT1}]$;
    $W_T \leftarrow \frac{1-u_0}{N_{\text{teeth}}}$;
    if $W_T \geq r - u_0$ **then**
        break loop;
    end
    $n_0(1) \leftarrow n_0(1) + 1$;
    if $N_{gen} - n_0(1) < (minDetInT1 - 1) \times N_{teeth}$ **then**
        /* Cannot compute with those parameters */
        Try with fewer teeth. ;
    end
end
for $t \leftarrow 2, N_{teeth}$ **do**
    $r \leftarrow u_0 + W_T \times (t-1)$;
    $n_0(t) \leftarrow FIND\_SAMPLE(r, P)$;
end
/* For convenience */
$n_0(N_{\text{teeth}} + 1) \leftarrow N_{\text{gen}}$;

Algorithm 20: Compute teeth weights and boundaries

Data: $p(e_I), n_0, W_T$
Result: w_i

$P_I \leftarrow \sum_{J \leq I} p(e_J)$;
$w_{I \leq n_0(1)} \leftarrow p\left(e_{I \leq n_0(1)}\right)$;
for $t \leftarrow 1, N_{teeth}$ **do**
    tooth_width $\leftarrow P_{n_0(t+1)} - P_{n_0(t)}$;
    for $I \leftarrow n_0(t) + 1, n_0(t+1)$ **do**
        $w_I \leftarrow p(e_I) \times \frac{W_T}{\text{tooth\_width}}$;
    end
end

Algorithm 21: Compute effective distribution function

- N_Q: is the number of tasks currently created. When the task queue is fully computed $N_Q = N_{\text{gen}}$.
- N_c: is the number of combs currently drawn.
- R: is an array of integer size N_{gen}, keeping track of available combs at any point of the computation. The collector node checks $R[j]$ when the first j tasks have been computed (for all j with increasing order). If $R[j] = c \neq 0$, all samples for comb c have just become available.

Algorithm 22 shows the computation of the task queue.

```
1  /* See section 6.5.3 for variables description                */
2  R ← array of size N_gen initialized to 0 ;
3  N_c ← 0 ;
4  N_Q ← n_0(1) ;
5  for I ← 1, n_0(1) do
6  |  d[I] ← TRUE ;
7  |  Q[I] ← I ;
8  end
9  /* N_gen is an upper bound of the maximum number of combs.    */
10 for i ← 1, N_gen do
11 |  u[i] ← random value in [0, 1)_R ;
12 end
13 F ← 0 ;
14 while N_Q < N_gen do
15 |  ADD_COMB shown as algorithm 23;
16 |  R[N_Q] ← N_c ;
17 |  FILL_TOOTH shown as algorithm 24 ;
18 end
19 /* For convenience, pretend the last comb is available when
      the last task is done (last task may be a tooth filling).
      */
20 if N_gen > 1 then
21 |  R[N_Q − 1] ← 0 ;
22 |  R[N_Q] ← N_c ;
23 end
```

Algorithm 22: Building the task queue.

Data: $N_c, u[N_c], F, d, N_Q, Q$ as in the scope of the calling function
Data: $\tilde{M}$ as in the scope of the calling function if defined, otherwise ignored
$N_c \leftarrow N_c + 1$;
for $t \leftarrow 0, N_{teeth} - 1$ **do**
  $v \leftarrow u_0 + W_T \times (t + u[N_c])$;
  $i \leftarrow FIND\_SAMPLE(v, W)$;
  $\tilde{M}_i \leftarrow \tilde{M}_i + 1$;
  if *not* $d[i]$ **then**
    $N_Q \leftarrow N_Q + 1$;
    $Q[N_Q] \leftarrow i$;
    $d[i] \leftarrow$ TRUE ;
  end
end

Algorithm 23: ADD_COMB, called by algorithm 22.

Data: F,d,N_Q,Q as in the scope of the calling function
while $F < N_{gen}$ **do**
  $F \leftarrow F + 1$;
  if *not* $d[F]$ **then**
    $N_Q \leftarrow N_Q + 1$;
    $Q[N_Q] \leftarrow F$;
    $d[F] \leftarrow$ TRUE ;
    break ;
  end
end

Algorithm 24: FILL_TOOTH, called in algorithm 22.

6.5.4 Computing the average and error

The algorithm for the master thread is shown as algorithm 25. The slave threads take a generator and a “fragment” index, and returns the result to the master thread.

```
1  n ← 1 ;
2  t ← 0 ;
3  U ← 0 ;
4  f integer array of size N_gen initialized with F the fragmentation. ;
5  d logical array of size N_gen + 1 initialized with FALSE ;
6  S and S^(2) float arrays size N_teeth + 1 initialized with 0 ;
7  error ← 0 ;
8  while n ≤ N_gen do
9      if f[Q[n]] = 0 then
10         d[Q[n]] ← TRUE ;
11         while d[U + 1] do
12             U ← U + 1 ;
13         end
14         /* Short-circuit boolean evaluation is required to
              prevent out of bound access to n_0                    */
15         while t ≤ N_teeth ∧ U ≥ n_0(t + 1) do
16             t ← t + 1 ;
17             E_0 ← Σ_{I<n_0(t)} e_I ;
18         end
19         if R[n] ≠ 0 then
20             c ← R[n] ;
21             /* Updating S and S^(2) is costly if done naively   */
22             S_* ← S_* + B_*(u[c]) ;
23             S_*^(2) ← S_*^(2) + B_*(u[c])^2 ;
24             E ← E_0 + S_t/c ;
25             if c>1 then
26                 error ← sqrt( (S_t^(2) − S_t^2)(c − 1)^{-1} ) ;
27                 exit on acceptable error ;
28             end
29         end
30         n ← n + 1 ;
31     else
32         retrieve I and e_I ;
33         store e_I ;
34         f[I] ← f[I] − 1 ;
35     end
36 end
37 /* Estimated energy E ± error                                    */
```

Algorithm 25: Master node in E_{PT2} computation

```
/* Done naively, updating S_* ← S_* + B_*(u) and S_*^(2) ← S_*^(2) + B_*(u)^2
   scales as O(N_teeth^2 log N_gen).                                        */
x ← 0 ;
for t ← N_teeth, 1 by −1 do
    I ← FIND_SAMPLE(u_0 + W_T × (u + t − 1), W) ;
    x ← x + W_T × e_I / w_I ;
    S_t ← S_t + x ;
    S_t^(2) ← S_t^(2) + x^2 ;
end
```

Algorithm 26: Update S and $S^{(2)}$ of algorithm 25.

6.6 Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory

THE JOURNAL OF CHEMICAL PHYSICS **147**, 034101 (2017)

Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory

Yann Garniron, Anthony Scemama,[a] Pierre-François Loos, and Michel Caffarel
Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, Toulouse, France

(Received 14 March 2017; accepted 26 June 2017; published online 17 July 2017)

A hybrid stochastic-deterministic approach for computing the second-order perturbative contribution $E^{(2)}$ within multireference perturbation theory (MRPT) is presented. The idea at the heart of our hybrid scheme—based on a reformulation of $E^{(2)}$ as a sum of elementary contributions associated with each determinant of the MR wave function—is to split $E^{(2)}$ into a stochastic and a deterministic part. During the simulation, the stochastic part is gradually reduced by dynamically increasing the deterministic part until one reaches the desired accuracy. In sharp contrast with a purely stochastic Monte Carlo scheme where the error decreases indefinitely as $t^{-1/2}$ (where t is the computational time), the statistical error in our hybrid algorithm displays a polynomial decay $\sim t^{-n}$ with $n = 3$–4 in the examples considered here. If desired, the calculation can be carried on until the stochastic part entirely vanishes. In that case, the exact result is obtained with no error bar and no noticeable computational overhead compared to the fully deterministic calculation. The method is illustrated on the F_2 and Cr_2 molecules. Even for the largest case corresponding to the Cr_2 molecule treated with the cc-pVQZ basis set, very accurate results are obtained for $E^{(2)}$ for an active space of (28e, 176o) and a MR wave function including up to 2×10^7 determinants. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4992127]

I. INTRODUCTION

Multireference (MR) approaches are based upon the distinction between non-dynamical (or static) and dynamical correlation effects. Though such a clear-cut distinction is questionable, it is convenient to discriminate between the so-called static correlation effects emerging whenever the description of the molecular system using a single configuration breaks down (excited-states, transition-metal compounds, systems far from their equilibrium geometry, etc.)[1] and the dynamical correlation effects resulting from the short-range part of the electron-electron repulsion.[2]

To quantitatively establish this distinction, the Hamiltonian is decomposed as

$$\hat{H} = \hat{H}^{(0)} + \hat{V}, \tag{1}$$

where the zeroth-order Hamiltonian $\hat{H}^{(0)}$ is chosen in conjunction with an MR wave function including the most chemically relevant configurations at the origin of static correlation effects, and

$$\hat{V} = \hat{H} - \hat{H}^{(0)} \tag{2}$$

is the residual part describing the bulk of dynamical correlation effects. The plethora of MR methods found in the literature results from the large freedom in choosing $\hat{H}^{(0)}$, and the fact that $\hat{V}$ may or may not be treated perturbatively. Among the non-perturbative approaches, let us cite the two most common ones, namely, the MR configuration interaction (MRCI)[1,3,4] and the MR coupled cluster (MRCC)[5–8] approaches. However, because of their high computational cost, these methods are usually limited to systems of moderate size.

To overcome the computational burden associated with these methods—yet still capturing the main physical effects—a natural idea is to treat the potential as a perturbation, entering the realm of MR perturbation theories (MRPTs). Several flavors of MRPT exist depending on the choice of $\hat{H}^{(0)}$ (Epstein-Nesbet decomposition,[9,10] Dyall Hamiltonian,[11,12] Fink's partitioning,[13,14] etc.). Among the most commonly used approaches, we have the CASPT2[15,16] and NEVPT2[11,12] methods. Regarding the construction of the zeroth-order part, CASSCF-type approaches are the most widely used schemes,[17–19] but other methods, such as Complete Active Space Configuration Interaction (CASCI), selected CI (see Refs. 20 and 21 and the references therein), Full Configuration Interaction Quantum Monte Carlo (FCIQMC),[22–24] or DMRG-type approaches[25–27] can also be employed.

In this work, we shall consider MRPTs limited to the second order in perturbation (MRPT2).[15] We address the important problem of calculating efficiently the second-order perturbative contribution $E^{(2)}$ in situations where standard calculations become challenging. Here, we suppose that the MR wave function has already been constructed by any method of choice.

Although the present method can be easily generalized to any externally decontracted MRPT approach (such as the recently introduced JM-MRPT2 method[28]), for the sake of simplicity, we shall restrict ourselves here to MR Epstein-Nesbet perturbation theory. Extension to externally contracted methods, such as CASPT2 or NEVPT2, is less obvious—although not impossible—since the excited contracted wave functions are non-orthogonal.

[a] Author to whom correspondence should be addressed: scemama@irsamc.ups-tlse.fr

The computational cost of MRPT2 can rapidly become unbearable when the number of electrons N_{el} and the number of one-electron basis functions N_{bas} become large. The cost is indeed proportional to the number of reference determinants N_{det} times the total number of singly and doubly excited determinants (scaling as $N_{\text{el}}^2 N_{\text{bas}}^2$). Because our main goal is to treat large, chemically relevant systems, the development of fast and accurate schemes for computing $E^{(2)}$ becomes paramount. Of course, in actual calculations, a trade-off must be found between the price to pay to build the MR wave function and the effort needed to evaluate $E^{(2)}$. Increasing N_{det} (i.e., improving the MR wave function) may appear as the natural thing to do as the magnitude of $E^{(2)}$ decreases and the contribution of the neglected higher orders is made smaller. However, its computational price (proportional to N_{det}) increases stiffly, and the calculation becomes rapidly unfeasible. Of course, this balance is strongly dependent on the method used to generate the MR wave function and on the ability to compute rapidly and accurately $E^{(2)}$.

In this work, we present a simple and efficient Monte Carlo (MC) method for computing the second-order perturbative contribution $E^{(2)}$. For all the systems reported here, the reference space is constructed using the Configuration Interaction using a Perturbative Selection done Iteratively (CIPSI) method,[20,21,29] a selected CI approach where important determinants are selected perturbatively. However, other variants of selected CI approaches or any other method for constructing the reference wave function may, of course, be used. Note that, in this study, the reported wall-clock times only refer to the computation of $E^{(2)}$, i.e., they do not take into account the preliminary calculation of the reference wave function.

A natural idea to evaluate $E^{(2)}$ with some targeted accuracy is to truncate the perturbational sum over excited determinants. However, since all the terms of the second-order sum have the same (negative) sign, the truncation will inevitably introduce a bias which is difficult to control. A way to circumvent this problem is to resort to a stochastic sampling of the various contributions. In this case, the systematic bias is removed at the price of introducing a statistical error. The key property is that this error can now be controlled, thanks to the central-limit theorem. However, in practice, to make the statistical average converge rapidly and to get statistical error small enough, care has to be taken in the way the statistical estimator is built and how the sampling is performed. The purpose of the present work is to propose a practical solution to this problem.

Note that the proposal of computing stochastically perturbative contributions is not new. In the context of second-order Møller-Plesset (MP2) theory, where the reference Hamiltonian reduces to the Hartree-Fock Hamiltonian, Hirata and coworkers have proposed a MC scheme for calculating the MP2 correlation energy.[30,31] However, we point out that this approach, based on a single-reference wave function, samples a 13-dimensional integral (in time and space) and has no direct relation with the present method. In a recent study, Sharma *et al.*[32] address the very same problem of computing stochastically the second-order perturbative contribution of Epstein-Nesbet MRPT. Similarly to what is proposed here, $E^{(2)}$ is recast as a sum over contributions associated with each reference determinant, and contributions are stochastically sampled. However, the definition of the quantities to be averaged and the way the sampling is performed are totally different. Finally, let us mention the recent work of Jeanmairet *et al.*[33] addressing a similar problem in a different way. Within the framework of the recently proposed linear CC MRPT, it is shown that both the zeroth-order and first-order wave functions can be sampled using a generalization of the FCIQMC approach. Here also, $E^{(2)}$ can be expressed as a stochastic average.

The present paper is organized as follows. In Sec. II, we report notations and basic definitions for MRPT2. Section III proposes an original reformulation of the second-order contribution allowing an efficient MC sampling. The expression of the MC estimator is given, and a hybrid stochastic-deterministic approach greatly reducing the statistical fluctuations is presented. In Sec. IV, some illustrative applications for the F_2 and Cr_2 molecules are discussed. Finally, some concluding remarks are given in Sec. V.

II. SECOND-ORDER MULTIREFERENCE PERTURBATION THEORY

A. Second-order energy contribution

In MR Epstein-Nesbet perturbation theory, the reference Hamiltonian is chosen to be

$$\hat{H}^{(0)} = E^{(0)} \left|\Psi\right\rangle \left\langle\Psi\right| + \sum_{\alpha \in \mathcal{A}} H_{\alpha\alpha} \left|\alpha\right\rangle \left\langle\alpha\right|, \tag{3}$$

where $H_{\alpha\alpha} = \langle\alpha|\hat{H}|\alpha\rangle$ and

$$\left|\Psi\right\rangle = \sum_{I \in \mathcal{D}} c_I \left|I\right\rangle \tag{4}$$

is the reference wave function expressed as a sum of N_{det} determinants belonging to the reference space

$$\mathcal{D} = \{\left|I\right\rangle, I = 1, \ldots, N_{\text{det}}\}, \tag{5}$$

and

$$E^{(0)} = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{6}$$

is the corresponding (variational) energy. The sum in Eq. (3) is over the set of determinants $|\alpha\rangle$ that do not belong to $\mathcal{D}$ but are connected to $\mathcal{D}$ via $\hat{H}$,

$$\mathcal{A} = \{\left|\alpha\right\rangle \notin \mathcal{D} \wedge (\exists \left|I\right\rangle \in \mathcal{D} \mid H_{\alpha I} \neq 0)\}. \tag{7}$$

Due to the two-body character of the interaction, the determinants $|\alpha\rangle$ are either singly or doubly excited with respect to (at least) one reference determinant.[34] However, several reference determinants can be connected to the same $|\alpha\rangle$.

Using such notations, the second-order perturbative contribution is written as

$$E^{(2)} = \sum_{\alpha \in \mathcal{A}} \frac{\left|\langle\alpha|\hat{H}|\Psi\rangle\right|^2}{\Delta E_\alpha}, \tag{8}$$

with $\Delta E_\alpha = E^{(0)} - H_{\alpha\alpha}$.

B. Partition of $\mathcal{A}$

The first step of the method—instrumental in the MC algorithm efficiency—is the partition of $\mathcal{A}$ into N_{det} subsets $\mathcal{A}_I$

associated with each reference determinant $|I\rangle$,

$$\mathcal{A} = \bigcup_{I=1}^{N_{\text{det}}} \mathcal{A}_I \quad \text{with} \quad \mathcal{A}_I \cap \mathcal{A}_J = \emptyset \quad \text{if} \quad I \neq J. \tag{9}$$

To define $\mathcal{A}_I$, the determinants $|I\rangle$ are first sorted in descending order according to the weight

$$w_I = \frac{c_I^2}{\langle \Psi | \Psi \rangle}. \tag{10}$$

The partition of $\mathcal{A}$ starts with $\mathcal{A}_1$ defined as the set of determinants $|\alpha\rangle \in \mathcal{A}$ connected to the first reference determinant (i.e., $I = 1$). Then, $\mathcal{A}_2$ is constructed as the set of determinants of $\mathcal{A}$ connected to the determinant corresponding to $I = 2$, but not belonging to $\mathcal{A}_1$. The process is carried on up to the last determinant. This partition is schematically illustrated in Fig. 1. Mathematically, it can be written as

$$\mathcal{A}_I = \{|\alpha\rangle \in \mathcal{A} \mid H_{\alpha I} \neq 0 \wedge (\forall J < I, |\alpha\rangle \notin \mathcal{A}_J)\}. \tag{11}$$

Because of the way they are constructed, the size of $\mathcal{A}_I$ is expected to decrease rapidly as a function of I, except for a possible transient regime for very small I.

A key point in the construction of the partition of $\mathcal{A}$ is to avoid both the computation of redundant contributions and the storage of unnecessary intermediates. First, when a determinant $|\alpha\rangle$ is generated by applying a single or double excitation operator to a reference determinant $|I\rangle$, one has to check that $|\alpha\rangle$ does not belong to $\mathcal{D}$. If the reference determinants are stored in a hash table, the presence of $|\alpha\rangle$ in $\mathcal{D}$ can be checked in constant time. Next, one has to know if $|\alpha\rangle$ has already been generated via another reference determinant $|J\rangle$. To do so, one must compute the number of holes and particles between $|\alpha\rangle$ and each determinant preceding $|I\rangle$ in $\mathcal{D}$. As soon as an excitation degree lower than 3 is found, the search can be aborted since the contribution is known to have been considered before. In the worst-case scenario, this step scales as $\mathcal{O}(N_{\text{det}})$, and the prefactor is very small since finding the excitation degree between two determinants can be performed in less than 20 CPU cycles[35] (comparable to a floating-point division). Furthermore, the asymptotic scaling can be further reduced by sorting the determinants in groups with the same spin string. Indeed, one only has to probe determinants $|J\rangle$ that are no more than quadruply excited with respect to $|I\rangle$, and if the search is restricted to groups with the same spin-up string, the asymptotic scaling reduces to $\mathcal{O}\left(\sqrt{N_{\text{det}}}\right)$. To provide a quantitative illustration of the computational effort associated with the construction of the partitioning, using 2×10^7 determinants (as in the case of Cr_2 presented below), this preliminary step is negligible: on a single 2.7 GHz core, the calculation takes $20\,\text{cycles} \times N_{\text{det}}^{3/2}/(2.7 \times 10^9\,\text{cycles/s}) \sim 663\,\text{s}$ (CPU time), while the total execution time (wall-clock time) of the entire run ranges from 14 min to 18.5 h using 800 cores (see Table I).

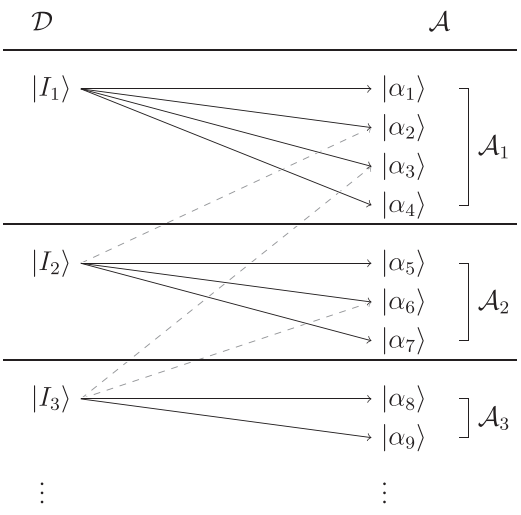


FIG. 1. Iterative construction of the subsets $\mathcal{A}_I$. Arrows indicate a non-zero matrix element $H_{I\alpha} = \langle I|\hat{H}|\alpha\rangle$. Solid arrows: the determinant $|\alpha\rangle$ is accepted as a member of the subset $\mathcal{A}_I$. Dotted arrows: the determinant $|\alpha\rangle$ already belongs to a previous subset $\mathcal{A}_{J<I}$ and is therefore not incorporated into $\mathcal{A}_I$.

TABLE I. Convergence of $E^{(2)}$ for the Cr_2 molecule with bond length 1.68 Å as a function of the wall-clock time for various basis sets (800 CPU cores).

Basis	$E^{(2)}$	Wall-clock time
cc-pVDZ	−0.068 3(1)	14 min
	−0.068 36(1)	55 min
	−0.068 361(1)	2.4 h
	−0.068 360 604	3 h
cc-pVTZ	−0.124 4(5)	19 min
	−0.124 7(1)	58 min
	−0.124 63(1)	3.5 h
	−0.124 642(1)	8.7 h
	...	~15 h (estimated)
cc-pVQZ	−0.155 8(5)	56 min
	−0.155 9(1)	2.5 h
	−0.155 95(1)	9.0 h
	−0.155 952(1)	18.5 h
	...	~29 h (estimated)

C. Partition of $E^{(2)}$

Thanks to the partition of $\mathcal{A}$ [see Eq. (11)], the sum (8) can be decomposed into a sum over the reference determinants $|I\rangle$,

$$E^{(2)} = \sum_{I=1}^{N_{\text{det}}} e_I, \tag{12}$$

where

$$e_I = \sum_{\alpha \in \mathcal{A}_I} \frac{|\langle \alpha|\hat{H}|\Psi\rangle|^2}{\Delta E_\alpha}. \tag{13}$$

Moreover, noticing that by construction, the determinants $|\alpha\rangle$ belonging to $\mathcal{A}_I$ are not connected to the part of the reference function expanded over the preceding reference determinants; we have

$$e_I = \sum_{\alpha \in \mathcal{A}_I} \frac{|\langle \alpha|\hat{H}|\Psi_I\rangle|^2}{\Delta E_\alpha}, \tag{14}$$

where

$$|\Psi_I\rangle = \sum_{J=I}^{N_{\text{det}}} c_J\,|J\rangle \tag{15}$$

is a truncated reference wave function. Our final working expression for the second-order contribution $E^{(2)}$ is thus

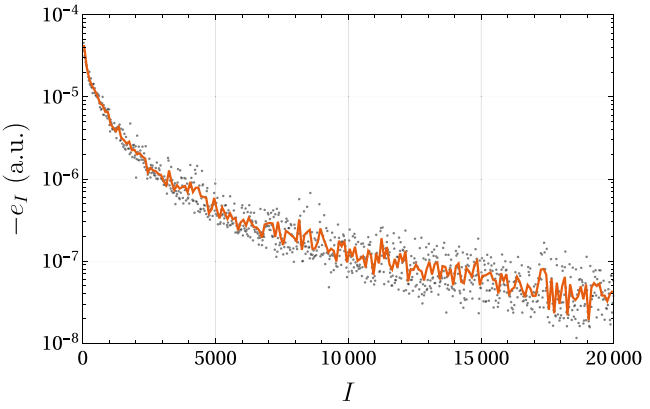


FIG. 2. $-e_I$ as a function of I for the first 20 000 determinants selected by the CIPSI method for the F_2 molecule at equilibrium geometry with the cc-pVQZ basis set. The two sets of data are obtained by averaging either by groups of 20 (point cloud) or 100 (solid line) values.

written as

$$E^{(2)} = \sum_{I=1}^{N_{\text{det}}} e_I = \sum_{I=1}^{N_{\text{det}}} \sum_{\alpha \in \mathcal{A}_I} \frac{|\langle \alpha | \hat{H} | \Psi_I \rangle|^2}{\Delta E_\alpha}. \tag{16}$$

A key property at the origin of the efficiency of the MC simulations presented below is that e_I's take their largest values at very small I. Then, they decay very rapidly as I increases.

This important property is illustrated in Fig. 2. The data have been obtained for the F_2 molecule at the equilibrium bond length of $R_{\text{F-F}} = 1.4119$ Å using Dunning's cc-pVQZ basis set.[36] The multideterminant reference space is built by selecting determinants using the CIPSI algorithm. Figure 2 displays e_I's for the first 20 000 selected determinants. As one can see, e_I's decay very rapidly with I. Of course, at the scale of individual determinants, there is no guarantee of a strictly monotonic decay, and it is indeed what we observe. By averaging groups of successive e_I's, the curve can be smoothed out. The two data sets presented in Fig. 2 have been obtained by averaging either by groups of 20 (point cloud) or 100 (solid line) values.

It is important to note that the rapid decay of e_I's is a direct consequence of the way we have chosen to decompose $\mathcal{A}$. To be more precise, we note that in Eq. (14), the decay has three different origins:

- the number of determinants involved in the sum over $|\alpha\rangle$ decreases as a function of I;
- the excitation energies ΔE_α increase with I;
- the norm of the truncated wave function Ψ_I decreases rapidly (as c_I^2) when I increases.

In addition, as a consequence of the first point, we note that the computation of e_I becomes faster when I increases.

III. MONTE CARLO METHOD

A. Monte Carlo estimator

To get an expression of $E^{(2)}$ suitable for MC simulations, the second-order contribution is recast as

$$E^{(2)} = \sum_{I=1}^{N_{\text{det}}} p_I \left(\frac{e_I}{p_I} \right) \tag{17}$$

and is thus rewritten as the following MC estimator:

$$E^{(2)} = \left\langle \frac{e_I}{p_I} \right\rangle_{p_I}. \tag{18}$$

Here, p_I is an arbitrary probability distribution. The optimal choice for p_I is given by the zero-variance condition, i.e.,

$$p_I^{\text{opt}} = \frac{e_I}{E^{(2)}}. \tag{19}$$

Note that e_I and $E^{(2)}$ being both negative, the probability distribution p_I is positive, as it should be.

To build a reasonable approximation of p_I, we note that the magnitude of e_I, as expressed in Eq. (14), is essentially given by the norm of the truncated wave function Ψ_I [see Eq. (15)]. Thus, a natural choice for the probability distribution is

$$p_I = \frac{\langle \Psi_I | \Psi_I \rangle}{\sum_{J=1}^{N_{\text{det}}} \langle \Psi_J | \Psi_J \rangle} = \frac{\sum_{J=I}^{N_{\text{det}}} c_J^2}{\sum_{J=1}^{N_{\text{det}}} \sum_{K=J}^{N_{\text{det}}} c_K^2}. \tag{20}$$

In our simulations, we have observed that summing totally or partially the squared coefficients in the numerator does not change significantly the statistical fluctuations. As a consequence, we restrict the summation in Eq. (20) to the leading term, i.e.,

$$p_I = \frac{c_I^2}{\sum_{J=1}^{N_{\text{det}}} c_J^2} = w_I. \tag{21}$$

Let us emphasize that performing a MC simulation in the e_I space is highly beneficial since the number of e_I is always small enough to make them all fit in memory. Hence, one can follow the so-called *lazy evaluation* strategy:[37] the value of e_I is computed only once when needed for the first time, and its value is then stored. If the same e_I is requested later, the stored value will be returned.

B. Improved Monte Carlo sampling

The stochastic calculation of $E^{(2)}$, Eq. (18), can be done in a standard way by sampling the probability distribution and averaging the successive values of e_I/p_I. In practice, the sampling can be realized by drawing, at each MC step, a uniform random number $u \in [0, 1]$ and selecting the determinant $|I\rangle$ verifying

$$R(I-1) \leq u \leq R(I), \tag{22}$$

where R is the cumulative distribution function of the probability distribution defined as

$$R(I) = \sum_{J=1}^{I} p_J, \tag{23}$$

with $R(0) = 0$.

At this stage, it is useful to take advantage of the fact that, thanks to the way e_I's have been constructed, the quantity to be averaged, e_I/p_I, is a slowly varying function of I (providing that the small-scale fluctuations present at the level of individual determinants have been averaged out). This property, which is well illustrated by Fig. 2, is shared by $p_I \sim c_I^2$, hence by the ratio e_I/p_I. Thus, an efficient way to reduce the statistical fluctuations consists in sampling *piece-wisely* $\mathcal{D}$ by decomposing it into subdomains where the integrand is a slowly varying function [see the justification of this statement after Eq. (30)].

To implement this idea, the interval $[0, 1]$ is divided into M equally spaced intervals $\mathcal{U}_k$ and a "comb" of correlated random numbers

$$u_k = \frac{k - 1 + u}{M}, \quad \text{for} \quad k = 1, \ldots, M, \tag{24}$$

covering uniformly $[0, 1]$ is created (where u is a single uniform random number). At each MC step, a M-tuple of determinants $(I_1, I_2, \ldots, I_M)$ verifying

$$R(I_k - 1) \leq u_k \leq R(I_k), \quad \text{for} \quad k = 1, \ldots, M \tag{25}$$

is drawn.

Defining $\mathcal{D}_k$ as the subset of determinants $|I_k\rangle$ satisfying $R(I_k) \in \mathcal{U}_k$, we introduce the following partition:

$$\mathcal{D} = \bigcup_{k=1}^{M} \mathcal{D}_k \quad \text{with} \quad \mathcal{D}_k \cap \mathcal{D}_l = \emptyset, \forall\, k \neq l \tag{26}$$

and express $E^{(2)}$ as a sum of M contributions associated with each $\mathcal{D}_k$,

$$E^{(2)} = \sum_{k=1}^{M} \sum_{I_k \in \mathcal{D}_k} e_{I_k}. \tag{27}$$

Using the process described above [Eqs. (24) and (25)], the second-order energy can be rewritten as the following MC estimator:

$$E^{(2)} = \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{e_{I_k}}{p_{I_k}} \right\rangle_{p(I_1,\ldots,I_M)}, \tag{28}$$

where $p(I_1, \ldots, I_M)$ denotes the normalized probability distribution corresponding to Eqs. (24) and (25). Equation (28) follows from the fact that, by construction, p_{I_k} is the kth marginal distribution of $p(I_1, \ldots, I_M)$,

$$\sum_{I_1} \cdots \sum_{I_{k-1}} \sum_{I_{k+1}} \cdots \sum_{I_M} p(I_1, \ldots, I_M) = M p_{I_k}, \tag{29}$$

with

$$\sum_{I_k \in \mathcal{D}_k} p_{I_k} = \frac{1}{M}. \tag{30}$$

By drawing determinants on separate subsets $\mathcal{D}_k$, the sum to be averaged in Eq. (28) is expected to fluctuate less than the very same sum computed by independently drawing determinants over $\mathcal{D}$. This remarkable property can be explained as follows. For large M, the fluctuations of the sum based on independent drawings behave as in any MC scheme, i.e., as $M^{-1/2}$. Using a comb covering evenly (with weight p_I) the determinant space, the situation is different since the sum can now be seen as a Riemann sum over $\mathcal{D}$ with a residual error behaving as M^{-1}. As a consequence, the overall reduction in statistical noise resulting from the use of the comb is expected to be of the order of $\sqrt{M}$. We emphasize that such an attractive feature is only observed because e_I/p_I is a slowly varying function of I (as mentioned above). In the opposite case, the gain would vanish. In the application on the F_2 molecule presented below (see Fig. 5), the numerical results confirm this: a decrease of about one order of magnitude in statistical error is obtained when using $M = 100$. Note that using a comb reduces the estimator's variance but does not change the typical inverse square root behavior of the statistical error with respect to the number of MC steps.

Note that Eq. (26) is actually not correct when some determinants (first and/or last determinant of a given subset) belong to more than one subset. Thus, special care has to be taken for determinants at the boundary of two subsets, but this difficulty can be easily circumvented by formally duplicating each of these determinants into copies with suitable weights.

C. Hybrid stochastic-deterministic scheme

In practice, because the first few determinants are responsible for the most significant contribution in Eq. (17), it is advantageous not to sample the entire reference space but to remove from the stochastic sampling the leading determinants. Consequently, $E^{(2)}$ is split into a deterministic $E_{\mathrm{D}}^{(2)}$ and a stochastic $E_{\mathrm{S}}^{(2)}$ component, such as

$$\begin{aligned} E^{(2)} &= E_{\mathrm{D}}^{(2)} + E_{\mathrm{S}}^{(2)} \\ &= \sum_{J \in \mathcal{D}_{\mathrm{D}}} e_J + \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{e_{I_k}}{p_{I_k}} \right\rangle_{p(I_1,\ldots,I_M)}, \end{aligned} \tag{31}$$

where $\mathcal{D}_{\mathrm{D}}$ is the set of determinants in the deterministic space, and $\mathcal{D}_{\mathrm{S}} = \mathcal{D} \setminus \mathcal{D}_{\mathrm{D}}$ is its stochastic counterpart.

At a given point of the simulation, some determinants have been drawn, and some have not. If we keep track of the list of the drawn determinants, we can check periodically, for each $\mathcal{D}_k$, whether or not all elements have been drawn at least once. If that is the case, the full set of determinants is moved to $\mathcal{D}_{\mathrm{D}}$ and the corresponding contribution $\sum_{I_k} e_{I_k}$ is added to $E_{\mathrm{D}}^{(2)}$. The statistical average and error bar are then updated accordingly. The expression of the $E^{(2)}$ estimator is now time-dependent and, at the mth MC step, the deterministic part is given by

$$E_{\mathrm{D}}^{(2)}(m) = \sum_{k=1}^{M} \Theta_k(m) \sum_{I_k} e_{I_k}, \tag{32}$$

where

$$\Theta_k(m) = \begin{cases} 1, & \text{if } \mathcal{D}_k \subset \mathcal{D}_{\mathrm{D}} \text{ at step } m, \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

On the other hand, the stochastic part is now given by

$$E_{\mathrm{S}}^{(2)}(m) = \frac{1}{M} \sum_{k=1}^{M} [1 - \Theta_k(m)] \sum_{I_k \in \mathcal{D}_k} w_{I_k}^{(m)} \frac{e_{I_k}}{p_{I_k}}, \tag{34}$$

where

$$w_{I_k}^{(m)} = \frac{n_{I_k}^{(m)}}{\sum_{J_k \in \mathcal{D}_k} n_{J_k}^{(m)}}, \tag{35}$$

and $n_{I_k}^{(m)}$ denotes the number of times the determinant I_k has been drawn at iteration m.

If desired, the calculation can be carried on until the stochastic part entirely vanishes. In that case, all the determinants are in $\mathcal{D}_{\mathrm{D}}$, and the exact value of $E^{(2)}$ is obtained with zero statistical fluctuations.

Finally, to make sure that a given set $\mathcal{D}_k$ does not stay in the stochastic part because a very small number of its determinants have not been drawn, we have implemented an additional step as follows. At each MC iteration (where a new comb is created), the contribution e_I of the first not-yet-sampled determinant (i.e., corresponding to the smallest I value in the sorted

determinant list) is calculated and stored. By doing this, the convergence of the hybrid stochastic-deterministic estimator is significantly improved. Moreover, after N_{det} MC steps, it is now guaranteed that the exact deterministic value is reached.

D. Upper bound on the computational time

In the present method, the vast majority of the computational time is spent calculating e_I's. A crucial point which makes the algorithm particularly efficient is the lazy evaluation of these quantities. This implies that, in practice, the stochastic calculation will never be longer than the time needed to compute all the individual e_I's (i.e., the time necessary to complete the fully deterministic calculation) due to the negligible time required by the MC sampling (drawing 100×10^6 random numbers takes less than 3 s on a single CPU core).

Finally, it is noteworthy that the final expression of $E^{(2)}$ can be very easily decomposed into (strictly) independent calculations. The algorithm presented here is thus embarrassingly parallel (see Sec. IV C).

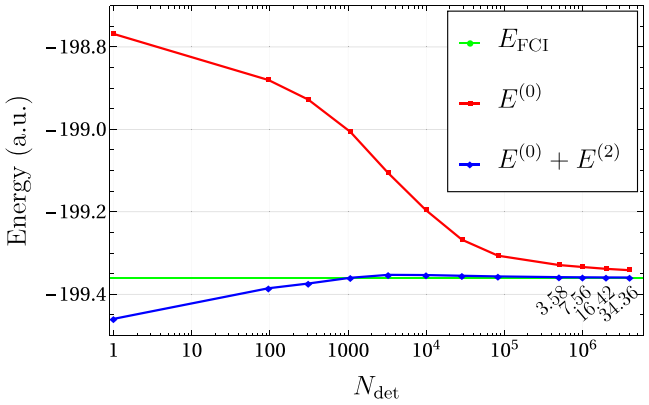


FIG. 3. F_2 molecule at equilibrium geometry. Convergence of the variational energy $E^{(0)}$ (red curve) as a function of the number of selected determinants N_{det} obtained with the CIPSI method and the cc-pVQZ basis set. The blue curve is obtained by adding the second-order energy contribution $E^{(2)}$ to the variational one $E^{(0)}$. The full CI (FCI) value (green curve) is reported as a reference. The wall-clock time (in minutes) needed to compute $E^{(2)}$ for various values of N_{det} is also reported (black numbers underneath the blue curve).

IV. NUMERICAL TESTS

The present algorithm has been implemented in our Quantum Package code.[38] The perturbatively selected CI algorithm CIPSI,[20,29] as described in Ref. 21, is used to build the multi-determinant reference space. In all the calculations performed in this section, we have chosen to use a comb with $M = 100$. All the simulations were performed on the Curie supercomputer (TGCC/CEA/GENCI) where each node is a dual socket Xeon E5-2680 at 2.70 GHz with 64 GB of RAM, interconnected with an Infiniband QDR network.

A. F_2 molecule

As a first illustrative example, we consider the calculation of $E^{(2)}$ for the F_2 molecule in its $^1\Sigma_g^+$ electronic ground state at equilibrium geometry. The two $1s$ core electrons are kept frozen, and Dunning's cc-pVQZ basis set is used. The Hilbert space is built by distributing the 14 active electrons within the 108 non-frozen molecular orbitals for a total of more than 10^{20} determinants.

Despite the huge size of the Hilbert space, the selected CI approach is able to reach the full CI (FCI) limit with a very good accuracy. The convergence of the variational energy $E^{(0)}$ and that of the total energy (given by the sum of the variational and second-order contribution $E^{(0)} + E^{(2)}$) with respect to the number of selected determinants are presented in Fig. 3. The maximum number of determinants we have selected is 4×10^6. For this value, $E^{(0)}$ is not converged but is already a reasonable approximation to the FCI energy with an error of about 18 mE_h. In sharp contrast, the total energy including the second-order correction converges very rapidly: millihartree accuracy is reached with about 2×10^6 determinants. For $N_{\text{det}} = 4 \times 10^6$, the best value obtained is −199.3594 a.u., in quantitative agreement with the estimated FCI value of −199.3598(2) a.u. obtained by Cleland *et al.* with FCIQMC.[23]

For this system and the maximum number of selected determinants considered, it is actually possible to calculate exactly $E^{(2)}$ by explicit evaluation of the entire sum (deterministic method). The corresponding wall-clock times (in minutes) using 50 nodes (800 cores) are reported directly in Fig. 3. For $N_{\text{det}} = 10^4$, the calculation takes a few seconds, while for the largest number of $N_{\text{det}} = 4 \times 10^6$ about 35 min are needed.

We now consider the hybrid stochastic-deterministic evaluation of $E^{(2)}$. The left graph of Fig. 4 shows the evolution of $E^{(2)}$ as a function of the wall-clock time (in minutes). Data are given for the cc-pVQZ basis and $N_{\text{det}} = 4 \times 10^6$. Similar curves are obtained with the two other basis sets. As one can see, the rate of convergence of the error is striking, and eventually, the exact value is obtained with very small fluctuations. If chemical accuracy is targeted (error of roughly 1 mE_h),

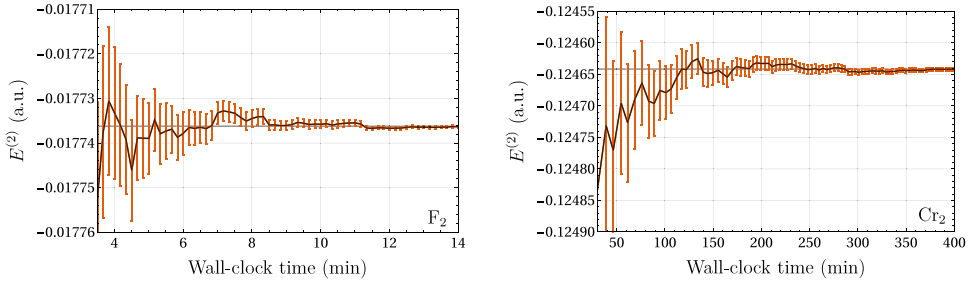


FIG. 4. Convergence of $E^{(2)}$ as a function of the wall-clock time for the F_2 molecule (left) with $N_{\text{det}} = 4 \times 10^6$ (cc-pVQZ basis set) and the Cr_2 molecule (right) with $N_{\text{det}} = 2 \times 10^7$ (cc-pVTZ basis set). Both graphs are obtained with 800 CPU cores. The grey line corresponds to the exact (deterministic) value for F_2 and to the value with the lowest statistical error for Cr_2. The error bars correspond to one standard deviation.

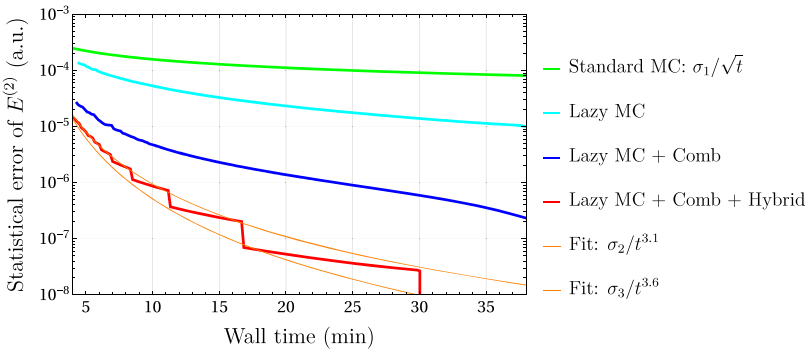


FIG. 5. Statistical error of $E^{(2)}$ as a function of the wall-clock time for the F_2 molecule obtained with the cc-pVQZ basis and $N_{\text{det}} = 4 \times 10^6$ with different schemes.

3 min are needed using 800 cores. This value has to be compared with the ~35 min needed to evaluate the exact value (see Fig. 3).

To have a better look at the fluctuations, the statistical error as a function of the wall-clock time is reported in Fig. 5. We have reported four curves to show the effects of the different strategies used in our algorithm. The first one (in green) is the curve one would typically obtain using a standard MC algorithm where the contributions are always recomputed (no lazy evaluation). Note that, for this particular curve, we have not performed the calculation, but we have plotted an arbitrary $\sigma_1/\sqrt{t}$ curve to illustrate its decay rate. The light blue curve is obtained using the MC estimator proposed in Sec. III A. The slope is steeper than that for the standard MC scheme, thanks to the lazy evaluation strategy. The introduction of the comb (Sec. III B) reduces the statistical error by an order of magnitude and produces the dark blue curve. Finally, incorporating the hybrid deterministic/stochastic scheme (Sec. III C) yields the red curve.

Quite remarkably, the overall convergence of the red curve is extremely rapid. Because of the irregular convergence, it is not easy to extract the exact mathematical form of the decay. However, it is clear that a typical polynomial decay is observed. Fitting the curve of the hybrid scheme gives a decrease of the error bar between $t^{-3.1}$ and $t^{-3.6}$, which is significantly faster than the $t^{-1/2}$ behavior of the standard MC algorithm. Note also that some discontinuities in the statistical error are regularly observed. Such sudden drops occur each time a subset $\mathcal{D}_k$ is entirely filled and its contribution is transferred to the deterministic part. Comparison with the standard MC algorithm illustrates that obtaining an arbitrary accuracy with a standard MC sampling can rapidly become prohibitively expensive. Most importantly, the wall-clock time would rapidly become larger than the time required to compute exactly (i.e., deterministically) $E^{(2)}$, which is not the case with the here-proposed method.

B. Cr_2 molecule

We now consider the challenging example of the Cr_2 molecule in its $^1\Sigma_g^+$ ground state. The internuclear distance is chosen to be close to its experimental equilibrium geometry, i.e., $R_{\text{Cr-Cr}} = 1.68$ Å. Full-valence calculations including 28 active electrons (two frozen neon cores) are performed. The cc-pVDZ, TZ, and QZ basis sets[39] are employed, and the associated active spaces corresponding to (28e, 76o), (28e, 126o), and (28e, 176o) include more than 10^{29}, 10^{36}, and 10^{42} determinants, respectively. For all the basis sets, the molecular orbitals (MOs) were obtained with the GAMESS[40] program using a CASSCF calculation with 12 electrons in 12 orbitals, and 2×10^7 determinants were selected in the FCI space with the CIPSI algorithm implemented in Quantum Package. In the cc-pVQZ basis set, we had to remove the h functions of the basis set since the version of GAMESS we used (prior to 2013) does not handle the corresponding two-electron integrals.

The right graph of Fig. 4 shows the convergence of $E^{(2)}$ as a function of the wall-clock time for the cc-pVTZ basis set and $N_{\text{det}} = 2 \times 10^7$. Again, similar curves are obtained with the two other basis sets. Similarly to F_2, the convergence is remarkably fast with a steep decrease of the statistical error with respect to the wall-clock time (for quantitative results, see Table I). Note that the maximum energy range in the right graph of Fig. 4 is only 0.35 mE_h.

Table II reports the quantitative results obtained with the three basis sets. One can observe that very accurate results for $E^{(2)}$ can be obtained even with the largest QZ basis set. For the three basis sets, the statistical error obtained is 10^{-6} E_h. However, it is clear that in practical applications, we do not need such high level of accuracy as the finite-size basis effects as well as the high-order perturbative contributions are much larger. If, more reasonably, we target an accuracy of about 0.1 mE_h, we see in Table I that the wall-clock time needed is

TABLE II. Variational ground-state energy $E^{(0)}$ and second-order contribution $E^{(2)}$ of the Cr_2 molecule with bond length 1.68 Å computed with various basis sets. For all basis sets, the reference is composed of 2×10^7 determinants selected in the valence FCI space (28 electrons).

Reference	Basis	Active space	$E^{(0)}$	$E^{(2)}$	$E^{(0)} + E^{(2)}$
CIPSI	cc-pVDZ	(28e, 76o)	−2087.227 883 3	−0.068 334(1)	−2087.296 217(1)
	cc-pVTZ	(28e, 126o)	−2087.449 781 7	−0.124 676(1)	−2087.574 423(1)
	cc-pVQZ	(28e, 176o)	−2087.513 373 3	−0.155 957(1)	−2087.669 330(1)

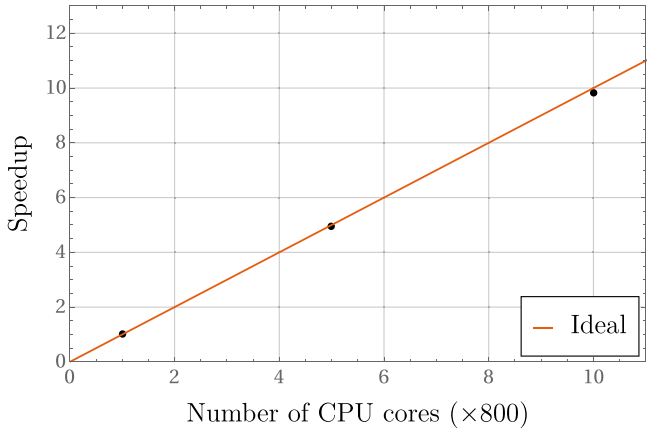


FIG. 6. Parallel speedup of our implementation using 800, 4000, and 8000 cores. The reference is the 800-core run.

about 14 min, 1 h, and 2.5 h with 800 CPU cores for the DZ, TZ, and QZ basis sets, respectively. Finally, we note that, in contrast with F_2, the absolute value of $E^{(2)}$ remains large even when relatively large MR wave functions are employed. This result clearly reflects the difficulty in treating accurately Cr_2. We postpone to a forthcoming paper the detailed analysis of this system and the calculation of the entire potential energy curve.

C. Parallel speedup

To measure the parallel speedup of the present implementation of our algorithm, we have measured the wall-clock time needed to reach a target statistical error of 10^{-6} a.u. with 800, 4000, and 8000 cores (50, 250, and 500 nodes) using the Cr_2/cc-pVQZ wave function with $N_{\text{det}} = 2 \times 10^7$. The speedup is calculated using the 800-core run as the reference, and the results are shown in Fig. 6. Going from 800 to 4000 cores gives a speedup of 4.95, and the 8000-core run exhibits a speedup of 9.82. These values are extremely close to the ideal values of 5 and 10. Therefore, we believe that this method is a good candidate for running on exascale machines in a near future.

V. CONCLUSIONS

In this work, a hybrid stochastic-deterministic algorithm to compute the second-order energy contribution $E^{(2)}$ within the Epstein-Nesbet MRPT has been introduced. Two main ideas are at the heart of the method. First, the reformulation of the standard expression of $E^{(2)}$, Eq. (8), into Eq. (16). Thanks to the unique property of the elementary contributions e_I; the latter expression turns out to be particularly well-suited for low-variance MC calculations. The second idea, which greatly enhances the convergence of the calculation, is to decompose $E^{(2)}$ as a sum of a deterministic and a stochastic part, the deterministic part being dynamically updated during the calculation.

We have observed that the size of the stochastic part (as well as the statistical error) decays in time with a polynomial behavior. If desired, the calculation can be carried on until the stochastic part entirely vanishes. In that case, the exact (deterministic) result is obtained with no error bar and no noticeable computational overhead compared to the fully deterministic calculation. Such a remarkable result is in sharp contrast with standard MC calculations where the statistical error decreases indefinitely as the inverse square root of the simulation time.

The numerical applications presented for the F_2 and Cr_2 molecules illustrate the great efficiency of the method. The largest calculation on Cr_2 (cc-pVQZ basis set) has an active space of (28e, 176o), corresponding to a Hilbert space consisting of approximately 10^{42} determinants and a multireference wave function containing 2×10^7 determinants. Even in this extreme case, $E^{(2)}$ can easily be calculated with sub-millihartree accuracy using a fully and massively parallel version of the algorithm.

As a final comment, we would like to mention that, although we have only considered two illustrative examples in the present manuscript, our method has been shown to be highly successful in all the cases we have considered.

ACKNOWLEDGMENTS

We thank the referees for their valuable comments on the first version of our manuscript. This work was performed using HPC resources from CALMIP (Toulouse) under Allocation No. 2016–0510 and from GENCI-TGCC (Grant No. 2016–08s015).

[1] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, Chem. Rev. **112**, 108 (2012).
[2] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2012).
[3] H. Lischka, R. Shepard, F. B. Brown, and I. Shavitt, Int. J. Quantum Chem. **20**, 91 (2009).
[4] H. Lischka, R. Shepard, R. M. Pitzer, I. Shavitt, M. Dallos, T. Müller, P. G. Szalay, M. Seth, G. S. Kedziora, S. Yabushita, and Z. Zhang, Phys. Chem. Chem. Phys. **3**, 664 (2001).
[5] J. Paldus and X. Li, Adv. Chem. Phys. **110**, 1 (1999).
[6] L. Kong, K. Shamasundar, O. Demel, and M. Nooijen, J. Chem. Phys. **130**, 114101 (2009).
[7] B. Jeziorski, Mol. Phys. **108**, 3043 (2010).
[8] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).
[9] S. Epstein, Phys. Rev. **28**, 695 (1926).
[10] R. K. Nesbet, Proc. R. Soc. A **230**, 312 (1955).
[11] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, J. Chem. Phys. **114**, 10252 (2001).
[12] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, J. Chem. Phys. **117**, 9138 (2002).
[13] R. F. Fink, Chem. Phys. Lett. **428**, 461 (2006).
[14] R. F. Fink, Chem. Phys. **356**, 39 (2009).
[15] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[16] K. Andersson, P.-A. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).
[17] P. Siegbahn, A. Heiberg, B. Roos, and B. Lévy, Phys. Scr. **21**, 323 (1980).
[18] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).
[19] P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, J. Chem. Phys. **74**, 2384 (1981).
[20] B. Huron, P. Rancurel, and J. P. Malrieu, J. Chem. Phys. **58**, 5745 (1973).
[21] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Using CIPSI nodes in diffusion Monte Carlo," in *Recent Progress in Quantum Monte Carlo* (American Chemical Society (ACS), 2016), Chap. 2, pp. 15–46.
[22] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[23] D. Cleland, G. H. Booth, C. Overy, and A. Alavi, J. Chem. Theory Comput. **8**, 4138 (2012).
[24] F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, Phys. Rev. Lett. **109**, 230201 (2012).
[25] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[26] S. R. White, Phys. Rev. B **48**, 10345 (1993).

[27]S. Sharma and G.-L. Chan, J. Chem. Phys. **136**, 124121 (2012).
[28]E. Giner, C. Angeli, Y. Garniron, A. Scemama, and J.-P. Malrieu, J. Chem. Phys. **146**, 224108 (2017).
[29]S. Evangelisti, J. P. Daudey, and J. P. Malrieu, Chem. Phys. **75**, 91 (1983).
[30]S. Y. Willow, K. S. Kim, and S. Hirata, J. Chem. Phys. **137**, 204122 (2012).
[31]S. Y. Willow, J. Zhang, E. F. Valeev, and S. Hirata, J. Chem. Phys. **140**, 031101 (2014).
[32]S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017), pMID: 28263594.
[33]G. Jeanmairet, S. Sharma, and A. Alavi, J. Chem. Phys. **146**, 044107 (2017).
[34]A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).
[35]A. Scemama and E. Giner, e-print arXiv:1311.6244 [physics.comp-ph] (2013).
[36]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[37]P. Hudak, ACM Comput. Surv. **21**(3), 359 (1989).
[38]A. Scemama, T. Applencourt, Y. Garniron, E. Giner, and M. Caffarel (2017). "Quantum Package," Zenodo V. 1.1, Dataset http://dx.doi.org/10.5281/zenodo.825876.
[39]N. B. Balabanov and K. A. Peterson, J. Chem. Phys. **123**, 064107 (2005).
[40]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

Chapter 7

Stochastic matrix dressing

Contents

7.1 Principle of matrix dressing

So far we have used the second order perturbation to build a zeroth-order wave function $|\Psi\rangle$ using CISPI, and estimate its distance to the FCI energy with a stochastic estimation of E_{PT2}.

In both cases we have computed the interaction between $|\Psi\rangle$ and the external space, but we have not let $|\Psi\rangle$ be revised under those interactions. This idea is based on the so-called B_k approximation proposed by Gershgorn and Shavitt.[64]

This can be achieved using the intermediate Hamiltonian theory,[65] intermediate Hamiltonians being a class of effective Hamiltonians[66] where not all roots are exact eigenvalues of the full Hamiltonian. The principle is to build a so-called intermediate Hamiltonian $\tilde{\mathbf{H}}$ which, when diagonalized, yields a wave function that takes into account the effect of an external space on the internal space.

Sticking to the state-specific case, the general principle can be understood as follows. This formulation, as it is limited to the state-specific case, is somewhat different

and wishes to be more intuitive than the one presented in the article presented in section 7.4. Fully taking into account the external space requires to solve

$$\begin{pmatrix} \mathbf{H}^{(0)} & \mathbf{h} \\ \mathbf{h}^\dagger & \mathbf{H}^{(1)} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{c}^\alpha \end{pmatrix} = E \begin{pmatrix} \mathbf{c} \\ \mathbf{c}^\alpha \end{pmatrix} \tag{7.1}$$

with $\mathbf{H}^{(0)}$ and $\mathbf{H}^{(1)}$ the zeroth and first-order Hamiltonians, $\mathbf{h}$ the coupling term between zeroth and first order spaces, $\mathbf{c}$ the coefficients for the zeroth-order space and $\mathbf{c}^\alpha$ the coefficients for the external space.

This diagonalization is normally not feasible due to the external space being too large. However, if we are willing to neglect the external-to-external and internal-to-external influences — in other words, if we freeze the external space — we can only solve the eigenequations associated with internal determinants. As usual, associating I and J to internal determinants and α to external ones, for line I we have

$$\left(H_{II}^{(0)} - E\right) c_I + \sum_{J \neq I} c_J H_{IJ}^{(0)} + \sum_\alpha c_\alpha h_{I\alpha} = 0. \tag{7.2}$$

Obviously, since we froze the external space, we need some way to estimate $\mathbf{c}^\alpha$. In the presented paper, we used a perturbative estimation consistently with what we used in CIPSI and E_{PT2}. Because of this, whenever $\mathbf{c}$ is revised, $\mathbf{c}^\alpha$ needs to be recomputed as well. This makes B_k an iterative method.

Because the c_α coefficients are frozen, $\sum_\alpha c_\alpha H_{I\alpha}$ is merely a constant added to the eigenequation of line I. Renaming this term δ_I, we can rewrite Eq.(7.2) as

$$\left(H_{II}^{(0)} - E + \frac{\delta_I}{c_I}\right) c_I + \sum_{J \neq I} c_J H_{IJ}^{(0)} = 0. \tag{7.3}$$

It appears solving this new system of linear equations is equivalent to diagonalizing

$$\tilde{\mathbf{H}} = \mathbf{H}^{(0)} + \mathbf{\Delta} \tag{7.4}$$

with $\mathbf{\Delta}$ a diagonal matrix

$$\begin{cases} \Delta_{II} = \frac{\delta_I}{c_I} \\ \Delta_{IJ} = 0 \quad \text{if } I \neq J. \end{cases} \tag{7.5}$$

Here $\tilde{\mathbf{H}}$ is the intermediate Hamiltonian and $\mathbf{\Delta}$ the so-called *dressing matrix*. Consequently $\tilde{\mathbf{H}}$ may be referred to as a *dressed Hamiltonian*. Note that the dressing matrix is diagonal because of an arbitrary rewriting of Eq. (7.2) into Eq. (7.3). We can actually build $\mathbf{\Delta}$ in any way that fulfills

$$\sum_J c_J \Delta_{IJ} = \delta_I. \tag{7.6}$$

While the choice of which elements of $\mathbf{\Delta}$ are non-zero is arbitrary, it can be of importance for numerical reasons (in addition to obvious storage reasons). However, because we diagonalize $\tilde{\mathbf{H}}$ using a Davidson diagonalization, this is of no concern to us. Indeed, Davidson's diagonalization only requires the knowledge of $\tilde{\mathbf{H}}\mathbf{c}$ and of the diagonal of $\tilde{\mathbf{H}}$

$$\tilde{\mathbf{H}}\mathbf{c} = \mathbf{H}^{(0)}\,\mathbf{c} + \mathbf{\Delta}\,\mathbf{c}. \tag{7.7}$$

Because of Eq. (7.6), we have by construction

$$\mathbf{\Delta}\,\mathbf{c} = \boldsymbol{\delta} \tag{7.8}$$

with $\boldsymbol{\delta}$ the vector $(\delta_1, \ldots, \delta_{N_{\text{det}}})$. Algorithmically, it boils down to computing $\boldsymbol{\delta}$, which is more expensive to compute than the product $\tilde{\mathbf{H}}\,\mathbf{U}$ needed for Davidson's diagonalization. But unlike $\tilde{\mathbf{H}}\,\mathbf{U}$ which is too large to be stored, and needs to be re-computed on the fly at each Davidson iteration, $\boldsymbol{\delta}$ is fixed and can easily be stored.

An improved version to this original idea was proposed by Davidson and co-workers under the name shifted-B_k ,[67, 68, 69, 70, 71, 72, 73] which is the one we implemented. Details about this improvement and on how it can be generalized to a multi-state case are available in the presented article (section 7.4).

7.2 Implementation

7.2.1 From E_{PT2} to matrix dressing

In some respect, computing the dressing matrix is akin to computing E_{PT2}. The dressing matrix can be decomposed as a sum of elementary dressing matrices $\boldsymbol{\delta}_\alpha$, each one associated with a particular $|\alpha\rangle$, just like E_{PT2} is a sum of e_α. It is possible to pack those elementary matrices together like we packed $|\alpha\rangle$ together for E_{PT2}.

$$\boldsymbol{\delta}_I = \sum_{\alpha \in \mathcal{A}_I} \boldsymbol{\delta}_\alpha \tag{7.9}$$

$$\boldsymbol{\delta} = \sum_I \boldsymbol{\delta}_I \tag{7.10}$$

So as we only need $\tilde{\mathbf{H}}\mathbf{c}$ for Davidson's algorithm, the quantity we sample is

$$\boldsymbol{\delta}_I = \mathbf{\Delta}_I\,\mathbf{c} = \sum_{\alpha \in \mathcal{A}_I} \boldsymbol{\delta}_\alpha\,\mathbf{c} \tag{7.11}$$

Thus, $\boldsymbol{\delta}_I$ is a sum over external determinants, and requires to find connections between those determinants and the wave function. Presumably, the elementary dressing vectors $\boldsymbol{\delta}_I$ will have a norm decreasing like e_I. Indeed,

$$e_I = \frac{\mathbf{c}^\dagger\,\mathbf{\Delta}_I\,\mathbf{c}}{\mathbf{c}^\dagger\mathbf{c}} = \mathbf{c}^\dagger\boldsymbol{\delta}_I \tag{7.12}$$

Using the Cauchy–Schwarz inequality

$$e_I \leq \sqrt{\left(\boldsymbol{\delta}_I^\dagger \boldsymbol{\delta}_I\right)\left(\mathbf{c}^\dagger \mathbf{c}\right)} \leq \|\boldsymbol{\delta}_I\|. \tag{7.13}$$

With that in mind, is seems possible, theoretically, to generalize our hybrid stochastic-deterministic PT2 for computing dressing vectors. However there are a few significant differences.

- We were estimating a scalar, now we are estimating a vector. How can we quickly estimate the running error? To address this problem, we decided to compute the statistical error associated with E_Δ, the energy contribution of $\mathbf{\Delta}$. Our dressed matrix being $\mathbf{H} + \mathbf{\Delta}$, the energy is

$$\frac{\langle \Psi | \widehat{H} + \widehat{\Delta} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} + \frac{\langle \Psi | \widehat{\Delta} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = E_{\text{var}} + E_\Delta \tag{7.14}$$

  where E_Δ is estimated the same way as E_{PT2} was, based on individual contributions e_I (see eq. (7.12)).

- In both cases we have N_{gen} samples, however in the case of E_{PT2} each sample is a scalar, here each sample is a vector of size N_{det}. It is easy to store N_{gen} scalars, not to store N_{gen} vectors of size N_{det}. In addition, these vectors must be transmitted from slave nodes to a master node, creating a potential network bottleneck, scaling with N_{det}.

- In the case of E_{PT2}, each connection found only requires an increment of some elements of $P(G_{pq})$. At no point two connections need to be known at the same time. This is different for methods implemented with matrix dressing. It is possible that one needs to know the detail of which internal determinants an $|\alpha\rangle$ connects to, in order to be able to compute $\boldsymbol{\delta}_\alpha$.

Implementationally speaking, just like the state-specific version requires computing a single $\boldsymbol{\delta}$ vector, the multi-state version requires a $\boldsymbol{\delta}^{(k)}$ vector to be computed for each desired state k. This, in principle, should come with minimal cost, since the loop over states can be set as the innermost one.

In practice, since the exact computation of $\boldsymbol{\delta}^{(k)}$ is as expensive as that of E_{PT2}, we need to use the same hybrid stochastic-deterministic approach. Unfortunately, using state-average coefficients for the sampling did not yield satisfying results, so we have to stick to state-specific sampling and thus compute each $\boldsymbol{\delta}^{(k)}$ individually. Therefore we will ignore the state from now on. The multi-state Davidson diagonalization, however, is done a single time per shifted-B_k iteration.

7.2.2 Reduction of the memory bottleneck

The core idea to reduce the required storage is that, in a Monte-Carlo scheme, even an "exotic" one like our hybrid approach, the estimated result has to be a linear combination of all samples. At any point m of the Monte-Carlo computation corresponding to M_m combs having been drawn, we can write our estimated dressing vector $\boldsymbol{\delta}^m$ as :

$$\boldsymbol{\delta}^m = \sum_{I=1}^{N_{\text{gen}}} \mu_I^m \boldsymbol{\delta}_I \tag{7.15}$$

The values for μ_I^m have no dependence on those of $\boldsymbol{\delta}_I$. They only depend on what samples have been drawn so far. Since we decide beforehand which combs are going to be drawn, we can compute the $\boldsymbol{\mu}$ vector for any point of the Monte-Carlo before any sample has been computed. The values we chose for M_m act as predetermined *checkpoints*.

They can be set at any arbitrary point, but they must be determined beforehand and cannot be changed during the computation ; we will only be able to get results at those points. For checkpoint m, we start with a zero-initialized vector for $\boldsymbol{\delta}^m$, and we increment it each time an elementary vector $\boldsymbol{\delta}_I$ is computed:

$$\boldsymbol{\delta}^m \leftarrow \boldsymbol{\delta}^m + \mu_I^m \boldsymbol{\delta}_I. \tag{7.16}$$

Once this has been done, $\boldsymbol{\delta}_I$ can be discarded. Indeed, when checkpoint m is reached, $\boldsymbol{\delta}^m$ has its final value, as obviously $\mu_I^m = 0$ for any $\boldsymbol{\delta}_I$ sample that hasn't yet been drawn at checkpoint m. For convenience, some parameters can be defined as functions of a checkpoint reached: $\dot{t}_m$, the first non-deterministic tooth when checkpoint m is reached, and

$$\dot{n}_0(m) = n_0(\dot{t}_m), \tag{7.17}$$

the size of the deterministic range when m is reached. μ_I^m is defined as follows

$$\mu_I^m = \begin{cases} 1 & \text{if } I \leq \dot{n}_0(m) \\ \frac{W_T \times M_{m,I}}{w_I \times M_m} & \text{if } I > \dot{n}_0(m) \end{cases} \tag{7.18}$$

with $M_{m,I}$ the number of times generator I has been drawn at checkpoint m.

The memory cost for a checkpoint m is $2 \times N_{\text{det}}$ floats, corresponding to the storage of $\boldsymbol{\mu}^m$ and $\boldsymbol{\delta}^m$. This cost is small enough to allow setting up quite a few checkpoints. However, in addition to this memory cost, comes some computational cost. If we set up N_{cp} checkpoints, it implies each time a sample is computed, we will have, theoretically, to increment N_{cp} vectors of size N_{det}. For quicker tasks, this price may not be negligible. It gets worse if, as was the case in our first implementation, a collector is in charge of updating checkpoints for multiple slaves.

We can drastically reduce the amount of writing required for each sample by rewriting $\boldsymbol{\delta}^m$. First, we define $\boldsymbol{\delta}^{D,t}$ as the total dressing contribution for tooth $\mathcal{T}_t$

$$\boldsymbol{\delta}^{D,t} = \sum_{I \in \mathcal{T}_t} \boldsymbol{\delta}_I. \tag{7.19}$$

We rewrite $\boldsymbol{\delta}^m$ as

$$\boldsymbol{\delta}^m = \sum_{t=0}^{i_m - 1} \boldsymbol{\delta}^{D,t} + \frac{1}{M_m} \sum_I \gamma_I^m \boldsymbol{\delta}_I. \tag{7.20}$$

The second term being $\boldsymbol{\delta}^m$ without its deterministic contribution, we can write, defining $\gamma_I^0 = 0$ for convenience,

$$\gamma_I^m = \begin{cases} 0 & \text{if } I \leq \dot{n}_0(m) \vee m = 0 \\ \mu_I^m \times M_m = \frac{W_T \times M_{m,I}}{w_I} & \text{if } I > \dot{n}_0(m) \wedge m \neq 0 \end{cases}. \tag{7.21}$$

We define

$$\boldsymbol{\delta}^{S,m} = \sum_I (\gamma_I^m - \gamma_I^{m-1}) \boldsymbol{\delta}_I, \tag{7.22}$$

we can rewrite the second term of Eq. (7.20)

$$\frac{1}{M_m} \sum_I \gamma_I^m \boldsymbol{\delta}_I = \frac{1}{M_m} \sum_{p=1}^{m} \boldsymbol{\delta}^{S,p} \tag{7.23}$$

and write the final form or $\boldsymbol{\delta}^m$ as

$$\boldsymbol{\delta}^m = \sum_{t=0}^{i_m - 1} \boldsymbol{\delta}^{D,t} + \frac{1}{M_m} \sum_{p=1}^{m} \boldsymbol{\delta}^{S,p}. \tag{7.24}$$

The vectors we need to store are $\boldsymbol{\delta}^{D,t}$ and $\boldsymbol{\delta}^{S,m}$. Each time we compute an elementary dressing vector $\boldsymbol{\delta}_I$, the need for update goes as follows

- $\boldsymbol{\delta}^{D,t}$ with $I \in \mathcal{T}_t$. This is exactly one write.
- $\boldsymbol{\delta}^{S,m}$ where $\gamma_I^m - \gamma_I^{m-1} \neq 0$. This is
  - No write if generator $|D_I\rangle$ is moved to $\mathcal{D}_D$ in the same checkpoint it is first drawn or computed for tooth filling.
  - Otherwise, one write per checkpoint in which I is drawn until the one where it is moved to $\mathcal{D}_D$, inclusive.

While this increases the theoretical maximum of writes to $N_{\text{cp}} + 1$ per sample, it is much lower in practice.

For convenience we define

$$\tilde{\mu}_I^m = \gamma_I^m - \gamma_I^{m-1}. \tag{7.25}$$

The task queue is built similarly to the one for E_{PT2} computation, with some differences.

- $M_{m,I}$ are evaluated for the computation of $\tilde{\mu}_I^m$.
- Instead of R we evaluate R^{-1}, based on indices of checkpoints rather than of combs. In the algorithm for E_{PT2}, when the first j tasks are completed, the $R[j]$-th comb is available. Here, when the first $R^{-1}[m]$ tasks have been completed, checkpoint m has just become computable.
- We create sample subsets $\mathcal{P}_m$ associated with checkpoints. $I \in \mathcal{P}_m$ iff

$$R^{-1}[m-1] < j \leq R^{-1}[m] \; ; R^{-1}[0] = 0 \tag{7.26}$$

with j the task index associated with I, i.e. $Q[j] = I$. These sets are tracked using an array P so that $P[I] = m$ iff $I \in \mathcal{P}_m$.

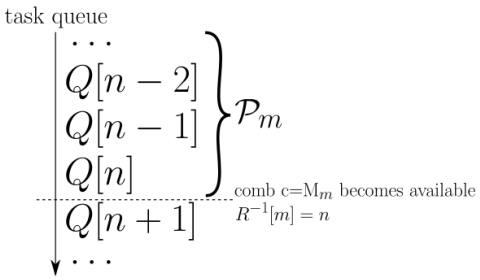


Figure 7.1: Task queue divided in checkpoints $\mathcal{P}_m$. The task array Q contains the indices of samples to be computed. When all tasks in sets $\mathcal{P}_{p \leq m}$ have been computed, checkpoint m is computable.

Algorithm 27 presents the computation of the task queue, which can be (optionally) optimized with algorithm 28, which ensures results for a checkpoint are available as quickly as possible, without altering them. Computation for other needed variables is shown as algorithm 29. Finally, algorithms 30 and 31 show the pseudo-code for the master and slave processes, respectively.

Data: M_m the desired numbers of combs for checkpoints. $M_0 = 0$ for convenience. $M_{n+1} > M_n$, $\exists M_n > N_{\text{gen}}$
Result: Q the task array, N_{cp} the number of checkpoints, P the array so that $P[I] = m$ iff $I \in \mathcal{P}_m$, R^{-1} the array so that checkpoint m becomes available when the first $R^{-1}[m]$ tasks have been computed.

$\tilde{M}$ array of size N_{det} initialized with 0;
u array of size N_{gen} initialized with random numbers $\in [0,1)_{\mathbb{R}}$;
d a boolean array of size N_{gen} initialized with FALSE.
$R^{-1}[0] \leftarrow 0$;
$N_c \leftarrow 0$;
$N_Q \leftarrow n_0(1)$;
for $i \leftarrow 1, N_Q$ **do**
    $d[i] \leftarrow$ TRUE ;
    $Q[i] \leftarrow i$;
end
$N_{\text{cp}} \leftarrow 0$;
$F \leftarrow N_Q + 1$;
while $N_Q < N_{gen}$ **do**
    ADD_COMB shown as algorithm 24 ;
    FILL_TOOTH shown as algorithm 23 ;
    if $M_{N_{cp}+1} = N_c$ **then**
        $N_{\text{cp}} \leftarrow N_{\text{cp}} + 1$;
        $R^{-1}[N_{\text{cp}}] \leftarrow N_Q$;
        $M_{N_{\text{cp}},*} \leftarrow \tilde{M}$;
    end
end
if $R^{-1}[N_{cp}] \neq N_Q$ **then**
    `/* Adds a final checkpoint. */`
    $N_{\text{cp}} \leftarrow N_{\text{cp}} + 1$;
    $R^{-1}[N_{\text{cp}}] = N_Q$;
end
optimize task queue with algorithm 28 ;
for $m \leftarrow 1, N_{cp}$ **do**
    for $i \leftarrow R^{-1}[m-1] + 1, R^{-1}[m]$ **do**
        $P[Q[i]] \leftarrow m$;
    end
end

Algorithm 27: Compute task queue and checkpoints. Tasks are fragmented as described in section 5.7 (not shown).

Result: Modifies Q the task array and R^{-1} the array defining the boundaries of checkpoints in Q.

for $m \leftarrow 1, N_{cp}$ **do**
  $Nmoved \leftarrow 0$;
  $firstTask \leftarrow R^{-1}[m-1]+1$;
  for $j \leftarrow firstTask, R^{-1}[m]$ **do**
    if $M_{m,Q[j]} = 0 \wedge Q[j] > \dot{n}_0(m)$ **then**
      /* Ensures moved tasks are at the end of the checkpoint once sorted. */
      $Q[j] \leftarrow Q[j] + N_{\text{gen}}$;
      $Nmoved \leftarrow Nmoved + 1$;
    end
  end
  Sort array Q from $firstTask$ to $R^{-1}[m]$, inclusive ;
  /* Moved tasks are sent to the next checkpoint. */
  $R^{-1}[m] \leftarrow R^{-1}[m] - Nmoved$;
  /* Restores the original values of moved tasks. */
  for $j \leftarrow R^{-1}[m]+1, R^{-1}[m]+Nmoved$ **do**
    $Q[j] \leftarrow Q[j] - N_{\text{gen}}$;
  end
end

Algorithm 28: Optimize checkpoints so that they are available faster.

- Because no result is available between two checkpoints, the order in which tasks are processed between two checkpoints is irrelevant for the result. So, as is usually the case with parallel tasks, we would like to do the longest tasks first, so that we don't get a load imbalance due to a massive task being done last. Therefore, tasks should always be in ascending order (descending computational time) between two checkpoints.
- Because of the "tooth filling", sometimes samples computed inside a checkpoint are not involved in its result. Since tooth filling picks the first non-computed tasks, they tend to be of high computational cost. The algorithm iterates over checkpoints in ascending order, each time moving such a sample to the next checkpoint. Thus, every sample is moved to the first checkpoint it is actually involved in, either deterministically or stochastically.

7.3 Conclusion

In this chapter we have implemented a stochastic version of the shifted-B_k algorithm, using the same algorithm we used for computation of E_{PT2}, and were able to get acceptable accuracies performing a few percent of the full computation. The additional difficulties of estimating a vector rather than a scalar were solved by setting a limited number of pre-determined checkpoints between which no result is available.

The fact that we implemented the shifted-B_k method boils down to our choice to build the external space using Epstein-Nesbet perturbative estimation for c_α. However, the proposed algorithm does not put any particular restriction on c_α. From this stemmed the idea of a more general framework to allow easy experimentation of the effect of different external spaces, which was used to design an MR-CCSD method as shown in chapter 8.

Data: Q, R^{-1}, n_0, N_{teeth}
Result: $\tilde{\mu}$, $\dot{t}$ and $\dot{n}_0$

$\tilde{\mu}_*^* \leftarrow 0$;
$F \leftarrow n_0(1) + 1$;
for $m \leftarrow 1, N_{cp}$ **do**
  for $i \leftarrow R^{-1}[m-1] + 1, R^{-1}[m]$ **do**
    $d[Q[i]] \leftarrow true$;
  end
  while $d(U+1)$ **do**
    $U \leftarrow U + 1$;
  end
  $\dot{t}_m \leftarrow N_{\text{teeth}} + 1$;
  $\dot{n}_0(m) \leftarrow N_{\text{gen}}$;
  for $t \leftarrow 2, N_{teeth} + 1$ **do**
    if $U < n_0(t)$ **then**
      $\dot{t}_m \leftarrow t - 1$;
      $\dot{n}_0(m) \leftarrow n_0(t-1)$;
      break loop ;
    end
  end
  for $I \leftarrow \dot{n}_0(m) + 1, N_{gen}$ **do**
    $\gamma_I^m \leftarrow \frac{W_T \times M_{m,I}}{w[I]}$;
  end
end
$\tilde{\mu}_*^1 \leftarrow \gamma_*^1$;
for $m \leftarrow 2, N_{cp}$ **do**
  $\tilde{\mu}_*^m \leftarrow \gamma_*^m - \gamma_*^{m-1}$;
end

Algorithm 29: Computation of $\tilde{\mu}$, $\dot{t}$ and $\dot{n}_0$.

S and $S^{(2)}$ float arrays size $N_{\text{teeth}} + 1$ initialized with 0 ;
$\dot{f}$ integer array of size N_{cp} initialized with $\dot{f}[m] = \sum_{I \in \mathcal{P}_{p \leq m}} F_I$;
$m \leftarrow 1$;
$c \leftarrow 1$;
$error \leftarrow 0$;
$e_* \leftarrow 0$;
while do
  if $\dot{f}[m] = 0$ **then**
    while $c \leq M_m$ **do**
      $S_* \leftarrow S_* + B_*(u[c])$;
      $S_*^{(2)} \leftarrow S_*^{(2)} + B_*(u[c])^2$;
      $c \leftarrow c + 1$;
    end
    `/* E for printing purpose */`
    $E \leftarrow \sum_{I \leq \dot{n}_0(m)} e_I + S_t / c$;
    $t \leftarrow \dot{t}_m$;
    if *c>1* **then**
      error $\leftarrow \sqrt{(S_t^{(2)} - {S_t}^2)(c-1)^{-1}}$;
    end
    exit loop if $m = N_{\text{cp}}$;
    `/* Choose to not exit if next checkpoint is available */`
    exit loop if acceptable error and $f[m+1] \neq 0$;
    $m \leftarrow m + 1$;
  else
    retrieve $(l, \breve{\boldsymbol{\delta}}^l, \breve{e}_{I \in \mathcal{P}_l}, \breve{f})$;
    increment $e_{I \in \mathcal{P}_l}$ with $\breve{e}_{I \in \mathcal{P}_l}$;
    increment $\boldsymbol{\delta}^l$ with $\breve{\boldsymbol{\delta}}^l$;
    decrement $\dot{f}[l]$ with $\breve{f}$;
  end
end
$\boldsymbol{\delta}^m$ is the estimated dressing vector ;

Algorithm 30: Master node for stochastic estimation of $\boldsymbol{\delta}$

```
1  cp_done, cp_sent: shared scalars initialized to 0 ;
2  cp_max: shared array of size N_proc initialized to 0 ;
3  f : shared array of size N_gen initialized to 0 ;
4  m ← 0 ;
5  /* Loop for each core i_proc                                          */
6  while cp_done > cp_sent ∨ m < N_cp + 1 do
7  |  Try to get task (I, s) from queue ;
8  |  if task was available then
9  |  |  m so that I ∈ P_m ;
10 |  else
11 |  |  m ← N_cp + 1 ;
12 |  end
13 |  will_send ← 0 ;
14 |  – OMP CRITICAL – ;
15 |  cp_max[i_proc] ← m ;
16 |  cp_done ← min(cp_max) − 1 ;
17 |  if cp_done > cp_sent then
18 |  |  cp_sent ← cp_sent + 1 ;
19 |  |  will_send ← cp_sent ;
20 |  end
21 |  – OMP END CRITICAL – ;
22 |  if will_send ≠ 0 then
23 |  |  Send δ^will_send (shown in algorithm 33)
24 |  end
25 |  if m < N_cp + 1 then
26 |  |  Perform task (I, s) (shown in algorithm 32) ;
27 |  end
28 end
```

Algorithm 31: Main matrix dressing code for a slave node, parallelized with OpenMP.

Data: global shared scope (at node level): $\boldsymbol{\delta}^{D,t}$ and $\boldsymbol{\delta}^{S,m}$ initialized with 0, f an array of size N_{gen} initialized with 0 (fragment count).
Data: from outer scope : s, I

m so that $I \in \mathcal{P}_m$;
t so that $I \in \mathcal{T}_t$;
Compute $\boldsymbol{\delta}_{I,s}$ (fragment s of $\boldsymbol{\delta}_I$);
`/* Lock global arrays before update */`
$\boldsymbol{\delta}^{D,t} \leftarrow \boldsymbol{\delta}^{D,t} + \boldsymbol{\delta}_{I,s}$;
for $p \leftarrow m, N_{cp}$ **do**
  if $\tilde{\mu}_I^p \neq 0$ **then**
    $\boldsymbol{\delta}^{S,p} \leftarrow \boldsymbol{\delta}^{S,p} + \tilde{\mu}_I^p \boldsymbol{\delta}_{I,s}$;
  end
end
$x \leftarrow \mathbf{c} \cdot \boldsymbol{\delta}_{I,s}$;
– OMP ATOMIC – ;
$e_I \leftarrow e_I + x$;
– OMP ATOMIC – ;
$f[I] \leftarrow f[I] + 1$;

Algorithm 32: Perform task (I, s), that is, update "node-local" partial values of $\boldsymbol{\delta}^{D,t}$ and $\boldsymbol{\delta}^{S,m}$.

Data: global shared scope (at node level): $\boldsymbol{\delta}^{D,t}$ and $\boldsymbol{\delta}^{S,m}$ initialized with 0, f an array of size N_{gen} initialized with 0 (fragment count).
Data: from outer scope : will_send

$m \leftarrow$ will_send ;
$\dot{f} \leftarrow \sum_{I \in \mathcal{P}_{p \leq m}} f[I]$;
if $\dot{f} = 0$ **then**
  return ;
end
/* Compute partial value of $\boldsymbol{\delta}^m$ with partial values of $\boldsymbol{\delta}^{D,t}$ and $\boldsymbol{\delta}^{S,m}$ */
$\boldsymbol{\delta}^m \leftarrow 0$;
/* Reverse loop for numerical precision */
for $p \leftarrow m, 1$ **by** -1 **do**
  $\boldsymbol{\delta}^m \leftarrow \boldsymbol{\delta}^m + \boldsymbol{\delta}^{S,p}$
end
$\boldsymbol{\delta}^m \leftarrow \frac{\boldsymbol{\delta}^m}{M_m}$;
for $t \leftarrow \dot{t}_m - 1, 0$ **by** -1 **do**
  $\boldsymbol{\delta}^m \leftarrow \boldsymbol{\delta}^m + \boldsymbol{\delta}^{D,t}$;
end
/* Sending partial information for checkpoint m */
send $(m, \boldsymbol{\delta}^m, e_{I \in \mathcal{P}_m}, \dot{f})$;

Algorithm 33: Build $\boldsymbol{\delta}^m$ and send it to the master process.

7.4 Selected configuration interaction dressed by perturbation

THE JOURNAL OF CHEMICAL PHYSICS **149**, 064103 (2018)

Selected configuration interaction dressed by perturbation

Yann Garniron,[1] Anthony Scemama,[1] Emmanuel Giner,[2] Michel Caffarel,[1] and Pierre-François Loos[1,a)]

[1]*Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, Toulouse, France*
[2]*Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, Sorbonne Université, CNRS, Paris, France*

(Received 13 June 2018; accepted 26 July 2018; published online 8 August 2018)

Selected configuration interaction (sCI) methods including second-order perturbative corrections provide near full CI (FCI) quality energies with only a small fraction of the determinants of the FCI space. Here, we introduce both a state-specific and a multi-state sCI method based on the configuration interaction using a perturbative selection made iteratively (CIPSI) algorithm. The present method revises the reference (internal) space under the effect of its interaction with the outer space via the construction of an effective Hamiltonian, following the shifted-Bk philosophy of Davidson and co-workers. In particular, the multi-state algorithm removes the storage bottleneck of the effective Hamiltonian via a low-rank factorization of the dressing matrix. Illustrative examples are reported for the state-specific and multi-state versions. *Published by AIP Publishing.* https://doi.org/10.1063/1.5044503

I. INTRODUCTION

Recently, selected configuration interaction (sCI) methods have demonstrated their ability to reach, for moderate size basis sets, near full CI (FCI) quality energies for small organic and transition metal-containing molecules.[1–13] Selecting iteratively the most relevant determinants of the FCI space is an old idea that, to the best of our knowledge, dates back to the pioneering studies of Bender and Davidson[14] and Whitten and Hackmeyer[15] in 1969. A few years later, Huron *et al.*[16] proposed the so-called CIPSI (Configuration Interaction using a Perturbative Selection made Iteratively) approach to complement the variational sCI energy with a second-order Epstein-Nesbet perturbative correction. This has demonstrated to be a particularly efficient way of approaching the FCI limit.[8,11–13,17,18] Over these last few years, we have witnessed a resurgence of sCI methods under various variants and acronyms. In short, their main differences lie in the way (i) the determinant selection is done and (ii) the second-order contribution is computed. The selection can be done purely stochastically as in FCIQMC[19] or deterministically as in CIPSI or other variants, such as heat-bath CI,[7–10] adaptive sampling CI (ASCI),[20–22] or iterative CI (ICI).[23] Similarly, the second-order correction can be computed either purely deterministically or semi-stochastically by a Monte Carlo (MC) sampling.[4,8,18] Here, we shall use the CIPSI method[16] to generate the model space, but any other sCI variants could be employed.

For a given electronic state k, the ensemble of determinants $|I\rangle$, which constitutes the zeroth-order (normalized) wave function

$$|\Psi_k^{(0)}\rangle = \sum_{I=1}^{N_{\text{det}}} c_{Ik}^{(0)} |I\rangle \tag{1}$$

of (variational) zeroth-order energy

$$E_k^{(0)} = \langle \Psi_k^{(0)} | \hat{H} | \Psi_k^{(0)} \rangle = {}^{\dagger}\boldsymbol{c}_k^{(0)} \boldsymbol{H}^{(0)} \boldsymbol{c}_k^{(0)} \tag{2}$$

(where ${}^{\dagger}\boldsymbol{c}_k^{(0)}$ are the transposed coefficients), defines the (zeroth-order) reference model space or internal space. The remaining determinants of the FCI space belong to the external space or outer space. In particular, the ensemble of determinants $|\alpha\rangle$ connected to $\Psi_k^{(0)}$, i.e., $\langle \alpha | \hat{H} | \Psi_k^{(0)} \rangle \neq 0$ and $\langle \alpha | \Psi_k^{(0)} \rangle = 0$—the so-called "perturbers"—defines the (first-order) perturbative space, such as

$$|\Psi_k^{(1)}\rangle = \sum_{\alpha} c_{\alpha k}^{(1)} |\alpha\rangle, \quad \boldsymbol{c}_k^{(1)} = (E_k^{(0)} \mathbf{1} - \boldsymbol{D}^{(1)})^{-1} \boldsymbol{h} \boldsymbol{c}_k^{(0)}, \tag{3}$$

where $\mathbf{1}$ is the identity matrix and $\boldsymbol{D}^{(1)}$ is a diagonal matrix with elements $D_{\alpha\alpha}^{(1)} = \langle \alpha | \hat{H} | \alpha \rangle$ and $h_{\alpha I} = \langle \alpha | \hat{H} | I \rangle$. Within CIPSI, the "distance" to the FCI solution is estimated via a second-order Epstein-Nesbet perturbative energy correction

$$E_k^{(2)} = \langle \Psi_k^{(0)} | \hat{H} | \Psi_k^{(1)} \rangle = {}^{\dagger}\boldsymbol{c}_k^{(0)} \, {}^{\dagger}\boldsymbol{h} \, \boldsymbol{c}_k^{(1)}. \tag{4}$$

The second-order correction (4) has obvious advantages and can be computed efficiently using diagrammatic[24] or hybrid stochastic-deterministic approaches.[8,17,18] However, it has also an obvious disadvantage: the internal space is not revised under the effect of its interaction with the outer space. Here, thanks to intermediate effective Hamiltonian theory,[25] we propose to build and diagonalize an effective Hamiltonian taking into account the effect of the perturbative space.[26,27] This idea is based on the so-called Bk method, originally proposed by Gershgorn and Shavitt[28] and later refined and rebranded shifted-Bk (sBk) by Davidson and co-workers.[29–38] (See also Refs. 39–42.) All these studies lie on the seminal idea of Löwdin on the partition of the FCI Hamiltonian matrix.[43] Initially, Gershgorn and Shavitt[28] introduced several approximations, two of them being denoted as Ak and Bk. Both use a partitioning of the CI matrix based on the selection of a dominant subset of (primary) configurations. The Ak method, which is related to earlier work by Claverie,

a)Author to whom correspondence should be addressed: loos@irsamc.ups-tlse.fr

Diner, and Malrieu,[44] estimates the contribution of the configurations left out of the CI expansion, an idea very similar to the computation of the second-order correction [see Eq. (4)].[14,45] Compared to the Ak method, the coefficients of the primary configurations are allowed to relax in the Bk method. The different flavours of Bk methods are usually due to the distinct partition of the Hamiltonian matrix and the reference energy used to define the perturbers [see Eq. (3) and discussion below].[26,27,29–42]

To the best of our knowledge, the shifted-Bk method has never been coupled with CIPSI-like sCI methods. Moreover, in addition to its convergence acceleration to the FCI limit, one of the interesting advantages of shifted-Bk is to provide an explicit revised wave function that one can use, for example, as a trial wave function within quantum Monte Carlo.[1,2,5,6,11,13] In the present manuscript, we propose both a state-specific and a multi-state formulation which remove the storage bottleneck of the effective Hamiltonian. Furthermore, the present computations are performed semi-stochastically as in our recently proposed hybrid stochastic-deterministic algorithm for the computation of $E^{(2)}$.[17] Unless otherwise stated, atomic units are used throughout (see Sec. III).

II. SHIFTED-Bk

A. State-specific shifted-Bk

For a given electronic state k, in order to solve the Schrödinger equation $\boldsymbol{H}\boldsymbol{c}_k = E_k\boldsymbol{c}_k$ in the FCI space, the eigenvalue problem may be partitioned as

$$\begin{pmatrix} \boldsymbol{H}^{(0)} & {}^{\dagger}\boldsymbol{h} & \boldsymbol{0} \\ \boldsymbol{h} & \boldsymbol{H}^{(1)} & {}^{\dagger}\boldsymbol{g} \\ \boldsymbol{0} & \boldsymbol{g} & \boldsymbol{H}^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{c}_k^{(0)} \\ \boldsymbol{c}_k^{(1)} \\ \boldsymbol{c}_k^{(2)} \end{pmatrix} - E_k \begin{pmatrix} \boldsymbol{c}_k^{(0)} \\ \boldsymbol{c}_k^{(1)} \\ \boldsymbol{c}_k^{(2)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \tag{5}$$

where $\boldsymbol{H}^{(2)}$ is the second-order Hamiltonian corresponding to the external configurations excluding the perturbers and $\boldsymbol{g}$ is the coupling matrix between first- and second-order spaces. Equation (5) can be recast as an "effective" Schrödinger equation $\boldsymbol{H}_k^{\text{eff}}\boldsymbol{c}_k^{(0)} = E_k\boldsymbol{c}_k^{(0)}$ with the effective Hamiltonian

$$\boldsymbol{H}_k^{\text{eff}} = \boldsymbol{H}^{(0)} + \boldsymbol{\Delta}_k, \tag{6}$$

and dressing matrix

$$\boldsymbol{\Delta}_k = {}^{\dagger}\boldsymbol{h}\left[(E_k\boldsymbol{1} - \boldsymbol{H}^{(1)}) - {}^{\dagger}\boldsymbol{g}(E_k\boldsymbol{1} - \boldsymbol{H}^{(2)})^{-1}\boldsymbol{g}\right]^{-1}\boldsymbol{h}. \tag{7}$$

Within the state-specific version of the Bk method introduced by Gershgorn and Shavitt,[28] for each target electronic state k, we (i) approximate $\boldsymbol{H}^{(1)}$ by its (diagonal) zeroth-order approximation $\boldsymbol{D}^{(1)}$ and (ii) neglect the influence of the second-order space $\boldsymbol{H}^{(2)}$. Hence, the state-specific Bk dressing matrix is defined as

$$\boldsymbol{\Delta}_k^{\text{Bk}} = {}^{\dagger}\boldsymbol{h}(E_k\boldsymbol{1} - \boldsymbol{D}^{(1)})^{-1}\boldsymbol{h}, \tag{8}$$

which naturally yields to a Brillouin-Wigner perturbation approximation.[28]

The shifted-Bk method of Davidson and co-workers[29–33] still approximates $\boldsymbol{H}^{(1)}$ by its diagonal $\boldsymbol{D}^{(1)}$, but *"shifts"* (hence the name) the energy at the denominator of Eq. (7) to take into account the influence of the second-order term ${}^{\dagger}\boldsymbol{g}(E_k\boldsymbol{1} - \boldsymbol{H}^{(2)})^{-1}\boldsymbol{g}$; in other words,

$$E_k\boldsymbol{1} - {}^{\dagger}\boldsymbol{g}(E_k\boldsymbol{1} - \boldsymbol{H}^{(2)})^{-1}\boldsymbol{g} \approx E_k^{(0)}\boldsymbol{1}. \tag{9}$$

Therefore, the state-specific shifted-Bk dressing matrix is

$$\boldsymbol{\Delta}_k^{\text{sBk}} = {}^{\dagger}\boldsymbol{h}(E_k^{(0)}\boldsymbol{1} - \boldsymbol{D}^{(1)})^{-1}\boldsymbol{h}, \tag{10}$$

which leads to the Epstein-Nesbet variant of Rayleigh-Schrödinger perturbation theory.[31,32] Compared to the Bk method, its shifted variant has the indisputable advantage of correcting some of the size-consistency error.[31] However, as expected, the present methodology is only nearly size-consistent. Note that the shifted-Bk method is an iterative method as, thanks to the influence of the entire external space, both the zeroth-order coefficients $\boldsymbol{c}_k^{(0)}$ and energy $E_k^{(0)}$ [given by Eq. (2)] are revised at each iteration.

For small CI expansions, it is possible to store the entire dressed Hamiltonian matrix $\boldsymbol{H}_k^{\text{eff}}$ of size $N_{\text{det}} \times N_{\text{det}}$. However, when the CI expansion gets large, $\boldsymbol{H}_k^{\text{eff}}$ becomes too large to be stored in memory. Thankfully, it is not necessary to explicitly build $\boldsymbol{H}_k^{\text{eff}}$. Indeed, for large CI expansions, we switch to a Davidson diagonalization procedure[46] which only requires the computation of the vectors $\boldsymbol{H}^{(0)}\boldsymbol{c}_k^{(0)}$ and $\boldsymbol{\Delta}_k^{\text{sBk}}\boldsymbol{c}_k^{(0)}$ of size N_{det}.

B. Multi-state shifted-Bk

In a multi-state calculation, one has to adopt a different strategy in order to dress the Hamiltonian for all the target states simultaneously. This is particularly important in practice, for instance, to determine accurate vertical transition energies. An unbalanced treatment of the ground and excited states, even for states with different spatial or spin symmetries, could have significant effects on the accuracy of these energy differences.[12]

For the sake of simplicity, let us assume that our aim is to calculate the dressed energy of the N_{st} lowest electronic states. For $1 \leq k \leq N_{\text{st}}$, we wish to find a multi-state effective Hamiltonian $\boldsymbol{H}^{\text{eff}}$ and a dressing matrix $\boldsymbol{\Delta}^{\text{sBk}}$, with $\boldsymbol{H}^{\text{eff}} = \boldsymbol{H}^{(0)} + \boldsymbol{\Delta}^{\text{sBk}}$, such that, when applied to the kth state coefficient vector $\boldsymbol{c}_k^{(0)}$, one recovers the kth state-specific dressing matrix $\boldsymbol{\Delta}_k^{\text{sBk}}$ times the same vector $\boldsymbol{c}_k^{(0)}$, i.e.,

$$\boldsymbol{\Delta}^{\text{sBk}}\boldsymbol{c}_k^{(0)} = \boldsymbol{\Delta}_k^{\text{sBk}}\boldsymbol{c}_k^{(0)}. \tag{11}$$

A solution obeying Eq. (11) is

$$\boldsymbol{\Delta}^{\text{sBk}} = \sum_{kl} \boldsymbol{\Delta}_k^{\text{sBk}}\boldsymbol{c}_k^{(0)}(\boldsymbol{S}^{-1})_{kl}\,{}^{\dagger}\boldsymbol{c}_l^{(0)}, \tag{12}$$

where $(\boldsymbol{S}^{-1})_{kl} = \langle \boldsymbol{c}_k^{(0)} | \boldsymbol{c}_l^{(0)} \rangle$. In contrast to the state-specific case, $\boldsymbol{H}^{\text{eff}}$ is non-Hermitian as a consequence of the non-orthogonality of the exact state projections on the model space.[25] In practice, we have found that a robust algorithm can be defined by symmetrizing the multi-state dressing matrix as

$$\tilde{\boldsymbol{\Delta}}^{\text{sBk}} = \left({}^{\dagger}\boldsymbol{\Delta}^{\text{sBk}} + \boldsymbol{\Delta}^{\text{sBk}}\right)/2. \tag{13}$$

The eigenstates being now orthonormal, the dressing matrix reduces to

$$\boldsymbol{\Delta}^{\text{sBk}} = \boldsymbol{\delta}^{\text{sBk}}\,{}^{\dagger}\boldsymbol{c}^{(0)}, \tag{14}$$

which is reminiscent of a low-rank factorization. Here,

$$\boldsymbol{c}^{(0)} = \left[\boldsymbol{c}_1^{(0)}, \ldots, \boldsymbol{c}_{N_{\rm st}}^{(0)}\right], \tag{15a}$$

$$\boldsymbol{\delta}^{\rm sBk} = \left[\boldsymbol{\Delta}_1^{\rm sBk}\boldsymbol{c}_1^{(0)}, \ldots, \boldsymbol{\Delta}_{N_{\rm st}}^{\rm sBk}\boldsymbol{c}_{N_{\rm st}}^{(0)}\right] \tag{15b}$$

are both of size $N_{\rm det} \times N_{\rm st}$.

Two key remarks are in order here: (i) at first order, the symmetrization error is strictly zero, i.e., ${}^{\dagger}\boldsymbol{c}_k^{(0)}(\boldsymbol{\Delta}^{\rm sBk} - \tilde{\boldsymbol{\Delta}}^{\rm sBk})\boldsymbol{c}_k^{(0)} = 0$, and (ii) the symmetrization error becomes vanishingly small for large CI expansions. Consequently, the symmetrization error can be safely neglected in practice. Our preliminary tests have corroborated these theoretical justifications. Also, it can be further estimated via second-order perturbation theory. However, it requires the energies and coefficients of the entire internal space which is only possible for relatively small CI expansions.

The energies of the first $N_{\rm st}$ states, $\boldsymbol{E} = (E_1, \ldots, E_{N_{st}})$, are obtained by a Davidson diagonalization of the multi-state effective Hamiltonian $\boldsymbol{H}^{\rm eff} = \boldsymbol{H}^{(0)} + \tilde{\boldsymbol{\Delta}}^{\rm sBk}$. Similar to the state-specific case, technically, one is able to store the vectors $\boldsymbol{\delta}^{\rm sBk}$ and $\boldsymbol{c}^{(0)}$, but $\tilde{\boldsymbol{\Delta}}^{\rm sBk}$ (or $\boldsymbol{\Delta}^{\rm sBk}$) is potentially too large to be stored in memory. Luckily, compared to a standard CI calculation, the Davidson diagonalization procedure only requires, at each iteration, the extra knowledge of

$$\tilde{\boldsymbol{\Delta}}^{\rm sBk}\boldsymbol{U} = \left(\boldsymbol{c}^{(0)\,\dagger}\boldsymbol{\delta}^{\rm sBk}\,\boldsymbol{U} + \boldsymbol{\delta}^{\rm sBk\,\dagger}\boldsymbol{c}^{(0)}\boldsymbol{U}\right)/2, \tag{16}$$

where $\boldsymbol{U}$ is a $N_{\rm det} \times N_{\rm dav}$ matrix gathering the $N_{\rm dav}$ vectors considered in the Davidson diagonalization algorithm at a given iteration (with $N_{\rm st} \leq N_{\rm dav} \ll N_{\rm det}$). Thanks to Eq. (14), this term can be efficiently evaluated in a $\mathcal{O}(N_{\rm det})$ computational cost and storage via two successive matrix multiplications, for instance,

$$\boldsymbol{c}^{(0)\,\dagger}\boldsymbol{\delta}^{\rm sBk}\,\boldsymbol{U} = \left[\boldsymbol{c}^{(0)} \times \left({}^{\dagger}\boldsymbol{\delta}^{\rm sBk} \times \boldsymbol{U}\right)\right].$$

A pseudo-code of our iterative multi-state dressing algorithm is presented in the supplementary material. For $N_{\rm st} = 1$, the present multi-state algorithm reduces to the state-specific version.

III. HYBRID STOCHASTIC/DETERMINISTIC DRESSINGS

In Ref. 17, we proposed to express

$$E^{(2)} = \sum_{I=1}^{N_{\rm det}} E_{[I]}^{(2)} \tag{17}$$

as a sum of $N_{\rm det}$ contributions $E_{[I]}^{(2)}$, each of them associated with a determinant of the model space, and to compute it efficiently via a Monte Carlo (MC) algorithm. Thanks to the relatively small size of the MC space ($N_{\rm det}$), one is able to store each single contribution. Hence, during the MC simulation, if the contribution of a determinant is required and has never been computed previously, it is computed and stored. Otherwise, the value is retrieved from memory. This technique, known as *memoization*, drastically accelerates the MC calculation as each contribution needs to be computed only once. Moreover, we decompose the energy into a deterministic part and a stochastic part, making the deterministic part grow along the calculation until one reaches the desired accuracy. If desired, the calculation can be carried on until the stochastic part entirely vanishes. In that case, the exact result is obtained with no error bar and no noticeable computational overhead compared to the fully deterministic calculation. To summarize, this algorithm allows us to compute a truncated sum with no bias, but with a statistical error bar instead.

This algorithm is very general and is not limited to the calculation of $E^{(2)}$. Similar to Eq. (17), we express the dressing matrix (14) as the sum of dressing matrices

$$\boldsymbol{\Delta}^{\rm sBk} = \sum_{I=1}^{N_{\rm det}} \boldsymbol{\Delta}_{[I]}^{\rm sBk}. \tag{18}$$

Because the matrices $\boldsymbol{\Delta}_{[I]}^{\rm sBk}$ are too large to fit in memory, we sample the vectors $\boldsymbol{\delta}_{[I]}^{\rm sBk}$ [see Eq. (15b)], which are required for the Davidson diagonalization. During the sampling, one can monitor the "dressed" energy as

$$E_k = \langle\Psi_k^{(0)}|\boldsymbol{H}_k^{\rm eff}|\Psi_k^{(0)}\rangle = E_k^{(0)} + {}^{\dagger}\boldsymbol{c}^{(0)}\langle\boldsymbol{\delta}^{\rm sBk}\rangle, \tag{19}$$

as well as its accuracy by computing the corresponding statistical error. In Sec. IV, all sBk calculations have been carried on until the statistical error is below 10^{-5} a.u. Let us emphasize once again that the primary purpose of the present MC algorithm is to accelerate the computation of the dressing matrix. The same results would have been obtained via its deterministic version.

IV. ILLUSTRATIVE CALCULATIONS

Unless otherwise stated, all the calculations presented here have been performed with the electronic structure software QUANTUM PACKAGE,[47] developed in our group and freely available. The sCI wave functions are generated with the CIPSI algorithm, as described in Refs. 1 and 3 in the frozen-core approximation. The extrapolated FCI results, labeled as exFCI, have been obtained via the method recently proposed by Holmes, Umrigar, and Sharma[9] in the context of the heat-bath method.[7–9] This method has been shown to be robust even for challenging chemical situations,[10–13] and we refer the interested readers to Ref. 11 for additional details.

A. State-specific example

To illustrate the improvement brought by the shifted-Bk approach in its state-specific version (see Sec. II A), we have computed the total electronic energy of the $^2\Pi_g$ ground state of $CuCl_2$ with the 6-31G basis set. The geometry has been taken from Ref. 2 where additional information can be found on this system. For this particular example, we have chosen a small basis set in order to be able to easily reach the FCI limit. A larger basis set will be considered in the next (multi-state) example (see Sec. IV B). The molecular orbitals have been obtained at the restricted open-shell Hartree-Fock (ROHF) level, and the 15 lowest doubly occupied orbitals have been frozen. This corresponds to a sCI calculation of 33 electrons in 38 orbitals. sCI-PT2 stands for a sCI calculation where we have added to the (zeroth-order) variational energy $E^{(0)}$ defined in Eq. (2) the value of the second-order correction $E^{(2)}$ given

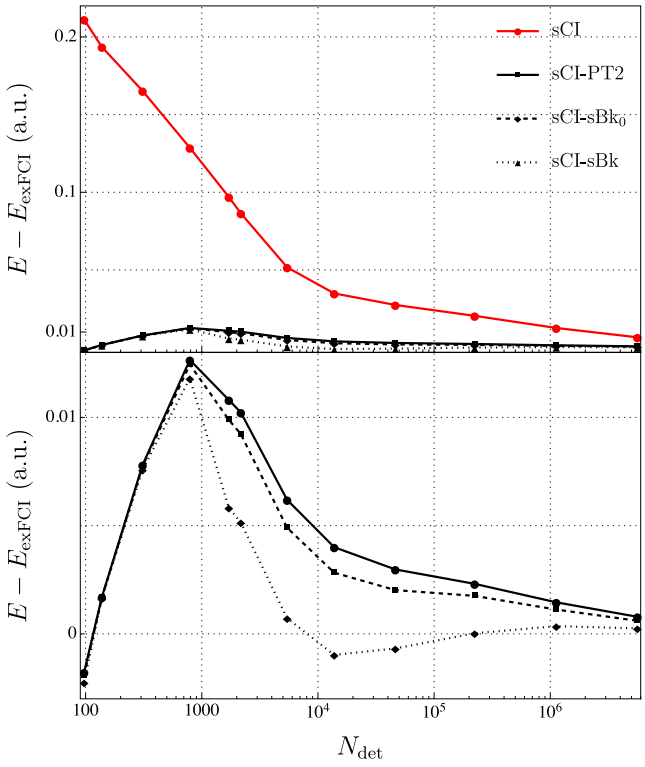


FIG. 1. Deviation from the extrapolated FCI energy E_{exFCI} of the total energy E of $CuCl_2$ (in hartree) as a function of the number of determinants N_{det} in the sCI wave function for various methods.

by Eq. (4). The one-shot non-iterative shifted-Bk procedure will be labeled as sCI-sBk$_0$, while its self-consistent version is simply labeled as sCI-sBk.

Figure 1 shows the convergence of the total energy of $CuCl_2$ as a function of the number of determinants N_{det} in the sCI wave function for the variational sCI results, as well as sCI-PT2, sCI-sBk$_0$, and sCI-sBk. The corresponding numerical values are reported in Table I. As expected, the sCI-PT2, sCI-sBk$_0$, and sCI-sBk energies are not variational as perturbative energies and energies obtained by projection are not guaranteed to be an upper bound of the FCI energy. Nonetheless, all of these corrections drastically improve the rate of convergence compared to the variational sCI results (note the logarithmic scale in Fig. 1). As shown in the bottom graph of Fig. 1, for small values of N_{det}, the three methods yield very similar total energies. However, for $N_{\text{det}} \gtrsim 10^3$, results start to deviate due to the inclusion of an important configuration corresponding to a ligand-to-metal charge transfer (LMCT) state.[48] This LMCT configuration induces a strong revision of the model space wave function $\Psi^{(0)}$. Because the LMCT configuration corresponds to a singly excited determinant with respect to the ROHF determinant, it is not included in the CIPSI expansion for small N_{det} values as it does not directly interact with the ROHF reference ($a^2b \rightarrow ab^2$ excitation for which Brillouin's theorem does apply[49]). Therefore, the double excitations which are strongly coupled with the ROHF configuration are first selected by the CIPSI algorithm. Then, the LMCT configuration is included via its connection with the doubles. In particular, the double excitations corresponding to a single excitation on top of the LMCT configuration have been found to strongly interact with it. The key observation here is that the sCI-sBk energy converges much faster to the FCI limit than the sCI-PT2 energy. Moreover, the significant difference between sCI-sBk and sCI-sBk$_0$ highlights the importance of the revision of the internal wave function brought by the self-consistent nature of the shifted-Bk method.

Table I also reports the overlap of the sCI and sCI-sBk wave functions with respect to the largest sCI wave function obtained for $N_{\text{det}} = 26\,493\,179$. These results also highlight the faster convergence of sCI-sBk and illustrate that the shifted-Bk

TABLE I. Deviation (in millihartree) from the extrapolated FCI energy ($E_{\text{exFCI}} = -2558.006\,880$ a.u.) for various methods as a function of the number of determinants N_{det} in the CIPSI expansion for the $CuCl_2$ molecule and the 6-31G basis set. The second-order correction $E^{(2)}$ is also reported. The error bar corresponding to one standard deviation is reported in parentheses. The exFCI energy has been obtained via a linear extrapolation using the energies of the two largest wave functions (see the supplementary material). The two rightmost columns report the overlap with respect to the largest sCI wave function.

		ΔE			Overlap	
N_{det}	$E^{(2)}$	sCI-PT2	sCI-sBk$_0$	sCI-sBk	sCI	sCI-sBk
97	−213.039(0)	−1.778(0)	−1.93(0)	−2.25(0)	0.9275	0.9275
138	−191.914(0)	+1.698(0)	+1.68(0)	+1.65(0)	0.9295	0.9295
309	−157.491(0)	+7.799(0)	+7.74(0)	+7.59(0)	0.9345	0.9345
789	−116.025(0)	+12.654(0)	+12.45(0)	+11.81(0)	0.9438	0.9447
1 708	−86.208(2)	+10.807(2)	+9.89(0)	+5.83(0)	0.9579	0.9671
2 167	−76.249(8)	+10.232(8)	+9.23(1)	+5.15(1)	0.9610	0.9700
5 428	−45.49(3)	+6.19(3)	+4.90(3)	+0.72(3)	0.9777	0.9854
13 803	−30.87(9)	+4.00(9)	+2.83(9)	−0.97(9)	0.9853	0.9912
46 327	−24.48(9)	+2.98(9)	+2.02(9)	−0.68(9)	0.9913	0.9952
223 089	−18.13(9)	+2.31(9)	+1.76(9)	+0.03(9)	0.9956	0.9975
1 125 547	−11.18(9)	+1.46(9)	+1.12(9)	+0.36(9)	0.9984	0.9990
5 615 264	−5.84(2)	+0.79(2)	+0.61(2)	+0.26(2)	0.9996	0.9997
26 493 179	−3.34(2)	+0.45(2)	...	...	1.0000	...

method could potentially provide better quality trial wave functions for quantum Monte Carlo.[1,2,5,6,11,13]

Although $\Psi_k^{(0)}$ may be an eigenfunction of $\widehat{S}^2$, the way $\Psi_k^{(1)}$ is built does not enforce this property. The expectation value of $\widehat{S}^2$ can be monitored by

$$\langle \Psi_k^{(0)} | \widehat{S}^2 | \Psi_k^{(1)} \rangle = {}^{\dagger}\boldsymbol{c}_k^{(0)}\, {}^{\dagger}(\boldsymbol{s}^2)\, \boldsymbol{c}_k^{(1)}. \tag{20}$$

As expected, the deviation from the eigenvalue is always small, with a maximum deviation of the order of 10^{-4} a.u. in the case of $CuCl_2$.

B. Multi-state example

We have chosen to illustrate the multi-state shifted-Bk algorithm presented in Sec. II B by computing the first singlet transition energy of two cyanine dyes: CN3 ($H_2N{-}CH{=}NH_2^+$) and CN5 ($H_2N{-}CH{=}CH{-}CH{=}NH_2^+$). These types of dyes are known to be particularly challenging for electronic structure methods and especially time-dependent density-functional theory.[50–53] The geometry of CN5 has been extracted from Ref. 51 and we have optimized CN3 at the same level of theory (PBE0/cc-pVQZ). Here, we use Dunning's aug-cc-pVDZ basis set which has been shown to be flexible enough to quantitatively model such transitions thanks to the weak basis dependency of this valence $\pi \rightarrow \pi^\star$ transition.[12,50] In order to treat the two singlet electronic states on equal footing, a common set of determinants is used for both states. In addition, state-averaged complete active space self-consistent field [CASSCF(2,2)] molecular orbitals, obtained with the GAMESS package,[54] are employed.

The difficulty of accurately modeling this vertical transition lies in the strong coupling between the σ and π spaces. To assess this peculiar effect, we have performed several calculations and our results are gathered in Table II. (The corresponding total energies can be found in the supplementary material.) For comparison purposes, Table II also reports reference calculations extracted from Ref. 50. First, we have performed CAS-CI calculations taking into account only the set of molecular orbitals with π symmetry. We refer to these calculations as CAS(π). For CN3 and CN5, there are, respectively, 4 and 6 electrons as well as 32 and 50 orbitals in the CAS(π) space. This results in multideterminant wave functions containing 11 296 and 670 630 determinants, respectively. To quantify the strong coupling between the σ and π space, we have also computed full-valence exFCI energies [denoted as exFCI($\sigma + \pi$)].[11,12] These values fit nicely with the exCC3($\sigma + \pi$) benchmark values reported by Send *et al.*,[50] in agreement with our previous study which shows that, at least for compact compounds, CC3 and exFCI yield similar excitation energies.[12]

The difference between CAS(π) and exFCI($\sigma + \pi$) is of the order of half an eV (slightly less for CN5), showing that the relaxation of the σ orbitals plays a central role here, this effect becoming less pronounced when the number of carbon atoms increases. Note that our CAS(π) excitation energies are extremely close to the CASSCF results reported in Table II. The diffusion Monte Carlo (DMC) estimates of Send *et al.*[50] are probably off by 0.2 eV due to the lack of direct $\sigma - \pi$ coupling in the active space, which is only partially recovered by the Jastrow factor and the orbital optimization.

TABLE II. Vertical excitation energy (in eV) of cyanines for various methods. The error bar corresponding to one standard deviation is reported in parentheses.

Method	CN3	CN5	References
CAS(π)[a]	7.62	5.27	This work
CAS(π) + PT2	7.43	5.02	This work
CAS(π) + sBk$_0$	7.40	4.98	This work
CAS(π) + sBk	7.17	4.77	This work
exFCI($\sigma + \pi$)[b]	7.17	4.89	This work
CASSCF(π)[c]	7.59	5.25	50
CASPT2(π)[d]	7.26	4.74	50
CC3($\sigma + \pi$)[e]	7.27	4.89	50
DMC[f]	7.38(2)	5.03(2)	50
exCC3($\sigma + \pi$)[g]	7.16	4.84	50

[a]CAS-CI/aug-cc-pVDZ calculations: CAS(4,32) and CAS(6,50) for CN3 and CN5, respectively.
[b]Extrapolated CIPSI/aug-cc-pVDZ calculations (see the supplementary material).
[c]CASSCF/ANO-L-VDZP calculations with optimal active spaces: CAS(4,6) and CAS(6,10) for CN3 and CN5, respectively.
[d]CASPT2/ANO-L-VDZP calculations with the standard IPEA Hamiltonian and optimal active spaces: CAS(4,6) and CAS(6,10) for CN3 and CN5, respectively.
[e]CC3/ANO-L-VDZP excitation energies.
[f]Diffusion Monte Carlo results based on optimal active space CASSCF trial wave functions obtained using the T′+ basis set and a Jastrow factor including electron-nuclear and electron-electron terms.
[g]Extrapolated CC3 excitation energies obtained by adding the difference between the CC3/ANO-L-VDZP and CC2/ANO-L-VDZP values to the CC2/ANO-L-VTZP results.

In CAS(π) + PT2, the second-order correction $E^{(2)}$, computed by taking into account all the determinants from the FCI space connected to the CAS(π) reference space, is added to the CAS(π) result. This correction goes in the right direction and recovers 0.19 and 0.25 eV for CN3 and CN5, respectively, bringing the excitation energies within 0.25 and 0.13 eV to the exFCI($\sigma + \pi$) values.

Similarly, CAS(π) + sBk$_0$ and CAS(π) + sBk correspond to sBk and sBk$_0$ calculations where the CAS(π) model space is renormalized by the effect of the perturbers. Like in the case of $CuCl_2$, CAS(π) + sBk$_0$ recovers slightly more than CAS(π) + PT2, while CAS(π) + sBk is spot on for CN3 and overshoots slightly the exFCI($\sigma + \pi$) values for CN5 with an error of 0.12 eV. These results show that the shifted-Bk method associated with a CIPSI-like sCI algorithm is able to recover a large fraction of the missing correlation energy, even with relatively small model spaces.

SUPPLEMENTARY MATERIAL

See supplementary material for the pseudo-code of the multi-state algorithm, total energies associated with Table II, and exFCI extrapolations.

ACKNOWLEDGMENTS

The authors would like to thank Jean-Paul Malrieu for stimulating discussions and the anonymous referees for valuable comments and suggestions. This work was performed using HPC resources from CALMIP (Toulouse) under

allocations Nos. 2018-0510 and 2018-18005 and from GENCI-TGCC (Grant No. 2018-A0040801738).

[1]E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[2]M. Caffarel, E. Giner, A. Scemama, and A. Ramírez-Solís, J. Chem. Theory Comput. **10**, 5286 (2014).
[3]E. Giner, A. Scemama, and M. Caffarel, J. Chem. Phys. **142**, 044115 (2015).
[4]M. Dash, S. Moroni, A. Scemama, and C. Filippi, "Perturbatively Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo," J. Chem. Theory Comput. (published online).
[5]M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, J. Chem. Phys. **144**, 151103 (2016).
[6]M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Using CIPSI nodes in diffusion Monte Carlo," in *Recent Progress in Quantum Monte Carlo* (ACS Publications, 2016), Chap. 2, pp. 15–46.
[7]A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).
[8]S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).
[9]A. A. Holmes, C. J. Umrigar, and S. Sharma, J. Chem. Phys. **147**, 164111 (2017).
[10]A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, J. Phys. Chem. A **122**, 2714 (2018).
[11]A. Scemama, Y. Garniron, M. Caffarel, and P. F. Loos, J. Chem. Theory Comput. **14**, 1395 (2018).
[12]P. F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, "A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks," J. Chem. Theory Comput. (published online).
[13]A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P. F. Loos, J. Chem. Phys. **149**, 034108 (2018).
[14]C. F. Bender and E. R. Davidson, Phys. Rev. **183**, 23 (1969).
[15]J. L. Whitten and M. Hackmeyer, J. Chem. Phys. **51**, 5584 (1969).
[16]B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).
[17]Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, J. Chem. Phys. **147**, 034101 (2017).
[18]N. S. Blunt, J. Chem. Phys. **148**, 221101 (2018).
[19]G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[20]F. A. Evangelista, J. Chem. Phys. **140**, 124114 (2014).
[21]J. B. Schriber and F. A. Evangelista, J. Chem. Phys. **144**, 161106 (2016).
[22]N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, J. Chem. Phys. **145**, 044112 (2016).
[23]W. Liu and M. R. Hoffmann, J. Chem. Theory Comput. **12**, 1169 (2016).
[24]R. Cimiraglia, Int. J. Quantum Chem. **60**, 167 (1996).
[25]J. P. Malrieu, P. Durand, and J. P. Daudey, J. Phys. A: Math. Gen. **18**, 809 (1985).
[26]E. Giner, C. Angeli, Y. Garniron, A. Scemama, and J.-P. Malrieu, J. Chem. Phys. **146**, 224108 (2017).
[27]S. Pathak, L. Lang, and F. Neese, J. Chem. Phys. **147**, 234109 (2017).
[28]Z. Gershgorn and I. Shavitt, Int. J. Quantum Chem. **2**, 751 (1968).
[29]L. E. Nitzsche and E. R. Davidson, J. Chem. Phys. **68**, 3103 (1978).
[30]L. E. Nitzsche and E. R. Davidson, J. Am. Chem. Soc. **100**, 7201 (1978).
[31]E. R. Davidson, L. E. McMurchie, and S. J. Day, J. Chem. Phys. **74**, 5491 (1981).
[32]D. C. Rawlings and E. R. Davidson, Chem. Phys. Lett. **98**, 424 (1983).
[33]D. C. Rawlings, E. R. Davidson, and M. Gouterman, Int. J. Quantum Chem. **26**, 237 (1984).
[34]P. Kozlowski and E. Davidson, Chem. Phys. Lett. **226**, 440 (1994).
[35]P. M. Kozlowski and E. R. Davidson, J. Chem. Phys. **100**, 3672 (1994).
[36]P. M. Kozlowski and E. R. Davidson, Chem. Phys. Lett. **222**, 615 (1994).
[37]P. M. Kozlowski, M. Dupuis, and E. R. Davidson, J. Am. Chem. Soc. **117**, 774 (1995).
[38]V. N. Staroverov and E. R. Davidson, Chem. Phys. Lett. **296**, 435 (1998).
[39]H. Nakano, J. Chem. Phys. **99**, 7983 (1993).
[40]H. Nakano, J. Nakatani, and K. Hirao, J. Chem. Phys. **114**, 1133 (2001).
[41]B. Kirtman, J. Chem. Phys. **75**, 798 (1981).
[42]G. Li Manni, D. Ma, F. Aquilante, J. Olsen, and L. Gagliardi, J. Chem. Theory Comput. **9**, 3375 (2013).
[43]P.-O. Löwdin, J. Chem. Phys. **19**, 1396 (1951).
[44]P. Claverie, S. Diner, and J. P. Malrieu, Int. J. Quantum Chem. **1**, 751 (1967).
[45]R. J. Buenker and S. D. Peyerimhoff, Theor. Chim. Acta **39**, 217 (1975).
[46]E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[47]A. Scemama, T. Applencourt, Y. Garniron, E. Giner, G. David, and M. Caffarel, Quantum package v1.0, https://github.com/LCPQ/quantum_package, 2016.
[48]E. Giner and C. Angeli, J. Chem. Phys. **143**, 124305 (2015).
[49]D. Feller and E. R. Davidson, J. Chem. Phys. **80**, 1006 (1984).
[50]R. Send, O. Valsson, and C. Filippi, J. Chem. Theory Comput. **7**, 444 (2011).
[51]D. Jacquemin, Y. Zhao, R. Valero, C. Adamo, I. Ciofini, and D. G. Truhlar, J. Chem. Theory Comput. **8**, 1255 (2012).
[52]P. Boulanger, D. Jacquemin, I. Duchemin, and X. Blase, J. Chem. Theory Comput. **10**, 1212 (2014).
[53]B. Le Guennic and D. Jacquemin, Acc. Chem. Res. **48**, 530 (2015).
[54]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su *et al.*, J. Comput. Chem. **14**, 1347 (1993).

Supplementary material for "Selected configuration interaction dressed by perturbation"

Yann Garniron,[1] Anthony Scemama,[1] Emmanuel Giner,[2] Michel Caffarel,[1] and Pierre-François Loos[1, a)]
[1)] *Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, France*
[2)] *Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, Sorbonne Université, CNRS, Paris, France*

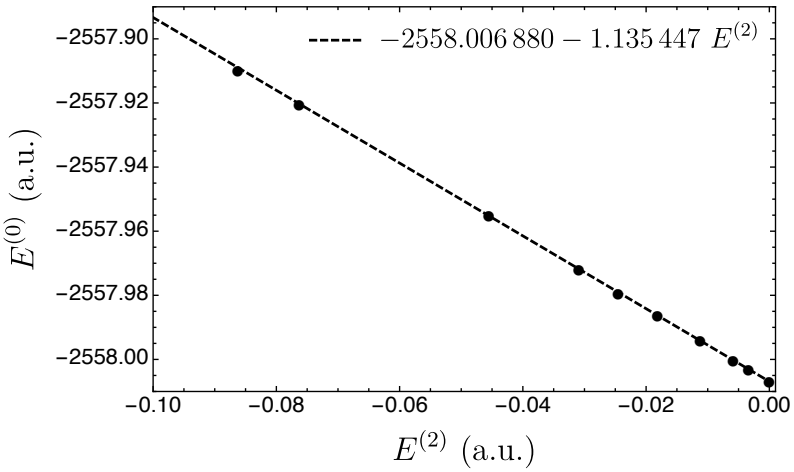


FIG. 1. Extrapolation of the sCI energies to the FCI limit (i.e. $E^{(2)} = 0$) for the ground state of the $CuCl_2$ molecule obtained with the 6-31G basis set. The last two points (corresponding to the the two largest wave functions, that is, having the smallest $E^{(2)}$ values) are taken into account in the linear extrapolation.

TABLE I. Total energies (in hartree) of cyanines for various methods. The error bar corresponding to one standard deviation is reported in parenthesis.

Method	Ground state		Excited state	
	CN3	CN5	CN3	CN5
CAS(π)[a]	−149.535 876	−226.477 375	−149.255 678	−226.283 870
CAS(π)+PT2	−150.050 696(0)	−227.231 239(3)	−149.777 783(0)	−227.046 760(4)
CAS(π)+sBk$_0$	−150.052 063(0)	−227.234 134(2)	−149.780 238(0)	−227.051 134(3)
CAS(π)+sBk	−150.056 70	−227.241 99	−149.793 39	−227.066 59
exFCI[b]	−150.019 253	−227.219 823	−149.755 922	−227.040 256

[a] CAS-CI/aug-cc-pVDZ calculations: CAS(4,32) and CAS(6,50) for CN3 and CN5, respectively.
[b] Extrapolated CIPSI/aug-cc-pVDZ calculations.

[a)]Corresponding author: loos@irsamc.ups-tlse.fr

TABLE II. Zeroth-order energy $E^{(0)}$ and second-order energy $E^{(2)}$ (both in hartree) of the ground and first excited states of CN3 and CN5 as a function of the number of determinants N_{det} in the sCI expansion. The excitation energies (in eV) are also reported. The error bar corresponding to one standard deviation is reported in parenthesis.

Molecule	N_{det}	Ground state		Excited state		Excitation energy (eV)
		$E^{(0)}$	$E^{(2)}$	$E^{(0)}$	$E^{(2)}$	
CN3	1 837	−149.496 568	−0.646 732(0)	−149.198 560	−0.720 103(0)	6.11
	3 654	−149.662 402	−0.386 269(0)	−149.368 197	−0.420 330(2)	7.08
	8 254	−149.746 405	−0.280 72(8)	−149.448 207	−0.318 53(6)	7.09
	19 311	−149.810 865	−0.207 1(2)	−149.516 943	−0.237 31(9)	7.18
	45 730	−149.860 116	−0.154 2(1)	−149.574 193	−0.174 5(2)	7.23
	108 321	−149.897 832	−0.115 52(8)	−149.616 166	−0.131 2(1)	7.24
	265 615	−149.923 376	−0.090 36(7)	−149.647 107	−0.100 27(8)	7.25
	713 756	−149.942 653	−0.071 80(7)	−149.669 387	−0.078 73(7)	7.25
	2 240 887	−149.958 296	−0.056 94(5)	−149.687 113	−0.061 93(6)	7.24
	8 287 086	−149.972 592	−0.043 27(4)	−149.702 834	−0.047 39(5)	7.23
CN5	4 453	−226.404 926	−1.013 276(0)	−226.193 101	−1.088 160(0)	3.73
	8 818	−226.591 687	−0.685 258(5)	−226.372 170	−0.743 445(6)	4.39
	21 356	−226.678 085	−0.565 791(9)	−226.458 189	−0.618 783(9)	4.54
	51 557	−226.751 503	−0.473 218(8)	−226.533 681	−0.519 42(1)	4.67
	124 732	−226.818 047	−0.394 998(1)	−226.599 394	−0.439 147(8)	4.75
	306 926	−226.879 535	−0.326 711(1)	−226.662 318	−0.366 750(1)	4.82
	763 320	−226.937 031	−0.265 417(4)	−226.722 332	−0.301 120(2)	4.87
	1 912 184	−226.988 127	−0.212 58(1)	−226.778 410	−0.242 12(2)	4.90
	4 880 107	−227.030 753	−0.170 4(1)	−226.827 065	−0.193 6(1)	4.91
	13 631 497	−227.063 119	−0.140 20(8)	−226.866 875	−0.155 6(1)	4.92

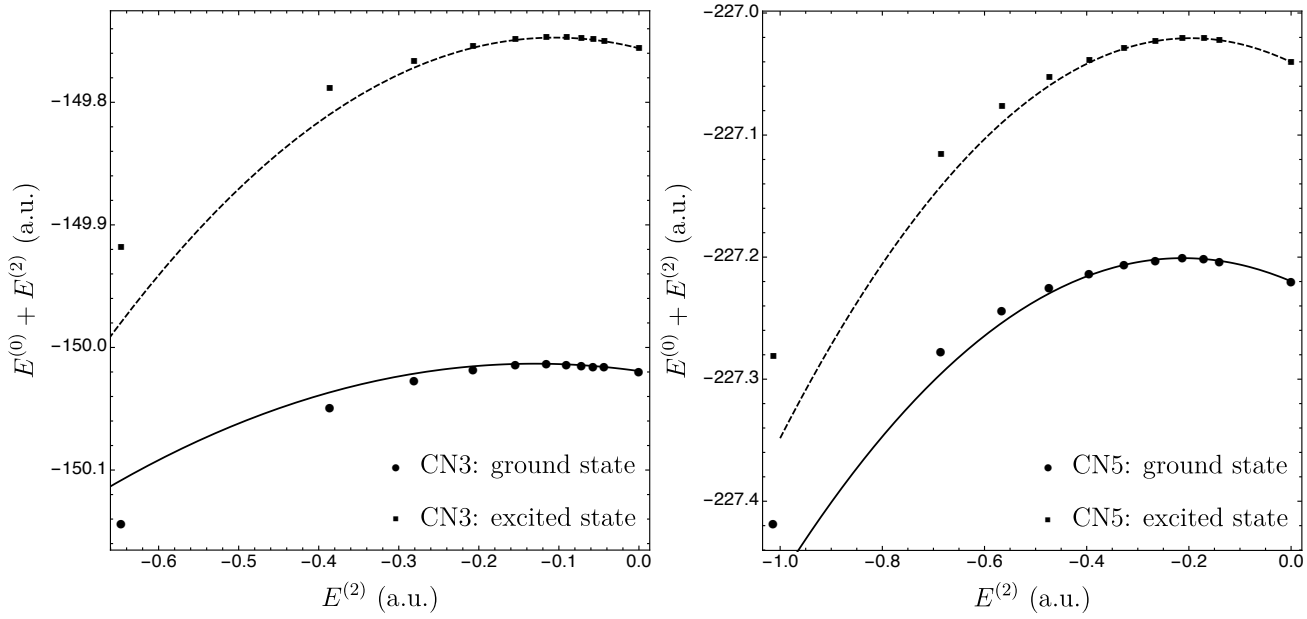


FIG. 2. Extrapolation of the sCI energies to the FCI limit (i.e. $E^{(2)} = 0$) for the ground state and the first singlet excited state of CN3 and CN5 obtained with the aug-cc-pVDZ basis set. The last five points (corresponding to the the five largest wave functions, that is, having the smallest $E^{(2)}$ values) are taken into account in the quadratic extrapolation.

procedure MS_sBk
  Perform CI calculation to get energies $E_k^{(0)}$ and coefficients $\boldsymbol{c}_k^{(0)}$ for $1 \leq k \leq N_{\text{st}}$
  Form $\boldsymbol{E}^{(0)} = \left(E_1^{(0)}, \ldots, E_{N_{\text{st}}}^{(0)}\right)$ and $\boldsymbol{c}^{(0)} = \left[\boldsymbol{c}_1^{(0)}, \ldots, \boldsymbol{c}_{N_{\text{st}}}^{(0)}\right]$
  $n \leftarrow 0$; $\boldsymbol{E}(n) \leftarrow \boldsymbol{E}^{(0)}$; $\Delta\boldsymbol{E} \leftarrow \infty$
  while $\max_k |\Delta E_k| > \tau$ **do** ▷ sBk iterations
    Build $\boldsymbol{\delta}^{\text{sBk}}$ using Eq. (15b) ▷ $[N_{\text{det}} \times N_{\text{st}}]$
    $\boldsymbol{U} \leftarrow$ guess vectors ▷ $[N_{\text{det}} \times N_{\text{dav}}]$
    for $k = 1, \ldots, N_{\text{st}}$ **do**
      $\boldsymbol{U}_k \leftarrow \boldsymbol{c}^{(0)}$
    end for
    $\boldsymbol{R} \leftarrow \infty$
    while $\max_k \|R_k\| > \tau'$ **do** ▷ Davidson iterations
      Orthonormalize $\boldsymbol{U}$
      $\boldsymbol{W} \leftarrow \boldsymbol{H}.\boldsymbol{U}$ ▷ $[N_{\text{det}} \times N_{\text{dav}}]$
      $\boldsymbol{T} \leftarrow {}^{\dagger}\boldsymbol{c}^{(0)}.\boldsymbol{U}$ ▷ $[N_{\text{st}} \times N_{\text{dav}}]$
      $\boldsymbol{W} \leftarrow \boldsymbol{W} + \frac{1}{2}\,\boldsymbol{\delta}^{\text{sBk}}.\boldsymbol{T}$
      $\boldsymbol{T}' \leftarrow {}^{\dagger}\boldsymbol{\delta}^{\text{sBk}}.\boldsymbol{U}$ ▷ $[N_{\text{st}} \times N_{\text{dav}}]$
      $\boldsymbol{W} \leftarrow \boldsymbol{W} + \frac{1}{2}\,\boldsymbol{c}^{(0)}.\boldsymbol{T}'$
      $\boldsymbol{h} \leftarrow {}^{\dagger}\boldsymbol{U}.\boldsymbol{W}$ ▷ $[N_{\text{dav}} \times N_{\text{dav}}]$
      Diagonalize $\boldsymbol{h}$ to get energies $\boldsymbol{E}$ and eigenvectors $\boldsymbol{y}$
      Compute the residual $\boldsymbol{R}$ ▷ $[N_{\text{det}} \times N_{\text{st}}]$
      Append correction vectors to $\boldsymbol{U}$
      $N_{\text{dav}} \leftarrow N_{\text{dav}} + N_{\text{st}}$
    end while
    $\boldsymbol{y} \leftarrow$ the N_{st} lowest eigenvectors in $\boldsymbol{y}$ ▷ $[N_{\text{det}} \times N_{\text{st}}]$
    $\boldsymbol{c}^{(0)} \leftarrow \boldsymbol{U}.\boldsymbol{y}$
    Compute $\boldsymbol{E}^{(0)}$ via Eq. (2) and set $\boldsymbol{E}(n) \leftarrow \boldsymbol{E}^{(0)}$
    Set $\Delta\boldsymbol{E} = \boldsymbol{E}(n) - \boldsymbol{E}(n-1)$ and $n \leftarrow n+1$
  end while
  return $\boldsymbol{E}$ and $\boldsymbol{c}^{(0)}$
end procedure

FIG. 3. Pseudo-code for the multi-state self-consistent shifted-Bk algorithm. The dimensions of the matrices are given as comments. τ and τ' are user-defined thresholds set as 10^{-5} and 10^{-10} respectively.

Chapter 8

Application of stochastic matrix dressing to MR-CCSD

Contents

8.1 Coupled-cluster approach

While CI methods are common ways to account for electron correlation, they suffer a severe size consistency problem.

The logics of the electronic Many-Body problem has been clarified a long time ago in the situations where the wave function may be generated from a single determinant (or single reference). Perturbative developments, translated in terms of diagrams, led to the formulation of the fundamental linked cluster theorem,[74] and clarified the defects of truncated Configuration Interaction methods. The conditions for a good scaling of the correlation energy and for the strict separability into closed shell fragments were established.

By strict separability (which is less ambiguous than the terms size-extensivity and size consistency) we mean that at the non-interacting limit of an $A \cdots B$ problem the energies are additive, $E_{AB} = E_A + E_B$, and that the amplitudes associated with the single and double excitation operators are the same as those obtained for the isolated

A and B problems. Why this is not the case in CI methods can be easily understood as the absence of some excitations occurring simultaneously on all subsystems.

Consider a supersystem AB made of two non-interacting subsystems A and B, and write its CID wave function (Hartree-Fock determinant and all its double excitations).

$$\Psi^{AB} = \Psi^{AB}_{\text{HF}} + \Psi^{AB}_{D} \tag{8.1}$$

with Ψ^{AB}_{HF} the Hartree-Fock determinant for system AB, and Ψ^{AB}_{D} the sum of all double excitation with respect to Ψ^{AB}_{HF}.

If we now write $\Psi^{A\cdots B}$ the product of the two separate CID wave functions for A and B

$$\Psi^{A\cdots B} = \Psi^{A} \times \Psi^{B} \tag{8.2}$$

$$= \Psi^{A}_{\text{HF}}\Psi^{B}_{\text{HF}} + \Psi^{A}_{\text{HF}}\Psi^{B}_{D} + \Psi^{A}_{D}\Psi^{B}_{\text{HF}} + \Psi^{A}_{D}\Psi^{B}_{D} \tag{8.3}$$

As can be seen, Ψ^{AB} isn't described as the product of Ψ^{A} and Ψ^{B} as it should, since simultaneous double excitations on A and B cannot be accounted for.

Some methods aim at partially or fully correcting this size-consistency error by eliminating the unlinked effects of the CISD: the so-called *Davidson corrections*, that are essentially correction to the energy[75, 76], and the so-called *Coupled Electron Pair Approximations (CEPA)* that correct the CI equations[77, 78, 79, 80, 81, 82]. Many different variations have been proposed, for a review see [83].

An alternative to CI approaches is the so-called Coupled-Cluster (CC) approach, that does not suffer this problem. The wave function is given an exponential structure

$$|\Psi\rangle = e^{\hat{T}} |\text{HF}\rangle \tag{8.4}$$

with $\hat{T}$ a so-called *cluster operator*. In the widely used *Coupled Cluster Single and Double (CCSD)* method, the cluster operators is

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \tag{8.5}$$

with $\hat{T}_1$ and $\hat{T}_2$ the one and two orbitals cluster operators

$$\hat{T}_1 = \sum_{pr} t^r_p \hat{T}^r_p \tag{8.6}$$

$$\hat{T}_2 = \sum_{pqrs} t^{rs}_{pq} \hat{T}^{rs}_{pq} \tag{8.7}$$

with the indices p, q running on occupied MOs and r, s running on virtual MOs. $\hat{T}^{rs}_{pq}$ is the usual excitation operator, and t^{rs}_{pq} the so-called *amplitude* associated with it.

Again considering a system made of two infinitely non-interacting sub-systems A and B, thanks to the exponential expression, if one uses local orbitals — that is, the

reference determinant $|\mathrm{HF}\rangle$ is factorizable in $|\mathrm{HF}_A\rangle$ and $|\mathrm{HF}_B\rangle$ determinants isolated on A and B — the energy and wave function on the isolated fragments will be the same as on the $A \cdots B$ supersystem. Indeed, amplitudes associated with excitations involving orbitals for both fragments will be zero due to the absence of interaction. Therefore $\hat{T}$ can be split in $\hat{T}_A$ and $\hat{T}_B$ the cluster operators involving one fragment.

$$|\Psi\rangle = e^{\hat{T}} |\mathrm{HF}\rangle = |\Psi\rangle = e^{\hat{T}_A + \hat{T}_B} |\mathrm{HF}\rangle = e^{\hat{T}_A} |\mathrm{HF}_A\rangle \, e^{\hat{T}_B} |\mathrm{HF}_B\rangle \tag{8.8}$$

The wave function of the supersystem is the product of the wave functions for the isolated subsystems.

In the article presented in section 8.2, we have proposed a definition for amplitudes for coupled cluster in a multi-reference context. The somewhat "brute force" initial implementation of our *Multi-Reference Coupled Cluster* (MR-CCSD) was re-written as a stochastic matrix dressing, based on our Shifted-B_k implementation.

8.2 Alternative definition of excitation amplitudes in multi-reference state-specific coupled cluster

THE JOURNAL OF CHEMICAL PHYSICS **146**, 154107 (2017)

Alternative definition of excitation amplitudes in multi-reference state-specific coupled cluster

Yann Garniron,[1] Emmanuel Giner,[2,3] Jean-Paul Malrieu,[1] and Anthony Scemama[1,a)]
[1]*Laboratoire de Chimie et Physique Quantiques, CNRS 5626, IRSAMC, Université Paul Sabatier, Toulouse, France*
[2]*Dipartimento di Scienze Chimiche e Farmaceutiche, Universita di Ferrara, Via Fossato di Mortara 17, I-44121 Ferrara, Italy*
[3]*Max Planck Institut for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany*

(Received 17 January 2017; accepted 30 March 2017; published online 19 April 2017)

A central difficulty of state-specific Multi-Reference Coupled Cluster (MR-CC) in the multi-exponential Jeziorski-Monkhorst formalism concerns the definition of the amplitudes of the single and double excitation operators appearing in the exponential wave operators. If the reference space is a complete active space (CAS), the number of these amplitudes is larger than the number of singly and doubly excited determinants on which one may project the eigenequation, and one must impose additional conditions. The present work first defines a state-specific reference-independent operator $\hat{\tilde{T}}^m$ which acting on the CAS component of the wave function $|\Psi_0^m\rangle$ maximizes the overlap between $(1+\hat{\tilde{T}}^m)|\Psi_0^m\rangle$ and the eigenvector of the CAS-SD (Singles and Doubles) Configuration Interaction (CI) matrix $|\Psi^m_{\text{CAS–SD}}\rangle$. This operator may be used to generate approximate coefficients of the triples and quadruples, and a dressing of the CAS-SD CI matrix, according to the intermediate Hamiltonian formalism. The process may be iterated to convergence. As a refinement towards a strict coupled cluster formalism, one may exploit reference-independent amplitudes provided by $(1+\hat{\tilde{T}}^m)|\Psi_0^m\rangle$ to define a reference-dependent operator $\hat{T}^m$ by fitting the eigenvector of the (dressed) CAS-SD CI matrix. The two variants, which are internally uncontracted, give rather similar results. The new MR-CC version has been tested on the ground state potential energy curves of 6 molecules (up to triple-bond breaking) and two excited states. The non-parallelism error with respect to the full-CI curves is of the order of 1 mE_h. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4980034]

I. INTRODUCTION

The single-reference Coupled Cluster (CC) formalism[1–4] is the standard technique in the study of the ground state of closed-shell molecules, i.e., those for which a mean-field treatment provides a reasonable zeroth-order single-determinant wave-function Φ_0. This method incorporates the leading contributions to the correlation energy in a given basis set; it is based on the linked-cluster theorem[5] and is size-extensive since it is free from unlinked contributions. The method generates an approximate wave function under the action of a wave operator $\hat{\Omega}$ acting on the single-determinant reference Φ_0, and assumes an exponential character to the wave operator

$$\Psi = \hat{\Omega}\Phi_0 = e^{\hat{T}}\Phi_0. \tag{1}$$

The most popular version only introduces single and double excitation operators in $\hat{T}$, and is known as the Coupled Cluster Singles and Doubles (CCSD) approximation. It incorporates the fourth-order correction of the quadruply excited determinants. The lacking fourth-order contribution concerns the triply excited determinants, which may be added in a perturbative manner. The CC equations, obtained by projecting the eigenequation on each of the Singles and Doubles (SD), lead to coupled quartic equations. In practice, guess values of the amplitudes of the $\hat{T}_{0\to i}$ operators appearing in the $\hat{T}$ operator may be taken as the coefficients of the singles and doubles $|i\rangle$ in the intermediate normalization of the SD Configuration Interaction (CI) vector. The solution of the CC equations may be obtained by treating the effect of the triples and quadruples as an iterative dressing of the SD CI matrix,[6] according to the Intermediate Effective Hamiltonian (IEH) theory.[7,8] The field of application of this method, which satisfies formal requirements and is numerically efficient, is however limited to the systems and the situations where a single-determinant zeroth-order description is relevant. This is no longer the case when chemical bonds are broken, creating open shells, as occurs in most of the chemical reactions. The magnetic systems generally present several open shells, and the low spin-multiplicity states are inherently of multiple-determinant character. Due to near degeneracies, most of the excited states are not only of multi-determinantal but of multi-configurational character. The conception of a multi-reference (MR) counterpart of the CCSD formalism is highly desirable, and has been the subject of intense research. The most comprehensive review has been given by Bartlett and his colleagues.[9] For formal reasons and in particular to treat correctly the breaking of bonds, the reference space, or model space, is usually taken as a Complete Active Space (CAS), i.e., the Full-CI (FCI) of a well-defined number of electrons (the active electrons) in a well-defined set of orbitals (the active MOs). The other MOs are called

a)Electronic mail: scemama@irsamc.ups-tlse.fr

inactive. Let us label $|I\rangle, |J\rangle, \ldots$ the reference determinants. The determinants $|i\rangle, |j\rangle, \ldots$ which interact with the reference space are obtained under purely inactive or semi-active single and double excitations; they generate the CAS-SD CI space, the diagonalization of which provides a size-inconsistent energy $E^m_{\text{CAS-SD}}$ and the corresponding eigenvector,

$$|\Psi^m_{\text{CAS-SD}}\rangle = |\Psi^m_0\rangle + |\Psi^m_{\text{SD}}\rangle = \sum_{I\in\text{CAS}} C^m_I |I\rangle + \sum_{i\notin\text{CAS}} c^m_i |i\rangle \quad (2)$$

with $\langle\Psi^m_{\text{CAS-SD}}|\Psi^m_{\text{CAS-SD}}\rangle = 1$.

One strategy, which is not very aesthetic since it breaks the symmetry between degenerate reference determinants, but which has given rather satisfactory results, consists in selecting (eventually in an arbitrary manner) a specific single reference and in introducing in the wave operator the multiple excitations which generate the other references (the other determinants of the model space).[10] A similar procedure was proposed by Li and Paldus which uses specific three and four body amplitudes issued from a MR-CISD function.[11] The other strategies consider all the references on an equal footing, and are really multi-reference. Let us call N the number of references, and n the number of SD determinants. If the treatment pretended to provide N eigenvectors simultaneously, one might define the $N \times n$ amplitudes sending from the references to the outer-space determinants, in a unique manner but this state-universal approach is not practicable when the model space is a CAS.

Most of the proposed formalisms are state-specific. In this case one faces the famous multi-parentage problem. This problem is recalled in Section II A. Sufficiency conditions have to be imposed.[12] One solution was proposed by Mukherjee and co-workers, and has been widely tested.[13–15] Another one had been proposed earlier by one of us (JPM) and co-workers.[16] It consists, for a given outer-space determinant, in scaling the amplitudes of the various excitation operators $\hat{T}_{I\to i}$ on the interaction between the outer-space determinant and its parents. A recent work has implemented this second solution of the state-specific Multi-Reference Coupled Cluster (MR-CC) problem and has tested its accuracy and robustness on a series of molecular benchmarks, comparing its results to the full-CI energies.[17] In the text, we will refer to this method as λ-MR-CCSD. The present work proposes an alternative process to define the amplitudes of the excitation operators, and this new method will be called μ-MR-CCSD.

The state-specific MR-CC formalisms are usually based on the Jeziorski-Monkhorst[18] splitting of the wave operator into a sum of operators acting individually on the various references

$$\hat{T}^m = \sum_I \hat{T}^m_I |I\rangle\langle I|. \quad (3)$$

We shall leave in a first time this assumption and define in Section II B a reference-independent operator $\hat{T}$ which acting on the component of the desired state in the model space, $|\Psi^m_0\rangle$, provides a vector as close as possible to the CAS-SD eigenvector. This solution, defining reference-independent amplitudes of the excitations, may be exploited directly to generate approximate values of the coefficients of the triply and quadruply excited determinants, according to the exponential structure of the wave operator. From these coefficients, one may dress the CAS-SD CI matrix, redefine amplitudes, and iterate the process to convergence. This solution, presented in Section II C, is not an MR-CC technique; one may call it an exponential dressing of the CAS-SD CI matrix. Section II D redefines reference-dependent excitation amplitudes from the reference-independent amplitudes by a fitting of the previous amplitudes on the coefficients of the singles and doubles of the (dressed) CAS-SD CI eigenvector. This represents an alternative solution to multi-parentage problem and opens the way to a strict MR-CC formalism. Section III presents a series of numerical tests on the bond breaking of single, double, and triple bonds in ground states of molecules as well as a few tests on excited states. The results are compared to our previous proposal and with Full Configuration Interaction (FCI) results.

II. FORMALISMS

In this section, all the presented formalisms are state-specific. To simplify the notations we will consider that the state superscript m is implicit for the wave functions ($\Psi^m \to \Psi$) and for the excitation operators ($\hat{T}^m \to \hat{T}$).

A. The multi-parentage problem in the Jeziorski-Monkhorst approach

Since one wants to produce a MR-CCSD method, one may start from a preliminary CAS-SD CI calculation which will help to fix guess values of the amplitudes of the excitation operators. Let us call $|I\rangle, |J\rangle, \ldots$ the determinants of the CAS, i.e., the so-called reference vectors, and $|i\rangle, |j\rangle, \ldots$ the singles and doubles which do not belong to the CAS and interact with them. The resulting approximate wave function of the targeted state $|\Psi\rangle$ is written as

$$|\Psi_{\text{CAS–SD}}\rangle = \sum_I C_I |I\rangle + \sum_i c_i |i\rangle. \quad (4)$$

Although this function is not size consistent, one may note that the coefficients on the CAS determinants are no longer those of the CAS-CI: they incorporate the effect of the dynamical correlation on the composition of the CAS component of the wave function.

In CC formalisms the wave operator $\hat{\Omega}$ is assumed to take an exponential form

$$\hat{\Omega} = \exp(\hat{T}) \quad (5)$$

and in our previous MR-CC formalism[17] the Jeziorski-Monkhorst multi-exponential structure of the wave operator was adopted, introducing reference-specific wave operators acting specifically on each reference vector (Eq. (3)). One may exploit the knowledge of the CAS-SD CI eigenvector to determine guess operators $\hat{T}_I$ defined in such a manner that

$$|\Psi_{\text{CAS–SD}}\rangle = \sum_I C_I \hat{T}_I |I\rangle. \quad (6)$$

Each of the $\hat{T}_I$ operators is a sum of single and double excitations $\hat{T}_{I\to i}$ possible on $|I\rangle$, multiplied by an amplitude $t_{I\to i}$,

$$\hat{T}_I = \sum_i t_{I\to i} \hat{T}_{I\to i}. \quad (7)$$

In the single-reference CC, the amplitudes of the excitation operators are obtained by projecting the eigenequation on the singly and doubly excited determinants; the number of unknowns is equal to the number of equations. This is no longer the case in the MR context: projecting the eigenequation on each of the singly or doubly excited vectors $|i\rangle$ is not sufficient to define the amplitudes $t_{I\to i}$ since for many classes of excitation, an outer-space determinant interacts with several references, $|i\rangle = \hat{T}_{I\to i}|I\rangle = \hat{T}_{J\to i}|J\rangle$. The condition

$$C_i = \sum_I t_{I\to i}\, C_I \tag{8}$$

is not sufficient to define the amplitudes, even if one restricts the excitation operators to single and double excitations. Additional constraints have to be introduced to fix the amplitudes, and this is the famous *multi-parentage problem*. The number of amplitudes is larger than the number of outer-space determinants so that one cannot determine directly the guess values of the amplitudes from Eq. (6). Different additional constraints have been proposed. One of them consists in scaling the amplitudes on the Hamiltonian interactions between the references and the outer space determinants,

$$\frac{t_{I\to i}}{t_{J\to i}} = \frac{\langle i|\hat{H}|I\rangle}{\langle i|\hat{H}|J\rangle}. \tag{9}$$

This constraint is expressed as

$$t_{I\to i} = \lambda_i \langle i|\hat{H}|I\rangle, \tag{10}$$

where

$$\lambda_i = \frac{c_i}{\langle i|\hat{H}|\Psi_0\rangle}. \tag{11}$$

This solution has been recently implemented[17] and shown to provide excellent agreements with full-CI results on a series of molecular problems. From now on, we will refer to this method as the λ-MR-CCSD.

When the term $\langle i|\hat{H}|\Psi_0\rangle$ is small, the λ-MR-CCSD presents minor stability problems which may introduce some jitter in the potential energy surfaces. A more important problem of the λ-MRCC is illustrated by considering the case of a non-interacting $A\cdots B$ system with localized MOs on A and B. The 2-hole 2-particle inactive double excitations of the type $\hat{T}_{i_A j_B\to r_A s_B}$ have zero amplitude in the λ-MRCC formalism since the integral $\langle i_A j_B|r_A s_B\rangle = 0$. The coefficient of the determinant $\hat{T}_{i_A j_B\to r_A s_B}|I\rangle$ is not zero but it is equal to the product $t_{i_A\to r_A} t_{j_B\to s_B} c_I$. In Sec. II B, we propose an alternative solution to the multi-parentage problem to define amplitudes which do not suffer from this pathological behavior.

B. Introduction of reference-independent amplitudes

The present method differs from the λ-MR-CCSD in the definition of the amplitudes, introduced in this section. The formalism will leave in the first step the Jeziorski-Monkhorst formulation of the wave operator and will consider the possibility to define a unique state-specific reference-independent operator $\hat{T}$, written as a sum of single and double excitation operators,

$$\hat{T} = \sum_{mnpq} t_{mn\to pq}\, a_p^\dagger a_q^\dagger a_n a_m + \sum_{mp} t_{m\to p}\, a_p^\dagger a_m, \tag{12a}$$

$$= \sum_{mnpq} t_{mn\to pq}\, \hat{T}_{mn\to pq} + \sum_{mp} t_{m\to p}\, \hat{T}_{m\to p}, \tag{12b}$$

where the indices p and q run on the virtual and active MOs and the indices m and n run on the inactive occupied and active MOs, excluding the possible occurrence of 4 active MOs. An operator of this kind (but keeping only the linearly independent combinations of the elementary operators) is used in the internally contracted MR-CC method (ic-MRCC) by Evangelista and Gauss,[19] and by Hanauer and Köhn.[20] A similar and more compact formulation was already suggested by Mahapatra *et al.*[21] Our formalism differs by both the determination of the amplitudes and by the way we use them, as will appear later. The ic-MRCC method determines the amplitudes of the excitations by solving the projected coupled cluster equations, where the amplitudes appear up to quartic terms. Hereafter we exploit the knowledge of the CAS-SD CI eigenvector to determine the guess values of the reference-independent amplitudes. These excitation amplitudes will be used later on to estimate the coefficients of the triples and quadruples, and perform an iterative dressing of the CAS-SD CI matrix introducing the coupling between the singles and doubles with the triples and quadruples.

We propose a criterion to fix the amplitudes $t = \{t_{mn\to pq}, t_{m\to p}\}$. Given the fact that we have at our disposal the CAS-SD wave function, a natural way to solve this overdetermined problem is to minimize the distance between the CAS-SD vector and the vector obtained by applying the $(1+\hat{T})$ operator on the CAS wave function

$$\arg\min_t \|(1+\hat{T})|\Psi_0\rangle - |\Psi_{\text{CAS-SD}}\rangle\| = \arg\min_t \|\hat{T}|\Psi_0\rangle - |\Psi_{\text{SD}}\rangle\|, \tag{13}$$

$\hat{T}|\Psi_0\rangle$ being normalized such that $\|\hat{T}|\Psi_0\rangle\| = \||\Psi_{\text{SD}}\rangle\|$.

To perform the minimization, we build the $N_{\text{SD}} \times N_{\text{t}}$ transformation matrix $A_{i,mn\to pq} = \langle i|\hat{T}_{mn\to pq}|\Psi_0\rangle$ which maps from the outer space of determinants $\{|i\rangle\}$ to the space of excited wave functions $\{\hat{T}_{mn\to pq}|\Psi_0\rangle\}$, and we search for the vector of amplitudes $\mathbf{t}$ which minimizes $\|\mathbf{A}.\mathbf{t}-\mathbf{c}\|$ by solving the normal equations,

$$(\mathbf{A}^\dagger\mathbf{A})\mathbf{t} = \mathbf{A}^\dagger\mathbf{c}. \tag{14}$$

Note that in the single-reference case, $\mathbf{A}$ is a permutation matrix and the CAS-SD wave function is exactly recovered.

The matrix $\mathbf{A}$ is usually so large that the use of standard singular value decomposition (SVD) routines to obtain the least squares solution is prohibitive.

Let us first consider the most numerous 2-hole-2-particle inactive double excitations $\hat{T}_{jk\to rs}$. These excitations consist in creating two holes in the doubly occupied orbitals and two particles in the unoccupied orbitals. For each excitation of this kind, as all the involved orbitals are outside of the active space, the number of determinants originating from this process is equal to the number of determinants in the reference. Moreover, each one of these excited determinants is doubly excited with respect to only one determinant $|I\rangle$ of the reference, and the excitation degree with respect to all other reference determinants is necessarily higher than two. Therefore, all excited determinants created by such a 2-hole-2-particle process have only one parent in the reference, and

the corresponding rows of **A** contain only one non-zero element located in the $jk \rightarrow rs$ column with value $A_{i,jk\rightarrow rs} = C_I$. The condition fixing the amplitude $t_{jk\rightarrow rs}$ is given by

$$\arg\min_{t_{jk\rightarrow rs}} \|\hat{T}_{jk\rightarrow rs}|\Psi_0\rangle t_{jk\rightarrow rs} - |\Psi_{\rm SD}\rangle\| \tag{15}$$

which is obtained by minimizing

$$\min_{t_{jk\rightarrow rs}} \left(\sum_I \left(C_I\, t_{jk\rightarrow rs} - \sum_i c_i \langle i|\hat{T}_{jk\rightarrow rs}|I\rangle \right) \right)^2 \tag{16}$$

using Eq. (14), and this condition turns out to be satisfied using only one non-zero coefficient c_i with

$$t_{jk\rightarrow rs} = \frac{\sum_I C_I\, c_i}{\sum_I C_I^2}. \tag{17}$$

One may notice that this is the weighted average of the ratios between the coefficients of the doubly excited determinants $|i\rangle$ and the coefficient of their unique reference generator,

$$t_{jk\rightarrow rs} = \frac{1}{\sum_I C_I^2} \left(\sum_I C_I^2 \left(\frac{c_i}{C_I} \right) \right). \tag{18}$$

The maximum number of non-zero elements per row of **A** is equal to the number of reference determinants since each excitation operator applied on a reference produces no more than one excited determinant. Hence, for all the remaining active excitations, **A** remains sparse and we solve Eq. (14) using Richardson's iterative procedure[22]

$$\begin{cases} \mathbf{t}_0 = \mathbf{A}^\dagger \mathbf{c}, \\ \mathbf{t}_{n+1} = \mathbf{A}^\dagger \mathbf{c} + \left(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A} \right) \mathbf{t}_n, \end{cases} \tag{19}$$

which may be implemented very efficiently using sparse matrix products.

There are cases where multiple amplitudes applied to different references lead to same determinant: $\hat{T}_{jk\rightarrow rs}|I\rangle = \hat{T}_{lm\rightarrow tv}|J\rangle = |i\rangle$. The linear system is underdetermined, so there are infinitely many possible amplitudes verifying the equations. Among the infinity of possibilities, the SVD picks one particular solution given by $\mathbf{A}^+\mathbf{c}$, where $\mathbf{A}^+$ is the pseudo-inverse of **A**. As this solution minimizes the norm of the amplitude vector,[23] the arbitrariness brought by the null space of **A** is minimized and one obtains the most sensible solution.

C. Evaluation of the coefficients of triples and quadruples and iterative dressing of the CAS-SD CI matrix

This section recalls the procedure described in our previous work.[17] The so-determined excitation operator $\hat{T}$ may be used to generate the approximate values of the coefficients of the triples and quadruples as obtained by the action of $\frac{1}{2}\hat{T}^2$. Actually one may assume, in the spirit of the internally contracted MR-CC methods, that the wave operator $\hat{\Omega}$ generating the correlated wave function Ψ from Ψ_0,

$$\Psi = \hat{\Omega}\Psi_0, \tag{20}$$

has an exponential structure,

$$\hat{\Omega} = \exp(\hat{T}). \tag{21}$$

But this form will be simply used to estimate the coefficients of the triply and quadruply excited determinants $\{|\alpha\rangle\}$, leaving the internally contracted structure of the outer-space. The coefficients of these determinants are estimated as

$$c_\alpha = \frac{1}{2}\langle\alpha|\hat{T}^2|\Psi_0\rangle. \tag{22}$$

All the determinants $\{|\alpha\rangle\}$ are generated by applying all the single and double substitutions on the singles and doubles, and filtering out the determinants which are already in the wave function. For each $|\alpha\rangle$ one searches for the reference determinants $\{|I\rangle\}_\alpha$ which differ by no more than 4 orbital substitutions from $|\alpha\rangle$ (its grand-parents). One then identifies the set of all possible complementary excitations as the products of excitations $\hat{T}_p$ and $\hat{T}_q$, which generate $|\alpha\rangle$ from every member $|I\rangle$ of the set $\{|I\rangle\}_\alpha$, i.e.,

$$\mathcal{S}_\alpha = \left\{ (p,q,I) : \forall |I\rangle \in \{|I\rangle\}_\alpha, \left(\hat{T}_p \hat{T}_q |I\rangle = |\alpha\rangle \right) \right\}. \tag{23}$$

It now straightforward to find the set of singles and doubles with which $|\alpha\rangle$ interacts through the matrix elements $\langle i|\hat{H}|\alpha\rangle$, namely,

$$\{|i\rangle\}_\alpha = \left\{ |i\rangle : \forall (p,\cdot,\cdot) \in \mathcal{S}_\alpha, \left(|i\rangle = \hat{T}_p^\dagger |\alpha\rangle \right) \right\}. \tag{24}$$

For each $|i\rangle$, in the eigenequation

$$\left(\langle i|\hat{H}|i\rangle - E \right) c_i + \sum_J \langle i|\hat{H}|J\rangle C_J + \sum_{j\neq i} \langle i|\hat{H}|j\rangle c_j + \sum_\alpha \langle i|\hat{H}|\alpha\rangle c_\alpha = 0, \tag{25}$$

the coefficient c_α is given by the genealogy of $|\alpha\rangle$,

$$c_\alpha = \sum_{(p,q,I)\in\mathcal{S}_\alpha} (-1)^{n(I\rightarrow\alpha)} t_p t_q C_I, \tag{26}$$

$n(I \rightarrow \alpha)$ being given by the number of permutations needed to go from $|I\rangle$ to $|\alpha\rangle$. One may replace the sum over the α by a dressing of the matrix elements between the determinant $|i\rangle$ and the references which are grand-parents of $|\alpha\rangle$,

$$\langle i|\hat{\Delta}|I\rangle = \sum_\alpha \langle i|\hat{H}|\alpha\rangle \left(\sum_{(p,q,J)\in\mathcal{S}_\alpha:(J=I)} (-1)^{n(I\rightarrow\alpha)} t_p t_q \right) \tag{27}$$

since

$$\sum_I \langle i|\hat{\Delta}|I\rangle C_I = \sum_\alpha \langle i|\hat{H}|\alpha\rangle c_\alpha. \tag{28}$$

The effect of the triples and quadruples is thus incorporated as a change of the columns of the CAS-SD CI matrix corresponding to the interaction between the references and the singles and doubles,

$$\left(\langle i|\hat{H}|i\rangle - E \right) c_i + \sum_J \langle i|(\hat{H}+\hat{\Delta})|J\rangle C_J + \sum_j \langle i|\hat{H}|j\rangle c_j. \tag{29}$$

This type of dressing was already employed in our previous MR-CC implementation.[17] One will find in the same reference the practical procedure to make the dressed matrix Hermitian without any loss of information. Of course the whole process may be iterated. The diagonalization of the dressed CAS-SD CI matrix provides new values of the coefficients, not only of the singles and doubles which no longer suffer from the truncation but also those of the references: *the method is*

fully decontracted. From the new wave function new amplitudes are obtained, a new dressing is defined and the process is repeated till convergence, which is usually rapidly obtained (3-4 iterations).

As opposed to the λ-MR-CCSD method which uses reference-specific amplitudes, the amplitudes introduced in Sec. II B are reference-independent. As a consequence, the formalism is not a strict MR-CC method since we exploit the CAS-SD CI function which slightly differs from the vector resulting from the action of $\hat{T}$ on the vector. Although the distance between these two vectors has been minimized, they are not identical, $(1+\hat{T})|\Psi_0\rangle \neq |\Psi_{\text{CAS–SD}}\rangle$.

Once the $\hat{T}$ operator has been obtained, one might imagine a contracted exponential formalism calculating $\hat{T}^2|\Psi_0\rangle$ and the interaction between $\hat{T}|\Psi_0\rangle$ and $\hat{T}^2|\Psi_0\rangle$, but this calculation requires to return to the determinants. This formalism would remain internally contracted and would be less accurate than the decontracted procedure we propose. Actually in this version, the deviations of the approximate reference-independent amplitudes from optimal ones, those which would generate the exact coefficients of the singles and doubles, only affect the evaluation of the coefficients of the triples and quadruples, and these deviations represent a minor source of error in the correction restoring the size extensivity. This reliability will be illustrated in the numerical tests.

D. State-specific MR-CC variant

In order to return to MR-CC formalism, one may simply exploit the reference-independent amplitudes as an initial guess to define reference-dependent amplitudes. Currently the determinant $|i\rangle$ belonging to the singles and doubles has a coefficient $\tilde{c}_i$ in $\hat{T}|\Psi_0\rangle$,

$$\tilde{c}_i = \langle i|\hat{T}|\Psi_0\rangle = \sum_{\{(p,I):(\hat{T}_p|I\rangle=|i\rangle)\}} t_p\, C_I, \tag{30}$$

which differs from the coefficient c_i in $|\Psi_{\text{SD}}\rangle$. One can define a parameter μ_i, specific of the determinant $|i\rangle$,

$$\mu_i = \frac{c_i}{\tilde{c}_i}, \tag{31}$$

which multiplying with $\tilde{c}_i$ will produce the exact coefficient c_i of $|i\rangle$ in the (dressed) CAS-SD CI eigenvector. So the previous reference-independent amplitudes have now become reference-dependent. The excitation $\hat{T}_p$ which excites $|I\rangle$ to $|i\rangle$ ($|i\rangle = \hat{T}_p|I\rangle$) receives a reference-dependent amplitude

$$t_{I\to i} = t_{p,I} = \mu_i t_p\,. \tag{32}$$

The same excitation will receive a somewhat different amplitude when it acts on another reference $t_{p,J} \neq t_{p,I}$. This version returns to the Jeziorski-Monkhorst formalism as the wave operator again is a sum of reference-specific operators. The so-obtained amplitudes may be exploited to generate the coefficients of the triples and quadruples, and one may follow the same strategy as in our previous formalism, with an iterative column dressing of the interactions between the singles and doubles and the references. In what follows, we will refer to this method as μ-MR-CCSD as it involves the μ_i (Eq. (31)).

As the overlap between $(1+\hat{T})|\Psi_0\rangle$ and $|\Psi_{\text{CAS–SD}}\rangle$ has been maximized, the coefficients $\tilde{c}_i$ and c_i are expected to be very close in particular if c_i is large, and the parameter μ_i should be close to 1, at least for the determinants which contribute significantly to the wave function. In practice we observe this tendency, but the smallest coefficients are sacrificed during the maximization of the overlap and their μ_i can be very far from 1. This introduces some instabilities in the iterations, so we chose to limit the values of μ_i in the $[-\mu_i^{\max}, \mu_i^{\max}]$ range, with

$$\mu_i^{\max} = 2 + 100 \times \exp\left(-20\frac{|c_i|}{\max_j |c_j|}\right). \tag{33}$$

In this way, when $|c_i|$ is large μ_i is constrained in the [−2,2] range, and when $|c_i|$ is small, μ_i is constrained in [−102,102]. The effect on the stability of the iterations is significant, and the effect on the energy differences is not noticeable, as seen in Sec. III.

Our procedure makes use of an (non-compulsory but convenient) approximation, namely, the fact that we have not subtracted the product of the single excitations (the $\hat{T}_1^2$ contributions) from coefficients of the doubles to fix the $\hat{T}_2$ amplitudes. From a perturbation expansion, one sees that this neglect only introduces fifth-order errors on the energy, which are responsible for small deviations from strict additivity of the energies. Notice that a correct treatment of the $\hat{T}_1^n$ operations, although tedious, is perfectly conceivable in our formalism and would insure a perfect MR-CC character.

III. NUMERICAL TESTS

In this section, we first numerically evaluate the errors made by the different approximations. Then, we compare the here-proposed dressed CAS-SD and μ-MR-CCSD to the λ-MR-CCSD presented in Ref. 17 on standard benchmark systems.[13,15,19,20,24–33]

The basis set used is Dunning's cc-pVDZ,[34] and the molecular orbitals were obtained using the CAS-SCF code present in GAMESS.[35] All the following calculations were made using the Quantum Package,[36] an open-source program developed in our group. Full-CI energies were obtained using the CIPSI algorithm.[37–39] In all the calculations (full-CI, CAS-SD, and MR-CC), only the valence electrons are correlated (frozen core approximation).

A. Approximations

1. $\hat{T}_1^2 \times \hat{T}_2$, $\hat{T}_1^3$, and $\hat{T}_1^4$

To estimate the errors due to the approximate treatment of the $\hat{T}_1^2 \times \hat{T}_2$, $\hat{T}_1^3$, and $\hat{T}_1^4$ operators, we chose a single-reference example in which the single excitations are important at the CISD level. In the single reference case, all the excitations are of the 2 hole-2 particle type, so the normal equations (Eq. (14)) are solved exactly and all the values of μ are equal to 1. The only difference to standard CCSD is the approximate treatment of the $\hat{T}_1^3$ and $\hat{T}_1^4$, as explained in Section II D.

We have calculated the energy of the FH molecule at a distance of 1.2 Å with the single-reference CCSD programs of GAMESS[40] and Gaussian 09,[41] and our μ-MR-CCSD implementation using the Hartree-Fock determinant as reference. The results are presented in Table I. The Hartree-Fock and

TABLE I. Comparison of the single-reference energies obtained with Gaussian 09, GAMESS, and the Quantum Package for FH at 1.2 Å, cc-pVDZ. All energies are converged below 10^{-10} a.u.

	Hartree-Fock	CISD	CCSD
Gaussian 09	−99.959 526 039	−100.170 216 059	−100.178 425 609
GAMESS	−99.959 526 065	−100.170 216 086	−100.178 425 629
Quantum package	−99.959 526 065	−100.170 216 097	−100.178 426 538

CISD energies agree up to 10^{-8} a.u., but our implementation differs from the CCSD by almost $\sim -10^{-6}$ a.u. We attribute this difference to the approximation in the $\hat{T}_1^2 \times \hat{T}_2$, $\hat{T}_1^3$, and $\hat{T}_1^4$ operators, and it represents a relative error of 4.4×10^{-6} on the correlation energy.

To measure the effect of this approximation on the size consistency, we have calculated the energy of the CH_3 radical, with C–H bond lengths of 1.103 Å and H–C–H bond angles of 107.69° in the 6-31G basis set. We have also calculated the energy of the dimer with an intermolecular distance of 100 Å. The active space of the monomer contains only the singly occupied orbital, and the dimer is an open shell singlet with a CAS(2,2) wave function. For the μ-MR-CCSD calculation, we have constrained the μ_i as in Eq. (33) or we have let it unconstrained. The results are given in Table II and show that the deviation to additivity of the energy is reduced by an order of magnitude going from the CAS-SD to the dressed CAS-SD, and by two orders of magnitude when including the μ_i factors. The constraint on the μ_i introduces a small error which is below 10^{-4} a.u.

B. Bond breaking

For all the applications we compare the dressed CAS-SD and μ-MR-CCSD with the λ-MR-CCSD and the CAS-SD values. Results are also given using the reference-independent dressing of the CAS-SD CI matrix. All the applications are presented as energy differences with respect to the full-CI energy estimated by a CIPSI calculation with a second-order perturbative correction. The smallest and largest values of the CIPSI perturbative corrections along the curves are given in Table III. We empirically estimate the error to the FCI energy to be in the order of 10% of the largest contribution. For ethane and twisted ethylene, which have the largest perturbative corrections, we have performed a larger CIPSI calculation at the point with the largest PT2 contribution. For ethylene the PT2 contribution was reduced to −5.6 mE_h, but the total energy changed by only 0.08 mE_h. In this case, we can consider that the CIPSI energy of ethylene is converged. In the case of ethylene, the largest calculation dropped the PT2 value to −6.7 mE_h and the total energies differed by 1.4 mE_h. So we estimate that the CIPSI curve of ethane has an accuracy less than 2 mE_h in total energy, and all the other curves have an accuracy below the mE_h. The non-parallelism errors (NPEs) of all the CIPSI curves are estimated with an error below the mE_h.

TABLE III. Second-order perturbative correction in the CIPSI calculations. Minimum and maximum values among all the points of the potential energy curve.

	E_{PT2} (E_h)	
	Smallest	Largest
C_2H_6	-13.4×10^{-3}	-17.8×10^{-3}
C_2H_4 twisted	-3.37×10^{-3}	-9.86×10^{-3}
C_2H_4	-2.41×10^{-3}	-6.85×10^{-3}
F_2 $^3\Sigma_u^+$	-0.31×10^{-3}	-1.42×10^{-3}
F_2	-0.13×10^{-3}	-0.47×10^{-3}
N_2	-61.1×10^{-6}	-0.41×10^{-3}
BeH_2	-12.3×10^{-6}	-35.4×10^{-6}
H_2O	-1.59×10^{-6}	-69.1×10^{-6}
FH	-0.23×10^{-6}	-55.1×10^{-6}
LiF	-0.17×10^{-6}	-12.3×10^{-6}

For the full series of compounds, Figure 1 shows the energy difference with respect to the full-CI along the reaction coordinate. Table IV summarizes the non-parallelism errors (NPEs) and the maximum of the error obtained along the curve. The MR-CC treatment reduces the average and maximum error of the CAS-SD with respect to the full-CI by a factor close to 4. The correction is larger when the system involves an important number of inactive electrons (F_2, C_2H_6) than when this number is small (BeH_2, N_2). One actually knows that the size-consistency error of the CAS-SD treatment increases with the number of inactive electrons; this error disappears in the MRCC treatment, which essentially misses some fourth-order connected effects of the triples.

1. Single-bond breaking

We present here the single bond breaking of the σ bonds of the C_2H_6 and F_2 molecules and of the π bond of ethylene. The active spaces were chosen with two electrons in two MOs, the minimal wavefunctions to describe properly the dissociation of the molecules. In the case of ethane, the NPE of the CAS-SD is 5.1 mE_h, and is reduced to 3.5 mE_h with the μ-MR-CCSD. The curve of the dressed CAS-SD has the lowest NPE (1.3 mE_h). The curves obtained by both MR-CCSD methods give equivalent results, with NPEs of 3.5 and 3.6 mE_h.

TABLE II. Evaluation of the size-consistency error in the dressed CAS-SD and the μ-MR-CCSD, 6-31G basis set.

	CAS-SCF	CAS-SD	Dressed CAS-SD	μ-MR-CCSD $\mu \in$ Eq. (33)	μ-MR-CCSD unconstrained μ
$\cdot CH_3$	−39.528 586 5	−39.622 437 9	−39.625 570 2	−39.625 570 2	−39.625 570 2
$\cdot CH_3$ ×2	−79.057 173 0	−79.244 875 8	−79.251 140 3	−79.251 140 3	−79.251 140 3
$H_3C\cdots CH_3$	−79.057 173 0	−79.237 098 2	−79.250 695 7	−79.251 039 3	−79.251 107 3
Error		7.78×10^{-3}	4.44×10^{-4}	1.01×10^{-4}	3.29×10^{-5}

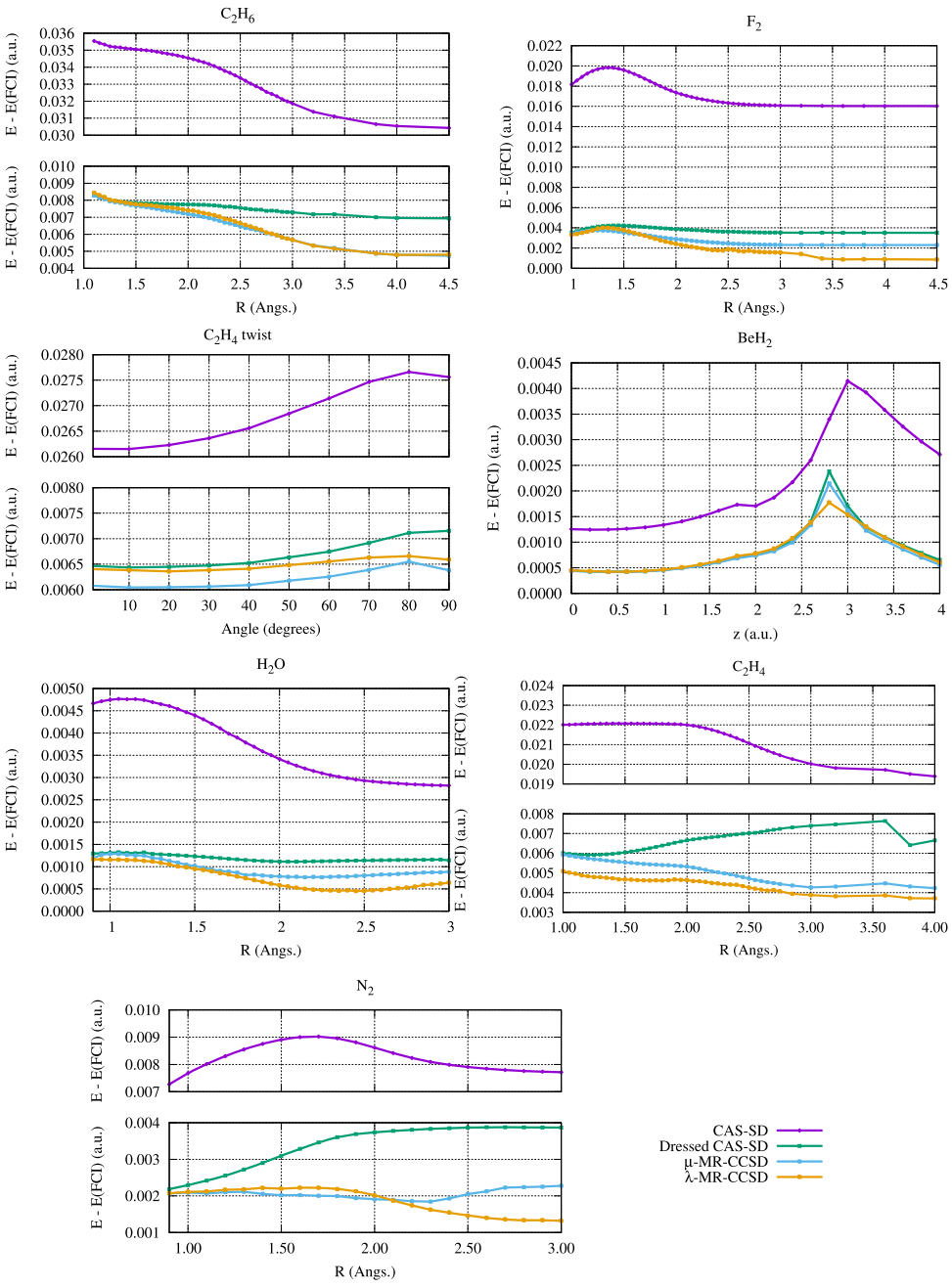


FIG. 1. Dissociation curves. Difference with respect to the full-CI energy using the MR-CCSD method presented in Ref. 17 and with the MR-CCSD method proposed in this work, as well as the CAS-SD and the dressed CAS-SD.

In the case of F_2 the NPE of the dressed CAS-SD is 0.9 mE_h and the NPE of the μ-MR-CCSD is 1.5 mE_h; both are better than the NPE of the λ-MR-CCSD which is 3.1 mE_h. Also, one can remark here some numerical instabilities in the λ-MR-CCSD where the curve is not smooth.

In the next example, the π bond of ethylene is broken by the rotation of the CH_2 fragments. The CAS-SD has an NPE of 1.5 mE_h, and using the dressed CAS-SD reduces the NPE to 0.7 mE_h. The μ-MR-CCSD gives an NPE of 0.5 mE_h, and the NPE obtained with the λ-MR-CCSD is slightly better with a value of 0.3 mE_h.

2. Insertion of Be in H_2

We present the results obtained by the insertion of a beryllium atom into the H_2 molecule, which is a popular benchmark for MR-CC methods. The reference is still a CAS(2,2) for comparison with the literature, even though this choice of reference is not the most appropriate for a correct description of the reaction. The geometries are given by the relation

$$z = 2.54 - 0.46x \quad \text{(a.u.)}, \tag{34}$$

where the beryllium atom is at the origin and the hydrogen atoms are at the coordinates $(x, 0, \pm z)$. In this particular case, the μ-MR-CCSD gives an NPE of 1.7 mE_h which is larger than the NPE of 1.3 mE_h obtained by the λ-MR-CCSD. This is due to only one point of the curve, the maximum which is higher by 0.3 mE_h; all the other points being very close by less than 0.1 mE_h. Here, the dressed CAS-SD and the μ-MR-CCSD are equivalent.

3. Two bond breaking

For breaking two bonds we have used CAS(4,4) wave functions as the reference space. The first example is the simultaneous breaking of the two O–H bonds of the water

TABLE IV. Non-parallelism errors (NPEs) and maximum errors with respect to the full-CI potential energy surface (mE_h).

	CAS-SD		λ-MR-CCSD		Dressed CAS-SD		μ-MR-CCSD	
	NPE	Max error	NPE	Max error	NPE	Max error	NPE	Max error
C_2H_6	5.1	35.5	3.6	8.4	1.3	8.3	3.5	8.3
F_2	3.8	19.8	3.1	4.0	0.9	4.2	1.5	3.8
C_2H_4 twist	1.5	27.7	0.3	6.7	0.7	7.1	0.5	6.5
BeH_2	2.9	4.1	1.3	1.8	2.0	2.4	1.7	2.1
H_2O	1.9	4.8	0.7	1.2	0.2	1.3	0.5	1.3
C_2H_4 stretch	2.7	20.0	1.6	5.2	1.7	6.2	1.8	6.0
N_2	1.7	9.0	0.9	2.2	1.7	3.8	0.3	2.3
F_2 $^3\Sigma_u^+(m_s = 1)$	2.6	18.6	1.3	3.3	1.2	3.5	1.2	3.3
F_2 $^3\Sigma_u^+(m_s = 0)$	2.6	18.6	1.2	1.8	1.3	3.5	1.1	3.3
FH (ground state)	2.6	14.6	1.8	4.0	2.1	4.5	1.8	4.0
FH (excited state)	3.3	20.9	8.8	8.5	10.5	10.1	7.1	8.3
F_2 (local)	3.8	19.8	1.2	3.2	1.5	3.5	0.9	3.0
N_2 (local)	1.7	9.0	3.6	5.0	1.1	3.5	0.4	1.8

molecule by stretching. Here, the CAS-SD exhibits an NPE of 1.9 mE_h which is significantly improved to 0.2 mE_h with the dressed CAS-SD. The μ-MR-CCSD, with an NPE of 0.5 mE_h, is slightly more parallel to the full-CI curve than the λ-MR-CCSD which has an NPE of 0.7 mE_h.

The second example is the double-bond breaking of ethylene by stretching. One should first clarify that the energy differences in the figure do not match those of the torsion along the bond because in the former example the reference was a CAS(2,2), and here it is a CAS(4,4). Dressing the CAS-SD reduces the NPE from 2.7 mE_h to 1.7 mE_h. One can remark a discontinuity in the curve at large distances. The μ-MR-CCSD and λ-MR-CCSD slightly improve the NPE to values of 1.7 mE_h and 1.8 mE_h.

4. Triple-bond breaking

N_2 is the typical benchmark for breaking a triple bond. Here, we have used a CAS(6,6) reference wave function. At the CAS-SD level, the NPE is 1.7 mE_h, and the dressed CAS-SD does not reduce the NPE. Here, it is necessary to use reference-dependant amplitudes to recover a low NPE: 0.9 mE_h with the λ-MR-CCSD, and 0.3 mE_h with the μ-MR-CCSD.

C. Excited states

1. Triplet state of F_2

We report here calculations on the triplet state $^3\Sigma_u^+$ of F_2. The reference wave function was prepared in two different ways, both using restricted open-shell Hartree-Fock molecular orbitals. The first reference wave function labeled $m_s = 1$ is a single open-shell determinant, and the second wave function is the triplet $m_s = 0$, made of two determinants $1/\sqrt{2}(\alpha\beta - \beta\alpha)$.

To ensure that the CAS-SD is a strict eigenfunction of the $\hat{S}^2$ operator, we have included in Ψ_{SD} all the determinants with the same space part as the singles and doubles with respect to the CAS. These determinants are treated in the same way as singles and doubles and are treated variationally in the diagonalizations. Of course, those which are triples or quadruples with respect to Ψ_{ref} are excluded from the set of the $\{\alpha\}$ and have no effect in the dressing.

To reduce the computational cost, the triples and quadruples were not augmented with all the determinants with the same space part. The absence of some determinants gives rise to a slight deviation ($<10^{-6}$ a.u.) of $\langle\hat{S}^2\rangle$ from the desired eigenvalue, and it is expected to have some impact on the iterative dressing. It is worth checking the effect of this deviation from the exact spin multiplicity. The first test concerns the comparison of the $m_s = 0$ and $m_s = 1$ components of a triplet state.

According to Figure 2, in all the cases, the NPE of the CAS-SD (2.6 mE_h) is improved to a value of 1.1–1.3 mE_h. As expected the dressed CASSD and the μ-MR-CCSD are strictly equivalent for $m_s = 1$. Indeed, for both variants, the usual single-reference amplitudes c_i/c_0 are recovered. The amplitudes of the λ-MR-CCSD lower the curve by 1 mE_h when going from $m_s = 1$ to $m_s = 0$. The dressed CAS-SD gives a slightly higher energy by only 0.3 mE_h, and introducing the reference-dependence via the μ_i reduces the difference to 0.2 mE_h.

If one considers the error on the singlet-triplet gap with respect to the full-CI reference, it appears clearly that the μ-MR-CCSD with $m_s = 0$ gives the most accurate results, with errors lying between 0 and 0.9 mE_h along the curve.

2. Avoided crossing in FH and LiF

We have calculated the potential energy surfaces of the two lowest $^1\Sigma^+$ states of FH, using as reference wave function the CAS(2,2) with state-averaged CAS-SCF molecular orbitals in the aug-cc-pVDZ basis set. Figure 3 shows the NPEs of the ground and excited states. In the ground state, the NPE is 1.8 mE_h for both MR-CCSD variants, but the λ-MR-CCSD shows some numerical instabilities, as opposed to the μ-MR-CCSD which gives a very smooth curve.

In the excited state, the situation is different: surprisingly the best NPE is obtained by the CAS-SD. The reason is the

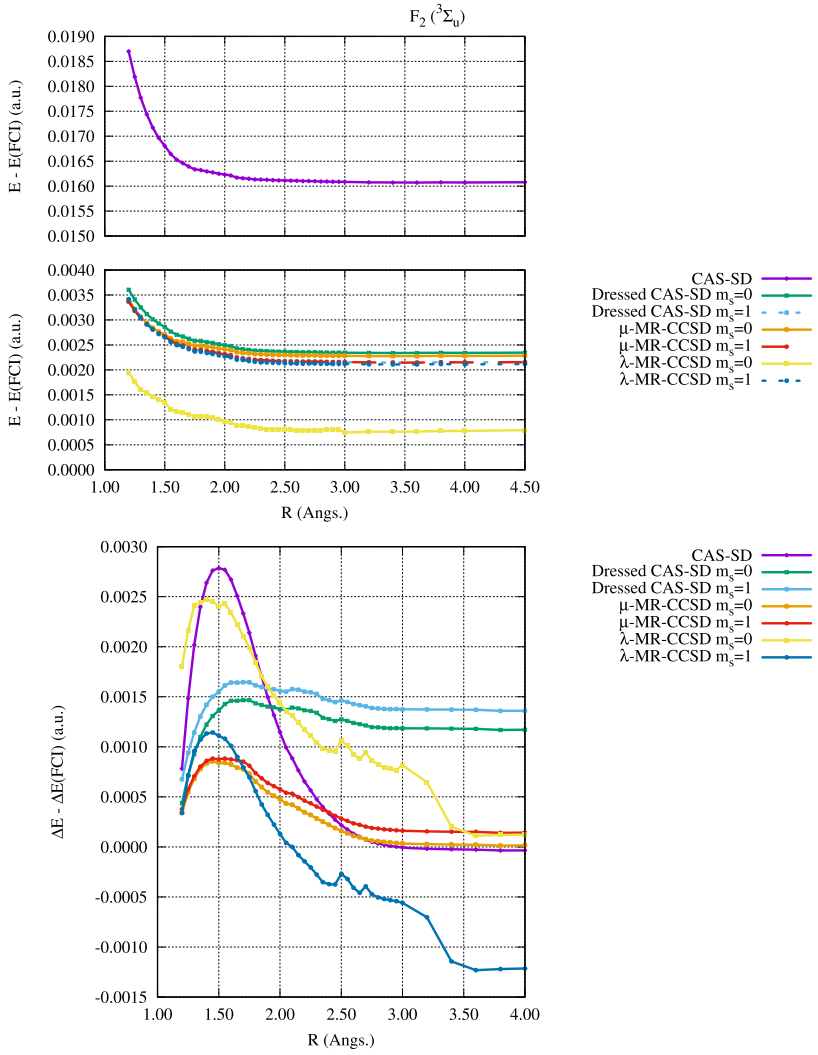


FIG. 2. F_2 $^3\Sigma_u^+$. Difference with respect to the full-CI energy for the $m_s = 0$ and $m_s = 1$ wave functions (top), and error on the singlet-triplet gap $\Delta E = E(^3\Sigma_u^+) - E(^1\Sigma_g^+)$ (bottom). On both graphics, the two curves of the dressed CAS-SD coincide.

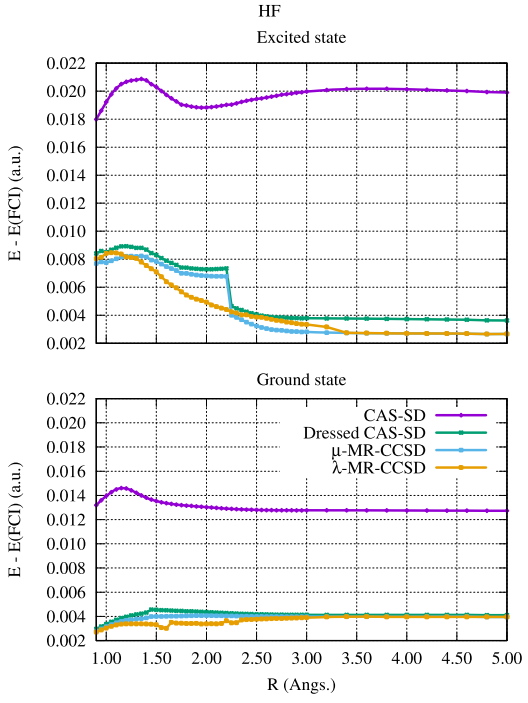


FIG. 3. Difference with respect to the full-CI energy for the two lowest $^1\Sigma^+$ states of FH.

following. At large distances, the CAS-SD description is correct, but the size-consistency error raises the energy. At short distances, the CAS-SD description of the excited state is not as accurate as for large distances since there are determinants with a large coefficient which are not in the active space: the CAS contributes by 0.85 to the norm of the FCI wave function, but in the case of the CAS-SD its contribution is 0.91. This bad description of the CAS-SD raises also the energy at short distances, and the errors compensate along the curve. All the other methods correct the size-consistency error, so the long range errors are corrected but not the error due to the incompleteness of the CAS at short range. This explains why the deviations to the FCI decrease with the distance, and why the NPEs are so large. The two variants of the MR-CCSD agree at short and long distances, but they differ significantly between 1.5 and 2.5 Å; the region of the avoided crossing. To understand these differences, we have plotted in Figure 4 the two eigenvalues of the two state-specific Hamiltonians—one dressed for the ground state and one dressed for the excited state. It appears that the state of interest is very well described, but the dressing for other root has a much lower quality. This strong state-specific character is due to the fitting procedure which is implicitly weighted by the state of interest. The large coefficients have a higher quality in the amplitudes, but the

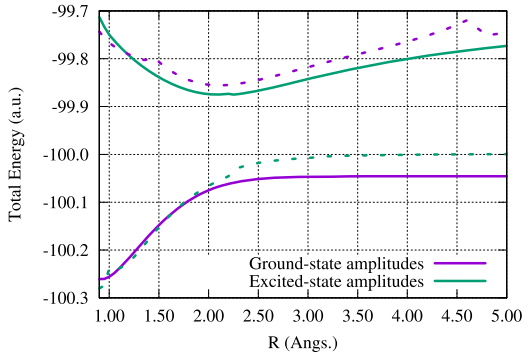


FIG. 4. Potential energy surfaces of the two lowest $^1\Sigma^+$ states of FH with the μ-MR-CCSD method. The energy of the state corresponding to the dressing is plotted in plain curves, and the energy of the other state is plotted in dashed curves.

important determinants of the second root are usually not the same as in the state of interest. The λ-MR-CCSD has amplitudes which depend less on the wave function, so the quality is comparable on both states, and the choice of these amplitudes is better suited for calculating excited states within the same symmetry, as will be confirmed by the next example.

In Figure 5 we have represented the avoided crossing of LiF, also calculated with the aug-cc-pVDZ basis set. The physical situation is similar to FH, but the energy difference between the ground and the excited states is much smaller. A striking result is that the λ-MR-CCSD, although being state-specific, is able to reproduce very well the whole potential energy surfaces of both states. The position of the avoided crossing is very well reproduced by the three methods: the CAS-SD crosses at 6.3 Å, the full-CI crosses at 6.8 Å, and the dressed CAS-SD and the two MR-CCSD variants cross at 6.9 Å. The μ-MR-CCSD and λ-MR-CCSD coincide in the short-range ($\leq$5 Å) and in the long range ($\geq$7.2 Å), but when the two states become very close in energy in the region of the crossing the dressed CAS-SD and the μ-MR-CCSD are unable to give reasonable values. This disappointing result motivates a future work on a multi-state μ-MR-CCSD.

D. Sensitivity to the choice molecular orbitals

The μ-MR-CCSD algorithm we propose is in the Jeziorski-Monkhorst framework, so it is not invariant with respect to the choice of molecular orbitals. In this section, we checked its sensitivity to the choice of the MO set by

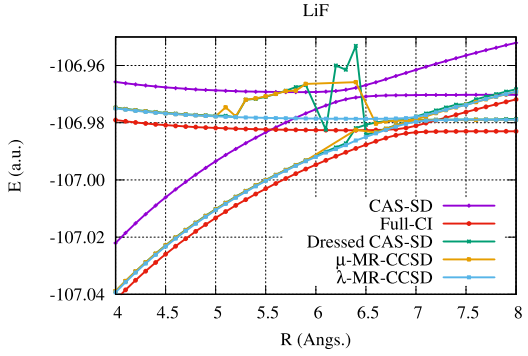


FIG. 5. Potential energy surfaces of the two lowest $^1\Sigma^+$ states of LiF.

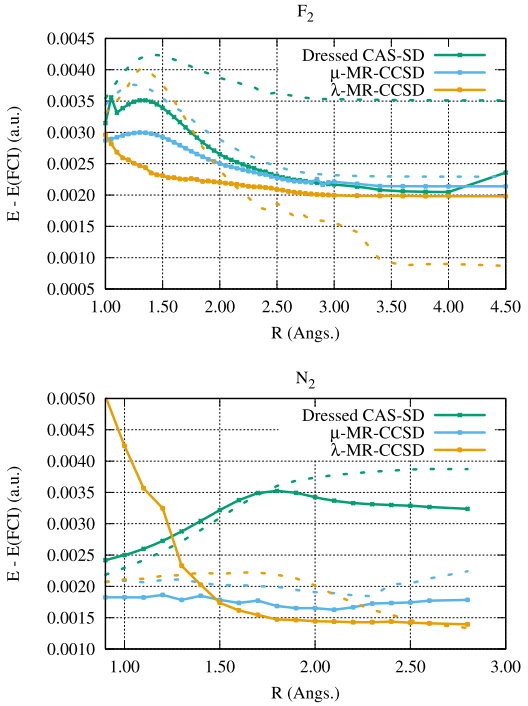


FIG. 6. Comparison between pseudo-canonical (dashed curves) and localized (plain curves) MOs in F_2 and N_2. Difference with respect to the full-CI energy.

comparing results obtained with pseudo-canonical CAS-SCF orbitals and with localized MOs in the F_2 and N_2 molecules (Figure 6).

In the F_2 molecule, using localized MOs is a better choice than the pseudo-canonical MOs. The best NPE is obtained by the μ-MR-CCSD method with a value of 0.9 mE_h. In the case of N_2, the situation is different: the NPE of the λ-MR-CCSD goes from 0.9 mE_h to 3.6 mE_h, and the NPE of the μ-MR-CCSD is relatively stable around 0.3–0.4 mE_h. On the other hand, the dressed CAS-SD gives a better NPE with local orbitals, going from 1.7 mE_h to 1.1 mE_h.

The fact that the μ-MR-CCSD is less sensitive to the MO set than the λ-MR-CCSD can be understood. By changing the MO set, a single excitation rotates into a combination of single and double excitations. In the λ-MR-CCSD method, the amplitudes are calculated by taking into account the matrix elements of the Hamiltonian, which are of different nature depending on the degree of excitation, so the amplitudes are expected to change significantly. In the μ-MR-CCSD variant, the amplitudes are adjusted in such a way that they fit the CAS-SD wave function, which is invariant by rotation of the MOs. Therefore, it is expected to be more robust with respect to the change of MO set.

IV. CONCLUSIONS

We have proposed a method to determine reference-independent amplitudes by fitting the CAS-SD CI vector. These amplitudes may be used to perform a state-specific iterative dressing of the CAS-SD Hamiltonian in order to take into account the effect of the triples and quadruples in the spirit of the coupled cluster formalism. Alternatively, these amplitudes may be rescaled to reproduce the exact coefficients of the singles and doubles to introduce a reference-dependent character.

In that case, the CAS-SD CI vector is recovered by the application of $(1 + \hat{T})$ on the reference wave function, so we reach here the Jeziorski-Monkhorst coupled cluster formalism.

The CAS-SD dressed with reference-independent amplitudes gives excellent results for single-bond breaking (F_2 and ethane) and the simultaneous breaking of the two O–H bonds of water, with a non-parallelism error lower than the milli-Hartree. When the active space becomes larger, it is necessary to go to the reference-dependent MR-CCSD introducing the μ factors in Eq. (31). In the case of ethylene and N_2, this keeps the NPE to a value close to the milli-Hartree.

We have shown numerically that the here-proposed amplitudes are not very sensitive to the value of m_s for open-shell systems, and to the choice of the molecular orbitals. This is clearly an improvement compared the amplitudes proposed earlier.[17] But we have also shown that the former amplitudes are a better choice when computing excited states of the same symmetry because the here-proposed amplitudes have a much more pronounced state-specific character which may be disadvantageous if the states are too close in energy. This problem can be cured by leaving the state-specific formalism for a multi-state formalism,[42] and this will be the object of a future work.

ACKNOWLEDGMENTS

This work has been made through generous computational support from CALMIP (Toulouse) under the Allocation No. 2015–0510, and GENCI under the Allocation No. x2015081738. We are also very grateful to the two anonymous reviewers for their comments, which helped us improve the manuscript significantly.

[1] F. Coester, Nucl. Phys. **7**, 421 (1958).
[2] F. Coester and H. Kümmel, Nucl. Phys. **17**, 477 (1960).
[3] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).
[4] R. J. Bartlett, J. Watts, S. Kucharski, and J. Noga, Chem. Phys. Lett. **165**, 513 (1990).
[5] J. Goldstone, Proc. R. Soc. A **239**, 267 (1957).
[6] I. Nebot-Gil, J. Sańchez-Mariń, J. P. Malrieu, J. L. Heully, and D. Maynau, J. Chem. Phys. **103**, 2576 (1995).
[7] J. P. Malrieu, P. Durand, and J. P. Daudey, J. Phys. A: Math. Gen. **18**, 809 (1985).
[8] B. Kirtman, J. Chem. Phys. **75**, 798 (1981).
[9] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).
[10] N. Oliphant and L. Adamowicz, J. Chem. Phys. **96**, 3739 (1992).
[11] X. Li and J. Paldus, J. Chem. Phys. **108**, 637 (1998).
[12] U. S. Mahapatra, B. Datta, and D. Mukherjee, Mol. Phys. **94**, 157 (1998).
[13] S. Das, D. Mukherjee, and M. Kállay, J. Chem. Phys. **132**, 074103 (2010).
[14] A. Szabados, J. Chem. Phys. **134**, 174113 (2011).
[15] U. S. Mahapatra, B. Datta, and D. Mukherjee, J. Chem. Phys. **110**, 6171 (1999).
[16] J. Meller, J. P. Malrieu, and R. Caballol, J. Chem. Phys. **104**, 4068 (1996).
[17] E. Giner, G. David, A. Scemama, and J. P. Malrieu, J. Chem. Phys. **144**, 064101 (2016).
[18] B. Jeziorski and H. J. Monkhorst, Phys. Rev. A **24**, 1668 (1981).
[19] F. A. Evangelista and J. Gauss, J. Chem. Phys. **134**, 114102 (2011).
[20] M. Hanauer and A. Köhn, J. Chem. Phys. **136**, 204107 (2012).
[21] U. S. Mahapatra, B. Datta, B. Bandyopadhyay, and D. Mukherjee, *Advances in Quantum Chemistry* (Elsevier BV, 1998), pp. 163–193.
[22] L. F. Richardson, Philos. Trans. R. Soc., A **210**, 307 (1911).
[23] G. Golub and W. Kahan, J. Soc. Ind. Appl. Math. Ser. B Numer. Anal. **2**, 205 (1965).
[24] A. Engels-Putzka and M. Hanrath, J. Mol. Struct.: THEOCHEM **902**, 59 (2009).
[25] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).
[26] G. D. Purvis, R. Shepard, F. B. Brown, and R. J. Bartlett, Int. J. Quantum Chem. **23**, 835 (1983).
[27] W. D. Laidig and R. J. Bartlett, Chem. Phys. Lett. **104**, 424 (1984).
[28] M. Kállay, P. G. Szalay, and P. R. Surján, J. Chem. Phys. **117**, 980 (2002).
[29] M. Hanrath, J. Chem. Phys. **123**, 084102 (2005).
[30] F. A. Evangelista, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **125**, 154113 (2006).
[31] F. A. Evangelista, E. Prochnow, J. Gauss, and H. F. Schaefer, J. Chem. Phys. **132**, 074107 (2010).
[32] M. Hanauer and A. Köhn, J. Chem. Phys. **134**, 204111 (2011).
[33] Y. A. Aoto and A. Köhn, J. Chem. Phys. **144**, 074103 (2016).
[34] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[35] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).
[36] A. Scemama, T. Applencourt, Y. Garniron, E. Giner, G. David, and M. Caffarel, Quantum package v1.0, 2016, https://github.com/LCPQ/quantum_package.
[37] B. Huron, P. Rancurel, and J. Malrieu, J. Chem. Phys. **58**, 5745 (1973).
[38] S. Evangelisti, J. Daudey, and J. Malrieu, Chem. Phys. **75**, 91 (1983).
[39] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[40] P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, Comput. Phys. Commun. **149**, 71 (2002).
[41] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision D.01, Gaussian, Inc., 2009.
[42] J.-P. Malrieu, Mol. Phys. **111**, 2451 (2013).

8.3 Computing c_α for matrix dressing

Whether the matrix dressing algorithm performs a Shifted-B_k , an MR-CCSD or some other method, depends on the external space, i.e. the c_α coefficients. Our goal is to set up a framework in which stochastic matrix dressing can be done efficiently using an external space only defined by $Z(\alpha, \Psi, \dots)$ a function that takes a determinant $|\alpha\rangle$ and the wave function $|\Psi\rangle$, and returns the c_α it should be associated with. The "…" notation indicates that the returned value may depend on any number of global parameters, such as approximation thresholds, etc…

Looking at the expression of c_α for Shifted-B_k , it looks like the computation required is the exact same as the one performed by our CIPSI

$$c_\alpha = \frac{\langle \alpha | \hat{H} | \Psi \rangle}{\Delta E_\alpha} = \sum_{I=1}^{N_{\text{det}}} c_I \frac{\langle \alpha | \hat{H} | D_I \rangle}{\Delta E_\alpha}, \tag{8.9}$$

but the expression of c_α for Shifted-B_k has a particularity that lifts a constraint compared to the general case: as can be seen, while for any $|\alpha\rangle$ we need to find all internal determinants it connects to, we do not need to know them *at the same time*, i.e. c_α can be incrementally built from terms that only involve one internal determinant at a time. This is not generally the case, and isn't the case for MR-CCSD.

Thus, matrix dressing generally requires the knowledge of all internal determinants a particular $|\alpha\rangle$ connects to, before c_α — and thus the associated increment to $\boldsymbol{\delta}$ — can be computed. For efficiency, as well as simplicity, this list must absolutely be computed upstream the computation of $Z(\alpha, \Psi, \dots)$, making it in practice $Z^\star(\alpha, \Psi_c, \dots)$ with Ψ_c the variational wave function stripped of all determinants that do not connect to $|\alpha\rangle$ (thus not normalized). In practice, it is a list of internal determinant indices.

This is a lot like the former implementation of CIPSI : considering one $|\alpha\rangle$ at a time, enumerate its connections to selectors. The new, more efficient algorithm, however, considers a batch of up to $N_{\text{virt}} = \left(N_{\text{orb}} - N_{\text{elec}}^{\uparrow}\right)^2$ rather than a single one. The solution is conceptually simple. Neglecting connections to determinants that are not selectors :

1. Loop over all G_{pq} batches in the same way as in the CIPSI algorithm (building the B_{rs} tag matrix).

2. For each batch, create $(2N_{\text{orb}})^2$ sets of internal determinant indices $\mathcal{C}_{rs}$, each associated with $\left|G_{pq}^{rs}\right\rangle$.

3. When a connection between a selector $|D_I\rangle$ and $|\alpha\rangle = \left|G_{pq}^{rs}\right\rangle$ is found, add I to $\mathcal{C}_{rs}$ (instead of incrementing $P(G_{pq})$).

4. When the computation of batch G_{pq} is completed, $\mathcal{C}_{rs}$ is the set of all indices of selectors connected to $|\alpha\rangle = \left|G_{pq}^{rs}\right\rangle$. For each untagged $|\alpha\rangle$:

   (a) $c_\alpha \leftarrow Z^\star(\alpha, \mathcal{C}_{rs}, \ldots)$.

   (b) For each index $K \in \mathcal{C}_{rs}$ increment δ_K with $c_\alpha \langle\alpha|\hat{H}|D_K\rangle$.

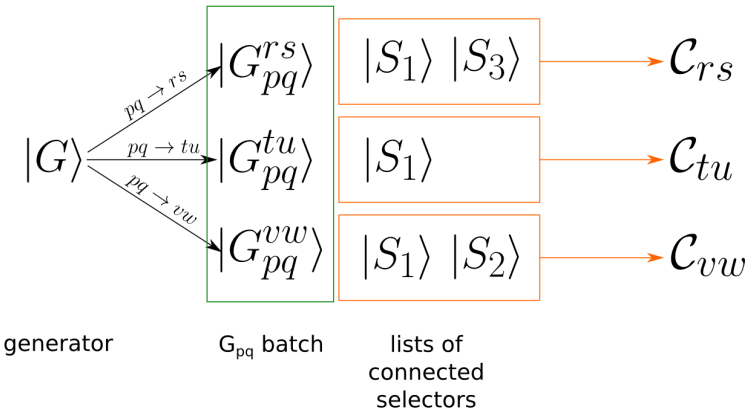


Figure 8.1: Build lists of connected selectors for unique $|\alpha\rangle$'s in batch G_{pq}.

As can be seen, this is very similar to the implementation of CIPSI, except we are building sets instead of incrementing scalars. This, of course, adds complexity to the implementation as the sizes of the $\mathcal{C}_{rs}$ sets aren't known in advance. More importantly, the storage space required may be prohibitive. Noticing that

- all $\left|G_{pq}^{rs}\right\rangle$ in a batch are connected to each other,
- there are ${N_{\text{virt}}}^2$ non-zero $\left|G_{pq}^{rs}\right\rangle$ in a batch,

and considering a particular case where

- G_{pq} is the first batch (all generated $|\alpha\rangle$'s are unique)
- half of $\left|G_{pq}^{rs}\right\rangle$ are internal determinants,

we have $\frac{1}{2}{N_{\text{virt}}}^2$ unique $|\alpha\rangle$'s in the batch each one connecting to at least $\frac{1}{2}{N_{\text{virt}}}^2$ internal determinants, for a total storage space of at least $\frac{1}{4}{N_{\text{virt}}}^4$. This case isn't unrealistic, with $|G\rangle$ the Hartree-Fock determinant and (p, q) the highest occupied spinorbitals.

Another issue is the high number of non-contiguous writes in memory, especially with the selectors that connect to all determinants of the batch ; they need to be added to ${N_{\text{virt}}}^2$ sets, which is ${N_{\text{virt}}}^2$ non-contiguous writes for a single selector.

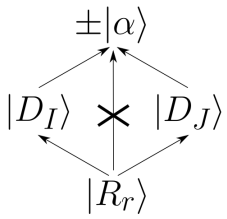
We can solve the storage issue and mitigate the number of non-contiguous writes, by creating sets of $|\alpha\rangle$ that are subsets of several $\mathcal{C}_{rs}$. Table 5.1 indicates which $|\alpha\rangle = \left|G^{rs}_{pq}\right\rangle$ of the current batch a selector $|S\rangle$ connects to. In some cases, there are "wild-card" indices X and Y. Instead of looping over the possible values for those wildcards and adding $|S\rangle$ to all the corresponding $\mathcal{C}_{rs}$ sets, we are going to give wildcard indices the special value 0 and build intermediate sets $\tilde{\mathcal{C}}_{rs}$. For example, in the case where both r and s are wildcards for a selector $|D_I\rangle$, instead of adding I to all $\mathcal{C}_{rs}$ sets, we will add it to a single set $\tilde{\mathcal{C}}_{00}$. When computation for the batch is completed, $\mathcal{C}_{rs}$ can be evaluated as

$$\mathcal{C}_{rs} \leftarrow \tilde{\mathcal{C}}_{rs} \cup \tilde{\mathcal{C}}_{r0} \cup \tilde{\mathcal{C}}_{s0} \cup \tilde{\mathcal{C}}_{00} \tag{8.10}$$

Among these intermediate sets only $\tilde{\mathcal{C}}_{r0}$ and $\tilde{\mathcal{C}}_{s0}$ may share common elements. Given its frequency, it is important that this computation is efficient. As is sometimes the case, efficiency implies $\mathcal{C}_{rs}$ are not computed individually, but become available inside a loop. An implementation is proposed as algorithm 34, which tries to reuse shared $\tilde{\mathcal{C}}_{rs}$ as much as possible.

8.4 Efficient application to multi-reference coupled cluster

The equation for c_α as well as the procedure for computing amplitudes is detailed in the article presented in section 8.2. In particular, Eq.(26) shows the formula to compute c_α as a sum over what can be described as "diamond" structures.



With $|\alpha\rangle$ the external determinant being considered, $|D_I\rangle$ and $|D_J\rangle$ internal determinants, and $|R_r\rangle$ a reference determinant. Parallel arrows indicate connections (excitation degree at most 2) by the same excitation, the vertical arrow indicates $|R_r\rangle$ and $|\alpha\rangle$ aren't connected (excitation degree at least 3). More generally, we only consider $|\alpha\rangle$ determinants that are not connected to any $|R_r\rangle$. All the determinants we are considering are ordered according to $\hat{\mathcal{O}}$.

Data: $\tilde{\mathcal{C}}_{rs}$ the intermediate sets assumed to be sorted in increasing order, B_{rs} the tag matrix

Result: $\mathcal{C}_{rs}$ is computed for all $\left|G^{rs}_{pq}\right\rangle$ of G_{pq} batch that are unique $|\alpha\rangle$

```
/* C̃_rs and C_rs are considered arrays, the syntax C_rs[i...j] is used
   to denote a segment of array, |C_rs| is the cardinality of C_rs
   */
L an array of determinants size N_sel ;
i_1 ← |C̃_00| ;
L[1...i_1] ← C̃_00[1...i_1] ;
/* B_r0 = TRUE if column r is entirely tagged                          */
forall r; ¬B_r0 do
    i_2 ← i_1 + |C̃_r0| ;
    L[i_1 + 1...i_2] ← C̃_r0 ;
    forall s; ¬(B_rs ∨ B_0s) do
        i_3 ← i_2 ;
        j ← 1 ;
        k ← 1 ;
        while k ≤ |C̃_s0| do
            if (j > |C̃_r0|) ∨ (C̃_s0[k] < C̃_r0[j]) then
                i_3 ← i_3 + 1 ;
                L[i_3] ← C̃_s0[k] ;
                k ← k + 1 ;
            else if C̃_s0[k] > C̃_r0[j] then
                j ← j + 1 ;
            else
                j ← j + 1 ;
                k ← k + 1 ;
            end
        end
        i_4 ← i_3 + |C̃_rs| ;
        L[i_3 + 1...i_4] ← C̃_rs ;
        /* L = C_rs                                                    */
    end
end
```

Algorithm 34: Build $\mathcal{C}_{rs}$ from $\tilde{\mathcal{C}}_{rs}$

Data: $|\alpha\rangle$ the considered external determinant
Data: $|D_I\rangle$ the list of N internal determinants connected to $|\alpha\rangle$, sorted with increasing integer value
Data: $|R_r\rangle$ the list of N_{ref} reference determinants sorted with increasing integer value and c_{R_r} the associated coefficients
Data: $t_{R_r \to D_I}$ the used amplitudes
Result: c_α

$c_\alpha \leftarrow 0$;
Discard R_r when EXC_DEGREE$(R_r, \alpha) > 4$;
If R_r so that EXC_DEGREE$(R_r, \alpha) \leq 2$ was found, discard $|\alpha\rangle$;
forall D_J **do**
    /* Note that $\Phi(|D_J\rangle \to |\alpha\rangle)$ may be pre-computed here */
    $\delta \leftarrow \alpha - D_J$;
    $i \leftarrow 1$;
    $r \leftarrow 1$;
    while $i \leq N \wedge r \leq N_{ref}$ **do**
        if $D_I - R_r > \delta$ **then**
            $r \leftarrow r + 1$;
        else if $D_I - R_r < \delta$ **then**
            $i \leftarrow i + 1$;
        else
            if $R_r \oplus D_I \oplus D_J \oplus \alpha = 0$ **then**
                /* diamond found */
                $\varphi \leftarrow \Phi\left(|D_J\rangle \to |\alpha\rangle\right) \times \Phi(|R_r\rangle \to |D_I\rangle)$;
                $c_\alpha \leftarrow c_\alpha + \varphi \times c_{R_r} \times \left(t_{R_r \to D_I}\right) \times \left(t_{R_r \to D_J}\right)$;
            end
            $r \leftarrow r + 1$;
            $i \leftarrow i + 1$;
        end
    end
end

Algorithm 35: Compute a c_α for MR-CCSD

The matrix dressing implementation supplies $\{\mathcal{D}\}$ the complete set of the N internal determinants that connect to $|\alpha\rangle$. We call $\{\mathcal{R}\}$ the set of the N_{ref} reference determinants.

A naive way to find those diamonds, would be to loop over all unordered triplets

$$(D_I \in \{\mathcal{D}\}, D_J \in \{\mathcal{D}\}, R_r \in \{\mathcal{R}\}) \tag{8.11}$$

so that

$$\hat{T}_{R_r \to D_I} \hat{T}_{R_r \to D_J} |R_r\rangle = \pm |\alpha\rangle \tag{8.12}$$

with $\hat{T}_{R_r \to D_I}$ the excitation operator so that $\hat{T}_{R_r \to D_I} |R\rangle_r = |D_I\rangle$.

Using excitation operators, a diamond can be identified by only verifying

$$\hat{T}_{R_r \to D_I} = \pm \hat{T}_{D_J \to \alpha} \tag{8.13}$$

or in other words, that $\hat{T}_{R_r \to D_I}$ and $\hat{T}_{D_J \to \alpha}$ involve the same holes and particles. We first set up a method to identify a diamond, then two methods to "locate" them.

Identifying diamonds For implementational efficiency, we are going to express excitations using $\hat{f}_p$ an operator that flips the occupation status of a spinorbital p.

$$\begin{cases} \hat{f}_p |A\rangle = \hat{\mathcal{O}} a_p |A\rangle & \text{if } a_p |A\rangle \neq 0 \\ \hat{f}_p |A\rangle = \hat{\mathcal{O}} a_p^\dagger |A\rangle & \text{if } a_p^\dagger |A\rangle \neq 0 \end{cases} \tag{8.14}$$

or in a less verbose formulation

$$\hat{f}_p = \hat{\mathcal{O}} a_p + \hat{\mathcal{O}} a_p^\dagger \tag{8.15}$$

This avoids the burden of making a distinction between annihilation and creation operator, and of checking whether they can be applied to a determinant. This also ignores phase factors.

The notation $\hat{f}_{ab\ldots}$ will be used as a shortcut for $\hat{f}_a \hat{f}_b \ldots$, and we define $\hat{F}_{A \to B} = \hat{F}_{B \to A}$ as the set of $\hat{f}$ operators that flips all spinorbitals whose occupation differ between $|A\rangle$ and $|B\rangle$.

Clearly a set of $\hat{f}$ operators does not univocally correspond to an excitation, but it is demonstrable that in this particular case, we can use $\hat{F}$ instead of the excitation operator $\hat{T}$, i.e. we can ensure (R_r, D_I, D_J, α) are forming a diamond by only verifying

$$\hat{F}_{R_r \to D_I} = \hat{F}_{D_J \to \alpha} \tag{8.16}$$

It is easy to understand from Eq. (8.13) how this is a necessary condition, it is less obvious that it is also a sufficient one, i.e. that this can only happen if there is a diamond. In fact, it is not generally true. It is demonstrable in this particular case

thanks to the known excitation degrees. It is known $|D_I\rangle$ and $|D_J\rangle$ are connected by at most a double excitation to $|\alpha\rangle$, therefore at most 4 orbitals have their occupation status flipping, two of them occupied and two unoccupied in $|\alpha\rangle$. For later clarity we use the dot ($\dot{a}$) notation to denote indices of spinorbitals that are occupied in $|\alpha\rangle$.

$$\hat{F}_{D_I\to\alpha} = \hat{F}_{R_r\to D_J} = \hat{f}_{\dot{a}bcd}\ ;\ \hat{F}_{D_J\to\alpha} = \hat{F}_{R_r\to D_I} = \hat{f}_{\dot{e}\dot{f}\dot{g}h} \tag{8.17}$$

The reference determinant $|R_r\rangle$ is reached from $|\alpha\rangle$ after chaining the two "flippings".

$$\hat{F}_{R\to\alpha} = (\hat{f}_{\dot{a}bcd})(\hat{f}_{\dot{e}\dot{f}\dot{g}h}) \tag{8.18}$$

If indices $\dot{a}, \dot{b}, c, d, \dot{e}, \dot{f}, g, h$ are all unique, $\hat{T}_{R\to D_I}$ and $\hat{T}_{D_J\to\alpha}$ are independent and thus can be chained, so the diamond is valid. If they are not, the diamond is still known to be valid thanks to our knowledge that $|\alpha\rangle$ is at least a triple excitation from $|R_r\rangle$, and therefore at least 6 orbitals must flip.

It is trivial that

$$\hat{f}_{aa} = \hat{1} \tag{8.19}$$

so only 2 among the 8 indices can refer to the same spinorbital x (which we arbitrarily choose unoccupied in $|\alpha\rangle$). For obvious reasons one is found among $(\dot{a}, \dot{b}, c, d)$ and the other among $(\dot{e}, \dot{f}, g, h)$.

$$\hat{F}_{R\to\alpha} = (\hat{f}_{\dot{a}\dot{b}cx})(\hat{f}_{\dot{e}\dot{f}\dot{g}x}) = (\hat{f}_{\dot{a}\dot{b}c})(\hat{f}_{\dot{e}\dot{f}\dot{g}}) \tag{8.20}$$

As can be seen, applying $(\hat{f}_{\dot{a}\dot{b}c})(\hat{f}_{\dot{e}\dot{f}\dot{g}})$ to $|\alpha\rangle$ will flip four occupied spinorbitals versus only two unoccupied ones ; this implies $|\alpha\rangle$ has two extra electrons compared to $|R_r\rangle$. Because we know this not to be the case, we know such a situation cannot happen.

Now that we know Eq. (8.16) is sufficient to identify a diamond, we have to write its implementational expression, which is straightforward. The set of spinorbitals whose occupation status differ between $|A\rangle$ and $|B\rangle$ can be computed as a bitstring as

$$A \oplus B \tag{8.21}$$

with A a single bitstring associated with $|A\rangle$ — the association between a spinorbital and a bit is arbitrary.

Eq. (8.16) is verified iff

$$R_r \oplus D_I = D_J \oplus \alpha \tag{8.22}$$

or alternatively, since $A \oplus A = 0$

$$R_r \oplus D_I \oplus D_J \oplus \alpha = 0 \tag{8.23}$$

Locating diamonds by binary search Now that we expressed the identification of a diamond in a simple way, we must figure a strategy to find them fast. Eq. (8.23) can be rewritten as

$$D_J = R_r \oplus D_I \oplus \alpha \tag{8.24}$$

which implies a simple algorithm that finds the diamonds in $\mathcal{O}(N \times N_{\text{ref}} \times \log(N))$:

1. Iterate over (R_r, D_I)
2. Binary search for $D_{J>I} = (R_r \oplus D_I \oplus \alpha)$ in $\{\mathcal{D}\}$ if $||D_J|| = N_{\text{elec}}$.

Depending on the sizes of the $\{D\}$ and $\{R\}$ sets, an alternative variant can also be used, giving a complexity of $\mathcal{O}(N^2 \times \log(N_{\text{ref}}))$:

1. Iterate over unordered pairs (D_I, D_J)
2. Binary search for $R_r = (D_I \oplus D_J \oplus \alpha)$ in $\{\mathcal{R}\}$ if $||R_r|| = N_{\text{elec}}$.

Those are desirable if one of $\{\mathcal{D}\}$ or $\{\mathcal{R}\}$ is relatively small, in particular, in case of single-reference computation, i.e. $N = 1$, the former variant is ideal. For larger sets, however, a more complex variant with a complexity $\mathcal{O}(N \times (N + N_{\text{ref}}))$ is introduced next.

Locating diamonds by rewriting excitations as additions Although the method could be chosen on a "by-α" basis depending on N and N_{ref}, only this one is currently used in the Quantum Package as it works for larger sets. Unlike the previous one, this method may yield false positives, thus it uses Eq. (8.23) for confirmation. The idea is to express an excitation as an addition. As was said, bitstrings can be interpreted as integers, so they can be added, substracted and compared as such. As in the case of the $\hat{f}$ operator, this approach entirely ignores phase factors. We can easily associate the integer value $T_{A\to B}$ to excitation $\pm\hat{T}_{A\to B}$.

$$T_{A\to B} = B - A \tag{8.25}$$
$$T_{A\to B} + A = B \tag{8.26}$$

With A the bitstring associated with $|A\rangle$ as a single integer — the association between a spinorbital and a bit is arbitrary. It is a necessary but not sufficient condition for a diamond that

$$T_{R\to D_I} = T_{D_J\to\alpha} \tag{8.27}$$

Indeed,

$$\pm\hat{T}\,|A\rangle = |B\rangle \implies T + A = B \tag{8.28}$$
$$T + A = B \not\implies \pm\hat{T}\,|A\rangle = |B\rangle \tag{8.29}$$

with T the integer associated with excitation $\pm\hat{T}$. If $\pm\hat{T}\,|A\rangle = 0$, in most cases $T + A$ will yield a bitstring with a wrong number of bits/electrons, but this is not guaranteed, hence the rare presence of false positives.

Eq. (8.13) implies

$$D_I - R_r = \alpha - D_J \tag{8.30}$$

Finding all (D_I, R_r) pairs that verify this, assuming $\{\mathcal{D}\}$ and $\{\mathcal{R}\}$ are sorted in ascending order, can be achieved with linear complexity.

1. Initialize I and r to 1
2. Loop while both I and r are not out of bounds
3. If $D_I - R_r < \alpha - D_J$, increment I and loop
4. If $D_I - R_r > \alpha - D_J$, increment r and loop
5. If $D_I - R_r = \alpha - D_J$, a (D_I, R_r) pair has been found. Check if a diamond can be formed using Eq. (8.23). Increment I and r, and loop.

The complexity is $\mathcal{O}(N \times (N + N_{\text{ref}}))$.

For implementational ease and efficiency, two things are worth noting:

- Integer overflows do not need to be handled. The size of integers being limited to 64 bits essentially means an unsigned integer I is represented as $I \mod 2^{64}$. Quite fortunately modulus has the properties of compatibility with addition and substraction.

$$a_1 + a_2 \equiv b_1 + b_2 \pmod{n} \tag{8.31}$$
$$a_1 - a_2 \equiv b_1 - b_2 \pmod{n} \tag{8.32}$$

  with $a_1 \equiv b_1 \pmod{n}$, and $a_2 \equiv b_2 \pmod{n}$. When it comes to addition and substraction, signed and unsigned integers are equivalent, so signed integers can be used as well.

- It is not required to handle bitstrings as actual arbitrary size integers when it comes to addition and substraction. While addition to a bitstring was introduced as being associated with an excitation, it can actually be associated with any combination of creation and annihilation operators. It is therefore valid to consider each 64-bit integer as an independent "sub-bitstring" on which a subset of the operators will be applied. In short, additions and substraction can be done integer-wise, without the overhead of handling carries.

When it comes to comparison, since the 64-bit integers corresponding to the lower orbitals will typically have more mobile electrons, they should be given more weight in order to resolve comparison as fast as possible.

Computing the phase factor So far we have ignored phase factors, but we need to know the sign of $|\alpha\rangle$ in each found diamond. This can be tricky, since confusion can easily arise from reference-dependent and reference-independent notations. To keep generality, we use reference-dependent notations, unlike Eq. (26) of the presented article.

Assuming

$$\hat{T}^{rs}_{pq}\,|I\rangle = \pm\,|i\rangle \tag{8.33}$$

reference-independent notation for amplitude $t_{pq\to rs}$ and reference-dependent notation $t_{I\to i}$ could be wrongfully assumed equivalent. This is not the case, since $t_{pq\to rs}$ is associated with excitation $\hat{T}^{rs}_{pq}$ as described in chapter 3 — with a particular ordering for creation an annihilation operators — while $t_{I\to i}$ is associated with $\hat{T}_{I\to i}$ so that

$$\hat{T}_{I\to i}\,|I\rangle = |i\rangle \tag{8.34}$$

preserving the phase factor of $|i\rangle$. Therefore

$$t_{I\to i} = \frac{t_{pq\to rs}}{\Phi(|I\rangle \to |i\rangle)} \tag{8.35}$$

In order to compute the contribution of a diamond to c_α, we are required to compute the phase factor

$$\Phi(|R_r\rangle \to |\alpha\rangle) = \Phi(|R_r\rangle \to |D_I\rangle)\,\Phi(|D_I\rangle \to |\alpha\rangle) \tag{8.36}$$

which is noted $(-1)^{n(I\to\alpha)}$ in Eq. (26) of the presented article, with I the reference determinant. The article uses the reference-independent notation for amplitudes. If we re-write the formula with reference-dependent notations, the contribution for each diamond becomes

$$c_r\Phi(|R_r\rangle \to |\alpha\rangle)\frac{t_{R_r\to D_I}}{\Phi(|R_r\rangle \to |D_I\rangle)}\frac{t_{D_I\to\alpha}}{\Phi(|D_I\rangle \to |\alpha\rangle)} = c_r t_{R_r\to D_I} t_{D_I\to\alpha} \tag{8.37}$$

with c_r the coefficient of $|R_r\rangle$. The amplitudes for $t_{D_I\to\alpha}$ of course cannot be stored and would be too expensive to compute on the fly. Because $\hat{T}_{D_I\to\alpha}$ involves the same orbitals as $\hat{T}_{R_r\to D_J}$, their associated amplitudes are the same up to the phase factor

$$t_{D_I\to\alpha} = t_{R_r\to D_J}\Phi\big(|R_r\rangle \to \big|D_J\big\rangle\big)\,\Phi(|D_I\rangle \to |\alpha\rangle) \tag{8.38}$$

So we can finally rewrite the contribution to c_α for each diamond

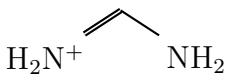
$$c_r t_{R_r\to D_I} t_{R_r\to D_J}\Phi\big(|R_r\rangle \to \big|D_J\big\rangle\big)\,\Phi(|D_I\rangle \to |\alpha\rangle) \tag{8.39}$$

Chapter 9

Performance measurements

Contents

In this chapter, we discuss the efficiency of the implementation. The system we chose for these numerical experiments is a cyanine dye,



in its ground state and in its first excited state. The geometry is the equilibrium geometry of the ground state, optimized at the PBE0/cc-pVQZ level. The ground state is a closed shell, well described by a single reference, and the excited state is singly excited and requires two determinants in the reference ($1/\sqrt{2}(a\bar{b} + b\bar{a})$). The calculations were performed in the aug-cc-pVDZ basis set with state-averaged natural orbitals obtained from an initial CIPSI calculation. The 1s orbitals of the carbon and the nitrogen atoms were frozen, so the FCI space which is explored is a CAS(18,111). The reference excitation energy, obtained at the CC3/ANO-L-VQZP level is 7.18 eV.[84] The measurements were made on the Olympe supercomputer (CALMIP). Each node is a dual-socket Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz with 192GiB of RAM, and contains 36 physical CPU cores. Parallel speedup curves are made with up to 1 800 cores for the four main parts presented in this manuscript, namely the Davidson diagonalization, the CIPSI selection, the hybrid stochastic/deterministic computation of E_{PT2} and the matrix dressing.

In figure 9.1, we plot the convergence of the energies of the ground and excited states as a function of the number of determinants, with and without the second order perturbative contribution. From these data, one can see that although E_{PT2} is still large ($\sim$ 0.02 au) the excitation energies both at the variational level and with the perturbative correction converge to a value of 7.20 eV compatible with the reference energy obtained in a larger basis set.

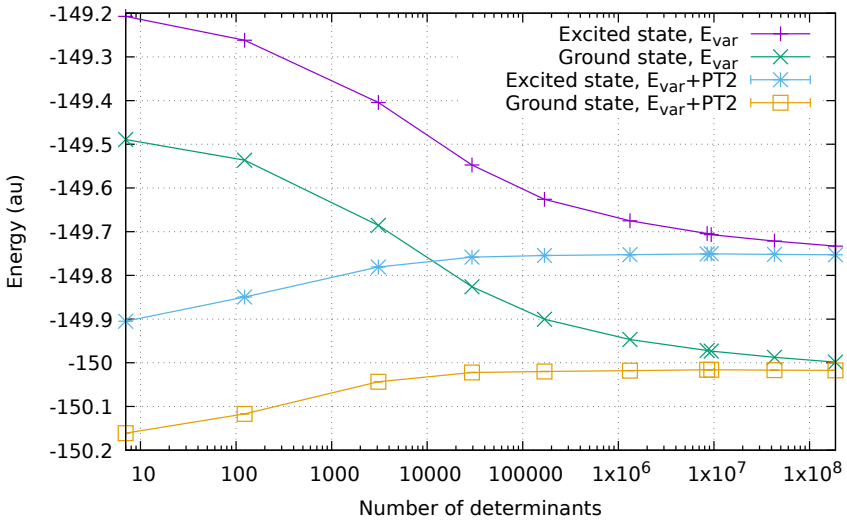


Figure 9.1: Convergence of the energy of the ground and excited states with respect to the number of determinants in the variational space.

Table 9.1: Energies and second-order perturbative corrections for increasingly large wave functions. ΔE is the energy difference between the ground state and the excited state.

N_{det}	Ground state	Excited state	ΔE (eV)
	E_{var}		
7	−149.489 186	−149.207 354	7.67
123	−149.536 265	−149.261 860	7.47
3 083	−149.685 606	−149.404 450	7.65
29 409	−149.826 151	−149.547 275	7.59
168 595	−149.900 352	−149.626 058	7.46
1 322 537	−149.946 655	−149.675 032	7.39
8 495 334	−149.972 032	−149.704 145	7.29
9 356 952	−149.973 375	−149.706 822	7.25
42 779 636	−149.987 370	−149.721 470	7.24
186 978 487	−149.998 582	−149.733 039	7.23
	$E_{\text{var}} + E_{\text{PT2}}$		
7	−150.161 107	−149.904 883	6.97
123	−150.116 958	−149.849 465	7.28
3 083	−150.043 5(2)	−149.780 8(2)	7.15
29 409	−150.022 2(2)	−149.758 3(2)	7.18
168 595	−150.019 9(1)	−149.754 5(1)	7.22
1 322 537	−150.017 89(7)	−149.752 55(7)	7.22
8 495 334	−150.015 97(4)	−149.750 87(5)	7.21
9 356 952	−150.015 89(3)	−149.750 66(3)	7.22
42 959 496	−150.016 75(2)	−149.751 88(2)	7.21
186 978 487	−150.017 51(2)	−149.752 90(2)	7.20

9.1 Davidson's diagonalization

First, we measure the time required to compute one iteration of the Davidson algorithm with increasingly large wave functions. The results are reported in table 9.2.

Plotting this data with a log-log scale (figure 9.2) shows an agreement with the predicted $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$ scaling.

Then, we took the two largest wave functions. One with 9 356 952 and one with 42 959 496 determinants, and measured the wall-clock time required to perform one iteration of the Davidson diagonalization, with an increasing number of compute nodes.

The timings are reported in table 9.3 and the parallel speedup curve is represented in figure 9.3. As the communication scales as $\mathcal{O}(N_{\text{det}})$ and the computation scales as $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$, the parallel efficiency increases together with N_{det}, as shown on figure 9.3. For the largest wave function a parallel efficiency of 76% is obtained on 50 nodes for the largest wave function.

Table 9.2: Wall-clock time (in seconds) to run one Davidson's iteration in parallel with increasingly large wave functions.

N_{det}	seconds
29 409	2.72
168 595	4.21
1 322 537	56.24
9 356 952	775.55
42 959 496	11 198.70

Table 9.3: Wall-clock time (in seconds) to run one Davidson's iteration in parallel on two different wave functions with an increasing number of 36-core compute nodes.

Nodes	9 356 952 determinants	42 959 496 determinants
1	775.55	11 198.70
5	169.88	2 288.58
10	93.22	1 213.95
20	56.86	626.41
30	43.76	445.65
40	36.18	350.25
50	33.67	295.25

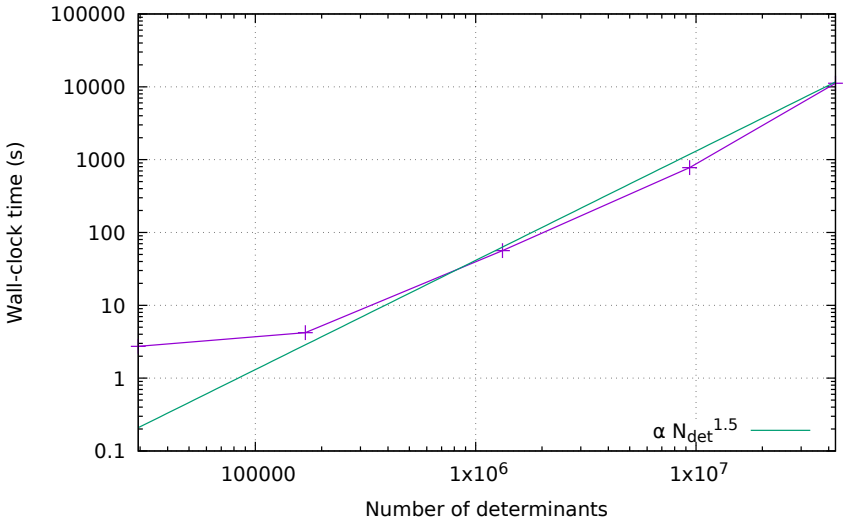


Figure 9.2: Wall-clock time of one Davidson iteration as a function of the number of determinants in the wave function.

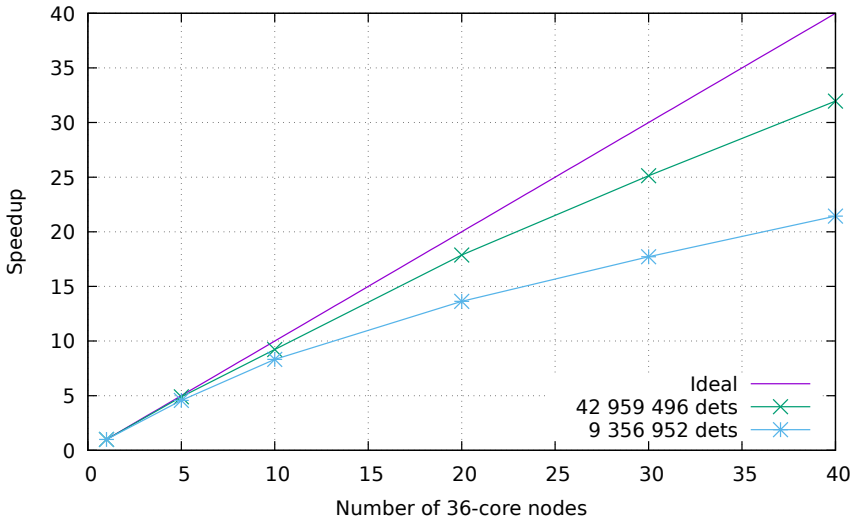


Figure 9.3: Speedup of one Davidson iteration as a function of the number of 36-core compute nodes.

9.2 Selection

We have measured the time necessary to realize a selection step, with an increasing number of determinants in the variational space. Figure 9.5 shows a near-linear scaling with the number of determinants. This is due to the threshold n_g which tends to make the number of external determinants constant when the wave function becomes large enough.

The parallel speedup was also measured with up to 50 nodes, showing an almost ideal speedup curve with 95% of parallel efficiency with 50 nodes. This is in part due to the tuning of the fragmentation of the tasks which gives a very well balanced task queue.

Table 9.4: Single-node (36-core) CIPSI selection for increasingly large wave functions. Time is given in seconds.

N_{det}	seconds
123	0.14
3 083	0.59
29 409	42.67
168 595	239.21
1 322 537	2 008.76
9 356 952	22 560.33

Table 9.5: Time (in seconds) to run parallel CIPSI selections on the 9 356 952-determinant wave function with an increasing number of 36-core compute nodes.

Nodes	seconds
1	22 560.33
5	4 468.28
10	2 245.00
20	1 137.67
30	769.58
40	582.62
50	472.33

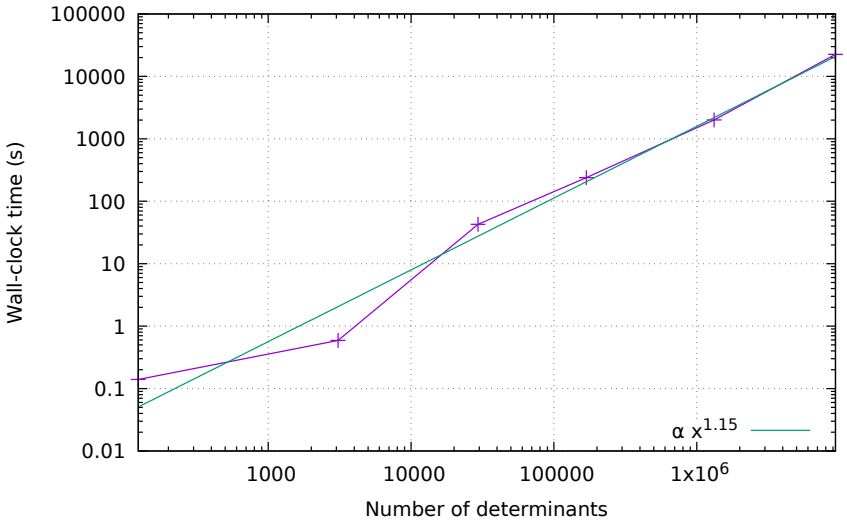


Figure 9.4: Wall-clock time of the selection as a function of the number of determinants in the wave function.

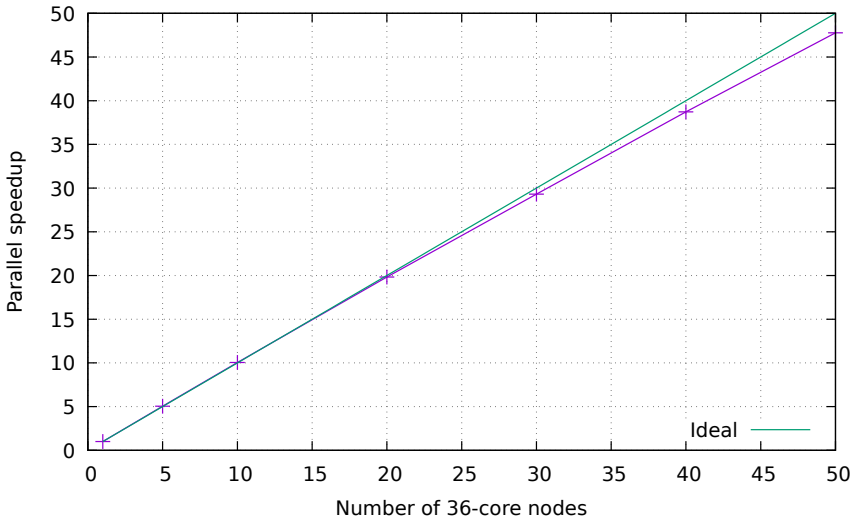


Figure 9.5: Parallel speedup of the CIPSI selection. The reference is a single 36-core node.

9.3 PT2 calculations

The algorithms for the computation E_{PT2} are very similar to those of the CIPSI selection. Therefore, we expect a similar behavior with N_{det} and with the number of nodes. The stopping criterion of the calculation of the E_{PT2} contribution was a relative statistical error below 1/1000-th. Hence, the error bar decreases with the E_{PT2} correction when the number of determinants in the internal space increases, but the fraction of the full deterministic calculation required to reach this criterion is relatively stable around 5%.

Table 9.6 reports the wall-clock time required to compute E_{PT2} on a single node. From these data, one can evaluate the scaling of the cost of E_{PT2} with the number of determinants, as plotted in figure 9.6, which is close to linear. This can be understood, since the number of $|\alpha\rangle$ determinants is proportional to the number of determinants in the variational wave function, and each $|\alpha\rangle$ needs to be compared to all the determinants in the computation of $\langle\alpha|H|\Psi\rangle$. But when the wave function becomes large enough, the second point is not true any more because only a limited number of determinants $|I\rangle$ of Ψ have a non-zero value $\langle\alpha|H|I\rangle$, and this number is bounded by the number of single and double excitations, characteristic of the basis set. Fitting the last points with a log-log plot shows an asymptotic scaling as $\mathcal{O}\left({N_{det}}^{1.07}\right)$, which is compatible with the expected linear scaling for very large numbers of determinants.

To analyze the parallel efficiency of the E_{PT2} calculation, we have made the parallel speedup curve using up to 50 nodes (1800 CPU cores), plotted in figure 9.7. With 50 nodes, one obtains a speedup with respect to the single-node reference of $40\times$ for both the ground and excited states. This corresponds to a parallel efficiency of 80%, which is less satisfactory than the almost ideal speedup obtained for the selection.

There are two reasons for this disappointing speedup. The first one is that the calculation is aborted when the target precision is below a given threshold. So when more compute nodes are used, more samples are gathered at the end of the calculation and the computation is more precise. In the limit of N_{det} CPU cores, only the deterministic calculation can be done, whatever the stopping criterion. So the computation of the speedup is not 100% fair. The second reason for the non-ideal speedup is the pre-computation of the combs on the master process which delays the start of the computation of the tasks.

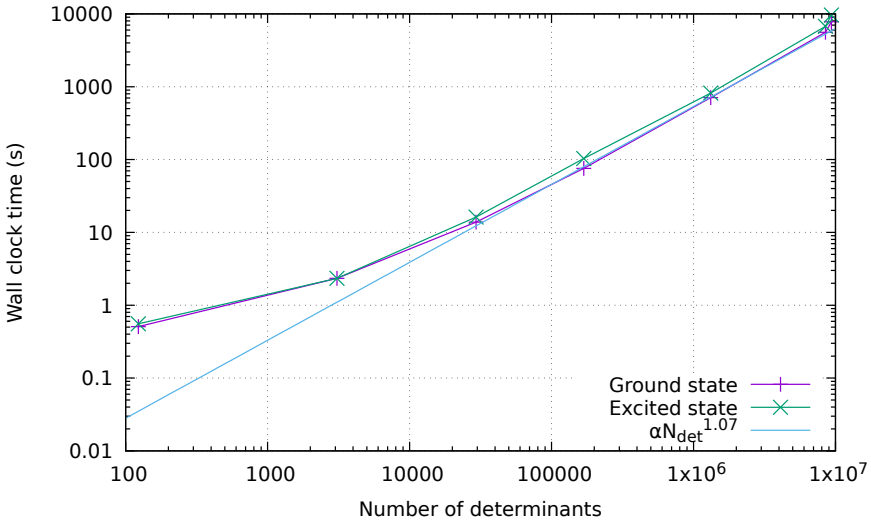


Figure 9.6: Wall clock time required to compute E_{PT2} for the ground and the excited states, as a function of the number of determinants in the wave function.

Table 9.6: Single-node (36-core) E_{PT2} calculations for increasingly large wave functions. Time is given in seconds.

N_{det}	Ground state	Excited state
123	0.51	0.56
3 083	2.33	2.34
29 409	13.83	16.21
168 595	75.62	103.26
1 322 537	708.53	818.58
8 495 334	5 578.76	6 751.25
9 356 952	7 883.74	9 829.19

Table 9.7: Time (in seconds) to run parallel E_{PT2} calculations on the largest wave function with an increasing number of 36-core compute nodes.

Nodes	Ground state	Excited state
1	7 883.74	9 829.19
5	1 629.06	2 022.36
10	832.89	1 029.91
20	440.76	537.37
30	303.31	378.69
40	246.12	296.31
50	201.84	241.55

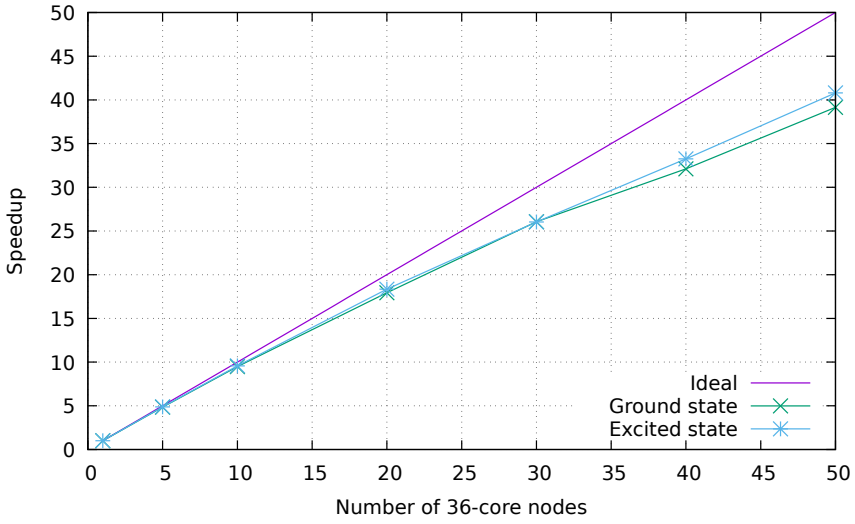


Figure 9.7: Parallel speedup for the calculation of the E_{PT2} contribution of the ground state using the largest wave function. Each node contains 36 physical CPU cores.

9.4 Matrix dressing

The algorithms for the matrix dressing are similar to those of E_{PT2}. Therefore, we expect a similar behavior with N_{det} and with the number of nodes.

The time necessary to build the dressing matrix in the Shifted-B_k method was measured on a single 36-core node with an increasing number of determinants (table 9.8 and figure 9.8). As for E_{PT2}, the stopping criterion was an estimated relative error on $\langle\Psi|\hat{\Delta}|\Psi\rangle$, the dressed energy, equal to 0.001.

Using the log-log plot, the scaling with the number of determinants was found to be $\mathcal{O}\left({N_{\text{det}}}^{1.15}\right)$. This scaling is slightly higher than the $\mathcal{O}\left({N_{\text{det}}}^{1.07}\right)$ measured for E_{PT2}, but close to linear, as expected. The additional overhead is due to the handling of the checkpoints which eliminates the communication bottleneck.

Then, the parallel speedup was measured using the wave function with 9 356 952 determinants. The results are plotted in figure 9.9.

Table 9.8: Single-node (36-core) Shifted-B_k iteration for increasingly large wave functions. Time is given in seconds.

N_{det}	Ground state	Excited state
123	0.80	0.78
3 083	4.06	5.38
29 409	20.58	28.81
168 595	188.58	204.48
1 322 537	1 871.66	2 123.38
9 356 952	19 881.30	22 082.60

Nodes	Ground state	Excited state
1	19 881.30	22 082.60
5	4 015.11	4 445.86
10	2 038.46	2 267.52
20	1 063.41	1 176.05
30	738.96	814.86
40	589.65	633.00
50	514.53	544.50

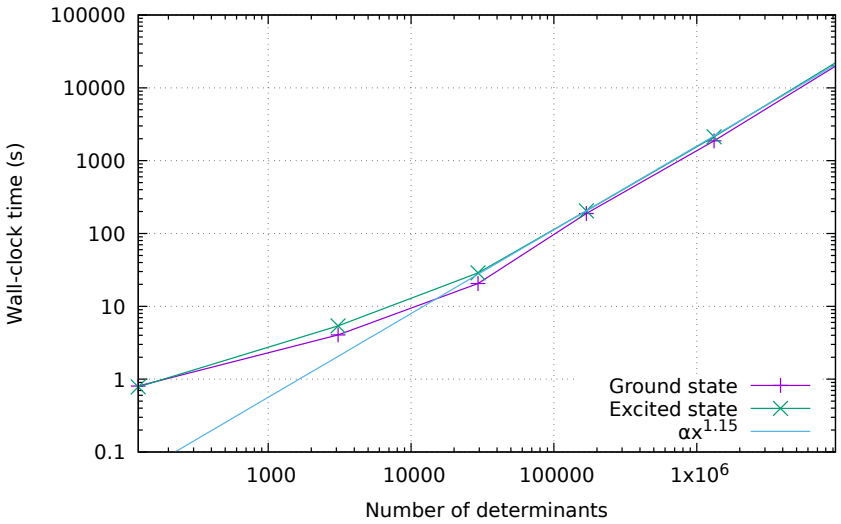


Figure 9.8: Wall clock time required to compute the Shifted-B_k dressing for the ground and the excited states, as a function of the number of determinants in the wave function.

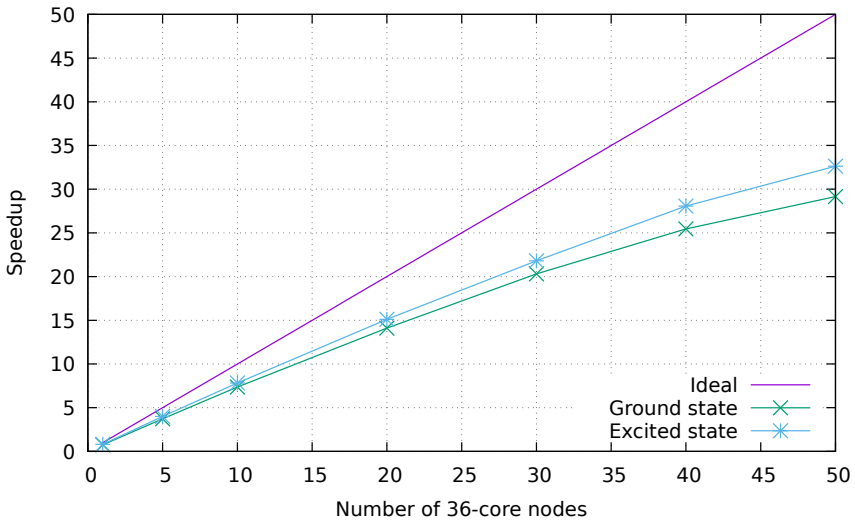


Figure 9.9: Parallel speedup for the calculation of the matrix dressing of the ground state using the largest wave function. Each node contains 36 physical CPU cores.

Chapter 10

Applications

Contents

In this chapter, we present applications that were made possible thanks by the optimized CIPSI implementation presented in this thesis, and the hybrid stochastic-deterministic algorithm. We first present a difficult benchmark of excited states, and then the use of CIPSI wave functions for quantum Monte Carlo calculations.

10.1 Excited states benchmark

On figure 9.1, one can see that both E_{var} and $E_{\text{var}} + E_{\text{PT2}}$ converge to the FCI energy when the number of determinants increases. A convenient extrapolation introduced by Holmes *et al*[85] is E_{var} as a function of E_{PT2}. Indeed, at the FCI limit $E_{\text{PT2}} = 0$ and $E_{\text{var}} = E_{\text{FCI}}$. Such an extrapolation was used to estimate the FCI energies of the ground and excited states of the molecules of the benchmark.

Another important point is to obtain a balanced description of both states, such that the errors due to the approximations cancel nicely when looking at energy differences. A way to achieve this goal is to select determinants for all states simultaneously in a state-averaged fashion. Here, our selection criterion for the external determinants was to take the maximum of the contribution ϵ_α among all the considered states.

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 4360−4379

pubs.acs.org/JCTC

A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks

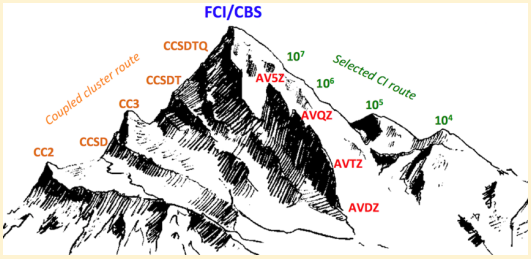
Pierre-François Loos,*[,†] Anthony Scemama,[†] Aymeric Blondel,[‡] Yann Garniron,[†] Michel Caffarel,[†] and Denis Jacquemin*[,‡]

[†]Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, 31013 Toulouse Cedex 6, France

[‡]Laboratoire CEISAM - UMR CNRS 6230, Université de Nantes, 2 Rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France

Supporting Information

ABSTRACT: Striving to define very accurate vertical transition energies, we perform both high-level coupled cluster (CC) calculations (up to CCSDTQP) and selected configuration interaction (sCI) calculations (up to several millions of determinants) for 18 small compounds (water, hydrogen sulfide, ammonia, hydrogen chloride, dinitrogen, carbon monoxide, acetylene, ethylene, formaldehyde, methanimine, thioformaldehyde, acetaldehyde, cyclopropene, diazomethane, formamide, ketene, nitrosomethane, and the smallest streptocyanine). By systematically increasing the order of the CC expansion, the number of determinants in the CI expansion as well as the size of the one-electron basis set, we have been able to reach near full CI (FCI) quality transition energies. These calculations are carried out on CC3/*aug*-cc-pVTZ geometries, using a series of increasingly large atomic basis sets systematically including diffuse functions. In this way, we define a list of 110 transition energies for states of various characters (valence, Rydberg, n → π^*, $\pi \rightarrow \pi^*$, singlet, triplet, etc.) to be used as references for further calculations. Benchmark transition energies are provided at the *aug*-cc-pVTZ level as well as with additional basis set corrections, in order to obtain results close to the complete basis set limit. These reference data are used to benchmark a series of 12 excited-state wave function methods accounting for double and triple contributions, namely ADC(2), ADC(3), CIS(D), CIS(D_∞), CC2, STEOM-CCSD, CCSD, CCSDR(3), CCSDT-3, CC3, CCSDT., and CCSDTQ. It turns out that CCSDTQ yields a negligible difference with the extrapolated CI values with a mean absolute error as small as 0.01 eV, whereas the coupled cluster approaches including iterative triples are also very accurate (mean absolute error of 0.03 eV). Consequently, CCSDT-3 and CC3 can be used to define reliable benchmarks. This observation does not hold for ADC(3) that delivers quite large errors for this set of small compounds, with a clear tendency to overcorrect its second-order version, ADC(2). Finally, we discuss the possibility to use basis set extrapolation approaches so as to tackle more easily larger compounds.



1. INTRODUCTION

Defining an effective method reliably providing accurate excited-state energies and properties remains a major challenge in theoretical chemistry. For practical applications, the most popular approaches are the complete active space self-consistent field (CASSCF)[1,2] and the time-dependent density functional theory (TD-DFT)[3,4] methods for systems dominated by static and dynamic electron correlation effects, respectively. When these schemes are not sufficiently accurate, one often uses methods including second-order perturbative corrections. For CASSCF, a natural choice is CASPT2,[5] but this method rapidly becomes impractical for large compounds. If a single-reference method is sufficient, the most popular second-order approaches are probably the second-order algebraic diagrammatic construction, ADC(2),[6] and the second-order coupled cluster, CC2, methods,[7,8] that both offer an attractive $O(N^5)$ scaling (where N is the number of basis functions) allowing applications up to systems comprising ca. 100 atoms. Compared to TD-DFT,[9] these approaches have the indisputable advantage of being free of the choice of a specific exchange-correlation functional. Using ADC(2) or CC2 generally provides more systematic errors with respect to reference values than TD-DFT, although the improvements in terms of error magnitude are often rather moderate (at least for valence singlet states).[10−12] Importantly, both ADC(n) and CCn offer a systematic pathway for improvement via an increase of the expansion order n. For example, using CCSD, CCSDT, CCSDTQ, etc., allows to check the quality of the obtained estimates. However, in practice, one can only contemplate such systematic approach and the ultimate choice of a method for excited-state calculations is often guided by previous benchmarks. These benchmark studies are either performed using experimental or theoretical reference values. While the former approach allows in principle to rely on an

Received: April 28, 2018
Published: July 2, 2018

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

almost infinite pool of reference data, most measurements are performed in solution and provide absorption bands that can be compared to theory only with the use of extra approximations for modeling environmental and vibronic effects. In addition, the most accurate experimental data are obtained for 0−0 energies, whereas obtaining trustworthy experimental estimates of vertical transition energies is an extremely difficult task, generally requiring to back-transform spectroscopic vibronic data through a numerical process,[13] an approach that is typically only applicable to diatomics. Consequently, it is easier to use first-principle reference values as benchmarks, as they allow to assess theoretical methods more consistently (vertical values, same geometries, no environmental effects, etc.). This is well illustrated by the recent contribution of Schwabe and Goerigk,[14] who decided to compute third-order response CC (CC3)[15,16] reference values instead of using the previously collected experimental values for the test set originally proposed by Gordon's group.[17]

While many benchmark sets have been proposed for excited states,[10,11,17−29] the most praised database of theoretical excited state energies is undoubtedly the one set up by Thiel and his co-workers. In 2008, they proposed a large set of theoretical best estimates (TBE) for 28 small and medium CNOH organic compounds.[30] More precisely, using some literature values but mainly their own CC3/TZVP and CASPT2/TZVP results computed on MP2/6-31G(d) geometries, these authors determined 104 singlet and 63 triplet reference excitation energies. The same group soon proposed *aug*-cc-pVTZ TBE for the same set of compounds,[31,32] though some CC3/*aug*-cc-pVTZ reference values were estimated by a basis set extrapolation technique. In their conclusion, they stated that they "*expect this benchmark set to be useful for validation and development purposes, and anticipate future improvements and extensions of this set through further high-level calculations*".[30] The first prediction was soon realized. Indeed, both the TZVP and *aug*-cc-pVTZ TBE were applied to benchmark various computationally effective methods, including semiempirical approaches,[33−35] TD-DFT,[24,25,36−46] the second-order polarization propagator approximation (SOPPA),[47] ADC(2),[48] the second order *N*-electron valence perturbation theory (NEVPT2),[49] the random phase approximation (RPA),[50] as well as several CC variants.[51−56] In contrast, even a decade after the original work appeared, the progresses aiming at improving and/or extending Thiel's set have been much less numerous. To the best of our knowledge, these extensions are limited to the more compact TZVP basis set,[48,52,57,58] but in one case.[59] This diffuse-less basis set offers clear computational advantages and avoids some state mixing. However, it has a clear tendency to overestimate transition energies, especially for Rydberg states, and it makes comparisons between methods more difficult as basis set dependencies are significantly different in wave function-based and density-based methods.[60]

Let us now briefly review these efforts. In 2013, Watson et al. obtained with the TZVP basis set and CCSDT-3—a method employing an iterative approximation of the triples — transition energies very similar to the CC3 values.[57] Nevertheless, as noted the same year by Nooijen and co-workers who also reported CCSDT-3/TZVP values,[52] "*the relative accuracy of EOM-CCSDT-3 versus CC3 compared to full CI (or EOM-CCSDT) is not well established*". In 2014, Dreuw and co-workers performed ADC(3) calculations on Thiel's set and concluded that "*based on the quality of the existing benchmark set it is practically not possible to judge whether ADC(3) or CC3 is more accurate*". The same year, Kannar and Szalay, revisited Thiel's set and proposed CCSDT/TZVP reference energies for 17 singlet states of six molecules.[58] Recently the same group reported CCSDT/*aug*-cc-pVTZ transition energies for valence and Rydberg states of five compact molecules,[59] and used these values to benchmark several simpler CC approaches. To the best of our knowledge, these stand as the highest-level values reported to date. However, it remains difficult to know if these CCSDT transition energies are significantly more accurate than their CC3, CCSDT-3 or ADC(3) counterparts. Indeed, for the $\pi \rightarrow \pi^*$ valence singlet excited state of ethylene, the CC3/TZVP, CCSDT/TZVP and CCSDTQ/TZVP estimates of 8.37, 8.38, and 8.36 eV (respectively) are nearly identical.[58]

Herein, we propose to continue the quest for ultra-accurate excited-state reference energies. First, although this prevents direct comparisons with previously published data, we decided to use more accurate CC3/*aug*-cc-pVTZ geometries for all the compounds considered here. Second, we employ only diffuse-containing Dunning basis sets to be reasonably close from the complete basis set limit. Third, we climb the mountain via two faces following: (i) the CC route (up to the highest computationally possible order) and (ii) the configuration interaction (CI) route with the help of selected CI (sCI) methods. By comparing the results of these two approaches, it is possible to get some reliable information about how far our results are from the full CI (FCI) ones. Fourth, in order not to limit our investigation to vertical absorption, we also report, in a few cases, fluorescence energies. Of course, such extreme choices impose drastic restrictions on the size of the molecules one can treat. However, we claim here that they allow to accurately estimate the FCI result for most excited states.

2. COMPUTATIONAL DETAILS

2.1. Geometries. All geometries are obtained at the CC3/*aug*-cc-pVTZ level without applying the frozen core approximation. These geometries are available in the Supporting Information. While several structures are extracted from ref 61 (acetylene, diazomethane, ethylene, formaldehyde, ketene, nitrosomethane, thioformaldehyde and streptocyanine-C1), additional optimizations are performed here following the same protocol as in that earlier work. First, we optimize the structures and compute the vibrational spectra at the CCSD/def2-TZVPP level[62] with Gaussian16.[63] These calculations confirm the minima nature of the obtained geometries.[64] We then reoptimize the structures at the CC3/*aug*-cc-pVTZ level[15,16] using Dalton[65] and/or CFOUR,[66] depending on the size and symmetry of the molecule. CFOUR advantageously provides analytical CC3 gradients for ground-state structures. For the CCSD calculations, the energy and geometry convergence thresholds are systematically tightened to 10^{-10}−10^{-11} au for the SCF energy, 10^{-8}−10^{-9} au for the CCSD energy, and 10^{-7}−10^{-8} au for the EOM-CCSD energy in the case of excited-state geometry optimizations. To check that the structures correspond to genuine minima, the (EOM-)CCSD gradients are differentiated numerically to obtain the vibrational frequencies. The CC3 optimizations are performed with the default convergence thresholds of Dalton or CFOUR without applying the frozen core approximation.

2.2. Coupled Cluster Calculations. Unless otherwise stated, the CC transition energies[67] are computed in the frozen-core approximation (large cores for Cl and S). We use several codes to achieve our objectives, namely CFOUR,[66]

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 1. Vertical Transition Energies for the Three Lowest Singlet and Three Lowest Triplet Excited States of Water (Top), the Four Lowest Singlet and the Lowest Triplet States of Ammonia (Center), and the Lowest Singlet State of Hydrogen Chloride (Bottom)[m]

water															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			lit.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[a]	th.[b]	th.[c]
$^1B_1(n \rightarrow 3s)$	7.51	7.50	7.53	7.53	7.53	7.60	7.59	7.62	7.62	7.65	7.64	7.68	7.41	7.81	7.57
$^1A_2(n \rightarrow 3p)$	9.29	9.28	9.31	9.32	9.32	9.38	9.37	9.40	9.41	9.43	9.41	9.46	9.20	9.30	9.33
$^1A_1(n \rightarrow 3s)$	9.92	9.90	9.94	9.94	9.94	9.97	9.95	9.98	9.99	10.00	9.98	10.02	9.67	9.91	9.91
$^3B_1(n \rightarrow 3s)$	7.13	7.11	7.14	7.14	7.14	7.23	7.22	7.24	7.25	7.28	7.26	7.30	7.20	7.42	7.21
$^3A_2(n \rightarrow 3p)$	9.12	9.11	9.14	9.14	9.14	9.22	9.20	9.23	9.24	9.26	9.25	9.28	8.90	9.42	9.19
$^3A_1(n \rightarrow 3s)$	9.47	9.45	9.48	9.49	9.49	9.52	9.50	9.53	9.54	9.56	9.54	9.58	9.46	9.78	9.50
hydrogen sulfide															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			lit.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[d]	exp.[e]	th.[f]
$^1A_2(n \rightarrow 4p)$	6.29	6.29	6.29	6.29	6.29	6.19	6.18	6.18	6.18	6.16	6.15	6.15			6.12
$^1B_1(n \rightarrow 4s)$	6.10	6.10	6.10	6.10	6.10	6.24	6.24	6.24	6.24	6.29	6.29	6.29	6.33		6.27
$^3A_2(n \rightarrow 4p)$	5.91	5.90	5.90	5.90	5.90	5.82	5.81	5.81	5.81	5.80	5.79	5.79		5.8	5.78
$^3B_1(n \rightarrow 4s)$	5.75	5.75	5.75	5.75	5.75	5.88	5.88	5.88	5.89	5.93	5.93	5.93		5.4	5.92
ammonia															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			lit.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[g]	exp.[h]	th.[i]
$^1A_2(n \rightarrow 3s)$	6.46	6.46	6.48	6.48	6.48	6.57	6.57	6.59	6.59	6.61	6.61	6.64	6.38	6.39	6.48
$^1E(n \rightarrow 3p)$	8.06	8.06	8.08	8.08	8.08	8.15	8.14	8.16	8.16	8.18	8.17	8.22	7.90	7.93	8.02
$^1A_1(n \rightarrow 3p)$	9.66	9.66	9.68	9.68	9.68	9.32	9.31		9.33	9.11	9.10	9.14	8.14	8.26	8.50
$^1A_2(n \rightarrow 4s)$	10.40	10.39	10.41	10.41	10.41	9.95	9.94		9.96	9.77	9.77				9.03
$^3A_2(n \rightarrow 3s)$	6.18	6.18	6.19	6.19	6.19	6.29	6.29	6.30	6.31	6.33	6.33	6.35	6.02[j]		
hydrogen chloride															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			lit.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	th.[k]		
$^1\Pi$ (CT)	7.82	7.81	7.82	7.82	7.82	7.84	7.83	7.84	7.84	7.89	7.88[l]	7.88	8.23		

[a]Energy loss experiment from ref 98. [b]MRCI+Q/*aug*-cc-pVTZ calculations from ref 99. [c]MRCC/*aug*-cc-pVTZ calculations from ref 100. [d]VUV experiment from ref 101. [e]Electron impact experiment from ref 102. [f]CASPT2/d-*aug*-cc-pVQZ results from ref 103. [g]Electron impact experiment from ref 104. [h]Electron impact experiment from ref 105. [i]EOM-CCSD($\tilde{T}$)/*aug*-cc-pVTZ with extra *diffuse* calculations from ref 106. [j]Deduced from the 6.38 eV value of the $^1A_2(n \rightarrow 3s)$ state and the −0.36 eV shift reported for the 0−0 energies compared to the corresponding singlet state in ref 107, a splitting consistent with an earlier estimate of −0.39 eV given in ref 108. [k]CC2/cc-pVTZ from ref 22.; [l]The CCSDTQ/*aug*-cc-pVQZ value is 7.88 eV as well. [m]All states of water and ammonia have a Rydberg character, whereas the lowest state of hydrogen chloride is a charge-transfer state. All values are in eV.

Dalton,[65] Gaussian16,[63] Orca,[68] MRCC,[69,70] and Q-Chem.[71] Globally, we use CFOUR for both CCSDT-3[72,73] and CCSDT[74] calculations, Dalton to perform the CIS(D),[75,76] CC2,[7,8] CCSD,[62] CCSDR(3),[77] and CC3[15,16] calculations, Gaussian for the CIS(D)[75,76] and CCSD,[62] Orca for the similarity-transformed EOM-CCSD (STEOM-CCSD)[56,78] calculations, Q-Chem for ADC(2) and ADC(3) calculations, and MRCC for the CIS(D_∞),[79] CCSDT,[74] and CCSDTQ[80] (and higher) calculations. As we mainly report transition energies, it is worth noting that the linear-response (LR) and equation-of-motion (EOM) formalisms provide identical results. Nevertheless, the oscillator strengths characterizing the excited states are obtained at the (LR) CC3 level with Dalton. Default program setting are generally applied, and when modified they are tightened. For the STEOM-CCSD calculations which relies on natural transition orbitals, it was checked that each state is characterized by an active character percentage of 98% or larger (states not matching this criterion are not reported). Nevertheless, the obtained results do slightly depend on the number of states included in the calculations, and we found typical variations of ±0.01−0.05 eV. For all calculations, we use the well-known Dunning's *aug*-cc-pVXZ (X = D, T, Q and 5) atomic basis sets, as well as some doubly- and triply augmented basis sets of the same series (d-*aug*-cc-pVXZ and t-*aug*-cc-pVXZ).

2.3. Selected Configuration Interaction Methods.

Alternatively to CC, we also compute transition energies using a selected CI (sCI) approach, an idea that goes back to 1969 in the pioneering works of Bender and Davidson,[81] and Whitten and Hackmeyer.[82] Recently, sCI methods have demonstrated their ability to reach near FCI quality energies for small organic and transition metal-containing molecules.[83−92] To avoid the exponential increase of the size of the CI expansion, we employ the sCI algorithm CIPSI[83,93,94] (Configuration Interaction using a Perturbative Selection made Iteratively) to retain only the energetically relevant determinants. To do so, the CIPSI algorithm uses a second-order energetic criterion to select perturbatively determinants in the FCI space.[83,85,87,92] In the numerical examples presented below, our CI expansions contain typically about a few millions of determinants. We refer the interested readers to refs 92 and

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

95 for more details about the general philosophy of sCI methods.

In order to treat the electronic states of a given spin manifold on equal footing, a common set of determinants is used for all states. Moreover, to speed up convergence to the FCI limit, a common set of natural orbitals issued from a preliminary (smaller) sCI calculation is employed. All sCI calculations have been performed in the frozen-core approximation. For a given basis set, we estimate the FCI limit using the approach introduced recently by Holmes et al.[90] in the context of the (selected) heat-bath CI method, and used with success, even for challenging chemical situations.[89,91,92] More precisely, we linearly extrapolate the sCI energy E_{sCI} as a function of E_{PT2}, which is an estimate of the truncation error in the sCI algorithm, i.e., $E_{PT2} \approx E_{FCI} - E_{sCI}$. When $E_{PT2} = 0$, the FCI limit has effectively been reached. Here, E_{PT2} is efficiently evaluated with a recently proposed hybrid stochastic-deterministic algorithm.[96] Note that we do not report error bars because the statistical errors originating from this algorithm are orders of magnitude smaller than the extrapolation errors. In practice, the extrapolation is based on the two largest sCI wave functions; i.e., we perform a two-point extrapolation, which is justified here because of the quasi-linear behavior of the sCI energy as a function of E_{PT2}. Estimating the extrapolation error is a complicated task with no well-defined method to do so. In practice, we have observed that this extrapolation procedure is robust and provides FCI estimates within ±0.02 eV. When the convergence to the FCI limit is too slow to provide reliable estimates, the number of significant digits reported has been reduced accordingly. From herein, the extrapolated FCI results are simply labeled exFCI. Several illustrative examples are reported in Supporting Information where we compare different types of extrapolations for several molecules (see Figure S1 and Table S11). In particular, diazomethane and streptocyanine-C1 can be considered as "difficult" cases (*vide infra*), and the results reported in Supporting Information show that, even in these challenging situations, the two-point linear extrapolation is fairly robust. Moreover, additional points do not significantly alter the exFCI estimates (typically 0.01 eV or less).

All the sCI calculations are performed with the electronic structure software QUANTUM PACKAGE, developed in Toulouse and freely available.[97] Additional information about the sCI wave functions, excitations energies as well as their extrapolated values can be found at the end of the Supporting Information.

3. RESULTS AND DISCUSSION

In the discussion below, we first discuss specific molecules of increasing size and compare the results obtained with exFCI and CC approaches, starting with the CC3 method for the latter. This first part is performed applying systematically the frozen-core approximation. We next define two series of TBE, one at the frozen-core *aug*-cc-pVTZ level, and one close to complete basis set limit by applying corrections for frozen-core and basis set effects. In a following stage, we assess the performances of several popular wave function methods using the former benchmark as reference. Finally, we discuss the performances of basis set extrapolation approaches starting from a compact basis. Unless otherwise stated, we considered the exFCI values as benchmarks.

3.1. Water, Hydrogen Sulfide, Ammonia, and Hydrogen Chloride. Because of its small size and ubiquitous role in life, water is often used as a test case for Rydberg excitations. Indeed, it is part of Head−Gordon's,[21] Gordon's[17] and Truhlar−Gagliardi's[29] data sets of compounds, and it has been investigated at many levels of theory.[99,100,103,109] Our results are collected in Table 1. With the *aug*-cc-pVDZ basis, there is an nearly perfect agreement between the exFCI values and the transition energies obtained with the two largest CC expansions, namely CCSDTQ and CCSDTQP. Indeed, the largest discrepancy is as small as 0.01 eV, and it is therefore reasonable to state that the FCI limit has been reached with that specific basis set. Compared to the exFCI results, the CCSDT values are systematically too low, with an average error of −0.03 eV. The same trend of underestimation is found with CC3, though with smaller absolute deviations for all states. Unsurprisingly, for Rydberg states, increasing the basis set size has a significant impact, and it tends to increase the computed transition energies in water. However, this effect is very similar for all methods listed in Table 1. This means that, on the one hand, the tendency of CCSDT to provide slightly too small transition energies pertains with both *aug*-cc-pVTZ and *aug*-cc-pVQZ, and, on the other hand, that estimating the basis set effect with a "cheap" method is possible. Indeed, adding to the exFCI/*aug*-cc-pVDZ energies, the difference between CC3/*aug*-cc-pVQZ and CC3/*aug*-cc-pVDZ results would deliver estimates systematically within 0.01 eV of the actual exFCI/*aug*-cc-pVQZ values. Such basis set extrapolation approach was already advocated for lower-order CC expansions,[31,110] and it is therefore not surprising that it can be applied with refined models. As it can be seen in Table S1 in the Supporting Information, further extension of the basis set or correlation of the 1s electron have small impacts, except for the Rydberg $^1A_1$ state. Eventually, as evidenced by the data from the rightmost columns of Table 1, the present estimates are in good agreement with previous MRCC values determined on the experimental geometry,[100] whereas the experimental values offer qualitative comparisons only, for reasons discussed elsewhere.[98] We underline that some of the 2013 measurements reported in Table 1 significantly differ from previous electron impact data,[111] that were used previously as references,[17] with, e.g., a 0.2 eV discrepancy between the two experiments for the lowest triplet state.

As water, hydrogen sulfide was also the subject of several high-level theoretical investigations,[103,112−114] which are necessary because there are either no (lowest $^1A_2$ state) or only a few experimental data available for the Rydberg states of H_2S,[101,102,115,116] especially as no accurate value could be measured for the first $^1A_2$ state. As can be seen in Table 1, for a given basis set all tested CC methods provide very similar results, systematically within 0.01 eV of the exFCI results. In contrast, the basis set has a significant impact, e.g., the two lowest singlet states switch order when going from *aug*-cc-pVDZ to *aug*-cc-pVTZ and the same is true for the two lowest triplet states. Our results are also very consistent with the CASPT2/d-*aug*-cc-pVQZ values given in ref 103, confirming that a near FCI limit has been reached.

Ammonia is also another popular molecule for evaluating Rydberg excitations, and it was previously investigated at several levels of theory.[14,21,106,117] As in the case of water, we note a nearly perfect match between the CCSDTQ and exFCI estimates with both the *aug*-cc-pVDZ and *aug*-cc-pVTZ atomic basis sets, indicating that the FCI limit is reached. Both CC3 and CCSDT are close to this limit, and the former model slightly outperforms the latter. For ammonia, the basis set

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 2. Vertical Transition Energies for Various Excited States of Dinitrogen (Top) and Carbon Monoxide (Bottom)[i]

dinitrogen															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			litt.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[a]	exp.[b]	th.[c]
$^1\Pi_g(n \rightarrow \pi^*)$	9.44	9.41	9.41	9.41	9.41	9.34	9.33	9.32	9.34	9.33	9.31	9.34	9.31	9.31	9.27
$^1\Sigma_u^- (\pi \rightarrow \pi^*)$	10.06	10.06	10.06	10.05	10.05	9.88	9.89	9.88	9.88	9.87	9.88	9.92	9.92	9.92	10.09
$^1\Delta_u(\pi \rightarrow \pi^*)$	10.43	10.44	10.43	10.43	10.43	10.29	10.30		10.29	10.27	10.28	10.31	10.27	10.27	10.54
$^1\Sigma_g^+$ (R)	13.23	13.20	13.18	13.18	13.18	13.01	13.00	12.97	12.98	12.90	12.89	12.89		12.2	12.20
$^1\Pi_u$ (R)	13.28	13.17	13.13	13.13	13.12	13.22	13.14	13.09	13.03	13.17		13.1[d]	12.78	12.90	12.84
$^1\Sigma_u^+$ (R)	13.14	13.13	13.11	13.11	13.11	13.12	13.12	13.09	13.09	13.09	13.09	13.2[d]	12.96	12.98	12.82
$^1\Pi_u$ (R)	13.64	13.59	13.56	13.56	13.56	13.49	13.45	13.42	13.46	13.42	13.37	13.7[d]	13.10	13.24	13.61
$^3\Sigma_u^+(\pi \rightarrow \pi^*)$	7.67	7.68	7.69	7.70	7.70	7.68	7.69	7.70	7.70	7.71	7.71	7.74	7.75	7.75	7.56
$^3\Pi_g(n \rightarrow \pi^*)$	8.07	8.06	8.05	8.05	8.05	8.04	8.03	8.02	8.01	8.04	8.04	8.03	8.04	8.04	8.05
$^3\Delta_u (\pi \rightarrow \pi^*)$	8.97	8.96	8.96	8.96	8.96	8.87	8.87	8.87	8.87	8.87	8.87	8.88	8.88	8.88	8.93
$^3\Sigma_u^- (\pi \rightarrow \pi^*)$	9.78	9.76	9.75	9.75	9.75	9.68	9.68	9.66	9.66	9.68		9.66	9.67	9.67	9.86
carbon monoxide															
	aug-cc-pVDZ					*aug*-cc-pVTZ				*aug*-cc-pVQZ			litt.		
state	CC3	CCSDT	CCSDTQ	CCSDTQP	exFCI	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[e]	th.[f]	th.[g]
$^1\Pi(n \rightarrow \pi^*)$	8.57	8.57	8.56	8.56	8.57	8.49	8.49	8.48	8.49	8.47	8.48	8.50	8.51	8.54	8.83
$^1\Sigma^- (\pi \rightarrow \pi^*)$	10.12	10.06	10.06	10.06	10.05	9.99	9.94	9.93	9.92	9.99	9.94	9.99	9.88	10.05	9.97
$^1\Delta (\pi \rightarrow \pi^*)$	10.23	10.18	10.17	10.17	10.16	10.12	10.08	10.07	10.06	10.12	10.07	10.11	10.23	10.18	10.00
$^1\Sigma^+$ (R)	10.92	10.94	10.93	10.92	10.94	10.94	10.99	10.96	10.95	10.90	10.95	10.96	10.78	10.98	
$^1\Sigma^+$ (R)	11.48	11.52	11.51	11.51	11.52	11.49	11.54	11.52	11.52	11.46	11.51	11.53	11.40		
$^1\Pi$ (R)	11.74	11.77	11.76	11.75	11.76	11.69	11.74	11.72	11.72	11.63	11.69	11.70	11.53		
$^3\Pi(n \rightarrow \pi^*)$	6.31	6.30	6.29	6.28	6.29	6.30	6.30	6.28	6.28	6.30	6.30	6.29	6.32		6.41
$^3\Sigma^+ (\pi \rightarrow \pi^*)$	8.45	8.43	8.44	8.44	8.46	8.45	8.42	8.44	8.45	8.48	8.45	8.49	8.51		8.39
$^3\Delta (\pi \rightarrow \pi^*)$	9.37	9.33	9.34	9.34	9.33	9.30	9.26	9.26	9.27	9.31	9.26	9.29	9.36		9.23
$^3\Sigma^- (\pi \rightarrow \pi^*)$	9.89				9.83	9.82			9.80	9.82		9.78	9.88		9.60
$^3\Sigma^+$ (R)	10.39	10.42	10.42	10.41	10.41	10.45	10.50	10.48	10.47	10.44	10.49		10.4[h]		

[a]Experimental vertical values given in ref 13 and computed from the spectroscopic constants of ref 118. [b]Experimental vertical values given in ref 119 and computed from the spectroscopic constants of ref 118. [c]MRCCSD/6-311G with one additional *d* calculations from ref 119. [d]CI convergence too slow to provide estimates to 0.01 eV. [e]Experimental vertical values given in ref 120 and computed from the spectroscopic constants of ref 118. [f]CCSDT/PVTZ+ results from ref 121. [g]CASSCF(10,10)/cc-pVTZ results from ref 122. [h]Only one digit reported for that state, see ref 120. [i]R stands for Rydberg states. All values are in eV.

J. Chem. Theory Comput. 2018, 14, 4360−4379 DOI: 10.1021/acs.jctc.8b00406

Table 3. Vertical (Absorption) Transition Energies for the Five Lowest Low-Lying Valence Excited States of Acetylene (Top) and the Three Lowest Singlet and Triplet Excited States of Ethylene (Bottom)[g]

acetylene										
	aug-cc-pVDZ				*aug*-cc-pVTZ			lit.		
state	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[a]	th.[b]	th.[c]
$^1\Sigma_u^-$ ($\pi \rightarrow \pi^*$)	7.21	7.21	7.21	7.20	7.09	7.09	7.10	7.1	6.96	7.10
$^1\Delta_u$($\pi \rightarrow \pi^*$)	7.51	7.52	7.52	7.51	7.42	7.43	7.44	7.2	7.30	7.43
$^3\Sigma_u^+$ ($\pi \rightarrow \pi^*$)	5.48	5.49	5.50	5.50	5.50	5.51	5.53	5.2	5.26	5.58
$^3\Delta_u$($\pi \rightarrow \pi^*$)	6.46	6.46	6.46	6.46	6.40	6.39	6.40	6.0	6.20	6.41
$^3\Sigma_u^-$ ($\pi \rightarrow \pi^*$)	7.13	7.14	7.14	7.14	7.07		7.08	7.1	6.90	7.05
^1A_u [F]($\pi \rightarrow \pi^*$)	3.70	3.72	3.70	3.71	3.64	3.66	3.64			
$^1A_2$ [F]($\pi \rightarrow \pi^*$)	3.92	3.94	3.93	3.93	3.84	3.86	3.85			
ethylene										
	aug-cc-pVDZ				*aug*-cc-pVTZ			lit.		
state	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[d]	th.[e]	
$^1B_{3u}$($\pi \rightarrow 3s$)	7.29	7.29	7.30	7.31	7.35	7.37	7.39	7.11	7.45	
$^1B_{1u}$($\pi \rightarrow \pi^*$)	7.94	7.94	7.93	7.93	7.91	7.92	7.93	7.60	8.00	
$^1B_{1g}$($\pi \rightarrow 3p$)	7.97	7.98	7.99	8.00	8.03	8.04	8.08	7.80	8.06	
$^3B_{1u}$($\pi \rightarrow \pi^*$)	4.53	4.54	4.54	4.55	4.53	4.53	4.54	4.36	4.55	
$^3B_{3u}$($\pi \rightarrow 3s$)	7.17	7.18	7.18	7.16	7.24	7.25	*f*	6.98	7.29	
$^3B_{1g}$($\pi \rightarrow 3p$)	7.93	7.94	7.94	7.93	7.98	7.99	*f*	7.79	8.02	

[a]Electron impact experiment from ref 129. Note that the 7.1 eV value for the Σ_u^- singlet and triplet states should be viewed as a tentative assignment. [b]LS-CASPT2/*aug*-ANO calculations from ref 124. [c]MR-AQCC/extrap. calculations from ref 126. [d]Experimental values collected from various sources from ref 116. (see discussions in refs 30, 130, and 131). [e]Best composite theory from ref 131, close to FCI. [f]CI convergence too slow to provide estimates reliable to 0.01 eV. [g]For acetylene, we also report the vertical emission (denoted [F]) obtained from the lowest *trans* and *cis* isomers. All values are in eV.

effects are particularly strong for the third and fourth singlet excited states but these basis set effects are nearly transferrable from one method to another. In fact, as hinted by the large differences between the *aug*-cc-pVTZ and *aug*-cc-pVQZ results in Table 1, these two high-lying states require the use of additional diffuse orbitals to attain convergence. The CC3/t-*aug*-cc-pVQZ values of 8.60 and 9.15 eV (see Table S1 in the Supporting Information), are close from the previous results of Bartlett and co-workers,[106] who also applied extra diffuse orbitals in their calculations relying on approximate triples (see the footnotes in Table 1). As in water, the experimental values do not provide sufficiently clear-cut results to ultimately decide which method is the most accurate. Indeed, the vertical experimental estimates reported in Table 1 differ significantly from the more trustworthy adiabatic values with variations of ca. 0.5 eV.[106] Consequently, a good match between an experimental measurement and a theoretical calculation determined with a compact basis set is, in the present case, a sign of lucky cancellation of errors.

Hydrogen chloride was less frequently used in previous benchmarks, but is included in Tozer's set as an example of charge-transfer (CT) state.[22] Again, the results listed at the bottom of Table 1 demonstrate a remarkable consistency between the various theories. Though large frozen cores are used during the calculations, this does not strongly impact the results, as can be deduced from the data of Table S1. As expected, the absorption band corresponding to this CT state is very broad experimentally (starting at 5.5 eV and peaking at 8.1 eV),[118] making direct comparisons tricky.

3.2. Dinitrogen and Carbon Monoxide. Dinitrogen is a simple diatomic compound for which the low-lying valence and Rydberg states have been investigated at several levels of theory.[13,22,119,121] With a numerical solution of the nuclear Schrödinger equation, it is possible to treat the experimental spectroscopic constants,[118] so as to obtain reliable vertical estimates, and this procedure was applied previously.[13,119,123] While such approach is supposedly providing experimental vertical excited-state energies with a ca. 0.01 eV error only, it remains that significant excitation energy differences have been reported for the two lowest $^1\Pi_u$ states (see Table 2). As in the previous cases, we find a remarkable agreement between the CCSDTQ and exFCI estimates for most cases in which both could be determined. The only exceptions are the two $^1\Pi_u$ states with the *aug*-cc-pVTZ basis, but in these two cases, the CC expansion is also converging more slowly than usual, which is consistent with the relatively small degree of single excitation character in these two states (82.9 and 87.4% according to CC3). In contrast to water and ammonia, CCSDT outperforms CC3 with respective mean absolute deviation (MAD) compared to exFCI of 0.02 and 0.04 eV, when using the *aug*-cc-pVDZ basis set. As it can be deduced from Table S2 in the Supporting Information, the basis set corrections are negligible for all valence states, but significant for some of the Rydberg states, especially $^1\Sigma_g^+$, that requires two sets of diffuse orbitals to be reasonably close from the basis set limit. Applying CC3/d-*aug*-cc-pV5Z corrections to the most accurate exFCI data, once can determine TBE values (*vide infra*) that deviate only by 0.02 eV on (absolute) average compared to the experimental estimates for the seven valence states of dinitrogen. Considering the expected inaccuracy of 0.01 eV of the reference values, chemical accuracy is obviously reached without any experimental input. The deviations are about twice larger for the Rydberg states. Nevertheless, for the two $^1\Pi_u$ states, our TBE values, determined on the basis of exFCI/*aug*-cc-pVTZ results are 12.73 and 13.27 eV (*vide infra*). This indicates that for the lowest $^1\Pi_u$ state the estimate of ref 13

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 4. Vertical (Absorption) Transition Energies for Various Excited States of Formaldehyde (Top), Methanimine (Center), and Thioformaldehyde (Bottom)[h]

formaldehyde										
	aug-cc-pVDZ				*aug*-cc-pVTZ			lit.		
state	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[a]	th.[b]	th.[c]
$^1A_2(n \rightarrow \pi^*)$	4.00	3.99	4.00	3.99	3.97	3.95	3.98	4.07	3.98	3.88
$^1B_2(n \rightarrow 3s)$	7.05	7.04	7.09	7.11	7.18	7.16	7.23	7.11	7.12	
$^1B_2(n \rightarrow 3p)$	8.02	8.00	8.04	8.04	8.07	8.07	8.13	7.97	7.94	8.11
$^1A_1(n \rightarrow 3p)$	8.08	8.07	8.12	8.12	8.18	8.16	8.23	8.14	8.16	
$^1A_2(n \rightarrow 3p)$	8.65	8.63	8.68	8.65	8.64	8.61	8.67	8.37	8.38	
$^1B_1(\sigma \rightarrow \pi^*)$	9.31	9.29	9.30	9.29	9.19	9.17	9.22		9.32	9.04
$^1A_1(\pi \rightarrow \pi^*)$	9.59	9.59	9.54	9.53	9.48	9.49	9.43		9.83	9.29
$^3A_2(n \rightarrow \pi^*)$	3.58	3.57	3.58	3.58	3.57	3.56	3.58	3.50		3.50
$^3A_1(\pi \rightarrow \pi^*)$	6.09	6.08	6.09	6.10	6.05	6.05	6.06	5.86		5.87
$^3B_2(n \rightarrow 3s)$	6.91	6.90	6.95	6.95	7.03	7.02	7.06	6.83		
$^3B_2(n \rightarrow 3p)$	7.84	7.82	7.86	7.87	7.92	7.90	7.94	7.79		
$^3A_1(n \rightarrow 3p)$	7.97	7.95	8.00	8.01	8.08	8.06	8.10	7.96		
$^3B_1(n \rightarrow 3d)$	8.48	8.47	8.48	8.48	8.41	8.40	8.42			
$^1A''[F](n \rightarrow \pi^*)$	2.87	2.84	2.86	2.86	2.84	2.82	2.80			
methanimine										
	aug-cc-pVDZ				*aug*-cc-pVTZ			lit.		
state	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	th.[d]	th.[e]	
$^1A''(n \rightarrow \pi^*)$	5.26	5.24	5.25	5.25	5.20	5.19	5.23	5.32	5.18	
$^3A''(n \rightarrow \pi^*)$	4.63	4.63	4.63	4.63	4.61	4.61	4.65			
thioformaldehyde										
	aug-cc-pVDZ				*aug*-cc-pVTZ			lit.		
state	CC3	CCSDT	CCSDTQ	exFCI	CC3	CCSDT	exFCI	exp.[a]	exp.[f]	
$^1A_2(n \rightarrow \pi^*)$	2.27	2.25	2.26	2.26	2.23	2.21	2.22		2.03	
$^1B_2(n \rightarrow 4s)$	5.80	5.80	5.82	5.83	5.91	5.89	5.96	5.85	5.84	
$^1A_1(\pi \rightarrow \pi^*)$	6.62	6.60	6.51	6.5[g]	6.48	6.47	6.4[g]	6.2	5.54	
$^3A_2(n \rightarrow \pi^*)$	1.97	1.96	1.96	1.97	1.94	1.93	1.94		1.80	
$^3A_1(\pi \rightarrow \pi^*)$	3.43	3.43	3.44	3.45	3.38	3.38	3.43	3.28		
$^3B_2(n \rightarrow 4s)$	5.64	5.63	5.65	5.66	5.72	5.71	5.6[g]			
$^1A_2[F](n \rightarrow \pi^*)$	2.00	2.00	1.98	1.98	1.97	1.98	1.95			

[a]Various experimental sources, summarized in ref 116. [b]MR-AQCC-LRT calculations from ref 134. [c]CC3/*aug*-cc-pVQZ calculations from ref 30. [d]DMC results form ref 135. [e]CCSDT/*aug*-cc-pVTZ calculations from ref 59. [f]0−0 energies collected in ref 136. [g]CI convergence too slow to provide reliable estimates. [h]All values are in eV.

(12.78 eV) is probably more accurate than the one of ref 119 (12.90 eV), whereas the opposite is likely true for the highest $^1\Pi_u$ state that was reported to be located at 13.10 and 13.24 eV in refs 13 and 119, respectively. One could argue that reaching agreement between CI and CC is particularly challenging for these two states. However, performing the basis set extrapolation starting from the CCSDTQP/*aug*-cc-pVDZ results would yield similar TBE of 12.77 and 13.22 eV.

For the isoelectronic carbon monoxide, experimental vertical energies deduced from rovibronic data[118] using a numerical approach are also available.[22,120] With the *aug*-cc-pVTZ (*aug*-cc-pVQZ) atomic basis set, the CCSDT and CC3 results are within 0.02 eV (0.03 eV) and 0.03 eV (0.03 eV) of the exFCI results, whereas the errors made by both CCSDTQ and CCSDTQP are again trifling. As for dinitrogen, all the valence states are rather close from the basis set limit with *aug*-cc-pVTZ, whereas larger basis sets are required for the Rydberg states (Table S2). By correcting the exFCI/*aug*-cc-pVQZ (exFCI/*aug*-cc-pVTZ for the highest triplet state) data with basis set effects determined at the CC3/d-*aug*-cc-pV5Z level, we obtain TBE values that can be compared to the experimental estimates. The computed MAD is 0.05 eV, the largest deviations being obtained for the Δ and Σ^- excited states of both spin symmetries. The agreement between theory and experiment is therefore very satisfying though slightly less impressive than for N_2. We note that the CC3/*aug*-cc-pVTZ C = O bond length (1.134 Å) is 0.006 Å larger than the experimental r_e value of 1.128 Å,[118] whereas the discrepancy is twice smaller for dinitrogen: 1.101 Å for CC3/*aug*-cc-pVTZ compared to 1.098 Å experimentally. This might partially explained the larger deviations noticed for carbon monoxide.

3.3. Acetylene and Ethylene. Acetylene is the smallest conjugated organic molecule possessing stable low-lying excited-state structures, therefore allowing to investigate vertical fluorescence. This molecule has been the subject of previous investigations at the CASPT2,[124] CCSD,[125] CCSDT,[59] and MR-AQCC[126] levels. Our results are collected in Table 3. With the double-ζ basis set, the differences between the CC3, CCSDT, and CCSDTQ results are negligible, and the latter estimates are also systematically within 0.02 eV of the exFCI results. In contrast to water and ammonia, both CC3 and CCSDT provide similar accuracies compared to higher

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 5. Vertical (Absorption) Transition Energies for Various Excited States of Diazomethane (Top) and Ketene (Bottom)[v]

		aug-cc-pVDZ			*aug*-cc-pVTZ			lit.	
molecule	state	CC3	CCSDT	exFCI	CC3	CCSDT	exFCI	exp.	th.
acetaldehyde	^{1}A″(n → π^*)	4.34	4.32	4.34	4.31	4.29	4.31	4.27[a]	4.29[b]
	^{3}A″(n → π^*)	3.96	3.95	3.98	3.95	3.94	4.0[c]	3.97[a]	3.97[b]
cyclopropene	^{1}B$_1$ (σ→ π^*)	6.72	6.71	6.7[c]	6.68	6.68	6.6[c]	6.45[d]	6.89[e]
	^{1}B$_2$ (π→ π^*)	6.77	6.78	6.82	6.73	6.75	6.7[c]	7.00[f]	7.11[e]
	^{3}B$_2$ (π→ π^*)	4.34	4.35	4.35	4.34		4.38	4.16[f]	4.28[g]
	^{3}B$_1$ (σ→ π^*)	6.43	6.43	6.43	6.40		6.45		6.40[g]
diazomethane	^{1}A$_2$ (π→ π^*)	3.10	3.10	3.09	3.07	3.07	3.14	3.14[h]	3.21[i]
	^{1}B$_1$ (π→ 3s)	5.32	5.35	5.35	5.45	5.48	5.54		5.33[i]
	^{1}A$_1$ (π→ π^*)	5.80	5.82	5.79	5.84	5.86	5.90	5.9[h]	5.85[i]
	^{3}A$_2$ (π→ π^*)	2.84	2.84	2.81	2.83	2.82	2.8[c]		2.92[j]
	^{3}A$_1$ (π→ π^*)	4.05	4.04	4.03	4.03	4.02	4.05		3.97[j]
	^{3}B$_1$ (π→ 3s)	5.17	5.20	5.18	5.31	5.34	5.35		
	^{3}A$_1$ (π→ 3p)	6.83	6.83	6.81	6.80		6.82		7.02[j]
	^{1}A″ [F] (π→ π^*)	0.68	0.67	0.65	0.68	0.67	0.71		
formamide	^{1}A″(n → π^*)	5.71	5.68	5.70	5.66	5.63	5.7[c]	5.8[k]	5.63[l]
	^{1}A′(n → 3s)	6.65	6.64	6.67	6.74	6.74		6.35[k]	6.62[l]
	^{1}A′ (π→ π^*)[m]	7.63	7.62	7.64	7.62		7.63	7.37[k]	7.22[l]
	^{1}A′(n → 3p)[m]	7.31	7.29		7.40	7.38		7.73[k]	7.66[l]
	^{3}A″(n → π^*)	5.42	5.39	5.42	5.38		5.4[c]	5.2[k]	5.34[l]
	^{3}A′ (π→ π^*)	5.83	5.81	5.82	5.82		5.7[c]	~6[k]	5.74[l]
ketene	^{1}A$_2$ (π→ π^*)	3.89	3.88	3.84	3.88	3.87	3.86	3.7[n]	3.74[o]
	^{1}B$_1$ (n → 3s)	5.83	5.86	5.88	5.96	5.99	6.01	5.86[n]	5.82[o]
	^{1}A$_2$ (π→ 3p)	7.05	7.09	7.08	7.16	7.20	7.18		7.00[o]
	^{3}A$_2$ (n → π^*)	3.79	3.78	3.79	3.78	3.78	3.77	3.8[p]	3.62[q]
	^{3}A$_1$ (π→ π^*)	5.62	5.61	5.64	5.61	5.60	5.61	5[p]	5.42[q]
	^{3}B$_1$ (n → 3s)	5.63	5.66	5.68	5.76	5.80	5.79	5.8[p]	5.69[q]
	^{3}A$_2$ (π→ 3p)	7.01	7.05	7.07	7.12	7.17	7.12		
	^{1}A″[F] (π→ π^*)	1.00	0.99	0.96	1.00	1.00	1.00		
nitrosomethane	^{1}A″(n → π^*)	2.00	1.98	1.99	1.96	1.95	2.0[c]	1.83[r]	1.76[s]
	^{1}A′ (n, n → π^*,π^*)	5.75	5.26	4.81	5.76	5.29	4.72		4.96[s]
	^{1}A′ (n → 3s/3p)	6.20	6.19	6.29	6.31	6.30	6.4[c]		6.54[s]
	^{3}A″(n → π^*)	1.13	1.12	1.15	1.14	1.13	1.16		1.42[t]
	^{3}A′ (π→ π^*)	5.54	5.54	5.56	5.51		5.60		5.55[t]
	^{1}A″ [F] (n → π^*)	1.70	1.69	1.70	1.69	1.66	1.7[c]		
streptocyanine-C1	^{1}B$_2$ (π→ π^*)	7.14	7.12	7.14	7.13	7.11	7.1[c]		7.16[u]
	^{3}B$_2$ (π→ π^*)	5.48	5.47	5.47	5.48	5.47	5.52		

[a]Electron impact experiment from ref 145. [b]NEVPT-PC from ref 127. [c]CI convergence too slow to provide more reliable estimates. [d]Maximum in the gas UV from ref 146. [e]CCSDT/TZVP from ref 58. [f]Electron impact experiment from ref 147. [g]CC3/*aug*-cc-pVTZ from ref 32. [h]VUV maxima from ref 148. [i]CCSD/6-311(3+,+)G(d) calculations from ref 149. [j]MR-CC/DZP calculations from ref 150. [k]EELS (singlet) and trapped electron (triplet) experiments from ref 151. [l]*n*R-SI-CCSD(T) results from ref 142. [m]Strong state mixing. [n]Electron impact experiment from ref 152. [o]CASPT2/6-311+G(d) results from ref 153. [p]Electron impact experiment from ref 116. [q]STEOM-CCSD/Sad+//CCSD/Sad+ results from ref 154. [r]Maximum in the gas UV from ref 155. [s]CASPT2/ANO results from ref 156. [t]CASSCF/cc-pVDZ results from ref 157. [u]exCC3//MP2 result from ref 128. [v]All values are in eV.

levels of theory. As expected, for valence states, going from double- to triple-ζ basis set tends to slightly decrease the computed energies (except for the lowest triplet). Nonetheless, as with the smaller basis set, the same near-perfect methodological match pertains with *aug*-cc-pVTZ. Estimating the exFCI/*aug*-cc-pVTZ results from the exFCI/*aug*-cc-pVDZ values and CC3 basis set effects would yield estimates with absolute errors of 0.00−0.02 eV. One also notices that the exFCI/*aug*-cc-pVTZ values are all extremely close to the previous MR-AQCC estimates, whereas the published CASPT2 values appear to be too low though closer from the electron impact experiment, underlying once more the difficulty to obtain very accurate experimental estimates for vertical energies. This underestimating trend of standard CASPT2 was reported before for other molecules.[127,128] Although our theoretical vertical energy estimates still slightly vary when passing from the *aug*-cc-pVDZ to *aug*-cc-pVTZ basis sets, we claim that these vertical energies are probably more trustworthy for further benchmarks than the available experimental values because basis set effects beyond *aug*-cc-pVTZ seem rather limited (Table S3).

Despite its small size, ethylene remains a challenging molecule and is included in many benchmark sets.[17,26,29,30,75,132] The assignments of the experimental data have been the subject of countless works, and we refer the interested readers to the discussions in refs 30, 91, 116, 130,

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

131, and 133. On the theoretical side, the most complete and accurate investigation dedicated to the excited states of ethylene is due to Davidson's group, who performed refined CI calculations.[131] They indeed obtained highly accurate transition energies for ethylene, including for the valence yet challenging $^1B_{1u}$ state. From our data, collected in Table 3, one notices that the differences between exFCI/*aug*-cc-pVDZ and CCSDTQ/*aug*-cc-pVDZ results are again trifling, the largest deviation being obtained for the $^3B_{3u}(\pi \rightarrow 3s)$ Rydberg state (0.02 eV). In addition, given the nice agreement between CC3, CCSDT, and exFCI values, one can directly compare our CC3/*aug*-cc-pV5Z results (Table S3) to the values of reported in ref 131: a mean absolute deviation (MAD) of 0.03 eV is obtained. The fact that our transition energies tend to be slightly smaller than Davidson's is likely due to geometrical effects. Indeed, our CC3/*aug*-cc-pVTZ C═C distance is 1.3338 Å, i.e., slightly longer than the best estimate provided in Davidson's work (1.3305 Å). Recently, a stochastic heat-bath CI (SHCI)/ANO-L-pVTZ work reported 4.59 and 8.05 eV values for the $^3B_{1u}$ and $^1B_{1u}$ states, respectively,[91] and we also ascribe the differences with our results to the use of a MP2 geometry in ref 91. Interestingly, these authors found quite large discrepancies between their SHCI and their CC results. Indeed, they reported CR-EOMCC(2,3)D estimates significantly larger than their SHCI results with +0.17 and +0.20 eV upshifts for the triplet and singlet states, respectively. This highlights that only high-level CC schemes are able to recover the exFCI (or SHCI) results for ethylene.

3.4. Formaldehyde, Methanimine, and Thioformaldehyde. Similarly to ethylene, formaldehyde is a very popular test molecule,[17,22,26,29,30,59,75,76,132,137−142] and stands as the prototype carbonyl dye with a low-lying $n \rightarrow \pi^*$ transition. Nevertheless, even for this particular valence state, well-separated from higher-lying excited states, the choice of an experimental reference remains difficult. Indeed, values of 3.94,[22] 4.00,[26,29,138] 4.07,[17,75,139] and 4.1 eV,[137,140] have been used in previous theoretical benchmarks. In contrast to their oxygen cousin, both methanimine and thioformaldehyde were the subject of less attention from the theoretical community.[135,143,144] The results obtained for these three molecules are collected in Table 4. Considering all transitions listed in this table, one obtains a MAD of 0.01 eV between the CCSDTQ/*aug*-cc-pVDZ and exFCI/*aug*-cc-pVDZ results, the largest discrepancies of 0.03 eV being observed for two states for which the differences between CCSDT and CCSDTQ are also large (0.05 eV). As in water, using the exFCI/*aug*-cc-pVDZ values as reference, we found that CC3 delivers slightly more accurate transition energies (MAD of 0.02 eV, maximal deviation of 0.06 eV) than CCSDT (MAD of 0.03 eV, maximal deviation of 0.07 eV). By adding the difference between CC3/*aug*-cc-pVTZ and CC3/*aug*-cc-pVDZ results to the exFCI/*aug*-cc-pVDZ values, we obtain good estimates of the actual exFCI/*aug*-cc-pVTZ data, with a MAD of 0.02 eV for formaldehyde. Compared to the CC3/*aug*-cc-pVQZ results of Thiel,[30] the transition energies reported in Table 4 are slightly larger, which is probably due to the influence of the ground-state geometry rather than to basis set effects (see Table S4). Indeed, the carbonyl bond is significantly more contracted with CC3/*aug*-cc-pVTZ (1.208 Å) than with MP2/6-31G(d) (1.221 Å). In particular, for the hallmark $n \rightarrow \pi^*$, our best estimate is 3.97 eV (*vide infra*), nicely matching a previous MR-AQCC value of 3.98 eV,[134] but significantly below the previous DMC/BLYP estimate of 4.24 eV.[135] The latter discrepancy is probably due to the use of both different structures and pseudopotentials within DMC calculations.

For methanimine and thioformaldehyde, the basis set effects are rather small for the states considered here (see Table S4) and the data reported in the present work are probably the most accurate vertical transition energies reported to date. For the latter molecule, these vertical estimates are systematically larger than the known experimental 0−0 energies,[136] which is the expected trend.

3.5. Larger Compounds. Let us now turn our attention to molecules that encompass three heavy (non-hydrogen) atoms. We have treated seven molecules of that family, and all were previously investigated at several levels of theory: acetaldehyde,[26,29,127,138−140,158,159] cyclopropene,[30−32,58,132,160] diazomethane,[149,150,158,161] formamide,[30−32,58,59,162,163] ketene,[150,153,154,164] nitrosomethane,[156,157,165,166] and the shortest streptocyanine.[128,167−170] The results are gathered in Table 5. Note that, for these molecules containing three heavy atoms, it is sometimes challenging to obtain reliable exFCI estimes, especially for the largest basis set.

Experimentally, the lowest singlet and triplet $n \rightarrow \pi^*$ transitions of acetaldehyde are located 0.3−0.4 eV above their formaldehyde counterparts,[116,145] and this trend is accurately reproduced by theory, which also delivers estimates very close to the NEVPT2 values given in ref 127.

For cyclopropene, the lowest singlet $\sigma \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ are close from one another, and both CCSDT and exFCI predict the former to be slightly more stabilized, which is consistent with the large basis set CC3 results obtained by Thiel.[32]

For the isoelectronic diazomethane and ketene molecules (see Table 5), one notes, yet again, consistent results with, however, differences between the exFCI/*aug*-cc-pVTZ and CCSDT/*aug*-cc-pVTZ results larger than 0.05 eV for the two lowest singlet states of diazomethane. There is also a reasonable match between our data and previous theoretical results reported for these two molecules.[149,150,153,154] The basis set effects are significant for the Rydberg transitions, especially for the $\pi \rightarrow 3s$ states of diazomethane (Table S5).

In formamide, we found strong state mixing between the lowest singlet valence and Rydberg states of A′ symmetry. This is consistent with the CCSDT/TZVP analysis of Kannar and Szalay,[58] who reported, for example, a larger oscillator strength for the lowest Rydberg state than for the $\pi \rightarrow \pi^*$ transition. This state-mixing problem pertains with *aug*-cc-pVTZ, making unambiguous assignments difficult. Consequently, we have decided to classify the three lowest $^1A'$ transitions according to their dominant orbital character, which gives a picture consistent with the computed oscillator strengths (*vide infra*) but yields state inversions compared to Thiel's and Szalay's assignments.[31,58] This strong state mixing also prevented the convergence of several state energies with the exFCI/*aug*-cc-pVTZ approach. Despite these uncertainties, we obtained transition energies for the Rydberg states that are much closer from experiment[151] as well as from previous multireference CC estimates,[142] than the TZVP ones.[58]

Nitrosomethane is an interesting test molecule for three reasons: (i) it presents very low-lying $n \rightarrow \pi^*$ states of A″ symmetry, close to ca. 2.0 eV (singlet) and 1.2 eV (triplet), among the smallest absorption energies found in a compact molecule;[171] (ii) it changes from an eclipsed to a staggered conformation of the methyl group when going from the ground to the lowest singlet state;[157,172,173] (iii) the lowest-lying singlet A′ state corresponds to an almost pure double

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 6. TBE (in eV) for Various States and Wave Function Approaches[e]

					corrected TBE		
	state	f	% T_1	TBE(FC) AVTZ	method	corr.	value
acetaldehyde	$^1A''$(V; n → π^*)	0.000	91.3	4.31	exFCI/AVTZ	AVQZ	4.31
	$^3A''$(V; n → π^*)		97.9	3.97[a]	exFCI/AVDZ	AVQZ	3.98
acetylene	$^1\Sigma_u^-$(V; π → π^*)		96.5	7.10	exFCI/AVTZ	dAV5Z	7.10
	$^1\Delta_u$(V; π → π^*)		93.3	7.44			7.44
	$^3\Sigma_u^+$(V; π → π^*)		99.2	5.53			5.56
	$^3\Delta_u$(V; π → π^*)		99.0	6.40			6.40
	$^3\Sigma_u^-$(V; π → π^*)		98.8	7.08			7.09
	^1A_u [F](V; π → π^*)		95.6	3.64			3.63
	$^1A_2$ [F](V; π → π^*)		95.5	3.85			3.85
ammonia	$^1A_2$ (R; n → 3s)	0.086	93.5	6.59	exFCI/AVQZ	dAV5Z	6.66
	^1E(R; n → 3p)	0.006	93.7	8.16			8.21
	$^1A_1$ (R; n → 3p)	0.003	94.0	9.33			8.65
	$^1A_2$ (R; n → 4s)	0.008	93.6	9.96	exFCI/AVTZ	dAV5Z	9.19
	$^3A_2$ (R; n → 3s)		98.2	6.31	exFCI/AVQZ	dAV5Z	6.37
carbon monoxide	$^1\Pi$(V; n → π^*)	0.084	93.1	8.49	exFCI/AVQZ	dAV5Z	8.48
	$^1\Sigma^-$(V; π → π^*)		93.3	9.92			9.98
	$^1\Delta$(V; π → π^*)		91.8	10.06			10.10
	$^1\Sigma^+$ (R)	0.003	91.5	10.95			10.80
	$^1\Sigma^+$ (R)	0.200	92.9	11.52			11.42
	$^1\Pi$(R)	0.053	92.4	11.72			11.55
	$^3\Pi$(V; n → π^*)		98.7	6.28			6.28
	$^3\Sigma^+$ (V; π → π^*)		98.7	8.45			8.49
	$^3\Delta$(V; π → π^*)		98.4	9.27			9.28
	$^3\Sigma^-$(V; π → π^*)		97.5	9.80			9.77
	$^3\Sigma^+$ (R)		98.0	10.47	exFCI/AVTZ	dAV5Z	10.37
cyclopropene	$^1B_1$ (V;σ → π^*)	0.001	92.8	6.68[b]	CCSDT/AVTZ	AVQZ	6.68
	$^1B_2$ (V; π → π^*)	0.071	95.1	6.79[a]	exFCI/AVDZ	AVQZ	6.78
	$^3B_2$ (V; π → π^*)		98.0	4.38	exFCI/AVTZ	AVQZ	4.38
	$^3B_1$ (V;σ → π^*)		98.9	6.45			6.45
diazomethane	$^1A_2$ (V; π → π^*)		90.1	3.14	exFCI/AVTZ	dAVQZ	3.13
	$^1B_1$ (R; π → 3s)	0.002	93.8	5.54			5.59
	$^1A_1$ (V; π → π^*)	0.210	91.4	5.90			5.89
	$^3A_2$ (V; π → π^*)		97.7	2.79[a]	exFCI/AVDZ	dAVQZ	2.80
	$^3A_1$ (V; π → π^*)		98.6	4.05	exFCI/AVTZ	dAVQZ	4.05
	$^3B_1$ (R; π → 3s)		98.0	5.35			5.40
	$^3A_1$ (R; π → 3p)		98.5	6.82			6.72
	$^1A''$ [F](V; π → π^*)		87.4	0.71			0.70
dinitrogen	$^1\Pi_g$(V; n → π^*)		92.6	9.34	exFCI/AVQZ	dAV5Z	9.33
	$^1\Sigma_u^-$(V; π → π^*)		97.2	9.88			9.91
	$^1\Delta_u$ (V; π → π^*)	0.000	95.9	10.29			10.31
	$^1\Sigma_g^+$ (R)		92.2	12.98			12.30
	$^1\Pi_u$ (R)	0.229	82.9	13.03	exFCI/AVTZ	dAV5Z	12.73
	$^1\Sigma_u^+$ (R)	0.296	92.8	13.09			12.95
	$^1\Pi_u$ (R)	0.000	87.4	13.46			13.27
	$^3\Sigma_u^+$ (V; π → π^*)		99.3	7.70	exFCI/AVQZ	dAV5Z	7.74
	$^3\Pi_g$ (V; n → π^*)		98.4	8.01			8.03
	$^3\Delta_u$ (V; π → π^*)		99.3	8.87			8.88
	$^3\Sigma_u^-$ (V; π → π^*)		98.8	9.66			9.65
ethylene	$^1B_{3u}$(R; π → 3s)	0.078	95.1	7.39	exFCI/AVTZ	dAV5Z	7.43
	$^1B_{1u}$(V; π → π^*)	0.346	95.8	7.93			7.92
	$^1B_{1g}$(R; π → 3p)		95.3	8.08			8.10
	$^3B_{1u}$(V; π → π^*)		99.1	4.54			4.54
	$^3B_{3u}$(R; π → 3s)		98.5	7.23[a]	exFCI/AVDZ	dAV5Z	7.28
	$^3B_{1g}$(R; π → 3p)		98.4	7.98[a]			8.00
formaldehyde	$^1A_2$ (V; n → π^*)		91.5	3.98	exFCI/AVTZ	dAV5Z	3.97
	$^1B_2$ (R; n → 3s)	0.021	91.7	7.23			7.30

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 6. continued

					corrected TBE		
	state	f	% T_1	TBE(FC) AVTZ	method	corr.	value
	$^1B_2$ (R; n → 3p)	0.037	92.4	8.13			8.14
	$^1A_1$ (R; n → 3p)	0.052	91.9	8.23			8.27
	$^1A_2$ (R; n → 3p)		91.7	8.67			8.50
	$^1B_1$ (V;$\sigma \rightarrow \pi^*$)	0.001	90.8	9.22			9.21
	$^1A_1$ (V; $\pi \rightarrow \pi^*$)	0.135	90.4	9.43			9.26
	$^3A_2$ (V; n → π^*)		98.1	3.58			3.58
	$^3A_1$ (V; $\pi \rightarrow \pi^*$)		99.0	6.06			6.07
	$^3B_2$ (R; n → 3s)		97.1	7.06			7.14
	$^3B_2$ (R; n → 3p)		97.4	7.94			7.96
	$^3A_1$ (R; n → 3p)		97.2	8.10			8.15
	$^3B_1$ (R; n → 3d)		97.9	8.42			8.42
	$^1A''$ [F] (V; n → π^*)		87.8	2.80			2.80
formamide	$^1A''$ (V; n → π^*)	0.000	90.8	5.65[a]	exFCI/AVDZ	AVQZ	5.63
	$^1A'$ (R; n → 3s)	0.001	88.6	6.77[a]			6.81
	$^1A'$ (V; $\pi \rightarrow \pi^*$)	0.251	89.3	7.63	exFCI/AVTZ	AVQZ	7.64
	$^1A'$ (R; n → 3p)	0.111	89.6	7.38[b]	CCSDT/AVTZ	AVQZ	7.41
	$^3A''$(V; n → π^*)		97.7	5.38[c]	exFCI/AVDZ	AVQZ	5.37
	$^3A'$ (V; $\pi \rightarrow \pi^*$)		98.2	5.81[c]			5.81
hydrogen chloride	$^1\Pi$(CT)	0.056	94.3	7.84	exFCI/AVQZ	dAV5Z	7.86
hydrogen sulfide	$^1A_2$ (R; n → 4p)		94.6	6.18	exFCI/AVQZ	dAV5Z	6.10
	$^1B_1$ (R; n → 4s)	0.063	94.3	6.24			6.29
	$^3A_2$ (R; n → 4p)		98.7	5.81			5.74
	$^3B_1$ (R; n → 4s)		98.4	5.88			5.94
ketene	$^1A_2$ (V; $\pi \rightarrow \pi^*$)		91.0	3.86	exFCI/AVTZ	dAVQZ	3.86
	$^1B_1$ (R; n → 3s)	0.035	93.9	6.01			6.06
	$^1A_2$ (R; $\pi \rightarrow$ 3p)		94.4	7.18			7.19
	$^3A_2$ (V; n → π^*)		91.0	3.77			3.77
	$^3A_1$ (V; $\pi \rightarrow \pi^*$)		98.6	5.61			5.60
	$^3B_1$ (R; n → 3s)		98.1	5.79			5.85
	$^3A_2$ (R; $\pi \rightarrow$ 3p)		94.4	7.12			7.14
	$^1A''$ [F] (V; $\pi \rightarrow \pi^*$)		87.9	1.00			1.00
methanimine	$^1A''$(V; n → π^*)	0.003	90.7	5.23	exFCI/AVTZ	dAVQZ	5.21
	$^3A''$(V; n → π^*)		98.1	4.65			4.64
nitrosomethane	$^1A''$(V; n → π^*)	0.000	93.0	1.96[a]	exFCI/AVDZ	AVQZ	1.95
	$^1A'$ (V; n, n → π^*,π^*)	0.000	2.5	4.72	exFCI/AVTZ	AVQZ	4.69
	$^1A'$ (R; n → 3s/3p)	0.006	90.8	6.40[a]	exFCI/AVDZ	AVQZ	6.42
	$^3A''$(V; n → π^*)		98.4	1.16			1.16
	$^3A'$ (V; $\pi \rightarrow \pi^*$)		98.9	5.60			5.61
	$^1A''$ [F] (V; n → π^*)		92.7	1.67[a]	exFCI/AVDZ	AVQZ	1.66
streptocyanine-C1	$^1B_2$ (V; $\pi \rightarrow \pi^*$)	0.347	88.7	7.13[a]	exFCI/AVDZ	AVQZ	7.12
	$^3B_2$ (V; $\pi \rightarrow \pi^*$)		98.3	5.52	exFCI/AVTZ	AVQZ	5.52
thioformaldehyde	$^1A_2$ (V; n → π^*)		89.3	2.22	exFCI/AVTZ	dAVQZ	2.20
	$^1B_2$ (R; n → 4s)	0.012	92.3	5.96			5.99
	$^1A_1$ (V; $\pi \rightarrow \pi^*$)	0.178	90.8	6.38[d]	CCSDTQ/AVDZ	dAVQZ	6.34
	$^3A_2$ (V; n → π^*)		97.7	1.94	exFCI/AVTZ	dAVQZ	1.94
	$^3A_1$ (V; $\pi \rightarrow \pi^*$)		98.9	3.43			3.44
	$^3B_2$ (R; n → 4s)		97.6	5.72[a]	exFCI/AVDZ	dAVQZ	5.76
	$^1A_2$ [F] (V; n → π^*)		87.2	1.95	exFCI/AVTZ	dAVQZ	1.94
water	$^1B_1$ (R; n → 3s)	0.054	93.4	7.62	exFCI/AVQZ	dAV5Z	7.70
	$^1A_2$ (R; n → 3p)		93.6	9.41			9.47
	$^1A_1$ (R; n → 3s)	0.100	93.6	9.99			9.97
	$^3B_1$ (R; n → 3s)		98.1	7.25			7.33
	$^3A_2$ (R; n → 3p)		98.0	9.24			9.30
	$^3A_1$ (R; n → 3s)		98.2	9.54			9.59

[a]exCI/*aug*-cc-pVDZ data corrected with the difference between CCSDT/*aug*-cc-pVTZ and CCSDT/*aug*-cc-pVDZ values. [b]CCSDT/*aug*-cc-pVTZ value. [c]exCI/*aug*-cc-pVDZ data corrected with the difference between CC3/*aug*-cc-pVTZ and CC3/*aug*-cc-pVDZ values. [d]CCSDTQ/*aug*-cc-

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 6. continued

pVDZ data corrected with the difference between CCSDT/*aug*-cc-pVTZ and CCSDT/*aug*-cc-pVDZ values. [e]For each state, we provide the oscillator strength and percentage of single excitations obtained at the CC3(FC)/*aug*-cc-pVTZ level. Unless otherwise stated, the TBE(FC)/*aug*-cc-pVTZ have been obtained directly from exFCI. For the basis-set-corrected TBE, we provide the method used to determine the starting value and the basis set used at the CC3(full) level to correct it. CC3(full)/*aug*-cc-pVTZ geometries and abbreviated forms of Dunning's basis set are systematically used. R, V and F stand for Rydberg, valence and fluorescence, respectively.

excitation of (n, n) → (π^*,π^*) nature.[156] Indeed, CC3 returns a 2.5% single excitation character only for this second transition, to be compared to more than 80% (and generally more than 90%) in all other states treated in this work (*vide infra*). For example, the notoriously difficult A_g dark state of butadiene has a 72.8% single character.[30] For the A″ state of nitrosomethane, CC3, CCSDT and exFCI yield similar results, and the corresponding transition energies are slightly larger than previous CASPT2 estimates.[156] In contrast, the CC approaches are expectedly far from the spot for the (n, n) → (π^*,π^*) transition: they yield values significantly blue-shifted and large discrepancies between the CC3 and CCSDT values are found. For this particular state, it is not surprising that the exFCI result is indeed closer to the CASPT2 value,[156] as modeling double excitations with single-reference CC models is not a natural choice.

Finally for the shortest model cyanine, a molecule known to be difficult to treat with TD-DFT,[170] all the theoretical results given in Table 5 closely match each other for both the singlet and triplet manifolds. For the former, the reported CASPT2 (with IPEA) value of 7.14 eV also fits these estimates.[128]

3.6. Theoretical Best Estimates. We now turn to the definition of theoretical best estimates. We decided to provide two sets for these estimates, one obtained in the frozen-core approximation with the *aug*-cc-pVTZ atomic basis set, and one including further corrections for basis set and "all electron" (full) effects. This choice allows further benchmarks to either consider a reasonably compact basis set, therefore allowing to test many levels of theory, or to rely on values closer to the basis set limit. For the former set, we systematically selected exFCI/*aug*-cc-pVTZ values except when explicitly stated. For the latter set, both the "all electron" correlation and the basis set corrections (see Supporting Information for complete data) were systematically obtained at the CC3 level of theory and used d-*aug*-cc-pV5Z for the nine smallest molecules, but slightly more compact basis sets for the larger compounds. At least for Rydberg states, the use of d-*aug*-cc-pVQZ apparently delivers results closer to basis set convergence than *aug*-cc-pV5Z, and the former basis set was used when technically possible. The interested readers may find in Supporting Information the values obtained with and without applying the frozen-core approximation for several basis sets. Clearly, the largest amount of the total correction originates from basis set effects. In other words, "full" and frozen-core transition energies are typically within 0.01−0.02 eV of each other for a given basis set. The results are listed in Table 6 and provide a total of 110 transition energies. This set of states is rather diverse with 61 singlet and 45 triplet states, 60 valence and 45 Rydberg states, 21 n → π^* and 38 π → π^* states, with an energetic span from 0.70 to 13.27 eV. Among these 110 excitation energies, only 13 are characterized by a single-excitation character smaller than 90% according to CC3. As expected,[30] the dominant single-excitation character is particularly pronounced for triplet excited states. Therefore, this set is adequate for evaluating single-reference methods, though a few challenging cases are incorporated. Consequently, we think that the TBE listed in Table 6 contribute to fulfill the need of more accurate reference excited state energies, as pointed out by Thiel one decade ago.[30] However, the focus on small compounds and the lack of charge-transfer states constitute significant biases in the present set of transition energies.

3.7. Benchmarks. We have used the TBE(FC)/*aug*-cc-pVTZ benchmark values to assess the performances of 12 wave function approaches, namely, ADC(2), ADC(3), CIS(D), CIS(D_∞), CC2, STEOM-CCSD, CCSD, CCSDR(3), CCSDT-3, CC3, CCSDT, and CCSDTQ. The complete list of results can be found in Table S6 in the Supporting Information. As expected, only the approaches including iterative triples, that is, ADC(3), CCSDT-3, CC3, and CCSDT are able to predict the presence of the doubly excited (n, n) → (π^*,π^*) transition in nitrosomethane (see Tables 5 and S6), but they all yield large quantitative errors. Indeed, the TBE value of 4.72 eV is strongly underestimated by ADC(3) (3.00 eV) and significantly overshot by the three CC models with estimates of 6.02 eV, 5.76, and 5.29 eV with CCSDT-3, CC3, and CCSDT, respectively. This 0.26 eV difference between the CCSDT-3 and CC3 values is also the largest discrepancy between these two models in the tested set. Obviously, from a general perspective, one should not use the standard single-reference wave function methods to describe double excitations. Therefore, the (n, n) → (π^*,π^*) transition of nitrosomethane was removed from our statistical analysis. Likewise, for the three lowest $^1A'$ excited states of formamide, strong state mixing — involving two or three states — are found at all levels of theory, making unambiguous assignments impossible. Consequently, they are also excluded from our statistics.

In Table 7, we report, for the entire set of compounds, the mean signed error (MSE), mean absolute error (MAE) root-mean-square deviation (RMS), as well as the positive [Max(+)] and negative [Max(−)] maximum deviations. A graphical representation of the errors obtained with all methods can be found in Figure 1. Note that only singlet states could be computed with the programs used for CCSDR(3) and CCSDT-3. As shown in Figure 1, CCSDTQ is on the spot with tiny MSE and MAE, which is consistent with the analysis carried out for individual molecules. With this method, the negative and positive maximum deviations are as small as −0.05 eV (singlet n → 4s Rydberg transition of thioformaldehyde) and +0.06 eV ($^1\Sigma_u^+$ Rydberg transition of dinitrogen), respectively. The three other CC models with iterative triples (CCSDT-3, CC3, and CCSDT) also deliver extremely accurate transition energies with MAE of 0.03 eV only. In agreement with the analysis of Watson and co-workers, we do not find any significant (statistical) differences between CCSDT-3 and CC3,[57] and although the former theory is formally closer to CCSDT, it does not seem more advantageous nor disadvantageous than CC3 in practice. The very good performance of CC3 is also consistent with the analysis of Thiel and co-workers, who reported a strong agreement with CASPT2,[32] as well as with the conclusion of Szalay's group

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Table 7. Mean Signed Error (MSE), Mean Absolute Error (MAE), Root-Mean Square Deviation (RMS), Positive [Max(+)] and Negative [Max(−)] Maximal Deviations with Respect to TBE(FC)/*aug*-cc-pVTZ for the Transition Energies Listed in Table S6[a]

method	no. of states	MSE	MAE	RMS	Max(+)	Max(−)
CIS(D)	106	0.10	0.25	0.32	−0.63	1.06
CIS(D_∞)	106	−0.01	0.21	0.28	−0.76	0.57
CC2	106	0.03	0.22	0.28	−0.71	0.63
STEOM-CCSD	102	0.01	0.10	0.14	−0.56	0.40
CCSD	106	0.05	0.08	0.11	−0.17	0.40
CCSDR(3)	59	0.01	0.04	0.05	−0.07	0.25
CCSDT-3	58	0.01	0.03	0.05	−0.07	0.24
CC3	106	−0.01	0.03	0.04	−0.09	0.19
CCSDT	104	−0.01	0.03	0.03	−0.10	0.11
CCSDTQ	73	0.00	0.01	0.02	−0.05	0.06
ADC(2)	106	−0.01	0.21	0.28	−0.76	0.57
ADC(3)	106	−0.15	0.23	0.28	−0.79	0.39

[a]All values are in eV and have been obtained with the *aug*-cc-pVTZ basis set.

who found it very close to CCSDT.[59] Nevertheless, CCSDT is not, on average, significantly more accurate than CC3 nor CCSDT-3. In other words, CCSDT is probably not a sufficiently accurate benchmark to estimate the accuracy of CCSDT-3 nor CC3. The perturbative inclusion of triples via CCSDR(3) stands as a good compromise between computational cost and accuracy with a MAE of 0.04 eV, a conclusion also drawn in the benchmark study performed by Sauer and co-workers.[51] These very small average deviations are related to the fact that the majority of our set is constituted of large single-excitation character transitions (see % T_1 in Table 6). Reasonably, we predict that they would slightly deteriorate for larger compounds.

For the second-order CC series, as expected, the errors increase when one uses more approximate models. Indeed, the MAE are 0.08, 0.10, and 0.22 eV with CCSD, STEOM-CCSD, and CC2, respectively. The magnitude of the CC2 average deviation is consistent with previous estimates obtained for Thiel's set (0.29 eV for singlets and 0.18 eV for triplets),[30] for fluorescence energies (0.21 eV for 12 small compounds),[174] as well as for larger compounds (0.15 eV for 0−0 energies of conjugated dyes).[11] Likewise, the fact that CCSD tends to overestimate the transition energies (positive MSE) was also reported previously in several works.[26,30,57−59,159,174] It can be seen that Nooijen's STEOM approach, which was much less benchmarked previously, delivers an accuracy comparable to CCSD, with a smaller MSE but a large dispersion. More surprisingly, we found a MAE smaller with CCSD than with CC2, which contrasts with the results reported for Thiel's set,[51] but is consistent with Kannar, Tajti and Szalay conclusion.[59] We attribute this effect to the small size of the compounds treated herein. Indeed, analyzing the TZVP values of ref 30., it appears clearly that CC2 more regularly outperforms CCSD for larger compounds.

As expected, the results for CIS(D_∞) and ADC(2), two closely related theories,[6,144] are nearly equivalent, with only 4 (out of 106) cases for which a difference of 0.01 eV could be evidenced (Table S6). In addition, Table 7 evidences that ADC(2) provides an accuracy similar to CC2 for a smaller computational cost, whereas CIS(D) is slightly less accurate. Both outcomes perfectly fit previous benchmarks.[10,11,48,144,174] Conversely, we found that ADC(3) results are rather poor with average deviations larger than the ones obtained with ADC(2) and a clear tendency to provide red-shifted transition energies with a MSE of −0.15 eV. This observation is in sharp contrast with a previous investigation which concluded that ADC(3) and CC3 have very similar performances,[48] though the ADC(3) excitation energies were also found to be, on average, smaller by 0.20 eV compared to their CC3 counterparts. At this stage, it is difficult to know if the large MAE of ADC(3) reported in Table 7 originates solely from the small size of the compounds treated herein. However, the fact that the CCSD MSE is relatively small compared to previous benchmarks hints that the choice of compact compounds has a non-negligible effect on the statistics.

Let us analyze the ADC(3) errors more thoroughly. First, ADC(3) deviations are quite large for all subsets (*vide infra*). Second, we have found that, for the 46 transition energies for which ADC(2) yields an absolute error exceeding 0.15 eV compared to our TBE, the signs of the ADC(2) and ADC(3) errors systematically differ (see Figure 2); i.e., ADC(3) goes in the right "direction" but has the tendency to overcorrect ADC(2). This is clearly reminiscent of the well-known oscillating behavior of the Møller−Plesset perturbative series for ground state properties. Third, this overestimation of the corrections pertains for the states in which the ADC(2) absolute error is smaller than 0.15 eV. Indeed, in those 60 cases, there are only 10 transitions for which the ADC(3) values are more accurate than their second-order counterpart. As a consequence, taking the average between the ADC(2) and ADC(3) transition energies yield rather accurate estimates with a MAE as small as 0.10 eV for the full set, half of the MAE obtained with the parent methods.

We provide a more detailed analysis for several subsets of states in Table S7 in the Supporting Information. Globally, we found no significant difference between the singlet and triplet transitions, though all CC models (except STEOM-CCSD) provide slightly smaller deviations for the latter transitions, in line with their larger single-excitation character. With the computationally lighter methods, CIS(D), CIS(D_∞), ADC(2), and CC2, the MAEs are significantly smaller for the valence transitions (0.20, 0.15, 0.15, and 0.18 eV, respectively) than for the Rydberg transitions (0.32, 0.29, 0.29, and 0.26 eV, respectively). We also found MSE of opposite sign for valence and Rydberg transitions with CC2, which fits the results of Kannar and co-workers.[59] Surprisingly, ADC(3) gives 0.28 and 0.17 eV MAE for valence and Rydberg, respectively. All CC methods including triples deliver similar deviations for both sets of states. All methods provide smaller (or equal) MAE for the n → π^* than for the $\pi \rightarrow \pi^*$ transitions, which was already found for Thiel's set.[30] The differences are particularly significant with CIS(D), CC2, STEOM-CCSD, and ADC(3) with errors twice larger for $\pi \rightarrow \pi^*$ than n → π^* states. Finally, when considering the few states with %T_1 smaller than 90%, we logically found larger statistical errors with, for example, MAE of, e.g., 0.03 eV for CCSDTQ, 0.04 eV for CC3, and 0.06 eV for CCSDT-3.

3.8. On the Use of a Compact Basis Set. In several of the molecules considered here, we have found that adding corrections for basis set effects determined at the CC3 level to exFCI/*aug*-cc-pVDZ results effectively provides accurate estimates of the exFCI values directly determined with larger bases. Nevertheless, the dreadful scalings of both exFCI and

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

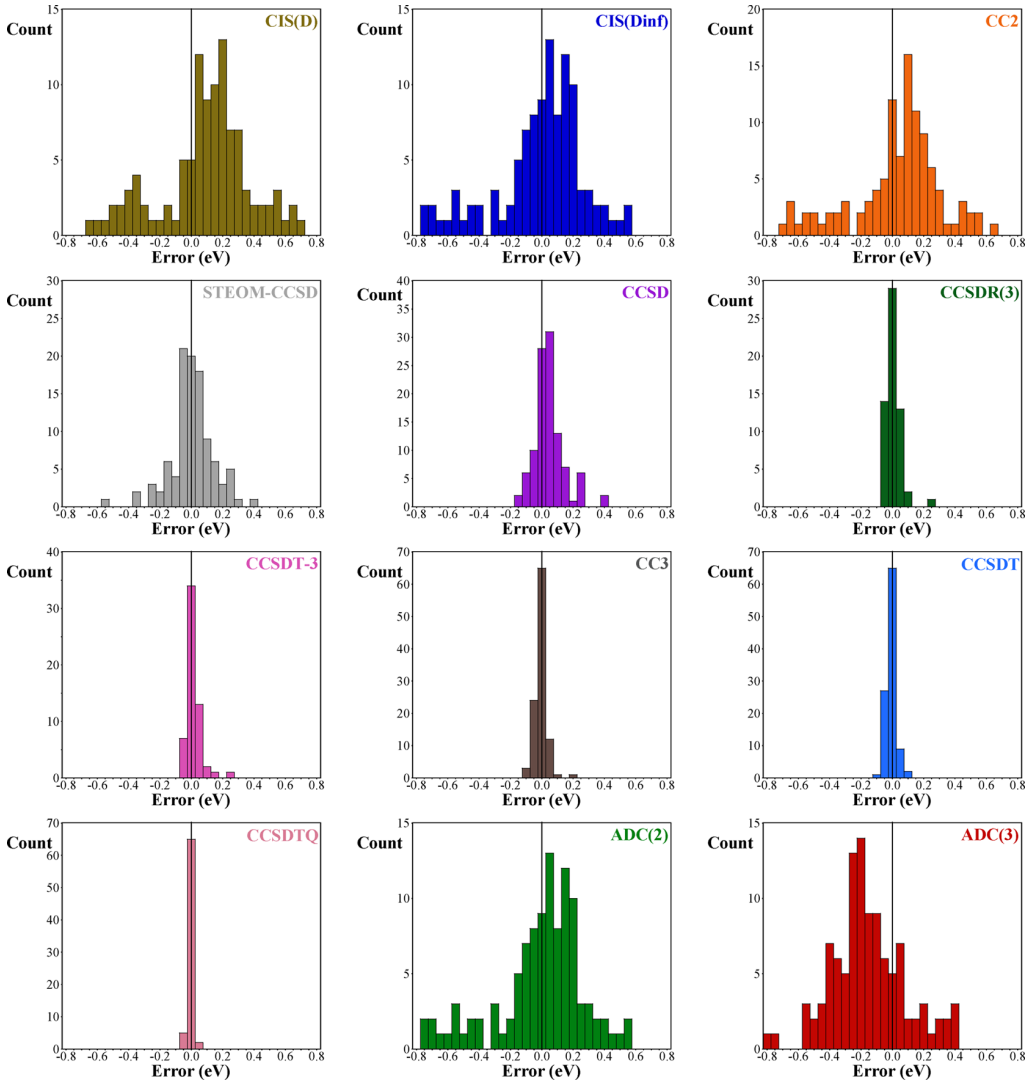


Figure 1. Histograms of the error patterns for several wave function methods compared to TBE(FC). Note the variation of scaling of the vertical axes.

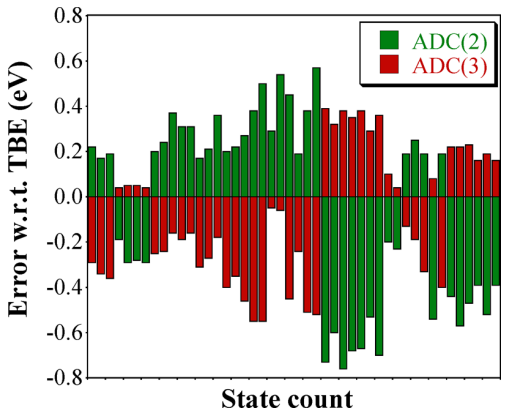


Figure 2. Comparison between the errors obtained with ADC(2) and ADC(3) [compared to TBE(FC)] for the 46 states for which ADC(2) yields an absolute deviation larger than 0.15 eV. All values are in eV.

CCSDTQ make the size of the atomic basis the central bottleneck. For this reason, we have tested the use of one of the most compact basis encompassing both diffuse and polarization functions, namely Pople's 6-31+G(d). We have performed CC3, CCSDT, and CCSDTQ calculations with this particular basis. The results are collected in the Supporting Information (Table S8). First, we compare the 6-31+G(d) results to those obtained with the same theoretical method in conjunction with the *aug*-cc-pVTZ basis set. As expected, large discrepancies are found with mean absolute deviation of 0.20, 0.19, and 0.25 eV, for CC3, CCSDT, and CCSDTQ, respectively.[175] Second, by adding the differences between the CC3/*aug*-cc-pVTZ and CC3/6-31+G(d) results to the CCSDT/6-31+G(d) and CCSDTQ/6-31+G(d) values, we obtained improved values. Such procedure yields very good estimates of the actual *aug*-cc-pVTZ results, as the MAE are down to 0.01 eV with no error larger than 0.04 eV for both CCSDT and CCSDTQ. This is a particularly remarkable result for Rydberg states that are extremely basis set dependent. For example, for the $^3A_2(n \rightarrow 3p)$ transition in water, the CCSDTQ/6-31+G(d) value of 10.34 eV is more than 1 eV

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

above its CCSDTQ/*aug*-cc-pVTZ counterpart (9.23 eV, see Table 1). Applying the CC3 basis set correction makes the final error as small as 0.03 eV. This composite methodology opens the way to calculations on larger systems without significant loss of accuracy.

4. CONCLUSIONS AND OUTLOOK

We have defined a set of more than 100 vertical transition energies, as close as possible to the FCI limit. To this end, we have used both the coupled cluster route up to the highest computationally possible order and the selected configuration interaction route up to the largest technically affordable number of determinants, that is here about few millions. These calculations have been performed on 18 compounds encompassing one, two or three non-hydrogen atoms, using geometries optimized at the CC3 level and a series of diffuse Dunning's basis sets of increasing size. It was certainly gratifying to find extremely good agreements between the results obtained independently with these two distinct approaches with typical differences as small as 0.01 eV between CCSDTQ and exFCI transition energies. In fact, during the course of this joint work, the two groups involved in this study were able to detect misprints or incorrect assignments in each others calculations even when the differences were apparently negligible. For the two diatomic molecules considered in this work, N_2 and CO, the mean absolute deviation between our theoretical best estimates and the "experimental" vertical transition energies deduced from spectroscopic measurements using a numerical solution of the nuclear Schrödinger equation is as small as 0.04 eV, and it was possible to resolve previous inconsistencies between these "experimental" values. A significant share of the remaining error is likely related to the use of theoretically determined geometries. Although, it is not possible to provide a definitive error bar for the 110 TBE listed in this work, our estimate, based on the differences between the two routes as well as the extrapolations used in the sCI procedure, is ±0.03 eV.

In another part of this work, we have used the TBE(FC)/*aug*-cc-pVTZ values to benchmark a series of 12 popular wave function approaches. For the computationally most effective approaches, CIS(D), CIS(D_∞), ADC(2), and CC2, we found average deviations of ca. 0.21−0.25 eV with strong similarities between the ADC(2) and CC2 results. Both conclusions are backed up by previous works. Likewise, we obtained the expected trend that CCSD overestimates the transition energies, though with an amplitude that is quite small here, likely due to the small size of the compounds investigated. More interestingly, we could demonstrate that STEOM-CCSD is, on average, as accurate as CCSD, and we were also able to benchmark the methods including contributions from triples using reliable theoretical references. Interestingly, we found no significant differences between CCSDT-3, CC3, and CCSDT, which all yield a MAE of 0.03 eV. In other words, we could not demonstrate that CCSDT is statistically more accurate than its approximated (and computationally more effective) forms, nor highlight significant differences between CCSDT-3 and CC3. We have observed that the use of perturbative triples, as in CCSDR(3), allows to correct most of the CCSD error. This evidences that CCSDR(3) is a computationally appealing method as it gives average deviations only slightly larger than with iterative triples. In contrast, for the present set of molecules, ADC(3) was found significantly less accurate than CC3, and it was showed that ADC(3) overcorrects ADC(2). Whether this surprising result is related to the size of the compounds or is a more general trend remains to be confirmed.

As stated several times throughout this work, the size of the considered molecules is certainly one of the main limitations of the present effort, as it introduces a significant bias, e.g., charge-transfer over several Å are totally absent of the set. Obviously, the respective $O(N^{10})$ and $O(e^N)$ formal scalings of CCSDTQ and FCI do not offer an easy pathway to circumvent this limit. Nevertheless, it appears that performing exFCI calculations with a relatively compact basis, e.g., *aug*-cc-pVDZ or even 6-31+G(d), and correcting the basis set effects with a more affordable approach, e.g., CC3, might be a valuable and efficient approach to reach accurate vertical excitations energies for larger molecules, at least for the electronic transitions presenting a dominant single excitation character. Indeed, we have shown here that such basis set extrapolation approach is trustworthy. We are currently hiking along that path.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00406.

Basis set and frozen-core effects, geometries used, full list of transition energies for the benchmark section, additional statistical analysis, 6-31+G(d) results, and additional information for selected CI calculations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(P.-F.L.) E-mail: loos@irsamc.ups-tlse.fr.
*(D.J.) E-mail: Denis.Jacquemin@univ-nantes.fr.

ORCID

Pierre-François Loos: 0000-0003-0598-7425
Denis Jacquemin: 0000-0002-4217-0708

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

D.J. acknowledges the *Région des Pays de la Loire* for financial support. This research used resources of (i) the GENCI-CINES/IDRIS (Grant 2016-08s015), (ii) CCIPL (*Centre de Calcul Intensif des Pays de Loire*), (iii) the Troy cluster installed in Nantes, and (iv) CALMIP under Allocations 2018-0510 and 2018-18005 (Toulouse).

■ REFERENCES

(1) Hegarty, D.; Robb, M. A. Application of Unitary Group Methods to Configuration Interaction Calculations. *Mol. Phys.* **1979**, *38*, 1795−1812.
(2) Taylor, P. R. Analytical MCSCF Energy Gradients: Treatment of Symmetry and CASSCF Applications to Propadienone. *J. Comput. Chem.* **1984**, *5*, 589−597.
(3) Casida, M. E.; Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.* **2012**, *63*, 287−323.
(4) Ullrich, C. *Time-Dependent Density-Functional Theory: Concepts and Applications*; Oxford Graduate Texts; Oxford University Press: New York, 2012.

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

(5) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory With a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483−5488.
(6) Dreuw, A.; Wormit, M. The Algebraic Diagrammatic Construction Scheme for the Polarization Propagator for the Calculation of Excited States. *WIREs Comput. Mol. Sci.* **2015**, *5*, 82−95.
(7) Christiansen, O.; Koch, H.; Jørgensen, P. The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409−418.
(8) Hättig, C.; Weigend, F. CC2 Excitation Energy Calculations on Large Molecules Using the Resolution of the Identity Approximation. *J. Chem. Phys.* **2000**, *113*, 5154−5161.
(9) Laurent, A. D.; Adamo, C.; Jacquemin, D. Dye Chemistry with Time-Dependent Density Functional Theory. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14334−14356.
(10) Winter, N. O. C.; Graf, N. K.; Leutwyler, S.; Hattig, C. Benchmarks for 0−0 Transitions of Aromatic Organic Molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 Compared to High-Resolution Gas-Phase Data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6623−6630.
(11) Jacquemin, D.; Duchemin, I.; Blase, X. 0−0 Energies Using Hybrid Schemes: Benchmarks of TD-DFT, CIS(D), ADC(2), CC2 and BSE/GW formalisms for 80 Real-Life Compounds. *J. Chem. Theory Comput.* **2015**, *11*, 5340−5359.
(12) Oruganti, B.; Fang, C.; Durbeej, B. Assessment of a Composite CC2/DFT Procedure for Calculating 0−0 Excitation Energies of Organic Molecules. *Mol. Phys.* **2016**, *114*, 3448−3463.
(13) Oddershede, J.; Grūner, N. E.; Diercksen, G. H. Comparison Between Equation of Motion and Polarization Propagator Calculations. *Chem. Phys.* **1985**, *97*, 303−310.
(14) Schwabe, T.; Goerigk, L. Time-Dependent Double-Hybrid Density Functionals with Spin-Component and Spin-Opposite Scaling. *J. Chem. Theory Comput.* **2017**, *13*, 4307−4323.
(15) Christiansen, O.; Koch, H.; Jørgensen, P. Response Functions in the CC3 Iterative Triple Excitation Model. *J. Chem. Phys.* **1995**, *103*, 7429−7441.
(16) Koch, H.; Christiansen, O.; Jorgensen, P.; Sanchez de Merás, A. M.; Helgaker, T. The CC3Model: An Iterative Coupled Cluster Approach Including Connected Triples. *J. Chem. Phys.* **1997**, *106*, 1808−1818.
(17) Leang, S. S.; Zahariev, F.; Gordon, M. S. Benchmarking the Performance of Time-Dependent Density Functional Methods. *J. Chem. Phys.* **2012**, *136*, 104101.
(18) Parac, M.; Grimme, S. Comparison of Multireference Möller-Plesset Theory and Time-Dependent Methods for the Calculation of Vertical Excitation Energies of Molecules. *J. Phys. Chem. A* **2002**, *106*, 6844−6850.
(19) Dierksen, M.; Grimme, S. The Vibronic Structure of Electronic Absorption Spectra of Large Molecules: A Time-Dependent Density Functional Study on the Influence of Exact Hartree-Fock Exchange. *J. Phys. Chem. A* **2004**, *108*, 10225−10237.
(20) Grimme, S.; Izgorodina, E. I. Calculation of 0−0 Excitation Energies of Organic Molecules by CIS(D) Quantum Chemical Methods. *Chem. Phys.* **2004**, *305*, 223−230.
(21) Rhee, Y. M.; Head-Gordon, M. Scaled Second-Order Perturbation Corrections to Configuration Interaction Singles: Efficient and Reliable Excitation Energy Methods. *J. Phys. Chem. A* **2007**, *111*, 5314−5326.
(22) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation Energies in Density Functional Theory: an Evaluation and a Diagnostic Test. *J. Chem. Phys.* **2008**, *128*, 044118.
(23) Jacquemin, D.; Perpète, E. A.; Scuseria, G. E.; Ciofini, I.; Adamo, C. TD-DFT Performance for the Visible Absorption Spectra of Organic Dyes: Conventional Versus Long-Range Hybrids. *J. Chem. Theory Comput.* **2008**, *4*, 123−135.
(24) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420−2435.
(25) Goerigk, L.; Moellmann, J.; Grimme, S. Computation of Accurate Excitation Energies for Large Organic Molecules with Double-Hybrid Density Functionals. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4611−4620.
(26) Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. Electronic Transition Energies: A Study of the Performance of a Large Range of Single Reference Density Functional and Wave Function Methods on Valence and Rydberg States Compared to Experiment. *J. Chem. Theory Comput.* **2010**, *6*, 370−383.
(27) Jacquemin, D.; Planchat, A.; Adamo, C.; Mennucci, B. A TD-DFT Assessment of Functionals for Optical 0−0 Transitions in Solvated Dyes. *J. Chem. Theory Comput.* **2012**, *8*, 2359−2372.
(28) Isegawa, M.; Peverati, R.; Truhlar, D. G. Performance of Recent and High-Performance Approximate Density Functionals for Time-Dependent Density Functional Theory Calculations of Valence and Rydberg Electronic Transition Energies. *J. Chem. Phys.* **2012**, *137*, 244104.
(29) Hoyer, C. E.; Ghosh, S.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration Pair-Density Functional Theory Is as Accurate as CASPT2 for Electronic Excitation. *J. Phys. Chem. Lett.* **2016**, *7*, 586−591.
(30) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: CASPT2, CC2, CCSD and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.
(31) Silva-Junior, M. R.; Sauer, S. P. A.; Schreiber, M.; Thiel, W. Basis Set Effects on Coupled Cluster Benchmarks of Electronically Excited States: CC3, CCSDR(3) and CC2. *Mol. Phys.* **2010**, *108*, 453−465.
(32) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks of Electronically Excited States: Basis Set Effecs Benchmarks of Electronically Excited States: Basis Set Effects on CASPT2 Results. *J. Chem. Phys.* **2010**, *133*, 174318.
(33) Silva-Junior, M. R.; Thiel, W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.* **2010**, *6*, 1546−1564.
(34) Domínguez, A.; Aradi, B.; Frauenheim, T.; Lutsker, V.; Niehaus, T. A. Extensions of the Time-Dependent Density Functional Based Tight-Binding Approach. *J. Chem. Theory Comput.* **2013**, *9*, 4901−4914.
(35) Voityuk, A. A. INDO/X: A New Semiempirical Method for Excited States of Organic and Biological Molecules. *J. Chem. Theory Comput.* **2014**, *10*, 4950−4958.
(36) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: Time-Dependent Density Functional Theory and Density Functional Theory Based Multireference Configuration Interaction. *J. Chem. Phys.* **2008**, *129*, 104103.
(37) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *J. Chem. Phys.* **2009**, *130*, 054112.
(38) Jacquemin, D.; Perpète, E. A.; Ciofini, I.; Adamo, C. Assessment of Functionals for TD-DFT Calculations of Singlet-Triplet Transitions. *J. Chem. Theory Comput.* **2010**, *6*, 1532−1537.
(39) Jacquemin, D.; Perpète, E. A.; Ciofini, I.; Adamo, C.; Valero, R.; Zhao, Y.; Truhlar, D. G. On the Performances of the M06 Family of Density Functionals for Electronic Excitation Energies. *J. Chem. Theory Comput.* **2010**, *6*, 2071−2085.
(40) Mardirossian, N.; Parkhill, J. A.; Head-Gordon, M. Benchmark Results for Empirical Post-GGA Functionals: Difficult Exchange Problems and Independent Tests. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19325−19337.
(41) Jacquemin, D.; Perpète, E. A.; Ciofini, I.; Adamo, C. Assessment of the ωB97 Family for Excited-State Calculations. *Theor. Chem. Acc.* **2011**, *128*, 127−136.
(42) Huix-Rotllant, M.; Ipatov, A.; Rubio, A.; Casida, M. E. Assessment of Dressed Time-Dependent Density-Functional Theory

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

for the Low-Lying Valence States of 28 Organic Chromophores. *Chem. Phys.* **2011**, *391*, 120−129.
(43) Della Sala, F.; Fabiano, E. Accurate Singlet and Triplet Excitation Energies Using the Localized Hartree-Fock Kohn-Sham Potential. *Chem. Phys.* **2011**, *391*, 19−26.
(44) Trani, F.; Scalmani, G.; Zheng, G. S.; Carnimeo, I.; Frisch, M. J.; Barone, V. Time-Dependent Density Functional Tight Binding: New Formulation and Benchmark of Excited States. *J. Chem. Theory Comput.* **2011**, *7*, 3304−3313.
(45) Peverati, R.; Truhlar, D. G. Performance of the M11 and M11-L Density Functionals for Calculations of Electronic Excitation Energies by Adiabatic Time-Dependent Density Functional Theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11363−11370.
(46) Maier, T. M.; Bahmann, H.; Arbuznikov, A. V.; Kaupp, M. Validation of Local Hybrid Functionals for TDDFT Calculations of Electronic Excitation Energies. *J. Chem. Phys.* **2016**, *144*, 074106.
(47) Sauer, S. P.; Pitzner-Frydendahl, H. F.; Buse, M.; Jensen, H. J. A.; Thiel, W. Performance of SOPPA-Based Methods in the Calculation of Vertical Excitation Energies and Oscillator Strengths. *Mol. Phys.* **2015**, *113*, 2026−2045.
(48) Harbach, P. H. P.; Wormit, M.; Dreuw, A. The Third-Order Algebraic Diagrammatic Construction Method (ADC(3)) for the Polarization Propagator for Closed-Shell Molecules: Efficient Implementation and Benchmarking. *J. Chem. Phys.* **2014**, *141*, 064113.
(49) Schapiro, I.; Sivalingam, K.; Neese, F. Assessment of *n*-Electron Valence State Perturbation Theory for Vertical Excitation Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3567−3580.
(50) Yang, Y.; Peng, D.; Lu, J.; Yang, W. Excitation Energies from Particle-Particle Random Phase Approximation: Davidson Algorithm and Benchmark Studies. *J. Chem. Phys.* **2014**, *141*, 124104.
(51) Sauer, S. P. A.; Schreiber, M.; Silva-Junior, M. R.; Thiel, W. Benchmarks for Electronically Excited States: A Comparison of Noniterative and Iterative Triples Corrections in Linear Response Coupled Cluster Methods: CCSDR(3) versus CC3. *J. Chem. Theory Comput.* **2009**, *5*, 555−564.
(52) Demel, O.; Datta, D.; Nooijen, M. Additional Global Internal Contraction in Variations of Multireference Equation of Motion Coupled Cluster Theory. *J. Chem. Phys.* **2013**, *138*, 134108.
(53) Piecuch, P.; Hansen, J. A.; Ajala, A. O. Benchmarking the Completely Renormalised Equation-Of-Motion Coupled-Cluster Approaches for Vertical Excitation Energies. *Mol. Phys.* **2015**, *113*, 3085−3127.
(54) Tajti, A.; Szalay, P. G. Investigation of the Impact of Different Terms in the Second Order Hamiltonian on Excitation Energies of Valence and Rydberg States. *J. Chem. Theory Comput.* **2016**, *12*, 5477−5482.
(55) Rishi, V.; Perera, A.; Nooijen, M.; Bartlett, R. J. Excited States from Modified Coupled Cluster Methods: Are They Any Better Than EOM CCSD? *J. Chem. Phys.* **2017**, *146*, 144104.
(56) Dutta, A. K.; Nooijen, M.; Neese, F.; Izsák, R. Exploring the Accuracy of a Low Scaling Similarity Transformed Equation of Motion Method for Vertical Excitation Energies. *J. Chem. Theory Comput.* **2018**, *14*, 72−91.
(57) Watson, T. J.; Lotrich, V. F.; Szalay, P. G.; Perera, A.; Bartlett, R. J. Benchmarking for Perturbative Triple-Excitations in EE-EOM-CC Methods. *J. Phys. Chem. A* **2013**, *117*, 2569−2579.
(58) Kánnár, D.; Szalay, P. G. Benchmarking Coupled Cluster Methods on Valence Singlet Excited States. *J. Chem. Theory Comput.* **2014**, *10*, 3757−3765.
(59) Kánnár, D.; Tajti, A.; Szalay, P. G. Accuracy of Coupled Cluster Excitation Energies in Diffuse Basis Sets. *J. Chem. Theory Comput.* **2017**, *13*, 202−209.
(60) Laurent, A.; Blondel, A.; Jacquemin, D. Choosing an Atomic Basis Set for TD-DFT, SOPPA, ADC(2), CIS(D), CC2 and EOM-CCSD Calculations of Low-Lying Excited States of Organic Dyes. *Theor. Chem. Acc.* **2015**, *134*, 76.
(61) Budzák, Š.; Scalmani, G.; Jacquemin, D. Accurate Excited-State Geometries: a CASPT2 and Coupled-Cluster Reference Database for Small Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 6237−6252.
(62) Purvis, G. P., III; Bartlett, R. J. A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910−1918.
(63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision A.03; Gaussian Inc.: Wallingford, CT, 2016.
(64) For formamide, the CCSD/def2-TZVPP vibrational frequency calculation returns one imaginary frequency, which disappears at the CCSD/*aug*-cc-pVTZ level.
(65) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. A.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton Quantum Chemistry Program System. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269−284.
(66) Stanton, J. F., Gauss, J., Cheng, L., Harding, M. E., Matthews, D. A., Szalay, P. G. *CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package*; with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W. J. Lauderdale, F. Lipparini, T. Metzroth, L. A. Mück, D. P. O'Neill, D. R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.
(67) Kállay, M.; Gauss, J. Calculation of Excited-State Properties Using General Coupled-Cluster and Configuration-Interaction Models. *J. Chem. Phys.* **2004**, *121*, 9257−9269.
(68) Neese, F. The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(69) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An Efficient Linear-Scaling CCSD(T) Method Based on Local Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.
(70) Kállay, M.; Rolik, Z.; Csontos, J.; Nagy, P.; Samu, G.; Mester, D.; Csóka, J.; Szabó, B.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.;

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

Petrov, K.; Farkas, M.; Mezei, P. D.; Hégely, B. MRCC, Quantum Chemical Program. 2017; See: www.mrcc.hu.
(71) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184−215.
(72) Watts, J. D.; Bartlett, R. J. Iterative and Non-Iterative Triple Excitation Corrections in Coupled-Cluster Methods for Excited Electronic States: the EOM-CCSDT-3 and EOM-CCSD($\tilde{T}$) Methods. *Chem. Phys. Lett.* **1996**, *258*, 581−588.
(73) Prochnow, E.; Harding, M. E.; Gauss, J. Parallel Calculation of CCSDT and Mk-MRCCSDT Energies. *J. Chem. Theory Comput.* **2010**, *6*, 2339−2347.
(74) Noga, J.; Bartlett, R. J. The Full CCSDT Model for Molecular Electronic Structure. *J. Chem. Phys.* **1987**, *86*, 7041−7050.
(75) Head-Gordon, M.; Rico, R. J.; Oumi, M.; Lee, T. J. A Doubles Correction to Electronic Excited States From Configuration Interaction in the Space of Single Substitutions. *Chem. Phys. Lett.* **1994**, *219*, 21−29.
(76) Head-Gordon, M.; Maurice, D.; Oumi, M. A Perturbative Correction to Restricted Open-Shell Configuration-Interaction with Single Substitutions for Excited-States of Radicals. *Chem. Phys. Lett.* **1995**, *246*, 114−121.
(77) Christiansen, O.; Koch, H.; Jørgensen, P. Perturbative Triple Excitation Corrections to Coupled Cluster Singles and Doubles Excitation Energies. *J. Chem. Phys.* **1996**, *105*, 1451−1459.
(78) Nooijen, M.; Bartlett, R. J. A New Method for Excited States: Similarity Transformed Equation-Of-Motion Coupled-Cluster Theory. *J. Chem. Phys.* **1997**, *106*, 6441−6448.
(79) Head-Gordon, M.; Oumi, M.; Maurice, D. Quasidegenerate Second-Order Perturbation Corrections to Single-Excitation Configuration Interaction. *Mol. Phys.* **1999**, *96*, 593−602.
(80) Kucharski, S. A.; Bartlett, R. J. Recursive Intermediate Factorization and Complete Computational Linearization of the Coupled-Cluster Single, Double, Triple, and Quadruple Excitation Equations. *Theor. Chim. Acta* **1991**, *80*, 387−405.
(81) Bender, C. F.; Davidson, E. R. Studies in Configuration Interaction: The First-Row Diatomic Hydrides. *Phys. Rev.* **1969**, *183*, 23−30.
(82) Whitten, J. L.; Hackmeyer, M. Configuration Interaction Studies of Ground and Excited States of Polyatomic Molecules. I. The CI Formulation and Studies of Formaldehyde. *J. Chem. Phys.* **1969**, *51*, 5584−5596.
(83) Giner, E.; Scemama, A.; Caffarel, M. Using Perturbatively Selected Configuration Interaction in Quantum Monte Carlo Calculations. *Can. J. Chem.* **2013**, *91*, 879−885.
(84) Caffarel, M.; Giner, E.; Scemama, A.; Ramírez-Solís, A. Spin Density Distribution in Open-Shell Transition Metal Systems: A Comparative Post-Hartree-Fock, Density Functional Theory, and Quantum Monte Carlo Study of the $CuCl_2$ Molecule. *J. Chem. Theory Comput.* **2014**, *10*, 5286−5296.
(85) Giner, E.; Scemama, A.; Caffarel, M. Fixed-Node Diffusion Monte Carlo Potential Energy Curve of the Fluorine Molecule F_2 Using Selected Configuration Interaction Trial Wavefunctions. *J. Chem. Phys.* **2015**, *142*, 044115.
(86) Garniron, Y.; Giner, E.; Malrieu, J.-P.; Scemama, A. Alternative Definition of Excitation Amplitudes in Multi-Reference State-Specific Coupled Cluster. *J. Chem. Phys.* **2017**, *146*, 154107.
(87) Caffarel, M.; Applencourt, T.; Giner, E.; Scemama, A. Toward an Improved Control of the Fixed-Node Error in Quantum Monte Carlo: The Case of the Water Molecule. *J. Chem. Phys.* **2016**, *144*, 151103.
(88) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674−3680.
(89) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595−1604.
(90) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited States Using Semistochastic Heat-Bath Configuration Interaction. *J. Chem. Phys.* **2017**, *147*, 164111.
(91) Chien, A. D.; Holmes, A. A.; Otten, M.; Umrigar, C. J.; Sharma, S.; Zimmerman, P. M. Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction. *J. Phys. Chem. A* **2018**, *122*, 2714−2722.
(92) Scemama, A.; Garniron, Y.; Caffarel, M.; Loos, P. F. Deterministic Construction of Nodal Surfaces Within Quantum Monte Carlo: The Case of FeS. *J. Chem. Theory Comput.* **2018**, *14*, 1395−1402.
(93) Huron, B.; Malrieu, J. P.; Rancurel, P. Iterative Perturbation Calculations of Ground and Excited State Energies from Multi-configurational Zeroth-Order Wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745−5759.
(94) Evangelisti, S.; Daudey, J.-P.; Malrieu, J.-P. Convergence of an Improved CIPSI Algorithm. *Chem. Phys.* **1983**, *75*, 91−102.
(95) Caffarel, M.; Applencourt, T.; Giner, E.; Scemama, A. *ACS Symp. Ser.* **2016**, *1234*, 15−46.
(96) Garniron, Y.; Scemama, A.; Loos, P.-F.; Caffarel, M. Hybrid Stochastic-Deterministic Calculation of the Second-Order Perturbative Contribution of Multireference Perturbation Theory. *J. Chem. Phys.* **2017**, *147*, 034101.
(97) Scemama, A.; Applencourt, T.; Garniron, Y.; Giner, E.; David, G.; Caffarel, M. *Quantum Package*, v1.0; 2016; https://github.com/LCPQ/quantum_package, https://github.com/LCPQ/quantum_package.
(98) Ralphs, K.; Serna, G.; Hargreaves, L. R.; Khakoo, M. A.; Winstead, C.; McKoy, V. Excitation of the Six Lowest Electronic Transitions in Water by 9−20 eV Electrons. *J. Phys. B: At., Mol. Opt. Phys.* **2013**, *46*, 125201.
(99) Cai, Z.-L.; Tozer, D. J.; Reimers, J. R. Time-Dependent Density-Functional Determination of Arbitrary Singlet and Triplet Excited-State Potential Energy Surfaces: Application to the Water Molecule. *J. Chem. Phys.* **2000**, *113*, 7084−7096.
(100) Li, X.; Paldus, J. General-Model-Space State-Universal Coupled-Cluster Method: Excitation Energies of Water. *Mol. Phys.* **2006**, *104*, 661−676.

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

(101) Masuko, H.; Morioka, Y.; Nakamura, M.; Ishiguro, E.; Sasanuma, M. Absorption Spectrum of the H_2S Molecule In The Vacuum Ultraviolet Region. *Can. J. Phys.* **1979**, *57*, 745−760.
(102) Abuain, T.; Walker, I. C.; Dance, D. F. Electronic Excitation of Hydrogen Sulphide by Impact With (Near-Threshold) Low-Energy Electrons. *J. Chem. Soc., Faraday Trans. 2* **1986**, *82*, 811−816.
(103) Páleníková, J.; Kraus, M.; Neogrády, P.; Kellö, V.; Urban, M. Theoretical Study of Molecular Properties of Low-Lying Electronic Excited States of H_2O and H_2S. *Mol. Phys.* **2008**, *106*, 2333−2344.
(104) Skerbele, A.; Lassettre, E. N. Electron-Impact Spectra. *J. Chem. Phys.* **1965**, *42*, 395−401.
(105) Harshbarger, W. R. Identification of the $\tilde{C}$ State of Ammonia by Electron Impact Spectroscopy. *J. Chem. Phys.* **1971**, *54*, 2504−2509.
(106) Bartlett, R. J.; Del Bene, J. E.; Perera, S.; Mattie, R. Ammonia: The Prototypical Lone Pair Molecule. *J. Mol. Struct.: THEOCHEM* **1997**, *400*, 157−168.
(107) Arfa, M. B.; Tronc, M. Lowest Energy Triplet States of Group Vb Hydrides: NH_3 (ND_3) and PH_3. *Chem. Phys.* **1991**, *155*, 143−148.
(108) Abuain, T.; Walker, I. C.; Dance, D. F. The Lowest Triplet State in Ammonia and Methylamine Detected by Electron-Impact Excitation. *J. Chem. Soc., Faraday Trans. 2* **1984**, *80*, 641−645.
(109) Rubio, M.; Serrano-Andrés, L.; Merchán, M. Excited States of the Water Molecule: Analysis of the Valence and Rydberg Character. *J. Chem. Phys.* **2008**, *128*, 104305.
(110) Jacquemin, D.; Duchemin, I.; Blase, X. Benchmarking the Bethe-Salpeter Formalism on a Standard Organic Molecular Set. *J. Chem. Theory Comput.* **2015**, *11*, 3290−3304.
(111) Chutjian, A.; Hall, R. I.; Trajmar, S. Electron-Impact Excitation of H_2O and D_2O at Various Scattering Angles and Impact Energies in the Energy-Loss Range 4.2−12 eV. *J. Chem. Phys.* **1975**, *63*, 892−898.
(112) Pitarch-Ruiz, J.; Sánchez-Marín, J.; Martín, I.; Velasco, A. M. Vertical Excitation Energies and Ionization Potentials of H_2S. A Size-Consistent Self-Consistent Singles and Doubles Configuration Interaction $(SC)^2$-MR-SDCI Calculation. *J. Phys. Chem. A* **2002**, *106*, 6508−6514.
(113) Velasco, A. M.; Martín, I.; Pitarch-Ruiz, J.; Sánchez-Marín, J. MRSDCI Vertical Excitation Energies and MQDO Intensities for Electronic Transitions to Rydberg States in H2S. *J. Phys. Chem. A* **2004**, *108*, 6724−6729.
(114) Gupta, M.; Baluja, K. L. Application of R-Matrix Method to Electron-H_2S Collisions in the Low Energy Range. *Eur. Phys. J. D* **2007**, *41*, 475−483.
(115) O'Brien Lantz, K.; Vaida, V. Direct Absorption Spectroscopy of the First Excited Electronic Band of Jet-Cooled H_2S. *Chem. Phys. Lett.* **1993**, *215*, 329−335.
(116) Robin, M. B. In *Higher Excited States of Polyatomic Molecules*; Robin, M. B., Ed.; Academic Press: 1985; Vol. *III*.
(117) Chantranupong, L.; Hirsch, G.; Buenker, R. J.; Kimura, M.; Dillon, M. A. Theoretical Study of the Electronic Spectrum of Ammonia: Generalized Oscillator Strength Calculations for the *A*-*X* Transition. *Chem. Phys.* **1991**, *154*, 13−21.
(118) Huber, K. P.; Herzberg, G. *Constants of Diatomic Molecules*; Molecular Spectra and Molecular Structure; Van Nostrand: Princeton, NJ, 1979; Vol. *4*.
(119) Ben-Shlomo, S. B.; Kaldor, U. N_2 Excitations Below 15 eV by the Multireference Coupled-Cluster Method. *J. Chem. Phys.* **1990**, *92*, 3680−3682.
(120) Nielsen, E. S.; Jorgensen, P.; Oddershede, J. Transition Moments and Dynamic Polarizabilities in a Second Order Polarization Propagator Approach. *J. Chem. Phys.* **1980**, *73*, 6238−6246.
(121) Kucharski, S. A.; Włoch, M.; Musiał, M.; Bartlett, R. J. Coupled-Cluster Theory for Excited Electronic States: The Full Equation-Of-Motion Coupled-Cluster Single, Double, and Triple Excitation Method. *J. Chem. Phys.* **2001**, *115*, 8263−8266.
(122) Dora, A.; Tennyson, J.; Chakrabarti, K. Higher Lying Resonances in Low-Energy Electron Scattering with Carbon Monoxide. *Eur. Phys. J. D* **2016**, *70*, 197.
(123) Stahel, D.; Leoni, M.; Dressler, K. Nonadiabatic Representations of the $^1\Sigma_u^+$ and $^1\Pi_u$ States of the N_2 Molecule. *J. Chem. Phys.* **1983**, *79*, 2541−2558.
(124) Malsch, K.; Rebentisch, R.; Swiderek, P.; Hohlneicher, G. Excited States of Acetylene: A CASPT2 Study. *Theor. Chem. Acc.* **1998**, *100*, 171−182.
(125) Zyubin, A. S.; Mebel, A. M. Accurate Prediction of Excitation Energies to High-Lying Rydberg Electronic States: Rydberg States of Acetylene as a Case Study. *J. Chem. Phys.* **2003**, *119*, 6581−6587.
(126) Ventura, E.; Dallos, M.; Lischka, H. The Valence-Excited States T_1−T_4 and S_1−S_2 of Acetylene: A High-Level MR-CISD and MR-AQCC Investigation of Stationary Points, Potential Energy Surfaces, and Surface Crossings. *J. Chem. Phys.* **2003**, *118*, 1702−1713.
(127) Angeli, C.; Borini, S.; Ferrighi, L.; Cimiraglia, R. Ab Initio *N*-Electron Valence State Perturbation Theory Study of the Adiabatic Transitions in Carbonyl Molecules: Formaldehyde, Acetaldehyde, and Acetone. *J. Chem. Phys.* **2005**, *122*, 114304.
(128) Send, R.; Valsson, O.; Filippi, C. Electronic Excitations of Simple Cyanine Dyes: Reconciling Density Functional and Wave Function Methods. *J. Chem. Theory Comput.* **2011**, *7*, 444−455.
(129) Dressler, R.; Allan, M. A Dissociative Electron Attachment, Electron Transmission, and Electron EnergyLoss Study of the Temporary Negative Ion of Acetylene. *J. Chem. Phys.* **1987**, *87*, 4510−4518.
(130) Serrano-Andrès, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. Towards an Accurate Molecular Orbital Theory for Excited States: Ethene, Butadiene, and Hexatriene. *J. Chem. Phys.* **1993**, *98*, 3151−3162.
(131) Feller, D.; Peterson, K. A.; Davidson, E. R. A Systematic Approach to Vertically Excited States of Ethylene Using Configuration Interaction and Coupled Cluster Techniques. *J. Chem. Phys.* **2014**, *141*, 104302.
(132) Shen, J.; Li, S. Block Correlated Coupled Cluster Method with the Complete Active-Space Self-Consistent-Field Reference Function: Applications for Low-Lying Electronic Excited States. *J. Chem. Phys.* **2009**, *131*, 174101.
(133) Angeli, C. On the Nature of the Ionic Excited States: The V State of Ethene as a Prototype. *J. Comput. Chem.* **2009**, *30*, 1319−1333.
(134) Müller, T.; Lischka, H. Simultaneous Calculation of Rydberg and Valence Excited States of Formaldehyde. *Theor. Chem. Acc.* **2001**, *106*, 369−378.
(135) Schautz, F.; Buda, F.; Filippi, C. Excitations in Photoactive Molecules from Quantum Monte Carlo. *J. Chem. Phys.* **2004**, *121*, 5836−5844.
(136) Paone, S.; Moule, D.; Bruno, A.; Steer, R. Vibronic Analyses of the Rydberg and Lower Intravalence Electronic Transitions in Thioacetone. *J. Mol. Spectrosc.* **1984**, *107*, 1−11.
(137) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a Systematic Molecular Orbital Theory for Excited States. *J. Phys. Chem.* **1992**, *96*, 135−149.
(138) Hadad, C. M.; Foresman, J. B.; Wiberg, K. B. Excited States of Carbonyl Compounds. 1. Formaldehyde and Acetaldehyde. *J. Phys. Chem.* **1993**, *97*, 4293−4312.
(139) Gwaltney, S. R.; Bartlett, R. J. An Application of the Equation-Of-Motion Coupled Cluster Method to the Excited States of Formaldehyde, Acetaldehyde, and Acetone. *Chem. Phys. Lett.* **1995**, *241*, 26−32.
(140) Wiberg, K. B.; Stratmann, R. E.; Frisch, M. J. A Time-Dependent Density Functional Theory Study of the Electronically Excited States of Formaldehyde, Acetaldehyde and Acetone. *Chem. Phys. Lett.* **1998**, *297*, 60−64.
(141) Wiberg, K. B.; de Oliveira, A. E.; Trucks, G. A Comparison of the Electronic Transition Energies for Ethene, Isobutene, Formaldehyde, and Acetone Calculated Using RPA, TDDFT, and EOM-CCSD. Effect of Basis Sets. *J. Phys. Chem. A* **2002**, *106*, 4192−4199.

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

(142) Li, X.; Paldus, J. Multi-Reference State-Universal Coupled-Cluster Approaches to Electronically Excited States. *J. Chem. Phys.* **2011**, *134*, 214118.
(143) Furche, F.; Ahlrichs, R. Adiabatic Time-Dependent Density Functional Methods for Excited States Properties. *J. Chem. Phys.* **2002**, *117*, 7433−7447.
(144) Hättig, C. Response Theory and Molecular Properties (A Tribute to jan Linderberg and Poul Jørgensen). *Adv. Quantum Chem.* **2005**, *50*, 37−60.
(145) Walzl, K. N.; Koerting, C. F.; Kuppermann, A. Electron-Impact Spectroscopy of Acetaldehyde. *J. Chem. Phys.* **1987**, *87*, 3796−3803.
(146) Robin, M. B.; Basch, H.; Kuebler, N. A.; Wiberg, K. B.; Ellison, G. B. Optical Spectra of Small Rings. II. The Unsaturated Three-Membered Rings. *J. Chem. Phys.* **1969**, *51*, 45−52.
(147) Sauers, I.; Grezzo, L. A.; Staley, S. W.; Moore, J. H. Low-Energy Singlet-Triplet and Singlet-Singlet Transitions in Cycloalkenes. *J. Am. Chem. Soc.* **1976**, *98*, 4218−4222.
(148) McGlynn, S. P.; Rabalais, J. W.; McDonald, J. R.; Scherr, V. M. Electronic Spectoscopy of Isoelectronic Molecules. II. Linear Triatomic Groupings Containing Sixteen Valence Electrons. *Chem. Rev.* **1971**, *71*, 73−108.
(149) Fedorov, I.; Koziol, L.; Li, G.; Parr, J. A.; Krylov, A. I.; Reisler, H. Theoretical and Experimental Investigations of the Electronic Rydberg States of Diazomethane: Assignments and State Interactions. *J. Phys. Chem. A* **2007**, *111*, 4557−4566.
(150) Rittby, M.; Pal, S.; Bartlett, R. J. Multireference Coupled-Cluster Method: Ionization Potentials and Excitation Energies for Ketene and Diazomethane. *J. Chem. Phys.* **1989**, *90*, 3214−3220.
(151) Gingell, J.; Mason, N.; Zhao, H.; Walker, I.; Siggel, M. Vuv Optical-Absorption and Electron-Energy-Loss Spectroscopy of Formamide. *Chem. Phys.* **1997**, *220*, 191−205.
(152) Frueholz, R. P.; Flicker, W. M.; Kuppermann, A. Excited Electronic States of Ketene. *Chem. Phys. Lett.* **1976**, *38*, 57−60.
(153) Xiao, H.; Maeda, S.; Morokuma, K. CASPT2 Study of Photodissociation Pathways of Ketene. *J. Phys. Chem. A* **2013**, *117*, 7001−7008.
(154) Nooijen, M. First-Principles Simulation of the UV Absorption Spectrum of Ketene. *Int. J. Quantum Chem.* **2003**, *95*, 768−783.
(155) Dixon, R. N.; Kroto, H. W. The Electronic Spectrum of Nitrosomethane, CH_3NO. *Proc. R. Soc. London, Ser. A* **1965**, *283*, 423−432.
(156) Arenas, J. F.; Otero, J. C.; Peláez, D.; Soto, J. CASPT2 Study of the Decomposition of Nitrosomethane and Its Tautomerization Reactions in the Ground and Low-Lying Excited States. *J. Org. Chem.* **2006**, *71*, 983−991.
(157) Dolgov, E. K.; Bataev, V. A.; Pupyshev, V. I.; Godunov, I. A. Ab Initio Description of the Structure and Dynamics of the Nitrosomethane Molecule in the First Excited Singlet and Triplet Electronic States. *Int. J. Quantum Chem.* **2004**, *96*, 589−597.
(158) Reisler, H.; Krylov, A. I. Interacting Rydberg and Valence States in Radicals and Molecules: Experimental and Theoretical Studies. *Int. Rev. Phys. Chem.* **2009**, *28*, 267−308.
(159) Jacquemin, D.; Duchemin, I.; Blase, X. Is the Bethe−Salpeter Formalism Accurate for Excitation Energies? Comparisons with TD-DFT, CASPT2, and EOM-CCSD. *J. Phys. Chem. Lett.* **2017**, *8*, 1524−1529.
(160) Coe, J. P.; Paterson, M. J. State-Averaged Monte Carlo Configuration Interaction Applied to Electronically Excited States. *J. Chem. Phys.* **2013**, *139*, 154103.
(161) Habas, M. P.; Dargelos, A. Ab Initio CI Calculations of Electronic and Vibrational Spectra of Diazomethane. *Chem. Phys.* **1995**, *199*, 177−182.
(162) Serrano-Andrés, L.; Fülscher, M. P. Theoretical Study of the Electronic Spectroscopy of Peptides. 1. The Peptidic Bond: Primary, Secondary, and Tertiary Amides. *J. Am. Chem. Soc.* **1996**, *118*, 12190−12199.
(163) Besley, N. A.; Hirst, J. D. Ab Initio Study of the Electronic Spectrum of Formamide with Explicit Solvent. *J. Am. Chem. Soc.* **1999**, *121*, 8559−8566.
(164) Szalay, P. G.; Császár, A. G.; Nemes, L. Electronic States of Ketene. *J. Chem. Phys.* **1996**, *105*, 1034−1045.
(165) Lacombe, S.; Loudet, M.; Dargelos, A.; Camou, J. Calculation of the Electronic and Photoelectronic Spectra of Nitroso Compounds: A Reinvestigation by Use of Configuration Interaction Methods. *Chem. Phys.* **2000**, *258*, 1−12.
(166) Dolgov, E. K.; Bataev, V. A.; Godunov, I. A. Structure of the Nitrosomethane Molecule (CH_3NO) in the Ground Electronic State: Testing of Ab Initio Methods for the Description of Potential Energy Surface. *Int. J. Quantum Chem.* **2004**, *96*, 193−201.
(167) Moore, B., II; Autschbach, J. Longest-Wavelength Electronic Excitations of Linear Cyanines: The Role of Electron Delocalization and of Approximations in Time-Dependent Density Functional Theory. *J. Chem. Theory Comput.* **2013**, *9*, 4991−5003.
(168) Boulanger, P.; Jacquemin, D.; Duchemin, I.; Blase, X. Fast and Accurate Electronic Excitations in Cyanines with the Many-Body Bethe-Salpeter Approach. *J. Chem. Theory Comput.* **2014**, *10*, 1212−1218.
(169) Zhekova, H.; Krykunov, M.; Autschbach, J.; Ziegler, T. Applications of Time Dependent and Time Independent Density Functional Theory to the First π to π^* Transition in Cyanine Dyes. *J. Chem. Theory Comput.* **2014**, *10*, 3299−3307.
(170) Le Guennic, B.; Jacquemin, D. Taking Up the Cyanine Challenge with Quantum Tools. *Acc. Chem. Res.* **2015**, *48*, 530−537.
(171) Tarte, P. Recherches Spectroscopiques sur les Composés Nitrosés. *Bull. Soc. Chim. Belg.* **1954**, *63*, 525−541.
(172) Ernsting, N. P.; Pfab, J.; Romelt, J. Geometry Changes Accompanying Electronic Excitation of Nitrosomethane in the 650 nm Region. *J. Chem. Soc., Faraday Trans. 2* **1978**, *74*, 2286−2294.
(173) Gordon, R. D.; Luck, P. Conformational Changes Accompanying Electronic Excitation of CD_3NO. *Chem. Phys. Lett.* **1979**, *65*, 480−483.
(174) Jacquemin, D. What is the Key for Accurate Absorption and Emission Calculations ? Energy or Geometry ? *J. Chem. Theory Comput.* **2018**, *14*, 1534−1543.
(175) The larger deviations with CCSDTQ is likely due to the larger number of Rydberg states, which are more basis set sensitive, in that set. Let us note also that several CCSDTQ/*aug*-cc-pVTZ reference values are obtained by correcting CCSDTQ/*aug*-cc-pVDZ values. However, such a procedure has been shown to be very robust above, so that it does not impact the present analysis.

DOI: 10.1021/acs.jctc.8b00406
J. Chem. Theory Comput. 2018, 14, 4360−4379

10.2 Application to QMC : the Fe–S molecule

In a diffusion quantum Monte Carlo calculation (DMC), a trial wave function Ψ is required. One is able to compute

$$E_{\mathrm{DMC}} = \frac{\langle \Phi^{\mathrm{FN}} | \hat{H} | \Psi \rangle}{\langle \Phi^{\mathrm{FN}} | \Psi \rangle} = \frac{\int \Phi^{\mathrm{FN}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \hat{H} \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \mathrm{d}\mathbf{r}_1 \ \ldots \ \mathrm{d}\mathbf{r}_N}{\int \Phi^{\mathrm{FN}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \mathrm{d}\mathbf{r}_1 \ \ldots \ \mathrm{d}\mathbf{r}_N} \tag{10.1}$$

$$= \frac{\int \Phi^{\mathrm{FN}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \frac{\hat{H}\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} \mathrm{d}\mathbf{r}_1 \ \ldots \ \mathrm{d}\mathbf{r}_N}{\int \Phi^{\mathrm{FN}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \mathrm{d}\mathbf{r}_1 \ \ldots \ \mathrm{d}\mathbf{r}_N} \tag{10.2}$$

as a stochastic average with the $3N$-dimensional density $\Psi \times \Phi^{\mathrm{FN}}$:

$$E_{\mathrm{DMC}} = \left\langle \frac{\hat{H}\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} \right\rangle_{\Psi \times \Phi^{\mathrm{FN}}}. \tag{10.3}$$

Φ^{FN} is an improved wave function of the form

$$\Phi^{\mathrm{FN}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \times w(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{10.4}$$

where w is a non-negative function, given by

$$w(\mathbf{R}) = \lim_{t \to \infty} \exp\left(\int_0^t \mathrm{d}t \frac{\hat{H}\Psi(\mathbf{R}(t))}{\Psi(\mathbf{R}(t))} \right). \tag{10.5}$$

This expression can't be computed exactly, but it can be sampled with a diffusion process with drift and branching.[86]

The non-negativity constraint of w implies that the nodal hyper-surfaces of Φ^{FN} coincide with those of Ψ, but not necessarily with those of the exact wave function. Hence, the Diffusion Monte Carlo (DMC) method suffers from the so-called *fixed-node approximation*. But if the trial wave function has nodes which coincide with those of the exact wave function, the exact energy is obtained.

There is no way to improve the nodes of the wave function by minimizing directly E_{DMC}. However, it has recently been shown[87] that for each atomic basis set there was an E_{DMC} associated with the FCI wave function, and increasing the size of the basis set enabled a smooth extrapolation to the exact energy. Hence, one expects that computing DMC energy differences with FCI wave functions will show an efficient compensation of errors.

In this work, we have introduced an extrapolation scheme, E_{DMC} as a function of E_{PT2} to estimate the E_{DMC} we would have obtained if the trial wave function was a FCI wave function. We have applied this scheme to the difficult case of the Fe–S

molecule, for which the nature of the ground state in not clear. Two states of different symmetries, $^5\Sigma$ and $^5\Delta$ are very close in energy, and the main methods of quantum chemistry disagree. State-of-the art QMC calculations were giving the $^5\Delta$ state as the ground state,[88] and our results agree with these conclusions.

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 1395−1402

pubs.acs.org/JCTC

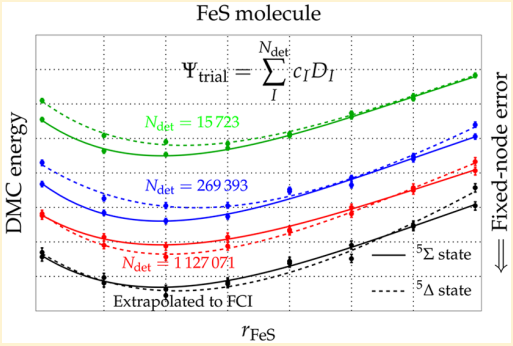
Deterministic Construction of Nodal Surfaces within Quantum Monte Carlo: The Case of FeS

Anthony Scemama,* Yann Garniron, Michel Caffarel, and Pierre-François Loos*

Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, 31013 Toulouse Cede, France

Supporting Information

ABSTRACT: In diffusion Monte Carlo (DMC) methods, the nodes (or zeroes) of the trial wave function dictate the magnitude of the fixed-node (FN) error. In standard DMC implementations, the nodes are optimized by *stochastically* optimizing a short multideterminant expansion in the presence of an explicitly correlated Jastrow factor. Here, following a recent proposal, we pursue a different route and consider the nodes of selected configuration interaction (sCI) expansions built with the CIPSI (Configuration Interaction using a Perturbative Selection made Iteratively) algorithm. By increasing the size of the sCI expansion, these nodes can be *systematically* and *deterministically* improved. The present methodology is used to investigate the properties of the transition metal sulfide molecule FeS. This apparently simple molecule has been shown to be particularly challenging for electronic structure theory methods due to the proximity of two low-energy quintet electronic states of different spatial symmetry and the difficulty to treat them on equal footing from a one-electron basis set point of view. In particular, we show that, at the triple-ζ basis set level, all sCI results—including those extrapolated at the full CI (FCI) limit—disagree with experiment, yielding an electronic ground state of $^5\Sigma^+$ symmetry. Performing FN-DMC simulation with sCI nodes, we show that the correct $^5\Delta$ ground state is obtained if sufficiently large expansions are used. Moreover, we show that one can systematically get accurate potential energy surfaces and reproduce the experimental dissociation energy as well as other spectroscopic constants.



1. INTRODUCTION

From an experimental point of view, transition metal sulfides have proven to be useful in a variety of fields including biological chemistry,[1] catalysis,[2] and electrochemistry.[3] From the computational side, the apparently simple FeS diatomic molecule turns out to be a challenging system for computational chemists. The major hurdle originates from the energetic proximity of two electronic states

$$^5\Delta: \sigma^2\pi^4\sigma^2\delta^3\sigma^1\pi^2 \qquad ^5\Sigma^+: \sigma^2\pi^4\sigma^2\delta^2\sigma^2\pi^2$$

with the same multiplicity competing for the ground state. To make things worse, the equilibrium bond lengths associated with these two states are extremely close to each other.

Experimentally, the ground state of FeS is assigned to be $^5\Delta$,[4,5] with an equilibrium bond length of r_e = 2.017 Å,[5] and a dissociation energy D_0 = 3.31(15) eV.[6] For this state, the harmonic frequency ω_e has been estimated to be 518 ± 5 cm^{-1}.[7] Very recently, a much more accurate value of the dissociation energy D_0 = 3.240(3) eV has been obtained by Matthew et al. using the predissociation threshold technique.[8]

FeS has been extensively studied by density functional theory (DFT) and post-Hartree−Fock methods. In short, most (but not all) DFT functionals correctly predict a $^5\Delta$ ground state,[9−13] while CAS-based multireference methods such as CASSCF/ACPF,[14] CASPT2,[15] or CASSCF/ICACPF[16] systematically predict $^5\Sigma^+$ lower than $^5\Delta$.

Here, we investigate this problem using quantum Monte Carlo (QMC). In recent years, QMC has been applied with great success to a large variety of main group compounds (see, e.g, refs 17−20 for recent applications). Transition metal systems are more challenging, but a number of successful studies have also been reported.[21−41]

When multireference effects are weak, QMC is seen as a very accurate method providing benchmark results of a quality similar or superior to the gold-standard CCSD(T). However, when multireference effects are dominant, as is usually the case for metallic compounds with partially filled d shells, the situation is more complicated, and one has to revert to multireference approaches.[42−44] Indeed, the results may depend significantly on the trial wave function Ψ_T used to guide the walkers through configuration space. In theory, QMC results should be independent of the choice of Ψ_T. However, it is not true in practice because of the fixed-node (FN) approximation, which imposes the Schrödinger equation to be solved with the additional constraint that the solution vanishes at the zeroes (nodes) of the trial wave function. Using an approximate Ψ_T leads to approximate nodes and, thus, to an approximate energy, known as the FN energy. The FN energy being an upper bound of the exact energy, this gives us a

Received: December 14, 2017
Published: January 27, 2018

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

practical and convenient variational criterion for characterizing the nodal quality. In situations where multireference effects are strong, getting accurate nodes may be difficult. As we shall see, this is the main challenge we are facing in the present work.

Most QMC studies for transition metal-containing systems have been performed with pseudopotentials. In this case, an additional source of error, the so-called localization error, is introduced. This error, specific to QMC, adds up to the standard error associated with the approximate nature of pseudopotentials. Similarly to the FN error, the localization error depends on Ψ_T and vanishes only for the exact wave function. Therefore, to get accurate and reliable QMC results, both sources of error have to be understood and controlled.

In 2011, Petz and Lüchow reported a FN diffusion Monte Carlo (FN-DMC) study of the energetics of diatomic transition metal sulfides from ScS to FeS using pseudopotentials and single-determinant trial wave functions.[34] The pseudopotential dependence was carefully investigated, and comparisons with both DFT and CCSD(T) as well as experimental data were performed. In short, it was found that FN-DMC shows a higher overall accuracy than both B3LYP and CCSD(T) for all diatomics except for CrS and FeS, which appeared to be particularly challenging.

Very recently, Haghighi-Mood and Lüchow had a second look at the difficult case of FeS.[41] In particular, they explored the impact of the level of optimization on the parameters of multideterminant trial wave functions (partial or full optimization of the Jastrow, determinant coefficients, and molecular orbitals) on both the FN and localization errors. Their main conclusions can be summarized as follows. Using a single-determinant trial wave function made of B3LYP orbitals or fully optimized orbitals in the presence of a Jastrow factor is sufficient to yield the correct state ordering. However, in both cases, the dissociation energy is far from the experimental value, and thus, multideterminant trial wave functions must be employed. Although a natural choice would be to take into account the missing static correlation via a CASSCF-based trial wave function, they showed that it is insufficient and that a full optimization is essential to get both the correct electronic ground state and reasonable estimates of the spectroscopic constants.

In the present study, we revisit this problem within the original QMC protocol developed in our group these past few years.[37,45−49] In the conventional protocol, prevailing in the QMC community and employed by Haghighi-Mood and Lüchow, the nodes of the Slater−Jastrow (SJ) trial wave function

$$\Psi_T^{SJ} = \exp(J)\Psi_{det} \quad (1)$$

are obtained by partially or fully optimizing the Jastrow factor J and the multiderminant expansion Ψ_{det} (containing typically a few hundreds or thousands of determinants). This step is performed in a preliminary variational Monte Carlo calculation by minimizing the energy, the variance of the local energy (or a combination of both), employing one of the optimization methods developed within the QMC context.[50−53] We note that, in practice, the optimization must be carefully monitored because of the large number of parameters (several hundreds or thousands), the nonlinear nature of most parameters (several minima may appear), and the inherent presence of noise in the function to be minimized.

Within our protocol, we rely on configuration interaction (CI) expansions in order to get accurate nodal surfaces, without resorting to the stochastic optimization step. Our fundamental motivation is to take advantage of all of the machinery and experience developed these last decades in the field of wave function methods. In contrast to the standard protocol described above, the CI nodes can be improved *deterministically* and *systematically* by increasing the size of the CI expansion. In the present work, we do not introduce any Jastrow factor, essentially to avoid the expensive numerical quadrature involved in the calculation of the pseudopotential and to facilitate control of the localization error. To keep the size of the CI expansion reasonable and retain only the most important determinants, we propose using selected CI (sCI) algorithms, such as CIPSI (Configuration Interaction using a Perturbative Selection made Iteratively).[45] Using a recently proposed algorithm to handle large numbers of determinants in FN-DMC[47] we are able to consider up to a few million determinants in our simulations.

Over the past few years, we have witnessed a rebirth of sCI methods.[36,37,45−48,54−77] Although these various approaches appear under diverse acronyms, most of them rely on the very same idea of selecting determinants iteratively according to their contribution to the wave function or energy, an idea that goes back to 1969 in the pioneering works of Bender and Davidson,[54] and Whitten and Hackmeyer.[55] Importantly, we note that any sCI variants can be employed here.

The price to pay for using sCI expansions instead of optimized SJ trial wave functions is the need to employ much larger multideterminant expansions in order to reach a comparable level of statistical fluctuations. In practice, a higher computational cost is thus required. Furthermore, because of the absence of an optimized Jastrow factor, systematic errors, such as the time step and basis set incompleteness errors, are larger. Then, in our procedure, it is particularly important to make use of extrapolation procedures for each systematic error. However, these disadvantages are compensated by the appealing features of sCI nodes: (i) they are built in a fully automated way; (ii) they are unique and reproducible; (iii) they can be systematically improved by increasing the level of selection and/or the basis set (with the possibility of complete basis set extrapolation[48]); and (iv) they easily produce smooth potential energy surfaces.[46]

2. COMPUTATIONAL DETAILS

All trial wave functions have been generated with the electronic structure software QUANTUM PACKAGE,[78] while the QMC calculations have been performed with the QMC = CHEM suite of programs.[79,80] Both softwares were developed in our laboratory and are freely available. For all calculations, we used the triple-ζ basis sets of Burkatzki et al.[81,82] (VTZ-ANO-BFD for Fe and VTZ-BFD for S) in conjunction with the corresponding Burkatzki−Filippi−Dolg (BFD) small-core pseudopotentials including scalar relativistic effects. For more details about our implementation of pseudopotentials within QMC, we refer the interested readers to ref 49. As pointed out by Hammond and co-workers,[83] when the trial wave function does not include a Jastrow factor, the nonlocal pseudopotential can be localized analytically and the usual numerical quadrature over the angular part of the nonlocal pseudopotential can be eschewed. In practice, calculation of the localized part of the pseudopotential represents only a small overhead (about 15%) with respect to a calculation without a pseudopotential (and the same number of electrons). To check that the BFD pseudopotentials do not introduce any serious artifact, we

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

have computed the nonparallelism error between the frozen-core FCI curves obtained with and without pseudopotentials. A nonparallelism error of 4 mE_h has been observed, which validates the accuracy of these pseudopotentials for the present study.

In order to compare our results for the dissociation energy of FeS with the experimental value of Matthew et al.[8] and the (theoretical) benchmark value of Haghighi-Mood and Lüchow,[41] we have taken into account the zero-point energy (ZPE) correction, the spin−orbit effects as well as the core−valence correlation contribution the same way as those in ref 41. For the $^5\Delta$ state, this corresponds to an increase of the dissociation energy by 0.06 eV and a 0.02 eV stabilization of $^5\Delta$ compared to $^5\Sigma^+$. Unless otherwise stated, atomic units are used throughout.

2.1. Jastrow-Free Trial Wave Functions. Within the spin-free formalism used in QMC, a CI-based trial wave function is written as

$$\Psi_T(\boldsymbol{R}) = \sum_{I=1}^{N_{det}} c_I D_I(\boldsymbol{R}) = \sum_{I=1}^{N_{det}} c_I D_I^{\uparrow}(\boldsymbol{R}^{\uparrow}) D_I^{\downarrow}(\boldsymbol{R}^{\downarrow}) \tag{2}$$

where $\boldsymbol{R} = (\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$ denotes the full set of electronic spatial coordinates, $\boldsymbol{R}^{\uparrow}$ and $\boldsymbol{R}^{\downarrow}$ are the two subsets of spin-up ($\uparrow$) and spin-down ($\downarrow$) electronic coordinates, and $D_I^{\sigma}(\boldsymbol{R}^{\sigma})$ ($\sigma = \uparrow$ or $\downarrow$) are spin-specific determinants.

In practice, the various products $D_I^{\uparrow}D_I^{\downarrow}$ contain many identical spin-specific determinants. For computational efficiency, it is then advantageous to group them and compute only once their contribution to the wave function and its derivatives.[47] Therefore, the Jastrow-free CI trial wave functions employed in the present study are rewritten in a "spin-resolved" form

$$\Psi_T(\boldsymbol{R}) = \sum_{i=1}^{N_{det}^{\uparrow}} \sum_{j=1}^{N_{det}^{\downarrow}} c_{ij} \mathcal{D}_i^{\uparrow}(\boldsymbol{R}^{\uparrow}) \mathcal{D}_j^{\downarrow}(\boldsymbol{R}^{\downarrow}) \tag{3}$$

where $\{\mathcal{D}_i^{\sigma}\}_{i=1,...,N_{det}^{\sigma}}$ denotes the set of all *distinct* spin-specific determinant appearing in eq 2.

2.2. Quantum Monte Carlo Calculations. To avoid handling too many determinants in Ψ_T, a truncation scheme has to be introduced. In most CI and/or QMC calculations, the expansion is truncated by introducing a cutoff either on the CI coefficients or on the norm of the wave function. Here, we use an alternative truncation scheme knowing that most of the computational effort lies in the calculation of the spin-specific determinants and their derivatives. Removing a product of determinants whose spin-specific determinants are already present in other products does not change significantly the computational cost. Accordingly, a natural choice is then to truncate the wave function by removing *independently* spin-up and spin-down determinants. To do so, we decompose the norm of the wave function as

$$\mathcal{N} = \sum_{i=1}^{N_{det}^{\uparrow}} \sum_{j=1}^{N_{det}^{\downarrow}} |c_{ij}|^2 = \sum_{i=1}^{N_{det}^{\uparrow}} \mathcal{N}_i^{\uparrow} = \sum_{j=1}^{N_{det}^{\downarrow}} \mathcal{N}_j^{\downarrow} \tag{4}$$

A determinant $\mathcal{D}_i^{\uparrow}$ is retained in Ψ_T if

$$\mathcal{N}_i^{\uparrow} = \sum_{j=1}^{N_{det}^{\downarrow}} |c_{ij}|^2 > \epsilon \tag{5}$$

where ϵ is a user-defined threshold. A similar formula is used for $\mathcal{D}_j^{\downarrow}$. When $\epsilon = 0$, the entire set of determinants is retained in the QMC simulation.

In order to treat the two electronic states ($^5\Sigma^+$ and $^5\Delta$) on equal footing, a common set of spin-specific determinants $\{\mathcal{D}_i^{\sigma}\}_{i=1,...,N_{det}^{\sigma}}$ is used for both states. In addition, a common set of molecular orbitals issued from a preliminary state-averaged CASSCF calculation is employed. These CASSCF calculations have been performed with the GAMESS, package[84] while for the atoms, we have performed ROHF calculations. The active space contains 12 electrons and 9 orbitals (3d and 4s orbitals of Fe and 3p orbitals of S). The multideterminant expansion (eq 2) has been constructed using the sCI algorithm CIPSI,[56,57] which uses a second-order perturbative criterion to select the energetically important determinants D_I in the FCI space.[36,37,45−48,67] An n_s-state truncated sCI expansion (here $n_s = 2$) is obtained via a natural generalization of the state-specific criterion introduced in eq 5: a determinant $\mathcal{D}_i^{\uparrow}$ is retained in Ψ_T if

$$\mathcal{N}_i^{\uparrow} = \frac{1}{n_s} \sum_{k=1}^{n_s} \sum_{j=1}^{N_{det}^{\downarrow}} |c_{ij}^{(k)}|^2 > \epsilon \tag{6}$$

with a similar formula for $\mathcal{D}_j^{\downarrow}$.

The characteristics of the various trial wave functions considered here (and their acronyms) at $r_{FeS} = 2.0$ Å are presented in Table 1. For other r_{FeS} values, the numbers of determinants are slightly different. Our largest sCI trial wave function contains 8 388 608 determinants and is labeled sCI(∞). The sCI(n) wave functions with $n = 4$, 5, and 6 are obtained by truncation of the sCI(∞) expansion setting $\epsilon = 10^{-n}$. They contain respectively 15 723, 269 393, and 1 127 071 determinants. At this stage, we are not able to use the entire 8 388 608 determinants of the sCI(∞) wave function within our FN-DMC simulations. In comparison, Haghighi-Mood and Lüchow's CASSCF-based trial wave function (labeled as HML in Table 2) only contains 630 and 500 determinants for the $^5\Sigma^+$ and $^5\Delta$ states, respectively.[41] However, as discussed in the Introduction, fully optimized SJ trial wave functions require much smaller multireference expansions.

Table 1. Characteristics of the Various sCI Expansions at r_{FeS} = 2.0 Å for Various Levels of Truncation along with Characteristics of the Extrapolated FCI (exFCI) Expansion

method	ϵ	N_{det}	$N_{det}^{\uparrow}$	$N_{det}^{\downarrow}$	acronym
sCI	10^{-4}	15 723	191	188	sCI(4)
	10^{-5}	269 393	986	1 191	sCI(5)
	10^{-6}	1 127 071	3883	4623	sCI(6)
	0	8 388 608	364 365	308 072	sCI(∞)
exFCI		$\sim 10^{27}$	$\sim 10^{16}$	$\sim 10^{11}$	exFCI

On the basis of these trial wave functions, we performed FN-DMC calculations with the stochastic reconfiguration algorithm developed by Assaraf et al.[85] One of the main advantages of this particular algorithm is that the number of walkers is constant during the simulation, hence avoiding the population control step. Here we have used 100 walkers in our simulations.

In order to remove the time step error, all of our FN-DMC results have been extrapolated to zero time step using a two-point linear extrapolation with $\tau = 2 \times 10^{-4}$ and 4×10^{-4}.[86]

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

Table 2. FN-DMC Energies E_{DMC} (in hartrees) at Equilibrium Geometry, Dissociation Energy D_0 (in eV), Equilibrium Distance r_e (in Å), and Harmonic Frequency ω_e (in cm^{-1}) for the $^5\Sigma^+$ and $^5\Delta$ of FeS Obtained with Various Trial Wave Functions Ψ_T[a]

	FeS ($^5\Sigma^+$)			FeS ($^5\Delta$)			Fe (^{5}D)	S (^{3}P)		
Ψ_T	E_{DMC}	r_e	ω_e	E_{DMC}	r_e	ω_e	E_{DMC}	E_{DMC}	D_0	ref
HML	−134.0571(4)	2.00(1)	518(7)	−134.0579(4)	2.031(7)	499(11)	−123.8126(4)	−10.1314(1)	3.159(15)	41
sCI(4)	−134.0101(8)	1.994(7)	532(20)	−134.0040(7)	2.029(7)	502(15)	−123.8028(9)	−10.1279(2)	2.055(20)	this work
sCI(5)	−134.0479(10)	1.992(8)	551(24)	−134.0402(10)	2.048(11)	489(21)	−123.8234(10)	−10.1312(2)	2.389(28)	this work
sCI(6)	−134.061 (14)	1.994(12)	497(35)	−134.0671(14)	2.004(11)	550(32)	−123.8300(12)	−10.1334(3)	3.062(39)	this work
exFCI	−134.0863(15)	1.990(12)	523(37)	−134.0885(18)	2.016(14)	525(40)	−123.8372(12)	−10.1336(3)	3.267(49)	this work
exp.					2.017	518(5)			3.240(3)	5, 7, 8

[a]The error bar corresponding to one standard error is reported in parentheses.

The behavior of the FN-DMC energy as a function of τ is depicted in Figure 1 for various time step values. Note that

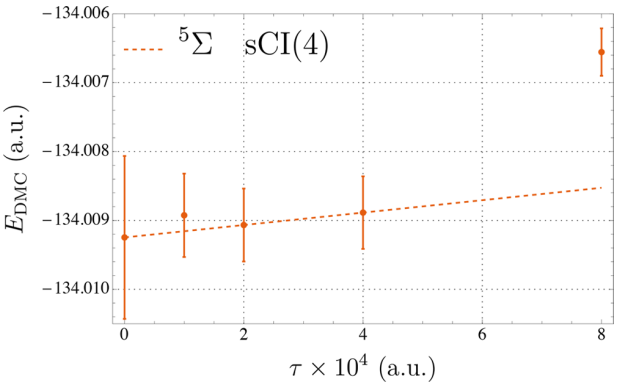


Figure 1. E_{DMC} (in hartrees) for the $^5\Sigma^+$ state of FeS as a function of the time step τ at r_{FeS} = 2.0 Å. The linear extrapolation between $\tau = 2 \times 10^{-4}$ and 4×10^{-4} is represented as a dashed read line. The error bar corresponds to one standard error.

because the variance of the local energy is larger than that in SJ calculations time step errors are enhanced and shorter time steps are required.

2.3. Extrapolation Procedure. In order to extrapolate our sCI results to the FCI limit, we have adopted the method recently proposed by Holmes, Umrigar, and Sharma[76] in the context of the (selected) heat-bath CI method.[72,75,76] It consists of extrapolating the sCI energy E_{sCI} as a function of the second-order Epstein−Nesbet energy

$$E_{PT2} = \sum_{\alpha} \frac{|\langle \alpha|\hat{H}|\Psi_{sCI}\rangle|^2}{E_{sCI} - \langle \alpha|\hat{H}|\alpha\rangle} \tag{7}$$

which is an estimate of the truncation error in the sCI algorithm, i.e, $E_{PT2} \approx E_{FCI} - E_{sCI}$.[56] In eq 7, the sum runs over all external determinants $|\alpha\rangle$ (i.e., not belonging to the sCI expansion) connected via $\hat{H}$ to the sCI wave function Ψ_{sCI}, i.e., $\langle \alpha|\hat{H}|\Psi_{sCI}\rangle \neq 0$. When $E_{PT2} = 0$, the FCI limit has effectively been reached. In our case, E_{PT2} is efficiently evaluated thanks to our recently proposed hybrid stochastic−deterministic algorithm,[67] which explains the presence of an error bar in the numerical values of E_{PT2} gathered in the tables reported in the Supporting Information. The extrapolated FCI results are labeled exFCI from hereon.

3. RESULTS AND DISCUSSION

In Table 2, we report FN-DMC energies at equilibrium geometry as well as other quantities of interest such as the dissociation energy D_0, the equilibrium distance r_e, and the harmonic frequency ω_e obtained with various trial wave functions. These values are obtained via the standard four-parameter Morse potential representation of the numerical values gathered in the Supporting Information. (The error bars have been obtained by fitting a large set of energy curves. Each of these curves is obtained from independent realizations of the statistical noise. Note that due to the absence of correlations in the statistical noise, the error bars obtained in this way are certainly overestimated.) For comparison purposes, Haghighi-Mood and Lüchow's results are also reported based on their best trial wave function.[41] When available, the experimental result is also reported.[5−8] The value of D_0 is always calculated with respect to the $^5\Delta$ state, adding the corresponding corrections for ZPE, spin−orbit effects, and core−valence correlation, as described above (see section 2.1). (The dissociation energies are calculated by treating the atoms and the molecule at the same level of theory, i.e., at the same truncation order.) The dissociation profile of FeS obtained with FN-DMC is depicted in Figure 2 for various trial wave functions.

The first observation we would like to make is that at the variational level the $^5\Delta$ state is never found lower in energy than the $^5\Sigma^+$ state, even after performing extrapolation to the FCI limit. This is illustrated by the left panel of Figure 3, which shows the behavior of the sCI energy as a function of E_{PT2} as well as the extrapolated FCI value The extrapolated value has been obtained via a three-point linear extrapolation of the sCI energy as a function of E_{PT2} using the sCI(5), sCI(6), and sCI(∞) results. (The raw data can be found in the Supporting Information.) It is clear from these results that the $^5\Sigma^+$ and $^5\Delta$ do not cross, even at the FCI limit. Because all post-Hartree−Fock methods are indeed an approximation of FCI, they are expected to predict a $^5\Sigma^+$ ground state for this particular basis set. This observation is in agreement with the CASPT2 results previously published in the literature.[14−16] Thus, one can attribute the wrong state ordering to basis set incompleteness, the only remaining approximation.

To obtain the FN-DMC curve with an *effective* FCI trial wave function, we have generalized the extrapolation procedure described in the previous section, and we have performed a three-point linear extrapolation of the FN-DMC energy as a function of $E_{exFCI} - E_{sCI}$ using the sCI(4), sCI(5), and sCI(6) results (see the right panel of Figure 3). Contrary to the sCI results, at the FN-DMC level, the $^5\Delta$ state does eventually become lower in energy than the $^5\Sigma^+$ state. However, one must include at least a few hundred thousand determinants in order to find the proper ground state. For larger ϵ values (10^{-4} and 10^{-5}), D_0 is underestimated due to the unbalanced treatment of the isolated atoms compared to the dimer at equilibrium

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

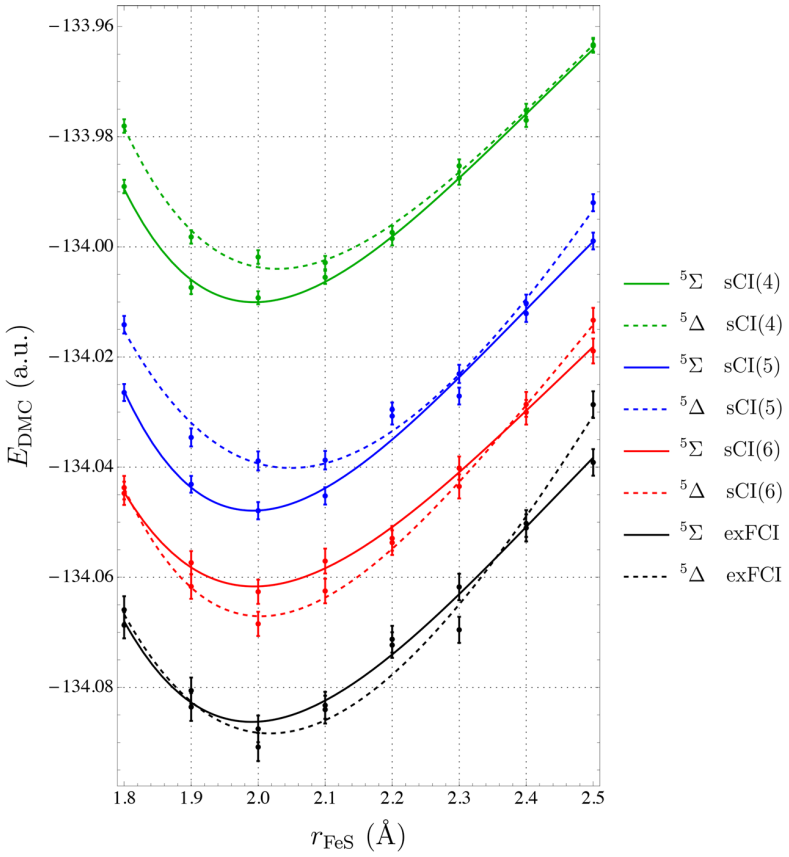


Figure 2. E_{DMC} (in hartrees) for the $^5\Sigma^+$ (solid) and $^5\Delta$ (dashed) states of FeS as a function of r_{FeS} (in Å) for various trial wave functions. The error bar corresponds to one standard error.

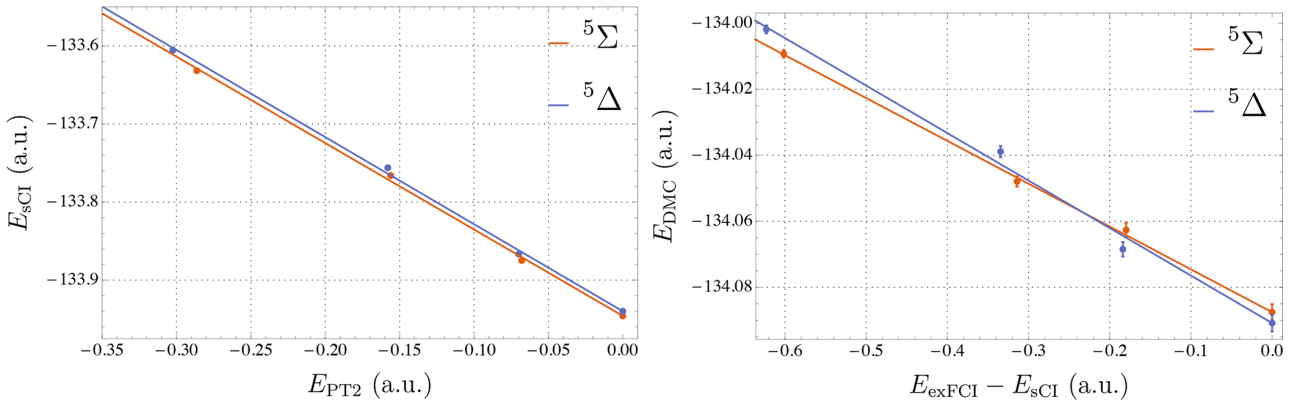


Figure 3. Three-point linear extrapolation of E_{sCI} (left) and E_{DMC} (right) to the FCI limit ($E_{PT2} = 0$ and $E_{exFCI} - E_{sCI} = 0$, respectively) for the $^5\Sigma^+$ (red) and $^5\Delta$ (blue) states of FeS at $r_{FeS} = 2.0$ Å. The error bar corresponds to one standard error.

geometry. Indeed, for a given number of determinants, the energy of the atomic species is much closer to the FCI limit than the energy of FeS.

For $\epsilon = 10^{-6}$, our approach correctly predicts a $^5\Delta$ ground state. However, although our FN-DMC energies are much lower than those obtained with the HML trial wave function, our estimate of the dissociation energy ($D_0 = 3.062(39)$ eV) is still below the experimental value. This underestimation of D_0 can be ultimately tracked to the lack of size-consistency of the truncated CI wave function. With more than 10^6 determinants in the variational space, the wave function is still 150 mE_h higher than the exFCI wave function, while the atoms are much better described by the sCI wave function. To remove the size-consistency error, we then extrapolate the FN-DMC energies to the (size-consistent) FCI limit of the trial wave function, as described above.

In that case, using the extrapolated FN-DMC energies of the molecule and isolated atoms reported in Table 2, we obtain a value of $D_0 = 3.267(49)$ eV, which nestles nicely between the experimental values of Matthew et al.[8] (3.240(3) eV) and Drowart et al.[6] (3.31(15) eV). As a final remark, we note that other spectroscopic constants, such as the equilibrium

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

geometry and the harmonic frequency, are fairly well reproduced by our approach. However, increasing the number of determinants in the trial wave function does not systematically improve the equilibrium distances. The same comment can be made for the harmonic frequencies. Overall, we found that our values of ω_e and r_e for the $^5\Delta$ state are closer to the experimental results[5,7] than HML's values.

4. CONCLUSIONS

In this article, the potential energy curves of two electronic states—$^5\Delta$ and $^5\Sigma^+$—of the FeS molecule have been calculated using the sCI algorithm CIPSI and the stochastic FN-DMC method. In all of our sCI calculations, $^5\Sigma^+$ is found to be the ground state, in disagreement with experiment. It is not only true for all CIPSI expansions with up to 8 million determinants but also for the estimated FCI limit obtained using the extrapolation procedure recently proposed by Holmes et al.[76]

This conclusion agrees with other high-level ab initio wave function calculations, which all wrongly predict a ground state of $^5\Sigma^+$ symmetry. FN-DMC calculations have been performed using CIPSI expansions including up to 1 127 071 selected determinants as trial wave functions. Contrary to standard QMC calculations, we do not introduce any Jastrow factor: the CI expansions have been used as they are (no optimization). It is found that, when the number of determinants in the trial wave function reaches a few hundred thousand, the FN-DMC ground state switches from the $^5\Sigma^+$ state to the correct $^5\Delta$ state, as predicted experimentally.

Generalizing the extrapolation procedure of Holmes et al.,[76] an estimate of the FN-DMC potential energy curves corresponding to the FCI nodes can be obtained. The resulting dissociation energy is found to be 3.267(49) eV, in agreement with the recent experimental value of Matthew et al. (3.240(3) eV).[8] As already observed in previous applications, the FN-DMC energy obtained with CIPSI nodes is found to *systematically* decrease as a function of the number of selected determinants.[36,37,45,46,48,49] For the largest expansion, our FN energies are lower than the values recently reported by Haghighi-Mood and Lüchow[41] using a fully optimized SJ trial wave function. This important result illustrates that "pure" sCI nodes are a realistic alternative to stochastically optimized SJ trial wave functions (although more computationally demanding), even for a challenging system such as FeS. A similar conclusion had already been drawn in our recent study of the water molecule.[48]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b01250.

sCI energies, second-order perturbation corrections, and FN-DMC energy data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: scemama@irsamc.ups-tlse.fr (A.S.).
*E-mail: loos@irsamc.ups-tlse.fr (P.-F.L.).

ORCID

Pierre-François Loos: 0000-0003-0598-7425

Funding

This work was performed using HPC resources from CALMIP (Toulouse) under Allocation 2016-0510 and from GENCI-TGCC (Grant 2016-08s015).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Arne Lüchow for numerous stimulating discussions.

■ REFERENCES

(1) Crack, J. C.; Green, J.; Thomson, A. J.; Brun, N. E. L. Iron-Sulfur Clusters as Biological Sensors: The Chemistry of Reactions with Molecular Oxygen and Nitric Oxide. *Acc. Chem. Res.* **2014**, *47*, 3196−3205.

(2) Stiefel, E. I. Transition Metal Sulfur Chemistry: Biological and Industrial Significance and Key Trends. *Trans. Metal Sulfur Chem.* **1996**, *653*, 2−38.

(3) Xiao, S.; Li, X.; Sun, W.; Guan, B.; Wang, Y. General and facile synthesis of metal sulfide nanostructures: In situ microwave synthesis and application as binder-free cathode for Li-ion batteries. *Chem. Eng. J.* **2016**, *306*, 251−259.

(4) Zhang, N.; Hayase, T.; Kawamata, H.; Nakao, K.; Nakajima, A.; Kaya, K. Photoelectron spectroscopy of iron-sulfur cluster anions. *J. Chem. Phys.* **1996**, *104*, 3413−3419.

(5) Takano, S.; Yamamoto, S.; Saito, S. The microwave spectrum of the FeS radical. *J. Mol. Spectrosc.* **2004**, *224*, 137−144.

(6) Drowart, J.; Pattoret, A.; Smoes, S. Mass Spectrometric Studies of the Vaporization of Refractory Compounds. *Proc. Br. Ceram. Soc.* **1967**, *8*, 67−88.

(7) Wang, L.; Huang, D.-l.; Zhen, J.-f.; Zhang, Q.; Chen, Y. Experimental Determination of the Vibrational Constants of FeS(X5) by Dispersed Fluorescence Spectroscopy. *Chin. J. Chem. Phys.* **2011**, *24*, 1−3.

(8) Matthew, D. J.; Tieu, E.; Morse, M. D. Determination of the bond dissociation energies of FeX and NiX (X = C, S, Se). *J. Chem. Phys.* **2017**, *146*, 144310.

(9) Bridgeman, A. J.; Rothery, J. Periodic trends in the diatomic monoxides and monosulfides of the 3d transition metals. *J. Chem. Soc., Dalton Trans.* **2000**, 211−218.

(10) Li, Y.-N.; Wang, S.; Wang, T.; Gao, R.; Geng, C.-Y.; Li, Y.-W.; Wang, J.; Jiao, H. Energies and Spin States of FeS0/, FeS20/, Fe2S20/, Fe3S40/, and Fe4S40/Clusters. *ChemPhysChem* **2013**, *14*, 1182−1189.

(11) Liang, B.; Wang, X.; Andrews, L. Infrared Spectra and Density Functional Theory Calculations of Group 8 Transition Metal Sulfide Molecules. *J. Phys. Chem. A* **2009**, *113*, 5375−5384.

(12) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. Density Functionals for Inorganometallic and Organometallic Chemistry. *J. Phys. Chem. A* **2005**, *109*, 11127−11143.

(13) Wu, Z. J.; Wang, M. Y.; Su, Z. M. Electronic structures and chemical bonding in diatomic ScX to ZnX (X = S, Se, Te). *J. Comput. Chem.* **2007**, *28*, 703−714.

(14) Hübner, O.; Termath, V.; Berning, A.; Sauer, J. A CASSCF/ACPF study of spectroscopic properties of FeS and FeS and the photoelectron spectrum of FeS. *Chem. Phys. Lett.* **1998**, *294*, 37−44.

(15) Clima, S.; Hendrickx, M. F. Photoelectron spectra of FeS explained by a CASPT2 ab initio study. *Chem. Phys. Lett.* **2007**, *436*, 341−345.

(16) Bauschlicher, C. W.; Maître, P. Theoretical study of the first transition row oxides and sulfides. *Theor. Chem. Acc.* **1995**, *90*, 189.

(17) Chen, J.; Zen, A.; Brandenburg, J. G.; Alfè, D.; Michaelides, A. Evidence for stable square ice from quantum Monte Carlo. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 220102.

(18) Dubecký, M.; Mitáš, L.; Jurečka, P. Noncovalent Interactions by Quantum Monte Carlo. *Chem. Rev.* **2016**, *116*, 5188−5215.

(19) Zhou, X.; Wang, F. Barrier heights of hydrogen-transfer reactions with diffusion quantum monte carlo method. *J. Comput. Chem.* **2017**, *38*, 798−806.
(20) Guareschi, R.; Zulfikri, H.; Daday, C.; Floris, F. M.; Amovilli, C.; Mennucci, B.; Filippi, C. Introducing QMC/MMpol: Quantum Monte Carlo in Polarizable Force Fields for Excited States. *J. Chem. Theory Comput.* **2016**, *12*, 1674−1683. PMID: 26959751.
(21) Christiansen, P. A. Relativistic effective potentials in transition metal quantum Monte Carlo simulations. *J. Chem. Phys.* **1991**, *95*, 361−363.
(22) Mitáš, L. In *Computer Simulations Studies in Condensed Matter V*; Landau, D. P., Mon, K. K., Schüttler, H. B., Eds.; Springer: Berlin, 1993; p 94.
(23) Belohorec, P.; Rothstein, S. M.; Vrbik, J. Infinitesimal differential diffusion quantum Monte Carlo study of CuH spectroscopic constants. *J. Chem. Phys.* **1993**, *98*, 6401−6405.
(24) Mitáš, L. Quantum Monte Carlo calculation of the Fe atom. *Phys. Rev. A: At., Mol., Opt. Phys.* **1994**, *49*, 4411−4414.
(25) Sokolova, S.; Lüchow, A. An ab initio study of TiC with the diffusion quantum Monte Carlo method. *Chem. Phys. Lett.* **2000**, *320*, 421−424.
(26) Wagner, L.; Mitáš, L. A quantum Monte Carlo study of electron correlation in transition metal oxygen molecules. *Chem. Phys. Lett.* **2003**, *370*, 412−417.
(27) Diedrich, C.; Lüchow, A.; Grimme, S. Performance of diffusion Monte Carlo for the first dissociation energies of transition metal carbonyls. *J. Chem. Phys.* **2005**, *122*, 021101.
(28) Caffarel, M.; Daudey, J.-P.; Heully, J.-L.; Ramírez-Solís, A. Towards accurate all-electron quantum Monte Carlo calculations of transition-metal systems: Spectroscopy of the copper atom. *J. Chem. Phys.* **2005**, *123*, 094102.
(29) Buendía, E.; Gálvez, F.; Sarsa, A. Correlated wave functions for the ground state of the atoms Li through Kr. *Chem. Phys. Lett.* **2006**, *428*, 241−244.
(30) Wagner, L. K.; Mitáš, L. Energetics and dipole moment of transition metal monoxides by quantum Monte Carlo. *J. Chem. Phys.* **2007**, *126*, 034105.
(31) Bande, A.; Lüchow, A. Vanadium oxide compounds with quantum Monte Carlo. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3371.
(32) Casula, M.; Marchi, M.; Azadi, S.; Sorella, S. A consistent description of the iron dimer spectrum with a correlated single-determinant wave function. *Chem. Phys. Lett.* **2009**, *477*, 255−258.
(33) Bouabça, T.; Braïda, B.; Caffarel, M. Multi-Jastrow trial wavefunctions for electronic structure calculations with quantum Monte Carlo. *J. Chem. Phys.* **2010**, *133*, 044111.
(34) Petz, R.; Lüchow, A. Energetics of Diatomic Transition Metal Sulfides ScS to FeS with Diffusion Quantum Monte Carlo. *ChemPhysChem* **2011**, *12*, 2031−2034.
(35) Buendía, E.; Gálvez, F.; Maldonado, P.; Sarsa, A. Quantum Monte Carlo ionization potential and electron affinity for transition metal atoms. *Chem. Phys. Lett.* **2013**, *559*, 12−17.
(36) Caffarel, M.; Giner, E.; Scemama, A.; Ramírez-Solís, A. Spin Density Distribution in Open-Shell Transition Metal Systems: A Comparative Post-Hartree-Fock, Density Functional Theory, and Quantum Monte Carlo Study of the CuCl2Molecule. *J. Chem. Theory Comput.* **2014**, *10*, 5286−5296.
(37) Scemama, A.; Applencourt, T.; Giner, E.; Caffarel, M. Accurate nonrelativistic ground-state energies of 3d transition metal atoms. *J. Chem. Phys.* **2014**, *141*, 244110.
(38) Trail, J. R.; Needs, R. J. Correlated electron pseudopotentials for 3d-transition metals. *J. Chem. Phys.* **2015**, *142*, 064110.
(39) Doblhoff-Dier, K.; Meyer, J.; Hoggan, P. E.; Kroes, G.-J.; Wagner, L. K. Diffusion Monte Carlo for Accurate Dissociation Energies of 3d Transition Metal Containing Molecules. *J. Chem. Theory Comput.* **2016**, *12*, 2583−2597.
(40) Krogel, J. T.; Santana, J. A.; Reboredo, F. A. Pseudopotentials for quantum Monte Carlo studies of transition metal oxides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 075143.
(41) Haghighi Mood, K.; Lüchow, A. Full Wave Function Optimization with Quantum Monte Carlo and Its Effect on the Dissociation Energy of FeS. *J. Phys. Chem. A* **2017**, *121*, 6165−6171.
(42) Giner, E.; David, G.; Scemama, A.; Malrieu, J. P. A simple approach to the state-specific MR-CC using the intermediate Hamiltonian formalism. *J. Chem. Phys.* **2016**, *144*, 064101.
(43) Giner, E.; Angeli, C.; Garniron, Y.; Scemama, A.; Malrieu, J.-P. A Jeziorski-Monkhorst fully uncontracted multi-reference perturbative treatment. I. Principles, second-order versions, and tests on ground state potential energy curves. *J. Chem. Phys.* **2017**, *146*, 224108.
(44) Garniron, Y.; Giner, E.; Malrieu, J.-P.; Scemama, A. Alternative definition of excitation amplitudes in multi-reference state-specific coupled cluster. *J. Chem. Phys.* **2017**, *146*, 154107.
(45) Giner, E.; Scemama, A.; Caffarel, M. Using perturbatively selected configuration interaction in quantum Monte Carlo calculations. *Can. J. Chem.* **2013**, *91*, 879−885.
(46) Giner, E.; Scemama, A.; Caffarel, M. Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F2 using selected configuration interaction trial wavefunctions. *J. Chem. Phys.* **2015**, *142*, 044115.
(47) Scemama, A.; Applencourt, T.; Giner, E.; Caffarel, M. Quantum Monte Carlo with very large multideterminant wavefunctions. *J. Comput. Chem.* **2016**, *37*, 1866−1875.
(48) Caffarel, M.; Applencourt, T.; Giner, E.; Scemama, A. Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule. *J. Chem. Phys.* **2016**, *144*, 151103.
(49) Caffarel, M.; Applencourt, T.; Giner, E.; Scemama, A. *ACS Symp. Ser.* **2016**, *1234*, 15−46.
(50) Umrigar, C. J.; Filippi, C. Energy and Variance Optimization of Many-Body Wave Functions. *Phys. Rev. Lett.* **2005**, DOI: 10.1103/PhysRevLett.94.150201.
(51) Toulouse, J.; Umrigar, C. J. Optimization of quantum Monte Carlo wave functions by energy minimization. *J. Chem. Phys.* **2007**, *126*, 084102.
(52) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions. *Phys. Rev. Lett.* **2007**, DOI: 10.1103/PhysRevLett.98.110201.
(53) Toulouse, J.; Umrigar, C. J. Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. *J. Chem. Phys.* **2008**, *128*, 174101.
(54) Bender, C. F.; Davidson, E. R. Studies in Configuration Interaction: The First-Row Diatomic Hydrides. *Phys. Rev.* **1969**, *183*, 23−30.
(55) Whitten, J. L.; Hackmeyer, M. Configuration Interaction Studies of Ground and Excited States of Polyatomic Molecules. I. The CI Formulation and Studies of Formaldehyde. *J. Chem. Phys.* **1969**, *51*, 5584−5596.
(56) Huron, B.; Malrieu, J. P.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth order wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745−5759.
(57) Evangelisti, S.; Daudey, J.-P.; Malrieu, J.-P. Convergence of an improved CIPSI algorithm. *Chem. Phys.* **1983**, *75*, 91−102.
(58) Cimiraglia, R. Second order perturbation correction to CI energies by use of diagrammatic techniques: An improvement to the CIPSI algorithm. *J. Chem. Phys.* **1985**, *83*, 1746−1749.
(59) Cimiraglia, R.; Persico, M. Recent advances in multireference second order perturbation CI: The CIPSI method revisited. *J. Comput. Chem.* **1987**, *8*, 39−47.
(60) Illas, F.; Rubio, J.; Ricart, J. M. Approximate natural orbitals and the convergence of a second order multireference many body perturbation theory (CIPSI) algorithm. *J. Chem. Phys.* **1988**, *89*, 6376−6384.
(61) Povill, A.; Rubio, J.; Illas, F. Treating large intermediate spaces in the CIPSI method through a direct selected CI algorithm. *Theor. Chem. Acc.* **1992**, *82*, 229−238.

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

(62) Abrams, M. L.; Sherrill, C. D. Important configurations in configuration interaction and coupled-cluster wave functions. *Chem. Phys. Lett.* **2005**, *412*, 121−124.
(63) Bunge, C. F.; Carbó-Dorca, R. Select-divide-and-conquer method for large-scale configuration interaction. *J. Chem. Phys.* **2006**, *125*, 014108.
(64) Bytautas, L.; Ruedenberg, K. A priori identification of configurational deadwood. *Chem. Phys.* **2009**, *356*, 64−75.
(65) Booth, G. H.; Thom, A. J. W.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **2009**, *131*, 054106.
(66) Knowles, P. J. Compressive sampling in configuration interaction wavefunctions. *Mol. Phys.* **2015**, *113*, 1655−1660.
(67) Garniron, Y.; Scemama, A.; Loos, P.-F.; Caffarel, M. Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory. *J. Chem. Phys.* **2017**, *147*, 034101.
(68) Evangelista, F. A. Adaptive multiconfigurational wave functions. *J. Chem. Phys.* **2014**, *140*, 124114.
(69) Liu, W.; Hoffmann, M. R. iCI: Iterative CI toward full CI. *J. Chem. Theory Comput.* **2016**, *12*, 1169−1178.
(70) Schriber, J. B.; Evangelista, F. A. Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy. *J. Chem. Phys.* **2016**, *144*, 161106.
(71) Tubman, N. M.; Lee, J.; Takeshita, T. Y.; Head-Gordon, M.; Whaley, K. B. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **2016**, *145*, 044112.
(72) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674−3680.
(73) Per, M. C.; Cleland, D. M. Energy-based truncation of multi-determinant wavefunctions in quantum Monte Carlo. *J. Chem. Phys.* **2017**, *146*, 164101.
(74) Ohtsuka, Y.; Hasegawa, J.-y Selected configuration interaction method using sampled first-order corrections to wave functions. *J. Chem. Phys.* **2017**, *147*, 034102.
(75) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595−1604.
(76) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.
(77) Zimmerman, P. M. Incremental full configuration interaction. *J. Chem. Phys.* **2017**, *146*, 104102.
(78) Scemama, A.; Applencourt, T.; Garniron, Y.; Giner, E.; David, G.; Caffarel, M. *Quantum Package* v1.0. https://github.com/LCPQ/quantum_package (2016).
(79) Scemama, A.; Giner, E.; Applencourt, T.; Caffarel, M. *QMC = Chem*. https://github.com/scemama/qmcchem (2017).
(80) Scemama, A.; Caffarel, M.; Oseret, E.; Jalby, W. Quantum Monte Carlo for large chemical systems: Implementing efficient strategies for petascale platforms and beyond. *J. Comput. Chem.* **2013**, *34*, 938−951.
(81) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-consistent pseudopotentials for quantum Monte Carlo calculations. *J. Chem. Phys.* **2007**, *126*, 234105.
(82) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-consistent small-core pseudopotentials for 3d-transition metals adapted to quantum Monte Carlo calculations. *J. Chem. Phys.* **2008**, *129*, 164115.
(83) Hammond, B. L.; Reynolds, P. J.; Lester, W. A. Valence quantum Monte Carlo with ab initio effective core potentials. *J. Chem. Phys.* **1987**, *87*, 1130−1136.
(84) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(85) Assaraf, R.; Caffarel, M.; Khelif, A. Diffusion Monte Carlo methods with a fixed number of walkers. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *61*, 4566−4575.
(86) Lee, R. M.; Conduit, G. J.; Nemec, N.; López Ríos, P.; Drummond, N. D. Strategies for improving the efficiency of quantum Monte Carlo calculations. *Phys. Rev. E* **2011**, *83*, 066706.

DOI: 10.1021/acs.jctc.7b01250
J. Chem. Theory Comput. 2018, 14, 1395−1402

Chapter 11

Summary and outlook

Significant improvements were brought to the QUANTUM PACKAGE . Some were single-core optimizations, and others were for adapting the algorithms for a better load balancing in the parallel regime. Today, the code has a parallel efficiency that enables routinely to realize runs on $\sim$ 2000 CPU cores, with hundreds of millions of determinants in the variational space, and such a gain in efficiency will lead to many more challenging chemical applications.

The Davidson diagonalization, which is at the center of variational methods, suffers from the impossibility to fully store the Hamiltonian in the memory of a single node. The solution we adopted was to resort to *direct methods*, recomputing the matrix elements on the fly at each iteration. While an extremely fast method was already available to detect zero matrix elements,[45] the former implementation still had to iterate over the $\sim {N_{\text{det}}}^2$ matrix elements to search the interacting determinant pairs. Now, determinants are split in disjoint sets, which are often identifiable as entirely disconnected from one another. Thus only a small fraction of the matrix elements need to be explored, and a linear-scaling algorithm was proposed. Although this algorithm was not kept as the actual implementation in the program because of it important memory footprint, the candidate we kept has a scaling in $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$, which is already a significant improvement. While the parallelization of this method was somewhat challenging, due to the elementary tasks being extremely unbalanced, a distributed implementation was realized with satisfying parallel speedups, typically $35\times$ for 50 nodes (1 800 cores) with respect to the 36-core single-node reference. Our implementation could be further improved by using together the linear-scaling and the present algorithms, keeping track of an estimate of the allocated memory and switching smoothly between the two variants.

The CIPSI selection algorithm, for which the previous implementation examined the external determinants one by one, was enormously improved, allowing for applications that were not feasible so far.[57, 89] Several different optimizations were used,

reducing the cost of finding connections between external and internal determinants (batch approach, filtering), as well as the cost of computing the corresponding matrix element in the Hamiltonian (phase mask, systematic determination of excitations). Again, a distributed implementation was realized after solving the problems related to load imbalance. For the selection algorithm, the speedup is almost ideal with 50 nodes. Despite the implementational improvements we have shown, there is still space for improvement from an algorithmic point of view. For instance, the selection step could be dramatically accelerated without reducing its quality by using a combination of CIPSI and of the Heat-Bath CI algorithm[90], by simply splitting the orbital space. The CIPSI selection, more expensive but more precise, would be used only where it is necessary, namely for excitations involving orbitals close to the Fermi level where the role of the denominator is crucial, and the Heat-Bath selection algorithm would be used the for the rest of the space.

A large improvement was also realized in the computation of the second order perturbative correction to the energy, E_{PT2}. The computation of E_{PT2} and the determinant selection were originally done with the same algorithm, but E_{PT2} did not allow for many approximations and thus was much more expensive. A natural idea to take into account a tremendous amount of tiny contributions was to imagine a stochastic approach. E_{PT2} being in itself an approximation for the Full-CI energy, an exact value with all the digits is indeed not required, as long as the value is unbiased and as the introduced statistical error is kept under control. The development of stochastic methods for quantum chemistry is one of the strengths of the LCPQ via Michel Caffarel's group, so we collaborated in the development of the hybrid stochastic/deterministic algorithm for the computation of E_{PT2}. Our scheme allows to compute E_{PT2} with an error bar smaller than the typical error of $E_{\mathrm{var}} + E_{\mathrm{PT2}}$ versus the Full-CI energy, for a few percent of the cost of the full deterministic computation. Now, the time to compute E_{PT2} is roughly similar to the time needed for the determinant selection. To push even further the efficiency of the CIPSI method, it would make sense to make the determinant selection along with the computation of E_{PT2}. In that case, the determinant selection would be free. Doing a stochastic selection has further advantages: at a given iteration, *any* external determinant can be potentially generated, so the algorithm will naturally converge to the unbiased Full-CI solution. This is not exactly the case with our current 3-class CIPSI-like implementation, where the determinants with small coefficients will never generate external determinants. This truncation is the source of a small but spurious size-consistence error.

To get the best of the data we were already able to compute, we implemented the shifted-B_k method, which uses the energy contributions computed by E_{PT2} to refine the wave function. It essentially creates an external space of determinants $|\alpha\rangle$ whose coefficients are perturbatively estimated, and allows them to act on the coefficients of the wave function in the internal space. This method requires the computation of

a dressing matrix, which we could estimate stochastically in a way similar to what we proposed for E_{PT2} (stochastic matrix dressing). The challenge in this case was to estimate, instead of a scalar, a vector of size N_{det}, which can be up to a few million elements. The storage and network issues could be solved by setting up a system of predefined checkpoints, a modest drawback being the impossibility to get an estimated dressing matrix outside of those checkpoints.

Finally, this stochastic matrix dressing computation was extended to another external space, that of the Multi-Reference Coupled-Cluster we had previously implemented deterministically,[91] which allowed to explore some possibilities of the bitstring-based determinant-driven approach. This was done by first setting up a framework to limit the implementational effort needed to build an external space ; essentially, one only needs to define a function mapping an external determinant to its desired coefficient. While this was proven convenient, in its current state it lacks some flexibility ; for instance, in the Multi-Reference Coupled-Cluster external space, external determinants generated from reference determinants are known to be of zero coefficient, but there is no simple way to inform the framework that a generator should be ignored. This issue, however minor, can be solved by a simple callback function to "ask" the programmer if the forthcoming generator is of interest. Setting up a few such callback functions could improve the efficiency for more specific situations. There is also no simple way to retrieve some custom information from a remote node ; the base algorithm only sends the partial $\boldsymbol{\delta}_I$ vectors, but the user may be interested in some other information. For instance the sum over generated ${c_\alpha}^2$ which hints the weight of the external space versus internal space. This functionality is currently being worked on.

During the multiple steps of evolution of the program, more and more applications were made possible.[1, 57, 89, 91, 92, 93, 94] This gave to the Quantum Package more visibility, and it was selected as a benchmark code for the choice of the new supercomputer of the CALMIP center. Moreover, different groups started to use it for applications and use it to develop new ideas. For example, Claudia Filippi's group in the Netherlands is now using CIPSI wave functions as a starting point for quantum Monte Carlo calculations,[95] and the Argonne group is currently implementing complex orbitals to adapt the Quantum Package to solids.[96]

Appendix A

A Jeziorski-Monkhorst fully uncontracted multi-reference perturbative treatment. I. Principles, second-order versions, and tests on ground state potential energy curves [1]

THE JOURNAL OF CHEMICAL PHYSICS **146**, 224108 (2017)

A Jeziorski-Monkhorst fully uncontracted multi-reference perturbative treatment. I. Principles, second-order versions, and tests on ground state potential energy curves

Emmanuel Giner,[1,a)] Celestino Angeli,[2] Yann Garniron,[3] Anthony Scemama,[3] and Jean-Paul Malrieu[3]

[1]*Max Planck Institue for Solid State Research, Heisenbergstraße 1, Stuttgart 70569, Germany*
[2]*Dipartimento di Scienze Chimiche e Famaceutiche, Universita di Ferrara, Via Fossato di Mortara 17, I-44121 Ferrara, Italy*
[3]*Laboratoire de Chimie et Physique Quantiques, UMR 5626 of CNRS, IRSAMC, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse Cedex, France*

(Received 13 February 2017; accepted 17 May 2017; published online 13 June 2017)

The present paper introduces a new multi-reference perturbation approach developed at second order, based on a Jeziorski-Mokhorst expansion using individual Slater determinants as perturbers. Thanks to this choice of perturbers, an effective Hamiltonian may be built, allowing for the dressing of the Hamiltonian matrix within the reference space, assumed here to be a CAS-CI. Such a formulation accounts then for the coupling between the static and dynamic correlation effects. With our new definition of zeroth-order energies, these two approaches are strictly size-extensive provided that local orbitals are used, as numerically illustrated here and formally demonstrated in the Appendix. Also, the present formalism allows for the factorization of all double excitation operators, just as in internally contracted approaches, strongly reducing the computational cost of these two approaches with respect to other determinant-based perturbation theories. The accuracy of these methods has been investigated on ground-state potential curves up to full dissociation limits for a set of six molecules involving single, double, and triple bond breaking together with an excited state calculation. The spectroscopic constants obtained with the present methods are found to be in very good agreement with the full configuration interaction results. As the present formalism does not use any parameter or numerically unstable operation, the curves obtained with the two methods are smooth all along the dissociation path. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4984616]

I. INTRODUCTION

The research of the ground-state wave function of closed-shell molecules follows well-established paths. The perturbative expansions from the mean-field Hartree-Fock single determinant usually converge and may be used as basic tools, especially when adopting a mono-electronic zero-order Hamiltonian known as the Møller-Plesset Hamiltonian.[1] In this approach, the wave function and the energy may be understood in terms of diagrams, which lead to the fundamental linked-cluster theorem.[2] The understanding of the size-consistency problem led to the suggestion of the coupled cluster approximation,[3–7] which is now considered as the standard and most efficient tool in the study of such systems in their ground state, especially in its CCSD(T) version where linked corrections by triple excitations are added perturbatively.[8] The situation is less evident when considering excited states, chemical reactions, and molecular dissociations, since it then becomes impossible to find a relevant single determinant zero-order wave function. These situations exhibit an intrinsic Multi-Reference (MR) character. A generalized linked-cluster theorem has been established by Brandow,[9] which gives a basis to the understanding of the size-consistency problem in this context, but the conditions for establishing this theorem are severe. They require a Complete Active Space (CAS) model space and a mono-electronic zero-order Hamiltonian. Consequently, the corresponding Quasi-Degenerate Perturbation Theory (QDPT) expansion cannot converge in most of the molecular MR situations.[10–12] The research of theoretically satisfying (size-consistent) and numerically efficient MR treatments remains a very active field in quantum chemistry, as summarized in recent review articles concerning either perturbative[13] or coupled-cluster[14] methods.

The present work concentrates on the search of a new MR perturbative approach at second order (MRPT2). Of course, pragmatic proposals have been rapidly formulated, consisting first in the identification of a reference model space, defined on the set of single determinants having large components in the desired eigenstates of the problem. Diagonalizing the Hamiltonian in this reference space delivers a zero-order wave function. Then one must define the vectors of the outer space to be used in the development and, in a perturbative context, choose a zero-order Hamiltonian. The simplest approach consists in using single determinants as outer-space eigenvectors, and this has been used in the CIPSI method[15,16] which is iterative, increasing the model space from the selection of the perturbing determinant of largest coefficients and their addition to the model space. From a practical point of view, this method is very efficient and is now employed to reach near

a) Author to whom correspondence should be addressed: E.Giner@fkf.mpg.de

exact Full Configuration Interaction (FCI) energies on small molecules[17,18] and also as trial wave function in the context of quantum Monte Carlo.[19–22] But the method suffers two main defects: (i) it is not size-consistent and (ii) it does not revise the model-space component of the wave function under the effect of its interaction with the outer-space. This last defect is avoided if one expresses the effect of the perturbation as a change of the matrix elements of the model space CI matrix, according to the Intermediate Effective Hamiltonian (IEH) theory,[23] as done in the state-specific[24] or multi-state[25] versions. Other methods which start from a CAS model space and use multi-determinantal outer-space vectors have been proposed later on and are now broadly used. The first one is the CASPT2 method,[26,27] which employs a mono-electronic zero-order Hamiltonian. The method suffers from intruder state problems, to be cured in a pragmatic manner through the introduction of some parameters, and is not strictly size-consistent. The NEVPT2 method[28–30] also uses multi-determinantal perturbers [defined in two different ways in its partially (pc-NEVPT2) and strongly contracted (sc-NRVPT2) versions], it makes use of a more sophisticated bi-electronic Hamiltonian (the Dyall Hamiltonian[31]) to define the zero-order energies of these perturbers, it is parameter-free, intruder-state free, and size-consistent. Both methods are implemented in several popular codes and use a contracted description of the model space component of the desired eigenfunction (fixed by the diagonalization of the Hamiltonian in the model space). Still in the spirit of the NEVPT2 approaches, a recent work of Sokolov and Chan[32] has allowed one to remove any contractions in the perturber space, thanks to the use of matrix product states and time-dependent perturbation theory (see Sec. III A for a comparison with the present work). Multi-state versions exist to give some flexibility to the model space component, in particular around weakly avoided crossings, but this flexibility is very limited.[33,34] If one returns to methods using single-determinant perturbers, the origin of their size-inconsistency problem has been identified as due to the unbalance between the multi-determinant character of the zero-order wave function and the single determinant character of the perturbers.[35] It is in principle possible to find size consistent formulations but they require rather complex formulations[36–39] and face some risk of numerical instabilities since they involve divisions by possibly small coefficients, the amplitudes of which may be small. Finally, one should mention the approaches based on the linearized internally contracted multi-reference coupled cluster (MRCC) theory using matrix product states[40,41] and stochastic techniques.[42] Such approaches use a much richer zeroth-order Hamiltonian (the Fink Hamiltonian) which provides very accurate results, to the price of a higher computational cost than the methods based on the Dyall Hamiltonian.

The present paper is composed as follows. In Sec. II, the here-proposed formalism is presented, whose main features are as follows:

1. it considers a CAS model space (to achieve the strict separability requirement), usually obtained from a preliminary CASSCF calculation;
2. the perturbers are single determinants (the method is externally non-contracted, according to the usual terminology);
3. it is state-specific and strictly separable when localized active MOs are used (see formal demonstration in the Appendix);
4. it makes use of the Dyall Hamiltonian to define the excitation energies appearing in the energy denominators;
5. it is based on a Jeziorski-Monkhorst[43] (JM) expression of the wave operator and proceeds through reference-specific partitionings of the zero-order Hamiltonian, as it has been previously suggested in the so-called Multi-Partitionning[44–46] (MUPA) and also in the UGA-SSMPRT2.[39] Consequently, it does not define a unique zero-order energy to the outer-space determinants (see a brief discussion in the Appendix);
6. it can be expressed either as a second-order energy correction or as a dressing of the CAS-CI matrix, which offers a full flexibility in the treatment of the feed-back effect of the post-CAS-CI correlation on the model space component of the wave function;
7. the contributions of the various classes of excitations are easily identified (as in the CASPT2 and NEVPT2 methods);
8. thanks to our definition of the zeroth-order energies, all processes involving double excitations can be treated by using only the one- and two-boy density matrices, avoiding to loop on the perturbers;
9. given a set of molecular orbitals, it is parameter free and does not contain any threshold to avoid numerical instabilities.

After having presented the working equations of the present formalism in Sec. II, Sec. III proposes a comparison with other existing MR approaches, such as some special cases of multi-reference coupled cluster (MRCC) and MRPT2. Section IV discusses the computational aspects of the two methods proposed here. Then, Sec. V presents the numerical results for the ground state potential energy curves of six molecules involving single, double, and triple bond breaking together with an excited state calculation with both the JM-MRPT2 and JM-HeffPT2 methods. A numerical test of size-extensivity is provided, together with the investigation of the dependency of the results on the locality of the active orbitals. Finally, Sec. VI summarizes the main results and presents its tentative developments. The reader can find in Sec. VI B a mathematical proof of strong separability of the JM-MRPT2 method.

II. WORKING EQUATIONS FOR THE PERTURBATION AND EFFECTIVE HAMILTONIAN AT SECOND ORDER

As demonstrated previously by one of the present authors and his collaborators,[35] the size-consistency problem in any multi-reference perturbative expansion using single Slater determinants as perturbers comes from the unbalanced zeroth order energies that occur in the denominators. More precisely, if the zeroth order wave function is a CAS-CI eigenvector, the zeroth order energy is stabilized by all the interactions within the active space. A perturber treated as a single Slater determinant does not take into account the correlation effects included in the zeroth order wave function, and consequently its zeroth order energies are unbalanced with respect to the one of the CAS-CI eigenvector. Nevertheless, if instead of a

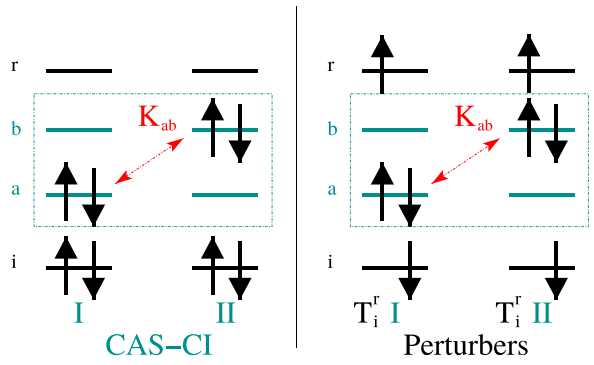


FIG. 1. Example of interactions: the two determinants of the CAS interact through a bi-electronic operator involving the two active orbitals a and b, just as the two perturber determinants generated by the same excitation operator T_i^r on the two CAS determinants.

unique perturber determinant, one considers the wave function created by the application of a given excitation operator on the whole CAS-CI wave function, most of the interactions found within the active space will also occur within this excited wave function (see Fig. 1 for a pictorial example). Therefore, the use of linear combinations of Slater determinants as perturbers together with a bi-electronic zeroth order operator, as it is the case in the NEVPT2 framework which uses the Dyall zeroth order operator, leads to balanced energy differences and removes the size-consistency problem.

On the basis of such considerations, the present work proposes an approach that uses single Slater determinants as perturbers and takes benefits of a new definition of energy denominators as expectation values of the Dyall zeroth order Hamiltonian over a specific class of linear combinations of Slater determinants. We first expose the definition of this perturbation theory, namely, the JM-MRPT2 method, which is strictly separable provided that local orbitals are used.

A large benefit from this new definition is that one may go beyond the sole calculation of the energy and improve the reference wave function by taking into account, in a strictly size-consistent way, the correlation effects brought by the perturbers on the reference space. In a second step, we reformulate the approach as a dressing of the Hamiltonian matrix within the set of Slater determinants belonging to the reference wave function, which is diagonalized. This approach will be referred to as the JM-HeffPT2 method.

A. The JM-MRPT2 method

1. First-order perturbed wave function and second-order energy

The formalism presented here is state specific and is not therefore restricted to ground state calculations. Nevertheless, for the sake of clarity and compactness, we omit the index referring explicitly to a specific eigenstate.

The zeroth order wave function $|\psi^{(0)}\rangle$ is assumed to be a CAS-CI eigenvector expanded on the set of reference determinants $|\mathrm{I}\rangle$,

$$|\psi^{(0)}\rangle = \sum_{\mathrm{I}\in\mathrm{CAS-CI}} c_{\mathrm{I}}\,|\mathrm{I}\rangle\,. \tag{1}$$

Such a wave function has a variational energy $e^{(0)}$,

$$e^{(0)} = \frac{\langle\psi^{(0)}|H|\psi^{(0)}\rangle}{\langle\psi^{(0)}|\psi^{(0)}\rangle}. \tag{2}$$

Starting from a normalized $|\psi^{(0)}\rangle$ (i.e., $\langle\psi^{(0)}|\psi^{(0)}\rangle = 1$), we assume that the exact wave function can be expressed as

$$|\Psi\rangle = |\psi^{(0)}\rangle + \sum_{\mu\notin\mathrm{CAS-CI}} c_\mu|\mu\rangle\,, \tag{3}$$

where $|\mu\rangle$ are all possible Slater determinants not belonging to the CAS-CI space. One should notice that such a form is in principle not exact, as some changes of the coefficients within the CAS-CI space can formally occur when passing from the CAS-CI eigenvector to the FCI one, but such an approximated form for the exact wave function is the basis of many MRPT2 approaches like NEVPT2, CASPT2, or CIPSI.

As in any projection technique, the exact energy can be obtained by projecting the Schrödinger equation on $|\psi^{(0)}\rangle$,

$$E = \langle\psi^{(0)}|H|\Psi\rangle = e^{(0)} + \sum_{\mu\notin\mathrm{CAS-\,CI}} c_\mu\,\langle\psi^{(0)}|H|\mu\rangle\,, \tag{4}$$

and one only needs to compute the coefficients of $|\mu\rangle$ that interact with $|\psi^{(0)}\rangle$, which consist in all individual Slater determinants being singly or doubly excited with respect to any Slater determinant in the CAS-CI space. From now on, we implicitly refer to $|\mu\rangle$ as any single Slater determinant belonging to such a space.

The coefficients c_μ are then written according to the JM ansatz,[43] whose general expression for wave function is not explicitly needed here, and will be therefore given in Sec. III B when the comparison of the present method with other multi-reference methodologies will be investigated. The JM ansatz introduces the genealogy of the coefficients c_μ with respect to the Slater determinants within the CAS-CI space

$$c_\mu = \sum_{\mathrm{I}} c_{\mathrm{I}}\; t_{\mathrm{I}\mu}, \tag{5}$$

where the quantity $t_{\mathrm{I}\mu}$ is the excitation amplitude related to the excitation process $T_{\mathrm{I}\mu}$ that leads from $|\mathrm{I}\rangle$ to $|\mu\rangle$,

$$T_{\mathrm{I}\mu}\,|\mathrm{I}\rangle = |\mu\rangle\,. \tag{6}$$

Here, we restrict $T_{\mathrm{I}\mu}$ to be a single or double excitation operator. Within this JM formulation of c_μ, a very general first order approximation of the amplitudes $t_{\mathrm{I}\mu}^{(1)}$ can be expressed as

$$t_{\mathrm{I}\mu}^{(1)} = \frac{\langle\mathrm{I}\,|H|\mu\rangle}{e^{(0)} - E_{\mathrm{I}\mu}^{(0)}} = \frac{\langle\mathrm{I}|H|\mu\rangle}{\Delta E_{\mathrm{I}\mu}^{(0)}}, \tag{7}$$

where the excitation energy $\Delta E_{\mathrm{I}\mu}^{(0)}$ depends explicitly on the couple $(|\mathrm{I}\rangle\,,\,|\mu\rangle)$. In that regard, a given Slater determinant $|\mu\rangle$ will have different zeroth-order energies according to the parent $|\mathrm{I}\rangle$ from which it is generated, implying that the zeroth-order Hamiltonian explicitly depends on the reference determinant $|\mathrm{I}\rangle$, just as was initially proposed in the MUPA approaches[44–46] (the interested reader could find in the Appendix a more detailed discussion of that aspect). Such a definition is different from other determinant-based MRPT2 like the CIPSI or shifted-B_k where the excitation energy does not depend on the couple $(|\mathrm{I}\rangle\,,\,|\mu\rangle)$ but only on $|\mu\rangle$. With this

definition of $t^{(1)}_{\mathrm{I}\mu}$, one can write the second-order correction to the energy $e^{(2)}$ as

$$e^{(2)} = \langle\psi^{(0)}|H|\psi^{(1)}\rangle = \sum_{\mu}\sum_{\mathrm{I}} c_{\mathrm{I}} \frac{\langle \mathrm{I}|H|\mu\rangle}{\Delta E^{(0)}_{\mathrm{I}\mu}} \langle\psi^{(0)}|H|\mu\rangle$$
$$= \sum_{\mu}\sum_{\mathrm{IJ}} c_{\mathrm{I}} \frac{\langle \mathrm{I}|H|\mu\rangle \langle\mu|H|\mathrm{J}\rangle}{\Delta E^{(0)}_{\mathrm{I}\mu}} c_{\mathrm{J}}, \quad (8)$$

and the total second-order energy $E^{(2)}$,

$$E^{(2)} = \langle\psi^{(0)}|H|\psi^{(0)}\rangle + \langle\psi^{(0)}|H|\psi^{(1)}\rangle = e^{(0)} + e^{(2)}. \quad (9)$$

2. Definition of the energy denominators

The first-order wave function can be written explicitly in terms of the excitation operators $T_{\mathrm{I}\mu}$,

$$|\psi^{(1)}\rangle = \sum_{\mu} c^{(1)}_{\mu}|\mu\rangle = \sum_{\mu}\sum_{\mathrm{I}} c_{\mathrm{I}} \frac{\langle \mathrm{I}|H\ T_{\mathrm{I}\mu}\,|\mathrm{I}\rangle}{\Delta E^{(0)}_{\mathrm{I}\mu}} T_{\mathrm{I}\mu}\,|\mathrm{I}\rangle\,. \quad (10)$$

However, one can notice that

1. the excitation operators $T_{\mathrm{I}\mu}$ do not explicitly depend on $|\mathrm{I}\rangle$ as they are general single or double excitation operators, just as in the Hamiltonian for instance;
2. a given excitation operator T contributes to the coefficients of several $|\mu\rangle$ ($T_{\mathrm{I}\mu} = T_{\mathrm{J}\nu} = T$);
3. the application of all the single and double excitation operators T on each $|\mathrm{I}\rangle$ generates the entire set of $|\mu\rangle$ as the reference is a CAS.

Therefore one can rewrite the first-order perturbed wave function directly in terms of excitation operators T applied on the each CAS-CI Slater determinant as

$$|\psi^{(1)}\rangle = \sum_{T} |\psi^{(1)}_{T}\rangle\,, \quad (11)$$

where $|\psi^{(1)}_{T}\rangle$ is the part of the first-order wave function associated with the excitation process T,

$$|\psi^{(1)}_{T}\rangle = \sum_{\mathrm{I}} c_{\mathrm{I}} \frac{\langle \mathrm{I}|H\ T|\mathrm{I}\rangle}{\Delta E^{(0)}_{\mathrm{I}\ T\mathrm{I}}} T|\mathrm{I}\rangle\,. \quad (12)$$

In order to fully define our perturbation theory and intermediate Hamiltonian theory, one needs to select an expression for the energy denominators occurring in the definition of $|\psi^{(1)}_{T}\rangle$. We propose to take a quantity that does not depend explicitly on the reference determinant $|\mathrm{I}\rangle$ but only depends on the excitation process T,

$$\Delta E^{(0)}_{\mathrm{I}\ T\mathrm{I}} = \Delta E^{(0)}_{T} \quad \forall\ \mathrm{I}. \quad (13)$$

Consequently, in the expression of $|\psi^{(1)}_{T}\rangle$ [see Eq. (12)], the energy denominator can be factorized

$$|\psi^{(1)}_{T}\rangle = \frac{1}{\Delta E^{(0)}_{T}} \sum_{\mathrm{I}} c_{\mathrm{I}}\ \langle \mathrm{I}|H\ T|\mathrm{I}\rangle\ T|\mathrm{I}\rangle = \frac{1}{\Delta E^{(0)}_{T}} |\tilde{\psi}^{(1)}_{T}\rangle\,, \quad (14)$$

where $|\tilde{\psi}^{(1)}_{T}\rangle$ is simply

$$|\tilde{\psi}^{(1)}_{T}\rangle = \sum_{\mathrm{I}} c_{\mathrm{I}}\ \langle \mathrm{I}|H\ T|\mathrm{I}\rangle\ T|\mathrm{I}\rangle\,. \quad (15)$$

Also, one can notice that as $|\tilde{\psi}^{(1)}_{T}\rangle$ and $|\psi^{(1)}_{T}\rangle$ differ by a simple constant factor, they have the same normalized expectation values

$$\frac{\langle\psi^{(1)}_{T}|H^{D}|\psi^{(1)}_{T}\rangle}{\langle\psi^{(1)}_{T}|\psi^{(1)}_{T}\rangle} = \frac{\langle\tilde{\psi}^{(1)}_{T}|H^{D}|\tilde{\psi}^{(1)}_{T}\rangle}{\langle\tilde{\psi}^{(1)}_{T}|\tilde{\psi}^{(1)}_{T}\rangle}. \quad (16)$$

Then, the excitation energy $\Delta E^{(0)}_{T}$ is simply taken as the difference of the normalized expectation values of the Dyall Hamiltonian H^{D} over $|\psi^{(0)}\rangle$ and $|\tilde{\psi}^{(1)}_{T}\rangle$,

$$\Delta E^{(0)}_{T} = \frac{\langle\psi^{(0)}|H^{D}|\psi^{(0)}\rangle}{\langle\psi^{(0)}|\psi^{(0)}\rangle} - \frac{\langle\tilde{\psi}^{(1)}_{T}|H^{D}|\tilde{\psi}^{(1)}_{T}\rangle}{\langle\tilde{\psi}^{(1)}_{T}|\tilde{\psi}^{(1)}_{T}\rangle}$$
$$= \frac{\langle\psi^{(0)}|H^{D}|\psi^{(0)}\rangle}{\langle\psi^{(0)}|\psi^{(0)}\rangle} - \frac{\langle\psi^{(1)}_{T}|H^{D}|\psi^{(1)}_{T}\rangle}{\langle\psi^{(1)}_{T}|\psi^{(1)}_{T}\rangle}. \quad (17)$$

This ensures the strong separability when localized orbitals are used, as will be illustrated numerically in Sec. V.

The Dyall Hamiltonian is nothing but the exact Hamiltonian over the active orbitals and a Møller-Plesset type operator over the doubly occupied and virtual orbitals. If one labels a, b, c, and d as the active spin-orbitals, i, j the spin-orbitals that are always occupied, and v, r the virtual spin-orbitals, the Dyall Hamiltonian can be written explicitly as

$$H^{D} = H^{D}_{iv} + H^{D}_{a}, \quad (18)$$

$$\begin{cases} H^{D}_{a} = \sum_{ab} h^{\mathrm{eff}}_{ab} a^{\dagger}_{a} a_{b} + \frac{1}{2}\sum_{abcd} (ad|bc)\ a^{\dagger}_{a} a^{\dagger}_{b} a_{c} a_{d} \\ H^{D}_{iv} = \sum_{i} \epsilon_{i}\ a^{\dagger}_{i} a_{i} + \sum_{v} \epsilon_{v}\ a^{\dagger}_{v} a_{v} + C \end{cases}, \quad (19)$$

where ϵ_i and ϵ_v are defined as the spin-orbital energies associated with the density given by $|\psi^{(0)}\rangle$, and the effective active one-electron operator $h^{\mathrm{eff}}_{ab} = \langle a|h + \sum_i (J_i - K_i)\,|b\rangle$. With a proper choice of the constant C in Eq. (19),

$$C = \sum_{i} \langle i|h|i\rangle + \frac{1}{2}\sum_{i,j} ((ii|jj) - (ij|ij))\,, \quad (20)$$

one has

$$\frac{\langle\psi^{(0)}|H^{D}|\psi^{(0)}\rangle}{\langle\psi^{(0)}|\psi^{(0)}\rangle} = \frac{\langle\psi^{(0)}|H|\psi^{(0)}\rangle}{\langle\psi^{(0)}|\psi^{(0)}\rangle} = e^{(0)}. \quad (21)$$

Because the Dyall Hamiltonian acts differently on the active and inactive-virtual orbitals, the excitation energy $\Delta E^{(0)}_{T}$ is the sum of an excitation energy $\Delta E^{(0)\,iv}_{T}$ associated with the inactive and virtual orbitals and of an excitation energy $\Delta E^{(0)\,a}_{T}$ associated with the active orbitals

$$\Delta E^{(0)}_{T} = \Delta E^{(0)\,a}_{T} + \Delta E^{(0)\,iv}_{T}. \quad (22)$$

Also, it is useful to differentiate the active part from the inactive-virtual part of the excitation T,

$$T = T_{a} T_{iv}. \quad (23)$$

The inactive-virtual excitation energy $\Delta E^{(0)\,iv}_{T}$ is simply

$$\Delta E^{(0)\,iv}_{T} = \sum_{i\,\in\,T} \epsilon_{i} - \sum_{v\,\in\,T} \epsilon_{v}, \quad (24)$$

where $i \in T$ and $v \in T$ refer to, respectively, the inactive and virtual spin-orbitals involved in the excitation operator T.

Conversely, the active excitation energy $\Delta E_T^{(0)\,a}$ has a more complex expression, namely,

$$\Delta E_T^{(0)\,a} = e^{(0)} - \frac{\sum_{\mathrm{I\,J}} \left(c_\mathrm{I} \langle \mathrm{I}|H\,T|\mathrm{I}\rangle\right) \langle \mathrm{I}|T_a^\dagger\, H^D\, T_a|\mathrm{J}\rangle \left(c_\mathrm{J} \langle \mathrm{J}|H\,T|\mathrm{J}\rangle\right)}{\sum_\mathrm{I} \left(c_\mathrm{I} \langle \mathrm{I}|H\,T|\mathrm{I}\rangle\right)^2 \langle \mathrm{I}|T^\dagger T|\mathrm{I}\rangle}. \tag{25}$$

3. Practical consequences: The difference between single and double excitation operators

From Eq. (25), one must differentiate the class of the pure single excitation operators from the pure double excitation operators. For the sake of clarity, we define the spin-adapted bielectronic integrals $((mn|pq))$ as

$$((mn|pq)) = \begin{cases} (mn|pq) & \text{if } \sigma(m,p) \neq \sigma(n,q) \\ (mn|pq) - (mp|nq) & \text{if } \sigma(m,p) = \sigma(n,q) \end{cases}, \tag{26}$$

where $\sigma(m,p)$ is the spin variable of the spin orbitals m and p. If one considers a given double excitation involving four different spin orbitals m, n, p, and q,

$$T_{mp}^{nq} = a_n^\dagger a_q^\dagger a_p a_m \qquad m \neq n \neq p \neq q, \tag{27}$$

one can notice that the Hamiltonian matrix elements associated with this double excitation only depend, up to a phase factor, on the four indices m, n, p, and q involved in T_{mp}^{nq}. Indeed, if T_{mp}^{nq} is possible on both $|\mathrm{I}\rangle$ and $|\mathrm{J}\rangle$, one has

$$\begin{aligned} \langle \mathrm{I}|H\ T_{mp}^{nq}|\mathrm{I}\rangle &= ((mn|pq))\ \langle \mathrm{I}|\left(T_{mp}^{nq}\right)^\dagger T_{mp}^{nq}|\mathrm{I}\rangle, \\ \langle \mathrm{J}|H\ T_{mp}^{nq}|\mathrm{J}\rangle &= ((mn|pq))\ \langle \mathrm{J}|\left(T_{mp}^{nq}\right)^\dagger T_{mp}^{nq}|\mathrm{J}\rangle, \end{aligned} \tag{28}$$

and as

$$\begin{aligned} \langle \mathrm{I}|\left(T_{mp}^{nq}\right)^\dagger T_{mp}^{nq}|\mathrm{I}\rangle &= \langle \mathrm{J}|\left(T_{mp}^{nq}\right)^\dagger T_{mp}^{nq}|\mathrm{J}\rangle \\ &= 1, \end{aligned} \tag{29}$$

it becomes

$$\langle \mathrm{J}|H\ T_{mp}^{nq}|\mathrm{J}\rangle = \langle \mathrm{I}|H\ T_{mp}^{nq}|\mathrm{I}\rangle. \tag{30}$$

Therefore, as the Hamiltonian matrix elements of type $\langle \mathrm{J}|H\ T_{mp}^{nq}|\mathrm{J}\rangle$ can be factorized both in the numerator and the dominator of the expression of the active part of the excitation energy [see Eq. (25)]. Finally, the expression of the active part of the excitation energy for a given double excitation T_{mp}^{nq} is simply

$$\begin{aligned} \Delta E_{T_{mp}^{nq}}^{(0)\,a} &= e^{(0)} - \frac{\sum_{\mathrm{I\,J}} c_\mathrm{I}\ \langle \mathrm{I}|T_a^\dagger\, H^D\, T_a|\mathrm{J}\rangle\ c_\mathrm{J}}{\sum_\mathrm{I} c_\mathrm{I}^2 \langle \mathrm{I}|T_a^\dagger\ T_a|\mathrm{I}\rangle} \\ &= e^{(0)} - \frac{\langle \psi^{(0)}|T_a^\dagger\, H^D\, T_a|\psi^{(0)}\rangle}{\langle \psi^{(0)}|T_a^\dagger T_a|\psi^{(0)}\rangle}. \end{aligned} \tag{31}$$

As a consequence, the amplitudes $t_{\mathrm{I}T_{mn}^{qp}\mathrm{I}}$ and $t_{\mathrm{J}T_{mn}^{qp}\mathrm{J}}$ associated with the same excitation T_{mn}^{qp} for different parents $|\mathrm{I}\rangle$ and $|\mathrm{J}\rangle$ are also equal,

$$\begin{aligned} t_{\mathrm{I}T_{mn}^{qp}\mathrm{I}} &= \frac{\langle \mathrm{I}|H\ T_{mn}^{qp}|\mathrm{I}\rangle}{\Delta E_{T_{mn}^{qp}}^{(0)}}, \\ t_{\mathrm{J}T_{mn}^{qp}\mathrm{J}} &= \frac{\langle \mathrm{J}|H\ T_{mn}^{qp}|\mathrm{J}\rangle}{\Delta E_{T_{mn}^{qp}}^{(0)}}, \end{aligned} \tag{32}$$

and one can define a unique excitation operator $\mathcal{T}_{mn}^{qp\,(1)}$ which does not depend on the reference determinant on which it acts. The explicit form of the reference-independent excitation operator $\mathcal{T}_{mn}^{qp\,(1)}$ is

$$\mathcal{T}_{mn}^{qp\,(1)} = \frac{((mq|np))}{\Delta E_{T_{mn}^{qp}}^{(0)}} a_q^\dagger a_p^\dagger a_n a_m. \tag{33}$$

In the case where T is a pure single excitation operator, the term $\langle \mathrm{I}|H\,T|\mathrm{I}\rangle$ may strongly depend on $|\mathrm{I}\rangle$ and Eq. (25) cannot be simplified.

4. Precaution for spin symmetry

As the formalism proposed here deals with Slater determinants, it cannot formally ensure to provide spin eigenfunctions. In order to ensure the invariance of the energy with the S_z value of a given spin multiplicity, we introduced a slightly modified version of the Dyall Hamiltonian which does not consider the following:

1. any exchange terms in the Hamiltonian matrix elements when active orbitals are involved,
2. any exchange terms involving two electrons of opposite spins (namely, $a_{b\alpha}^\dagger a_{a\beta}^\dagger a_{b\beta} a_{a\alpha}$ and $a_{b\beta}^\dagger a_{a\alpha}^\dagger a_{b\alpha} a_{a\beta}$).

B. The JM-HeffPT2 method

An advantage of a determinant-based multi-reference perturbation theory is that it can be easily written as a dressing of the Hamiltonian matrix within the reference space. Starting from the Schrödinger equation projected on a given reference determinant $|\mathrm{I}\rangle$, one has

$$c_\mathrm{I} \langle \mathrm{I}|H|\mathrm{I}\rangle + \sum_{\mathrm{J}\neq I} c_\mathrm{J} \langle \mathrm{I}|H|\mathrm{J}\rangle \sum_\mu c_\mu^{(1)} \langle \mathrm{I}|H|\mu\rangle = E^{(2)} c_\mathrm{I}. \tag{34}$$

Using the expression for the first order coefficients $c_\mu^{(1)}$, it becomes

$$\begin{aligned} &c_\mathrm{I} \left(\langle \mathrm{I}|H|\mathrm{I}\rangle + \sum_\mu \frac{\langle \mathrm{I}|H|\mu\rangle^2}{\Delta E_{\mathrm{I}\mu}^{(0)}} \right) \\ &\quad + \sum_{\mathrm{J}\neq I} c_\mathrm{J} \left(\langle \mathrm{I}|H|\mathrm{J}\rangle + \frac{\langle \mathrm{I}|H|\mu\rangle \langle \mu|H|\mathrm{J}\rangle}{\Delta E_{\mathrm{J}\mu}^{(0)}} \right) = E^{(2)} c_\mathrm{I}. \end{aligned} \tag{35}$$

Therefore, one can define an non-Hermitian operator $\Delta H^{(2)}$,

$$\langle \mathrm{I}|\Delta H^{(2)}|\mathrm{J}\rangle = \sum_\mu \frac{\langle \mathrm{I}|H|\mu\rangle \langle \mu|H|\mathrm{J}\rangle}{\Delta E_{\mathrm{J}\mu}^{(0)}}, \tag{36}$$

and a dressed Hamiltonian $\mathcal{H}$ as

$$\langle \mathrm{I}|\mathcal{H}^{(2)}|\mathrm{J}\rangle = \langle \mathrm{I}|H|\mathrm{J}\rangle + \langle \mathrm{I}|\Delta H^{(2)}|\mathrm{J}\rangle, \tag{37}$$

such that Eq. (35) becomes a non-symmetric linear eigenvalue equation within the CAS-CI space

$$c_\mathrm{I} \langle \mathrm{I}|\mathcal{H}^{(2)}|\mathrm{I}\rangle + \sum_{\mathrm{J}\neq I} c_\mathrm{J} \langle \mathrm{I}|\mathcal{H}^{(2)}|\mathrm{J}\rangle = E^{(2)} c_\mathrm{I}. \tag{38}$$

The second-order correction to the energy $e^{(2)}$ can be simply obtained as the expectation value of $\Delta H^{(2)}$ over the zeroth-order wave function

$$\begin{aligned} e^{(2)} &= \langle \psi^{(0)}|\Delta H^{(2)}|\psi^{(0)}\rangle \\ &= \sum_\mu \sum_{\mathrm{I\,J}} c_\mathrm{I}\ \frac{\langle \mathrm{I}|H|\mu\rangle \langle \mu|H|\mathrm{J}\rangle}{\Delta E_{\mathrm{J}\mu}^{(0)}}\ c_\mathrm{J}. \end{aligned} \tag{39}$$

Finally, one can define a Hermitian operator $\tilde{H}^{(2)}$,

$$\langle \mathrm{I}|\tilde{H}^{(2)}|\mathrm{J}\rangle = \frac{1}{2}\left(\langle \mathrm{I}|\mathcal{H}^{(2)}|\mathrm{J}\rangle + \langle \mathrm{J}|\mathcal{H}^{(2)}|\mathrm{I}\rangle\right), \tag{40}$$

and a corresponding eigenpair $(|\tilde{\Psi}_2\rangle, \tilde{E}^{(2)})$ verifying

$$\tilde{H}^{(2)}|\tilde{\Psi}_2\rangle = \tilde{E}^{(2)}|\tilde{\Psi}_2\rangle. \tag{41}$$

The diagonalization of such a Hamiltonian allows then to improve the CAS-CI wave function by treating the coupling that can exist between the correlation effects within and outside the CAS-CI space.

III. LINKS WITH OTHER MULTI-REFERENCE METHODS

A. MRPT2 based on the Dyall Hamiltonian

It is interesting to understand the similarities and differences between the present JM-MPRT2 and other strictly size-consistent MRPT2 methods based on the Dyall zeroth-order Hamiltonian. The most flexible solution of such MRPT2 makes use of the exact solution for the Dyall Hamiltonian with $N+1$, $N+2$, $N-1$, and $N-2$ electrons in the active space (where N is the number of electrons in the active space) as perturbers. Such a formulation is totally uncontracted in the perturber space, which implies a high computational cost, but a solution as been recently proposed by Sokolov and Chan.[32] using a time-dependent formulation and matrix product state techniques. Then, one can use the partially contracted NEVPT2 (pc-NEVPT2) which is computationally less demanding and provides very similar results, as shown by Sokolov and Chan.[32] Finally, when comparing JM-MRPT2, the nearest version of NEVPT2 is certainly the strongly contracted one and the present paragraph focusses on their differences and similarities.

JM-MRPT2 uses perturbers that are individual Slater determinants, whereas all versions of NEVPT2 use linear combinations of Slater determinants. However, in SC-NEVPT2, the contraction coefficients are closely related to the Hamiltonian matrix elements, just as in the JM-MRPT2 method. In order to better understand the differences between SC-NEVPT2 and JM-MRPT2, let us take a practical example. Here, i, j are the inactive spin-orbitals, a, b are the active spin-orbitals, and r, s are the virtual spin-orbitals. Considering a given semi-active double excitation $T_{ij}^{av} = a_a^\dagger a_v^\dagger a_j a_i$, the first-order amplitude $t_{ij}^{av(1)}$ associated with T_{ij}^{av} in the JM-MRPT2 formalism is given by

$$t_{ij}^{av(1)} = \frac{((ia|jv))}{\epsilon_i + \epsilon_j - \epsilon_v + \Delta E^{(0)}_{a_a^\dagger}}, \tag{42}$$

where the active part of the excitation energy $\Delta E^{(0)}_{a_a^\dagger}$ directly comes from Eq. (31)

$$\Delta E^{(0)}_{a_a^\dagger} = e^{(0)} - \frac{\langle \psi^{(0)}|a_a H^D a_a^\dagger|\psi^{(0)}\rangle}{\langle \psi^{(0)}|a_a a_a^\dagger|\psi^{(0)}\rangle}. \tag{43}$$

Note that such a quantity can be thought as an approximation of the electron affinity of the molecule, as it is the change in energy when one introduces "brutally" an electron in spin orbital a without relaxing the wave function. Consequently, as it has been emphasized in Sec. II A [see Eq. (33)], one can consider the part of the first-order perturbed wave function generated by the excitation T_{ij}^{av},

$$|\psi^{(1)}_{T_{ij}^{av}}\rangle = \sum_{\mathrm{I}} c_{\mathrm{I}}\, t_{ij}^{av(1)}\, T_{ij}^{av}|\mathrm{I}\rangle, \tag{44}$$

which turns out to be

$$\begin{aligned}|\psi^{(1)}_{T_{ij}^{av}}\rangle &= \frac{((ia|jv))}{\epsilon_i + \epsilon_j - \epsilon_v + \Delta E^{(0)}_{a_a^\dagger}} \sum_{\mathrm{I}} c_{\mathrm{I}}\, T_{ij}^{av}\, |\mathrm{I}\rangle \\ &= \frac{((ia|jv))}{\epsilon_i + \epsilon_j - \epsilon_v + \Delta E^{(0)}_{a_a^\dagger}} T_{ij}^{av}|\psi^{(0)}\rangle. \end{aligned} \tag{45}$$

In the SC-NEVPT2 framework, one does not consider explicitly a given T_{ij}^{av} but has to consider a unique excitation $\mathcal{T}_{ij}^{v}$ which is a linear combination of all possible T_{ij}^{av} for all active spin orbitals a, with proper contraction coefficients. To be more precise, the first-order perturbed wave function associated with $\mathcal{T}_{ij}^{v}$ is

$$|\psi^{(1)}_{\mathcal{T}_{ij}^{v}}\rangle = \frac{1}{\Delta E^{(0)}_{\mathcal{T}_{ij}^{v}}} \sum_a ((ia|jv))\, T_{ij}^{av}\, |\psi^{(0)}\rangle, \tag{46}$$

where the excitation energy $\Delta E^{(0)}_{\mathcal{T}_{ij}^{v}}$ associated with $\mathcal{T}_{ij}^{v}$ is unique for all the excitation operators T_{ij}^{av} and can be thought as an average excitation energy over all a. Consequently, one can express the part of $|\psi^{(1)}_{\mathcal{T}_{ij}^{v}}\rangle$ that comes from the T_{ij}^{av} as

$$|\psi^{(1)}_{T_{ij}^{av}}\rangle^{(\text{SC- NEVPT2})} = \frac{((ia|jv))}{\Delta E^{(0)}_{\mathcal{T}_{ij}^{v}}} T_{ij}^{av}\, |\psi^{(0)}\rangle, \tag{47}$$

which we can compare to Eq. (45) in the case of the JM-MRPT2 method. Then, the only difference between SC-NEVPT2 and JM-MRPT2 is the definition of the excitation energy occurring in Eqs. (45) and (47). In the SC-NEVPT2 method, the excitation energy $\Delta E^{(0)}_{\mathcal{T}_{ij}^{v}}$ is closely related to the excitation energy defined in JM-MRPT2

$$\begin{aligned}\Delta E^{(0)}_{\mathcal{T}_{ij}^{v}} &= e^{(0)} - \frac{\langle \psi^{(1)}_{\mathcal{T}_{ij}^{v}}|\, H^D\, |\psi^{(1)}_{\mathcal{T}_{ij}^{v}}\rangle}{\langle \psi^{(1)}_{\mathcal{T}_{ij}^{v}}|\psi^{(1)}_{\mathcal{T}_{ij}^{v}}\rangle} \\ &= \epsilon_i + \epsilon_j - \epsilon_v + \Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger}, \end{aligned} \tag{48}$$

where the quantity $\Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger}$ is the same for all active orbitals and defined as

$$\begin{aligned}&\Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger} \\ &= e^{(0)} - \frac{\sum_a \sum_b ((ia|jv))((ib|jv))\, \langle \psi^{(0)}|a_b\, H^D\, a_a^\dagger|\psi^{(0)}\rangle}{\sum_a ((ia|jv))^2\, \langle \psi^{(0)}|a_a a_a^\dagger|\psi^{(0)}\rangle}. \end{aligned} \tag{49}$$

Under this perspective, one sees that the quantity $\Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger}$ is related to $\Delta E^{(0)}_{a_a^\dagger}$ defined in Eq. (43):

- in the JM-MRPT2 method, the quantity $\Delta E^{(0)}_{a_a^\dagger}$ explicitly refers to the "brutal" addition of an electron in orbital a, whatever the inactive orbitals i, j or virtual orbitals v involved in T_{ij}^{av};

- the quantity $\Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger}$ involved in SC-NEVPT2 is an average electronic affinity over all possible excitation processes $a_a^\dagger$ within the active space, but keeping a trace of the inactive and virtual excitation processes involved in T^{av}_{ij}, thanks to the interaction $(ia|jv)$.

Consequently, the quantity $\Delta E^{(0)\text{SC–NEVPT2}}_{a^\dagger}$ contains also the interactions between various $a_a^\dagger\,|\psi^{(0)}\rangle$. To summarize, on one hand, JM-MRPT2 gives a different but rather crude excitation energy for each T^{av}_{ij}, and on the other hand, SC-NEVPT2 has a unique and sophisticated excitation energy for all T^{av}_{ij}. Of course, one can extend this comparison to all the other classes of double excitations.

Finally, one should notice that the effective Hamiltonian formulation of JM-MRPT2 leads to the revision of the zeroth-order wave function, which is not allowed by the NEVPT2 framework, whatever its degree of contraction in the perturber space.

B. Multi-reference coupled cluster methods

The present formalism has also several links with other multi-reference methods. First of all, as it uses a JM genealogical definition for the coefficients $c_\mu^{(1)}$ [see Eqs. (5) and (7)], the wave function corrected at first order $|\Psi^{(1)}\rangle$ can be written as

$$\begin{aligned}|\Psi^{(1)}\rangle &= |\psi^{(0)}\rangle + |\psi^{(1)}\rangle \\ &= \sum_{\text{I}} c_{\text{I}}\,|\text{I}\rangle + \sum_{\mu}\sum_{\text{I}} c_{\text{I}}\; t^{(1)}_{\text{I}\mu}\; T_{\text{I}\mu}|\text{I}\rangle \\ &= \sum_{\text{I}} c_{\text{I}} \left(1 + \sum_{\mu} t^{(1)}_{\text{I}\mu}\; T_{\text{I}\mu}\right)|\text{I}\rangle\,. \end{aligned} \tag{50}$$

By introducing the excitation operator $T^{(1)}_{\text{I}}$ acting only on $|\text{I}\rangle$ as

$$T^{(1)}_{\text{I}} = \sum_{\mu} t^{(1)}_{\text{I}\mu}\; T_{\text{I}\mu}, \tag{51}$$

the expression of $|\Psi^{(1)}\rangle$ in Eq. (50) becomes

$$|\Psi^{(1)}\rangle = \sum_{\text{I}} c_{\text{I}} \left(1 + T^{(1)}_{\text{I}}\right)|\text{I}\rangle\,. \tag{52}$$

Such a parameterization for the first-order corrected wave function $|\Psi^{(1)}\rangle$ recalls immediately a first-order Taylor expansion of the general JM-MRCC ansatz

$$|\text{JM}-\text{MRCC}\rangle = \sum_{\text{I}} c_{\text{I}}\; e^{T_{\text{I}}}\;|\text{I}\rangle. \tag{53}$$

Nevertheless, based on the JM-MRPT2 expression for the amplitudes, one might imagine to divide the general T_{I} operator into a reference-dependent single excitation operator and a reference-independent double excitation operator. The JM-MRPT2 amplitudes might be used as a guess to start the iterative research of the MRCC equations.

Also, within the present formalism, the class of the double excitations can be factorized as shown in Sec. II A [see Eq. (33)]. Therefore, using the reference-independent amplitudes defined in Eq. (33), one can define a unique double excitation operator $\mathcal{T}_D^{(1)}$ as

$$\mathcal{T}_D^{(1)} = \sum_{m,n,p,q} \mathcal{T}^{qp(1)}_{mn}, \tag{54}$$

recalling thus the formalism of the internally contracted-MRCC[58–61] (ic-MRCC) which uses a unique excitation operator $\mathcal{T}$ as in the single-reference coupled-cluster

$$|\text{ic}-\text{MRCC}\rangle = e^{\mathcal{T}}|\psi^{(0)}\rangle = e^{\mathcal{T}}\sum_{\text{I}} c_{\text{I}}\;|\text{I}\rangle\,. \tag{55}$$

In such a perspective, as the energy provided by the JM-MRPT2 equations is size-extensive, it can be seen as a linearized coupled cluster version using a hybrid parameterization of the wave function: internally contracted ansatz for the double excitation operators and JM ansatz for the single excitation operators.

C. Determinant-based multi-reference perturbation theories

JM-MRPT2 presented here can be directly compared to the CIPSI method, just as the JM-HeffPT2 can be directly compared to the shifted-B_k method.[48–50] Indeed, by using the following amplitudes:

$$t^{\text{CIPSI}}_{\text{I}\mu} = \frac{\langle \text{I}|H|\mu\rangle}{e^{(0)} - \langle \mu\,|H|\,\mu\rangle}, \tag{56}$$

in the equation of the second-order correction on the energy [see Eq. (8)], one obtains the CIPSI energy, and by introducing $t^{\text{CIPSI}}_{\text{I}\mu}$ in the definition of the dressed Hamiltonian $\tilde{H}^{(2)}$, one obtains

$$\langle \text{I}|H^{(2)}_{\text{Shifted}-B_k}|\text{J}\rangle = \langle \text{J}\,|H|\,\text{J}\rangle + \sum_{\mu} \frac{\langle \text{I}|H|\mu\rangle\,\langle \mu\,|H|\,\text{J}\rangle}{e^{(0)} - \langle \mu\,|H|\,\mu\rangle}, \tag{57}$$

which defines the shifted-B_k Hamiltonian and corresponding energy once $H^{(2)}_{\text{Shifted}-B_k}$ is diagonalized. As mentioned previously, it has been shown that the size-consistency error of these methods comes from the unbalanced treatment between the variational energy of a multi-reference wave function such as $|\psi^{(0)}\rangle$ and the variational energy of the single Slater determinant $|\mu\rangle$. Such an error is not present within the definitions of the excitation energies in the JM-MRPT2 method as the latter introduces expectation values of the Hamiltonian over linear combinations of perturber Slater determinants.

In a similar context, one can compare the JM-HeffPT2 method to the Split-GAS[47] of Li Manni *et al.* whose definition of the amplitude is

$$t^{\text{Split}-\text{GAS}}_{\text{I}\mu} = \frac{\langle \text{I}|H|\mu\rangle}{e^{(0)} + e^{(2)} - \langle \mu\,|H|\,\mu\rangle}. \tag{58}$$

In the Split-GAS framework, the correlation energy $e^{(2)}$ brought by the perturbers is included in the energy denominator, which introduces self consistent equations as in the Brillouin-Wigner perturbation theory.[51] However, the size-consistency error in such a method is even more severe than in the shifted-B_k as the excitation energies are much larger due to the presence of the total correlation energy $e^{(2)}$.

IV. COMPUTATIONAL COST

A. Mathematical complexity and memory requirements

Compared to other size-extensive MRPT2 methods, a clear advantage of JM-MRPT2 is its simplicity. The NEVPT2 approach requires to handle the four-body density matrix and

the CASPT2 needs to handle the three-body density matrix. Both of these computationally intensive phases can be skipped in our formalism as one only needs to compute expectation values whose number is relatively small compared to NEVPT2 and CASPT2. The most involved quantity to be computed is

$$\Delta E^{(0)}_{ir} = e^{(0)} - \frac{\sum_{\mathrm{I}}\sum_{\mathrm{J}} c_{\mathrm{I}} \langle \mathrm{I}|H\ a_r^\dagger a_i|\mathrm{I}\rangle \langle \mathrm{I}|H|\mathrm{J}\rangle \langle \mathrm{J}|H\ a_r^\dagger a_i|\mathrm{J}\rangle c_{\mathrm{J}}}{\left(\sum_{\mathrm{I}} c_{\mathrm{I}} \langle \mathrm{I}|H\ a_r^\dagger a_i\,|\mathrm{I}\rangle\right)^2}, \tag{59}$$

for all pairs (i, r) where i is an inactive orbital and r is a virtual orbital. These quantities need to be only computed once since they can all fit in memory. Each $\Delta E^{(0)}_{ir}$ is, from the computational point of view, equivalent to an expectation value over the CASSCF wave function. As all $\Delta E^{(0)}_{ir}$ are independent, the computation of these quantities can be trivially parallelized. Regarding the memory footprint of the JM-MRPT2 method, it scales as $\mathcal{O}(n_a^3)$ (n_a being the number of active orbitals) for the storage of the $\Delta E^{(0)}_{a_a^\dagger a_b^\dagger a_c}$ and $\Delta E^{(0)}_{a_a^\dagger a_b a_c}$ quantities.

Regarding the complexity of the equations for the amplitudes, it is clear that once computed the active part of the denominator, JM-MRPT2 is just a simple sum of contributions. This is in contrast with the UGA-SSMRPT2 equations which involve the handling of coupled amplitude equations.

B. Removal of the determinant-based computational cost

The present formalisms are formally determinant-based methods, which implies that the computational cost should be proportional to the number of perturbers $|\mu\rangle$ that one has to generate to compute the corrections to the energy or the dressing of the Hamiltonian matrix, just as in the CIPSI, shifted-B_k, or UGA-SSMRPT2 methods. To understand the main computational costs, one can divide the excitation classes according to the difference dedicated CI (DDCI) framework,[52] which classifies the Slater determinants in terms of numbers of holes in the doubly occupied orbitals and particles in the virtual orbitals. If N_{CAS} is the number of Slater determinants of the CAS-CI zeroth order wave function, n_o, n_a, and n_v, respectively, the number of doubly occupied, active and virtual orbitals, one can then classify each excitation class according to the number of perturbers needed to compute their contribution to the second-order perturbation correction to the energy:

1. the *two-holes-two-particles* excitation class (2h2p) which scales as $N_{\mathrm{CAS}} \times n_o^2 \times n_v^2$;
2. the *one-hole-two-particles* excitation class (1h2p) which scales as $N_{\mathrm{CAS}} \times n_o \times n_a \times n_v^2$;
3. the *two-holes-one-particle* excitation class (2h1p) which scales as $N_{\mathrm{CAS}} \times n_o^2 \times n_a \times n_v$;
4. the *two-particles* excitation class (2p) which scales as $N_{\mathrm{CAS}} \times n_v^2$;
5. the *two-holes* excitation class (2h) which scales as $N_{\mathrm{CAS}} \times n_o^2$;
6. the *one-hole-one-particle* excitation class (1h1p) which scales as $N_{\mathrm{CAS}} \times n_o \times n_v$;
7. the *one-particle* excitation class (1p) which scales as $N_{\mathrm{CAS}} \times n_v$;
8. the *one-hole* excitation class (1h) which scales as $N_{\mathrm{CAS}} \times n_o$.

Nevertheless, our formalism presents several mathematical simplifications that allow one to basically remove any browsing over the Slater determinants $|\mu\rangle$, and once more there is a difference between the single and double excitations processes.

C. Factorization of the most numerous double excitation processes

As the five most computationally demanding excitation classes involve only double excitation operators in their equations, their contribution can be formalized directly, thanks to the one- and two-body density matrices of the zeroth-order wave function. To understand how one can write the second-order correction to the energy as

$$\begin{aligned} e^{(2)}_{\text{double exc.}} &= \sum_{m,n,p,q} \sum_{\mathrm{I}} c_{\mathrm{I}} \langle \psi^{(0)}|H\ a_q^\dagger a_p^\dagger a_n a_m|\mathrm{I}\rangle \frac{((mq|np))}{\Delta E^{(0)}_{a_q^\dagger a_p^\dagger a_n a_m}} \\ &= \sum_{m,n,p,q} \sum_{\mathrm{I,\,J}} c_{\mathrm{I}} c_{\mathrm{J}} \langle \mathrm{J}|H\ a_q^\dagger a_p^\dagger a_n a_m|\mathrm{I}\rangle \frac{((mq|np))}{\Delta E^{(0)}_{a_q^\dagger a_p^\dagger a_n a_m}}. \end{aligned} \tag{60}$$

Consequently, as $\langle \mathrm{J}\,|H\ a_q^\dagger a_p^\dagger a_n a_m|\mathrm{I}\rangle$ is necessarily of type

$$\langle \mathrm{J}|H\ a_q^\dagger a_p^\dagger a_n a_m|\mathrm{I}\rangle = ((ef|gh))\,\langle \mathrm{J}|a_f^\dagger a_h^\dagger a_g a_e\ a_q^\dagger a_p^\dagger a_n a_m|\mathrm{I}\rangle\,, \tag{61}$$

one can reformulate the second-order correction to the energy in terms of many-body density matrices

$$\begin{aligned} e^{(2)}_{\text{double exc.}} = \sum_{m,n,p,q,e,f,g,h} & \langle \psi^{(0)}|a_f^\dagger a_h^\dagger a_g a_e\ a_q^\dagger a_p^\dagger a_n a_m|\psi^{(0)}\rangle \\ & \times \frac{((mq|np))\ ((ef|gh))}{\Delta E^{(0)}_{a_q^\dagger a_p^\dagger a_n a_m}}. \end{aligned} \tag{62}$$

Such a formulation avoids completely to run over Slater determinants and consequently kills the prefactor in N_{CAS} involved in each of the excitation classes, just as in the internally contracted formalisms. Of course, because of the restrictions in terms of holes and particles in the inactive and virtual orbitals, the handling of the four-body density matrix never occurs in our formalism. We report here the explicit equations for the energetic corrections of the five most numerous double excitation classes

$$e^{(2)}_{\mathrm{2h2p}} = \frac{1}{2} \sum_{i,j,v,r} \frac{3(iv|jr)^2 + (ir|jv)^2 - 2(iv|jr)(ir|jv)}{\epsilon_i + \epsilon_j - \epsilon_v - \epsilon_r}, \tag{63}$$

$$e^{(2)}_{\mathrm{1h2p}} = \frac{1}{2} \sum_{i,v,r,a,b} \langle \psi^{(0)}|a_a a_b^\dagger|\psi^{(0)}\rangle \frac{((ir|av))((ir|bv))}{\epsilon_i + \Delta E^{(0)}_{a_a^\dagger} - \epsilon_r - \epsilon_v}, \tag{64}$$

$$e^{(2)}_{\mathrm{2h1p}} = \frac{1}{2} \sum_{i,j,r,a,b} \langle \psi^{(0)}|a_a^\dagger a_b|\psi^{(0)}\rangle \frac{((ir|aj))((ir|bj))}{\epsilon_i + \epsilon_j + \Delta E^{(0)}_{a_a} - \epsilon_r}, \tag{65}$$

$$e^{(2)}_{\mathrm{2p}} = \frac{1}{2} \sum_{r,v,a,b,c,d} \langle \psi^{(0)}|a_a^\dagger a_b^\dagger a_c a_d|\psi^{(0)}\rangle \frac{((ar|bv))((cr|dv))}{\Delta E^{(0)}_{a_c a_d} - \epsilon_r - \epsilon_v}, \tag{66}$$

TABLE I. Geometries used for the ethane and ethylene molecules.

Geometrical parameters	C_2H_6	C_2H_4
C–H (Å)	1.103	1.089
H–C–C (°)	111.2	120.0
H–C–H (°)	107.6	120.0
H–C–C–H (°)	180.0	180.0

$$e_{2h}^{(2)} = \frac{1}{2} \sum_{i,j,a,b,c,d} \langle \psi^{(0)} | a_a a_b a_c^\dagger a_d^\dagger | \psi^{(0)} \rangle \frac{((ai|bj))((ci|dj))}{\epsilon_i + \epsilon_j + \Delta E^{(0)}_{a_c^\dagger a_d^\dagger}}. \tag{67}$$

D. Simplification for the 1h1p excitation class

Thanks to the factorization of the most numerous double excitations processes, the remaining main computational cost comes from the single excitations involved in the 1h1p excitation class. In the case of single excitation processes, the factorization cannot be applied as the Hamiltonian matrix elements depend on the Slater determinant on which the single excitation is applied. The total energetic correction brought by the single excitation processes involved in the 1h1p excitation class can be expressed as follows:

$$\begin{aligned} e_{1h1p}^{(2)\text{ Single exc.}} &= \sum_{i,r} \sum_{\mathrm{I}} \langle \psi^{(0)} | H\ a_r^\dagger a_i | \mathrm{I} \rangle\, c_\mathrm{I} \frac{\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle}{\Delta E^{(0)}_{ir}} \\ &= \sum_{i,r} \sum_{\mathrm{I,\,J}} c_\mathrm{J} \langle \mathrm{J} | H\ a_r^\dagger a_i | \mathrm{I} \rangle\, c_\mathrm{I} \frac{\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle}{\Delta E^{(0)}_{ir}}. \end{aligned} \tag{68}$$

As the Hamiltonian matrix elements $\langle \mathrm{J} | H\ a_r^\dagger a_i | \mathrm{I} \rangle$ are simply

$$\langle \mathrm{J} | H\ a_r^\dagger a_i | \mathrm{I} \rangle = ((ir|ab)) \langle \mathrm{J} | a_b^\dagger a_a | \mathrm{I} \rangle, \tag{69}$$

one can reformulate the sum as

$$e_{1h1p}^{(2)\text{ Single exc.}} = \sum_{\mathrm{I,\,J}} c_\mathrm{J} c_\mathrm{I} \sum_{a,b} \mathcal{F}^\mathrm{I}_{ab} \langle \mathrm{J} | a_b^\dagger a_a | \mathrm{I} \rangle, \tag{70}$$

where the quantity $\mathcal{F}^\mathrm{I}_{ab}$ is the effective Fock operator associated with the Slater determinant $|\mathrm{I}\rangle$ involving the active orbitals a and b,

$$\mathcal{F}^\mathrm{I}_{ab} = \sum_{i,r} ((ir|ab)) \frac{\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle}{\Delta E^{(0)}_{ir}}. \tag{71}$$

Of course, as $\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle$ depends on the occupation of $|\mathrm{I}\rangle$, there is one effective Fock operator for each reference determinant $|\mathrm{I}\rangle$ which would suggest to compute explicitly these quantities for each Slater determinant within the CAS-CI space. Nevertheless, one can notice that $\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle$ is just a sum of terms

$$\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle = \sum_{m \text{ occupied in } |\mathrm{I}\rangle} ((ir|mm)). \tag{72}$$

Considering that the inactive orbitals are always doubly occupied in $|\mathrm{I}\rangle$, this sum can be split into an inactive and an active contribution, namely,

$$\langle \mathrm{I} | H\ a_r^\dagger a_i | \mathrm{I} \rangle = F^{\text{c.s.}}_{ir} + F^\mathrm{I}_{ir}, \tag{73}$$

where F^{cs}_{ir} and F^I_{ir} are defined as

$$F^{\text{cs}}_{ir} = \sum_{j \text{ doubly occupied in } |\mathrm{I}\rangle} 2(ir|jj) - (rj|ij), \tag{74}$$

$$F^\mathrm{I}_{ir} = \sum_{c \text{ occupied in} |\mathrm{I}\rangle} ((ir|cc)). \tag{75}$$

Therefore, one can first compute the effective Fock operator associated with the closed shell orbitals

$$\mathcal{F}^{\text{cs}}_{ab} = \sum_{i,r} ((ir|ab)) \frac{F^{\text{cs}}_{ir}}{\Delta E^{(0)}_{ir}}, \tag{76}$$

TABLE II. Non-parallelism errors and spectroscopic constants computed from the potential energy curves obtained at different computational levels for the F_2, C_2H_6, and FH molecules. NPE and D_0 are reported in mH, R_{eq} in Å, and k in hartree/Å^2.

	F_2				C_2H_6				FH			
	NPE	D_0	R_{eq}	k	NPE	D_0	R_{eq}	k	NPE	D_0	R_{eq}	k
CASSCF	30.7	22.1	1.53	0.43	27.7	154.0	1.55	0.99	35.3	180.0	0.92	2.15
JM-MRPT2	6.7	46.3	1.44	0.85	2.5	179.0	1.53	1.06	9.7	220.4	0.93	2.11
JM-MRPT2 (deloc)	11.4	51.1	1.43	0.93	4.5	181.6	1.54	1.06	13.1	224.3	0.93	2.13
SC-NEVPT2	8.5	48.1	1.44	0.88	2.6	179.2	1.54	1.07	9.5	220.5	0.93	2.11
PC-NEVPT2	8.5	48.2	1.44	0.88	2.5	179.2	1.54	1.07	9.5	220.5	0.93	2.11
CASPT2 (IPEA = 0)	2.6	44.1	1.46	0.74	3.6	175.0	1.53	1.08	3.1	214.1	0.92	2.16
CASPT2 (IPEA = 0.25)	3.9	44.3	1.46	0.75	3.4	177.8	1.53	1.09	4.0	214.5	0.92	2.17
Mk-MRPT2[a]	...	47.2	1.44	0.60	...	...	...	...	...	...	...	...
Mk-MRPT2 (deloc)[a]	...	48.4	1.44	0.71	...	...	...	...	...	...	...	...
JM-HeffPT2	7.4	50.1	1.45	0.87	2.4	179.4	1.53	1.05	8.9	221.9	0.93	2.11
JM-HeffPT2 (deloc)	14.2	56.2	1.44	0.94	5.4	182.2	1.54	1.05	14.5	226.1	0.94	2.14
Shifted B_k	6.6	50.1	1.48	0.80	8.6	136.4	1.64	0.75	44.3	216.4	0.93	2.11
Shifted B_k (deloc)	5.7	91.8	1.41	1.26	4.7	220.5	1.53	1.09	26.7	236.1	0.94	2.31
FCI[b]	...	45.1	1.46	0.77	...	177.7	1.53	1.06	...	214.4	0.92	2.16

[a]Results from Ref. 62.

[b]Results obtained with CIPSI calculations converged up to a second-order perturbative correction lower than 10^{-4} hartree.

which is common for all the Slater determinants $|\mathrm{I}\rangle$ within the CAS-CI space. Then, what differentiates the effective Fock operator between two different determinants $|\mathrm{I}\rangle$ and $|\mathrm{J}\rangle$ is the active part

$$F^{\mathrm{I}}_{ab} = \sum_{i,r} ((ir|ab)) \frac{F^{\mathrm{I}}_{ir}}{\Delta E^{(0)}_{ir}}. \tag{77}$$

One can then notice that the active part of the Fock operator F^{I}_{ir} is just a sum over all active orbitals occupied in $|\mathrm{I}\rangle$ of quantities that only depend on the active orbitals

$$F^{\mathrm{I}}_{ab} = \sum_{c \text{ occupied in } |\mathrm{I}\rangle} F^{c}_{ab}, \tag{78}$$

where F^{c}_{ab} is nothing but

$$F^{c}_{ab} = \sum_{i,r} ((ir|ab)) \frac{((ir|cc))}{\Delta E^{(0)}_{ir}}. \tag{79}$$

Therefore, by computing and storing all possible F^{c}_{ab} together with $\mathcal{F}^{\mathrm{cs}}_{ab}$, one can then easily rebuild the total effective operator of a given Slater determinant $|\mathrm{I}\rangle$,

$$\mathcal{F}^{\mathrm{I}}_{ab} = \mathcal{F}^{\mathrm{cs}}_{ab} + \sum_{c \text{ occupied in } |\mathrm{I}\rangle} F^{c}_{ab}, \tag{80}$$

and consequently compute the total second-order correction to the energy $e^{(2)\ \text{Single exc.}}_{\text{1h1p}}$ as a simple expectation value. To summarize, a computational step scaling as $N^2_{\mathrm{CAS}} \times n_o \times n_v$ [see Eq. (68)] is replaced by a first calculation scaling as $n^3_{act} \times n_o \times n_v$ [see Eq. (79)], followed by the computation of an expectation value scaling as N^2_{CAS}, independent of the number of doubly occupied and virtual orbitals.

V. NUMERICAL RESULTS

The present section spells out the numerical results obtained for the potential energy curves and corresponding spectroscopic constants of six molecules involving a single, double, and triple bond breaking, which are F_2, FH, C_2H_6, C_2H_4, H_2O, and N_2. We also report the computation of the $^1A_g \rightarrow {}^1B_{1u}$ excitation energy of the ethylene molecule and compare it to the near FCI value obtained by Daday and coworkers[53] with the FCI-Quantum Monte Carlo (FCI-QMC) approach. A numerical test of strong separability is also provided in the case of the $F_2 \cdots FH$ molecule.

A. General computational details

The cc-pVDZ basis set has been used in all cases, except for the FH molecule for which the aug-cc-pVDZ basis set was retained, and pure spherical harmonics were used for all calculations. The frozen core approximation has been used, and consequently the 1*s* electrons were systematically frozen for all non-hydrogen atoms. The near FCI reference values were obtained using the CIPSI algorithm developed in the program *Quantum Package*[54] and all calculations were converged below 0.1 mH. The shifted-B_k, JM-MRPT2, and JM-HeffPT2 have been implemented in the *Quantum Package*, and all CASSCF calculations were performed using the GAMESS(US)[55] software. The CASPT2 calculations were performed with MOLCAS 7.8,[56] while the NEVPT2 results were obtained using stand-alone codes developed at the University of Ferrara and interfaced with MOLCAS 7.8. The geometrical parameters used for the C_2H_6 and C_2H_4 molecules can be found in Table I, and the H–O–H angle of the H_2O molecule has been set to 110.6°. Concerning the excitation energy calculation of the ethylene molecule, we used the experimental geometry and the ANO-L-VDZ basis set[57] in order to compare to one of the values obtained within the FCI-QMC method in Ref. 53.

In order to compare the performance of the here proposed formalisms with other determinant-based MPRT2 methods, we have also performed calculations using the shifted-B_k method using an Epstein-Nesbet zeroth order Hamiltonian, and we also report results obtained at the Mk-MPRT2[62] and UGA-SSMRPT2[39] level of theories when available. For

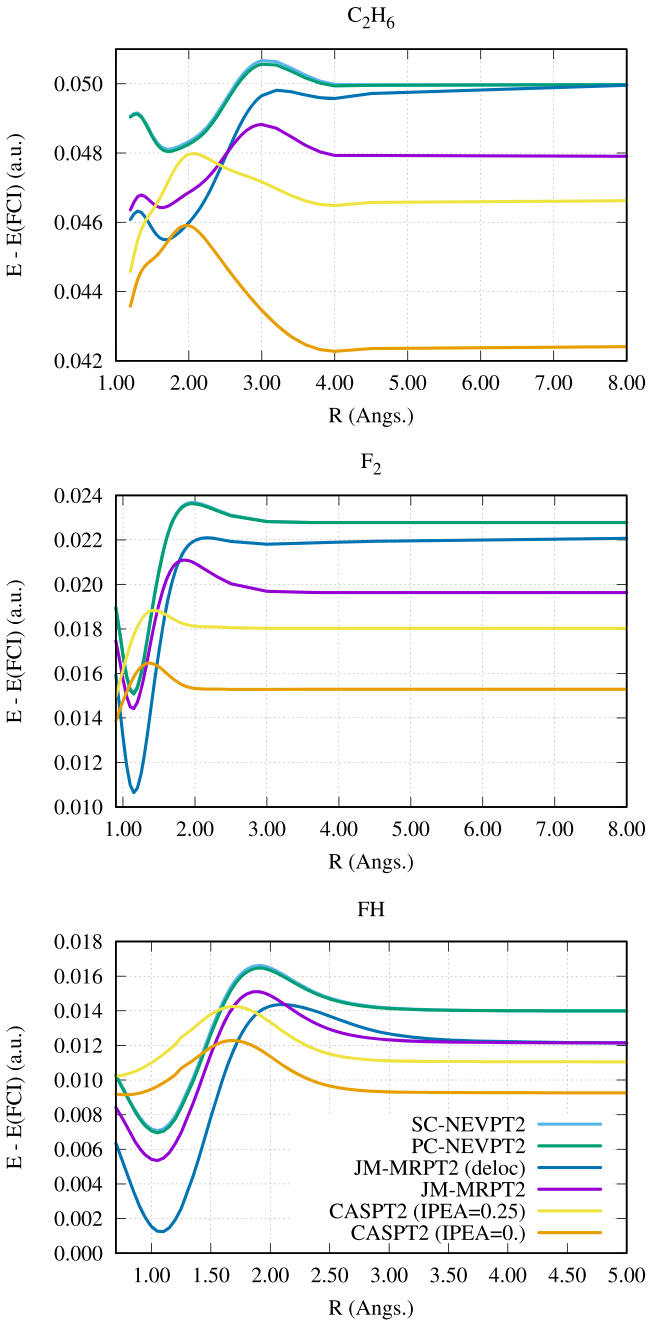


FIG. 2. Comparison of different MR-PT2 schemes with the FCI energy along the potential energy curves of C_2H_6, F_2, with the cc-pVDZ basis set, and FH with the aug-cc-pVDZ basis set. Energy differences in atomic units.

the sake of comparison with other state-of-the-art methods, we also report the spectroscopic constants and the error with respect to FCI obtained at the strongly contracted (SC-NEVPT2) and partially contracted (PC-NEVPT2) NEVPT2 using delocalized orbitals, together with CASPT2 with two different IPEA values. The IPEA values were chosen as 0 as in the original formulation of CASPT2, and 0.25 corresponding to the nowadays standard CASPT2 method.

B. Definition of the active spaces and localized orbitals

All MRPT2 calculations started with a minimal valence CASSCF involving the bonding and anti-bonding orbitals of each bond being broken along the potential energy curve. In the case of the single bond breaking, it simply implies a CASSCF(2,2) with the σ and σ^* orbitals. The following minimal valence active spaces are used for the three systems involving multiple bond breaking: for the H_2O molecule, a CASSCF(4,4) with four orbitals of valence character (using the C_{2v} symmetry point group, two orbitals of the A_1 irrep and two orbitals of the B_2 irrep having a C–H bonding character); for the C_2H_4 molecule, a CAS(4, 4) has been performed using the bonding and anti-bonding orbitals of both the σ and π C–C bonds; and for the N_2 molecule, a CAS(6, 6) has been used with the bonding and anti-bonding orbitals of the σ and the two π bonds.

Nevertheless, as it is the case for many multi-reference perturbation theories, our formalism is not invariant through orbital rotations within each orbital space (active, inactive, and virtual). Therefore one can choose to use delocalized orbitals, as the canonical ones, or localized orbitals. The present formalism is strictly separable when localized orbitals are used, so it seems therefore natural to use localized active orbitals rather than the canonical ones. In the case, F_2, N_2, C_2H_6, and C_2H_4, these orbitals are simply obtained by a rotation of $\pi/4$ between the bonding and anti-bonding active orbitals (σ and σ^* for the σ bond, π and π^* for the π bonds, and so on). In the case of the FH and H_2O molecules, the active orbitals were obtained, thanks to a rotation of the canonical active MOs in order to maximize the overlap with reference localized orbitals following chemical intuition: for the FH molecule, they consist in the $2p_z$ atomic orbital of the fluorine and $1s$ atomic orbital of the hydrogen atom, and for the H_2O molecule they consist in the two $1s$ atomic orbitals of the hydrogen atoms and of two simple linear combinations of the $2p_x$ and $2p_y$ orbitals, each one pointing to a given hydrogen atom.

Even if the present formalism is strictly separable only using localized orbitals, we nevertheless investigate the dependency of the choice of the active orbitals for the three molecules involving a single bond breaking (F_2, FH and C_2H_6) for which we report calculations both with canonical delocalized active orbitals (which are referred as "deloc") and localized active orbitals.

C. Single bond breaking

Table II presents the spectroscopic constants, namely, equilibrium distance (R_{eq}), the bond energy (D_0), and the second derivative (k) at R_{eq}, for the F_2, C_2H_6, and FH molecules at different computational levels. Also, we represent in Fig. 2 the difference of the FCI energy along the potential energy curves of those systems. From these data, several trends can be observed, both regarding the quality of the potential energy curves and the dependency on the choice of the active orbitals.

1. Dependency on the locality of the active orbitals

From the error of the potential energy curve to the FCI reference, it appears that the JM-MRPT2 method gives systematically better spectroscopic constants and a lower error with respect to the full-CI energy when localized orbitals are chosen. This is consistent with the fact that these methods are strictly separable when localized orbitals are used. Therefore, from now on we shall only refer to the results obtained with localized orbitals. One can remark that in the case of the F_2 molecule where Mk-MRPT2 calculations are available in the literature,[62] the JM-MRPT2 method gives very similar results.

2. Quality of the potential energy curves

From Table II, one can observe that the results obtained with the JM-MRPT2 method are comparable to those obtained

TABLE III. Non-parallelism errors and spectroscopic constants computed from the potential energy curves obtained at different computational levels for the H_2O, C_2H_4, and N_2 molecules. NPE and D_0 are reported in mH, R_{eq} in Å and k in hartree/Å^2.

	H_2O				C_2H_4				N_2			
	NPE	D_0	R_{eq}	k	NPE	D_0	R_{eq}	k	NPE	D_0	R_{eq}	k
CASSCF	40.9	289.3	0.96	3.74	26.2	252.6	1.36	2.03	18.2	313.7	1.11	5.34
JM-MRPT2	3.0	332.7	0.96	3.89	3.7	279.5	1.35	2.07	3.4	316.9	1.12	5.05
SC-NEVPT2	2.4	329.2	0.96	3.81	2.4	278.2	1.36	2.09	2.3	317.2	1.12	5.10
PC-NEVPT2	2.5	329.5	0.96	3.81	3.2	279.3	1.35	2.10	1.3	318.2	1.12	5.10
CASPT2 (IPEA = 0)	5.5	325.4	0.96	3.86	6.0	271.9	1.35	2.10	9.6	310.2	1.12	5.07
CASPT2 (IPEA = 0.25)	3.0	327.9	0.96	3.86	4.5	278.0	1.35	2.11	4.4	318.8	1.12	5.14
JM-HeffPT2	4.8	333.9	0.96	3.85	4.0	280.2	1.35	2.08	4.5	317.1	1.12	4.99
Shifted B_k	30.8	304.3	0.98	3.37	7.6	238.5	1.40	1.73	5.9	277.7	1.14	4.42
FCI[a]	...	330.3	0.96	3.89	...	277.0	1.35	2.09	...	319.4	1.12	5.04

[a]Results obtained with CIPSI calculations converged up to a second-order perturbative correction lower than 10^{-4} hartree.

with the well-established CAS-PT2 and NEVPT2 methods. The largest deviation on D_0 is of 6 mH for the FH molecule, representing less than 3% of error on the total binding energy, whereas it is of 1.2 mH and 1.3 mH which represents an error of less than 3% and 1% on the binding energy for the F_2 and C_2H_6 molecules, respectively. The equilibrium geometries obtained at the JM-MRPT2 level are always within 1% of error with respect to the FCI estimates, and so are the k values except for the F_2 molecule for which a significant deviation of 10% is observed. Except for the quality of the results, one can observe a systematic overestimation of the binding energy at the JM-MRPT2 level.

The non-parallelism error (NPE) is, within the computed points, the difference between the maximum and minimum absolute errors with respect to FCI energies. In addition to the the spectroscopic constants, the NPE is also a good indicator of the quality of the results of a given method. Using localized orbitals, the NPE obtained at JM-MRPT2 is of 6.7 mH for the F_2 molecule, 2.5 mH for C_2H_6, and 9.7 mH for the FH molecule. The maximum NPE is then for the FH molecule, which has also the largest energetic variation among the three molecules studied here.

D. Numerical results for double and triple bond breaking

Table III presents the spectroscopic constants obtained for the H_2O, C_2H_4, and N_2 molecules and Fig. 3 shows the difference of the FCI energy along the potential energy curves. From Table III, it appears that the results obtained with the JM-MRPT2 method follow a trend similar to what has been observed with the study of the three molecules involving a single bond breaking: the spectroscopic constants obtained at this level of theory are globally in good agreement with the FCI ones, D_0 obtained at the JM-MRPT2 level tends to be overestimated. Also, the absolute error on D_0 obtained at JM-MRPT2 is quite constant: 2.4 mH, 2.4 mH, and 2.5 mH, representing 0.7%, 0.9%, and 0.8% of the total binding energy for the H_2O, C_2H_4, and N_2 molecules, respectively.

Regarding the curves displaying errors with respect to the FCI energies, it appears that the JM-MRPT2 curves are smooth and do not present any intruder state problems, with an NPE between 3 and 4 mH.

E. Comparison of JM-HeffPT2 with shifted-B_k

Figure 4 shows the difference of the FCI for all the previously studied systems, for the JM-HeffPT2 and the shifted-B_k methods. It is clear that in all the cases, the potential energy curves obtained with the JM-HeffPT2 are much more parallel to the FCI curve than the shifted-B_k ones. Also, it is worth mentioning that the JM-HeffPT2 curves are smooth and do not present any intruder state problems. The spectroscopic constants and NPEs calculated with both methods are given in Tables II and III.

In general, the energetic values obtained after diagonalizing the effective Hamiltonian are not better than those obtained with the JM-MRPT2 method. But the main advantage of JM-HeffPT2 over JM-MRPT2 is that it provides improved CI-coefficients on the reference space, like the shifted-B_k method. To illustrate the quality of the improved wave functions, we report in Table IV the ratios c_i/c_n where c_i and c_n are the CI-coefficients of the determinants relative to the ionic and neutral structures of F_2 obtained at the CAS-CI, JM-HeffPT2, shifted-B_k, and CIPSI levels. As a reference, CIPSI calculations were carried out in the frozen-core FCI space, and the number of determinants (N_{det}) selected in the variational wave function are given in Table IV. For such large wave functions, the CI-coefficients on the reference determinants are expected to be very close to the FCI limit. Both the JM-HeffPT2 and shifted-B_k methods show a significant improvement of the wave function, and JM-HeffPT2 is in very good agreement with the FCI especially at the equilibrium distance. Similarly, we report in Table V computations of the dipole moment along the internuclear axis for the FH molecule and compare it to

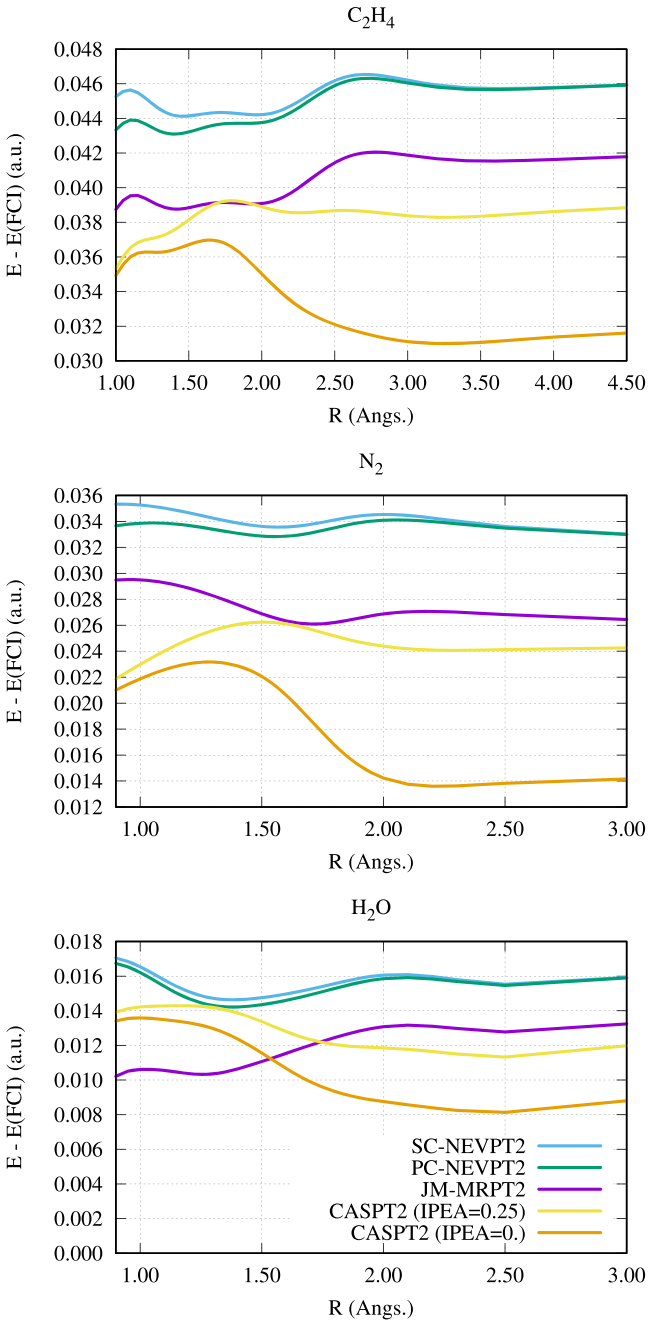


FIG. 3. Comparison of different MR-PT2 schemes with the FCI energy along the potential energy curves of C_2H_4, N_2, and H_2O with the cc-pVDZ basis set. Energy differences in atomic units.

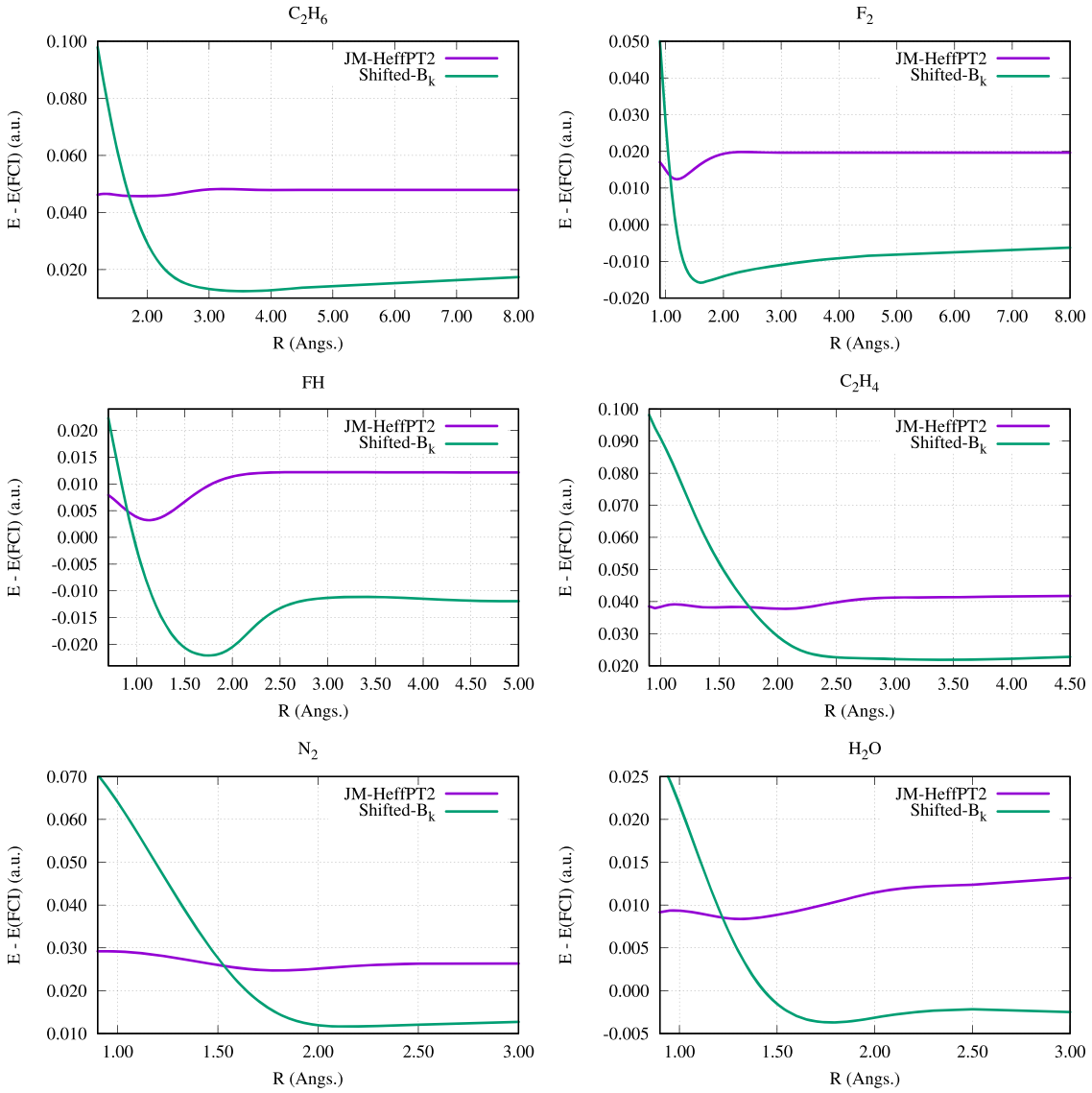


FIG. 4. Energy difference of JM-HeffPT2 and shifted-B_k with respect to the FCI energy along the potential energy curves of C_2H_6, F_2, C_2H_4, N_2, and H_2O with the cc-pVDZ basis set, and FH with the aug-cc-pVDZ basis set. Energy differences in atomic units.

values obtained by projecting and normalizing large CIPSI wave functions on the CAS-CI space (referred hereafter as CIPSI-proj-CAS). Therefore, the dipole moment computed with a given method only depends on the relative coefficients of the four Slater determinants belonging to the CAS-CI space. From these results, it clearly appears that the JM-HeffPT2 method allows one to obtain values for the dipole moment that are in excellent agreement with that obtained at the CIPSI-proj-CAS level of theory. Also, one can notice a significant improvement of the description of the dipole moment going from the CAS-CI wave function to the JM-HeffPT2 wave function, implying that the diagonalization of the dressed Hamiltonian leads to coefficients within the CAS-CI space that are closer to the ones of the FCI wave function, which is not the case for the shifted-B_k method.

TABLE IV. Ratios c_i/c_n at different internuclear distances, for the F_2 molecule (cc-pVDZ). The last row indicates in italics the number of Slater determinants in the CIPSI wave functions.

F2	1.4119 Å	2 Å	3 Å
CAS-CI	0.572	0.212	0.024
JM-HeffPT2	0.646	0.273	0.033
Shifted-B_k	0.707	0.274	0.030
CIPSI	0.638	0.259	0.030
N_{det}	*6 321 822*	*7 889 806*	*12 748 141*

TABLE V. Dipole moment (reported in a.u.[2]) along the internuclear axis obtained at various computational levels for the FH molecule (aug-cc-pVDZ). The last row indicates in italics the number of Slater determinants in the CIPSI wave functions.

FH	0.95 Å	1.4 Å	1.9 Å
CAS-CI	1.07	1.87	3.20
JM-HeffPT2	1.04	1.70	2.88
Shifted-B_k	1.01	1.47	2.19
CIPSI-proj-CAS	1.05	1.71	2.92
N_{det}	*2 677 789*	*2 545 448*	*2 153 580*

TABLE VI. Excitation energy of the ethylene molecule for the $^1B_{1u}$ singlet state computed in the ANO-L-VDZ basis set.

	ΔE (eV)
CASSCF(2,2)	8.83
JM-MRPT2 (loc)	8.38
JM-MRPT2 (deloc)	8.47
JM-HeffPT2 (loc)	8.42
JM-HeffPT2 (deloc)	8.53
Shifted-B_k (loc)	7.97
Shifted-B_k (deloc)	7.62
FCI-QMC[a]	8.25(1)

[a]Results obtained from Ref. 53.

F. The excited state $^1B_{1u}$ of the ethylene molecule

The excited state of the ethylene molecule $^1B_{1u}$ of singlet spin symmetry has been the subject of intense debates, both from a theoretical and experimental point of views. The excited state $^1B_{1u}$ resulting from the singlet coupling of the single $\pi \rightarrow \pi^*$ excitation has a strong ionic character. Consequently, the electronic correlation effects are much larger in such a state than in the ground state where the neutral forms dominate, explaining the high dependency of the excitation energy to the level of treatment of electronic correlation.[63,64] In order to test the applicability of the JM-MRPT2 method for the computation of excited states, we performed state-specific calculations on both the ground and the singlet $^1B_{1u}$ states and compare it to the near FCI values obtained by Daday *et al.*[53] As $^1B_{1u}$ is the lowest singlet of the B_{1u} symmetry (using the D_{2h} point group), we optimize the orbitals of both the ground and excited states at the CASSCF level using two electrons in the two π and π^* orbitals. We used the ANO-L-VDZ basis set[57] and performed the calculation both with symmetry adapted and localized active orbitals. The results are reported in Table VI. From this table, one can notice that the maximum deviation from the FCI-QMC result is of 0.31(1) eV using JM-HeffPT2 with delocalized orbitals, whereas the minimum deviation of 0.13(1) is obtained with JM-MRPT2 with localized orbitals.

G. Numerical evidence of strong separability

A given method based on the definition of active orbitals is said to be strongly separable when the energy computed for a system composed of two non interacting fragments A $\cdots$ B with active orbitals both on system A and B coincides with the sum of the energies of each sub system computed individually with the corresponding active orbitals on the fragments A and B. The present definitions of JM-MRPT2 and JM-HeffPT2 respect the property of strong separability when localized orbitals are used. A formal proof of the strict separability is given in the Appendix. In order to give a numerical example of the strong separability property, we report in Table VII calculations on F_2 (F–F = 1.45 Å), FH (F–H = 0.90 Å), and on the super-system of $F_2 \cdots$ FH at an intermolecular distance of 100 Å. As the two subsystems are different, the orbitals obtained by the CASSCF method are localized on each system, which is a necessary condition for the strong separability in our formalism. From Table VII, it appears that the deviations on the computed correlated energy $e^{(2)}$-JM-MRPT2 [see Eq. (8)] between the super system with non-interacting fragments and the sum of the two systems is lower than 10^{-13} hartree, which is actually smaller than the non-additivity of the CASSCF energies. For JM-HeffPT2, the relative error remains in the same order of magnitude than that for the CASSCF. This shows that the effective Hamiltonian does not introduce non-separability error. Finally, one should notice the strong non-separability error of the shifted-B_k approach.

VI. CONCLUSIONS AND PERSPECTIVES

A. Summary of the main results

The present work has presented a new MRPT2 approach, the JM-MRPT2 method, that uses individual Slater determinants as perturbers and allows for an intermediate Hamiltonian formulation, which is the JM-HeffPT2 approach. These methods are strictly size-consistent when localized orbitals are used, as has been numerically illustrated here. The link of these two new methods with other existing multi-reference theories has been established, specially in the case of the SC-NEVPT2 level of theory. The accuracy of the methods has been investigated on a series of ground state potential energy curves up to the full dissociation limit for a set of six molecules involving single (F_2, FH, and C_2H_6), double (H_2O, C_2H_4), and triple bond breaking (N_2), using the cc-pVDZ basis set and the aug-cc-pVDZ basis set in the case of FH. The two methods proposed here have been compared to near FCI energies, thanks to large CIPSI calculations converged bellow 0.1 mH, whose values can be found in the supplementary material. The quality of the results has been investigated by means of the non-parallelism error and three spectroscopic constants (R_{eq}, D_0, and k) together with absolute errors with respect to FCI energies along the whole potential energy curves. Among the six molecules studied here, the largest error found on the

TABLE VII. Total energies (a.u.) for the numerical separability check on $F_2 \cdots$ FH.

	CASSCF	Shifted-B_k	$e^{(2)}$-JM-MRPT2	JM-HeffPT2
F_2	−198.746 157 368 569	−199.122 170 300	−0.337 009 510 134 933	−199.085 305 155 169 4
FH	−100.031 754 985 880	−100.289 784 498	−0.230 422 886 638 017	−100.262 424 667 296 7
F_2 + FH	−298.777 912 354 448	−299.411 954 798	−0.567 432 396 772 949	−299.347 729 822 466 0
$F_2 \cdots$ FH	−298.777 912 354 443	−299.396 752 116	−0.567 432 396 773 035	−299.347 729 822 461 6
Absolute error (a.u.)	5.0×10^{-12}	1.5×10^{-2}	8.6×10^{-14}	4.4×10^{-12}
Relative error	1.7×10^{-14}	5.1×10^{-5}	1.5×10^{-13}	1.4×10^{-14}

binding energy at the JM-MRPT2 level of theory is of 6 mH for the FH molecule, representing a deviation lower than 3% with respect to the FCI value. In all other cases, the errors on D_0 are much smaller, ranging from 1.3 mH to 2.5 mH, which represents deviations between 1% and 3% with respect to the FCI estimates. The equilibrium distance is also found to be always within 1% of the FCI values. These results are very encouraging, specially considering the simplicity of this second-order perturbation theory, and its low computational cost. Regarding the JM-HeffPT2 method, its intermediate Hamiltonian formulation allows one to take into account the dominant part of the coupling between the static and dynamic correlation effects. From what has been observed in the present calculations, the diagonalization of the symmetrized intermediate Hamiltonian yields improved CI-coefficients on the reference determinants, together with a very small NPE compared to the shifted-B_k method.

B. Perspectives

Due to its flexibility, the present formalism offers a broad field of perspectives. First, the JM-MRPT2 and JM-HeffPT2 methods can be formalized with a zeroth order wave function that does not need to be a CAS-CI eigenvector. This opens the way of treating much larger active spaces as one can select the dominant configurations of a given CAS-CI space, thanks to the use of a perturbative criterion (as in the CIPSI algorithm) or by using localized orbitals. Second, the reasons of the systematic slight overestimation of the binding energy at the JM-MRPT2 level of theory can also be investigated, taking benefit from localized active orbitals and of the clear reading of the reference wave function that they offer. Moreover, this allows one to use as zeroth-order wave function quasi diabatic states obtained, for instance, by a unitary transformation of a few CI eigenvectors[65] (either of a CAS-CI or from a more general CI). Also, as it has been shown that the present formulation is connected to multi-reference coupled-cluster formalisms, it is possible to derive the working equations starting from the JM-MRCC ansatz. This will allow one to obtain higher order terms which may correct the slight overestimation of the binding energies. The coupling of the present formalism with multi-reference coupled cluster models follows naturally. For instance, the treatment of the most numerous excitation classes at the JM-MRPT2 level can easily be combined with the recently introduced JM-MRCC ansatz of some of the present authors.[66] This will allow for a drastic lowering of the computational costs of the JM-MRCC ansatz, and opens the way to the treatment of larger systems at high level of *ab initio* theory.

SUPPLEMENTARY MATERIAL

See supplementary material for all the near FCI energies obtained with the CIPSI calculations.

ACKNOWLEDGMENTS

This work was performed using HPC resources from CALMIP (Toulouse) under Allocation No. 2016-0510 and from GENCI (Grant No. 2016-081738).

APPENDIX: PROOF OF STRONG SEPARABILITY AND LINK WITH THE MUPA APPROACH

The present appendix provides analytical derivations in order to demonstrate analytically the size consistency property of the JM-MRPT2 and JM-HeffPT2 methods (see part 1), and also to show the link of these two methods with the multi-partitioning of the Hamiltonian (see part 2).

1. Proof of separability

The present section proposes an analytical proof of strong separability of the JM-MRPT2 method. In a MRPT2 framework, the strong separability requires that an excitation T_A located on a system A gives the same contribution to the correlation energy with or without the presence of another system B whose zeroth-order wave function contains correlation effects. To be more specific, let us define the zeroth-order wave function and energy of a system A,

$$|\psi^{(0)A}\rangle = \sum_{\mathrm{I_A}} c_{\mathrm{I_A}} \, |\mathrm{I_A}\rangle \,, \tag{A1}$$

$$E^{(0)A} = \frac{\langle \psi^{(0)A} \, | H_A \, | \psi^{(0)A} \rangle}{\langle \psi^{(0)A} \, | \psi^{(0)A} \rangle}, \tag{A2}$$

and the same quantities for the system B,

$$|\psi^{(0)B}\rangle = \sum_{\mathrm{I_B}} c_{\mathrm{I_B}} \, |\mathrm{I_B}\rangle \,, \tag{A3}$$

$$E^{(0)B} = \frac{\langle \psi^{(0)B} | H_B | \psi^{(0)B} \rangle}{\langle \psi^{(0)B} \, | \psi^{(0)B} \rangle}. \tag{A4}$$

Let us consider now a given excitation T_A acting only on a system A. According to the definition of Eq. (14), the corresponding contribution to the first-order perturbed wave function is

$$|\psi^{(1)A}_{T_A}\rangle = \frac{1}{\Delta E^{(0)A}_{T_A}} \; |\tilde{\psi}^{(1)A}_{T_A}\rangle \,, \tag{A5}$$

$$|\tilde{\psi}^{(1)A}_{T_A}\rangle = \sum_{\mathrm{I_A}} \mathrm{c_{I_A}} \; \langle \mathrm{I_A} \, | H_A \; T_A \, | \mathrm{I_A} \rangle \; T_A \, | \mathrm{I_A} \rangle \,, \tag{A6}$$

and the excitation energy $\Delta E^{(0)A}_{T_A}$ characteristic of the excitation T_A is defined according to Eq. (17) as

$$\Delta E^{(0)A}_{T_A} = E^{(0)A} - \frac{\langle \tilde{\psi}^{(1)A}_{T_A} | H_A \, | \tilde{\psi}^{(1)A}_{T_A} \rangle}{\langle \tilde{\psi}^{(1)A}_{T_A} | \tilde{\psi}^{(1)A}_{T_A} \rangle}. \tag{A7}$$

Therefore, its contribution to the correlation energy of A is

$$e^{(2)A}_{T_A} = \langle \psi^{(0)A} | H | \psi^{(1)A}_{T_A} \rangle = \frac{\langle \psi^{(0)A} | H_A \, | \tilde{\psi}^{(1)A}_{T_A} \rangle}{\Delta E^{(0)A}_{T_A}}. \tag{A8}$$

A necessary and sufficient mathematical condition for the strong separability property of the energy is that a given excitation process T_A involving only the orbitals of the system A gives the same contribution to the energy when it is considered on the sole system A or on the super non interacting system $A \cdots B$. To reach such a condition, one first needs that the zeroth-order wave function be the product of the zeroth-order wave function of the two sub-systems A and B,

$$|\psi^{(0)A+B}\rangle = |\psi^{(0)A}\rangle \otimes |\psi^{(0)B}\rangle \,, \tag{A9}$$

which ensures that its corresponding zeroth-order energy is the sum of zeroth-order energies of the sub-systems A and B,

$$\begin{aligned} E^{(0)A+B} &= \frac{\langle\psi^{(0)A+B}|H_A+H_B|\psi^{(0)A+B}\rangle}{\langle\psi^{(0)A+B}|\psi^{(0)A+B}\rangle} \\ &= \frac{\langle\psi^{(0)A}|H_A|\psi^{(0)A}\rangle\langle\psi^{(0)B}|\psi^{(0)B}\rangle}{\langle\psi^{(0)A}|\psi^{(0)A}\rangle\langle\psi^{(0)B}|\psi^{(0)B}\rangle} \\ &\quad + \frac{\langle\psi^{(0)B}|H_B|\psi^{(0)B}\rangle\langle\psi^{(0)A}|\psi^{(0)A}\rangle}{\langle\psi^{(0)B}|\psi^{(0)B}\rangle\langle\psi^{(0)A}|\psi^{(0)A}\rangle} \\ &= E^{(0)A}+E^{(0)B}, \end{aligned} \tag{A10}$$

as the total Hamiltonian can be written as the sum of H_A acting only on the orbitals of A and the corresponding H_B acting only on the orbitals of B. A CAS-CI wave function respects of course the property of the additivity of the energy.

Starting from $|\psi^{(0)A+B}\rangle$, one can generate the contribution to the first-order perturbed wave function $|\psi^{(1)A}_{T_A}\rangle$ associated with T_A in the super-system $A\cdots B$,

$$|\psi^{(1)A+B}_{T_A}\rangle = \frac{1}{\Delta E^{(0)A+B}_{T_A}}\ |\tilde{\psi}^{(1)A+B}_{T_A}\rangle, \tag{A11}$$

$$\begin{aligned} |\tilde{\psi}^{(1)A+B}_{T_A}\rangle = \sum_{\mathrm{I_A}\,\mathrm{I_B}} c_{\mathrm{I_A}}c_{\mathrm{I_B}}\ T_A\ |\mathrm{I_B}\rangle\otimes|\mathrm{I_A}\rangle \\ \langle\mathrm{I_A}|\otimes\langle\mathrm{I_B}|(H_A+H_B)\ T_A\,|\mathrm{I_B}\rangle\otimes|\mathrm{I_A}\rangle, \end{aligned} \tag{A12}$$

with the following excitation energy $\Delta E^{(0)A+B}_{T_A}$:

$$\Delta E^{(0)A+B}_{T_A} = E^{(0)A+B} - \frac{\langle\tilde{\psi}^{(1)A+B}_{T_A}|H_A+H_B|\tilde{\psi}^{(1)A+B}_{T_A}\rangle}{\langle\tilde{\psi}^{(1)A+B}_{T_A}|\tilde{\psi}^{(1)A+B}_{T_A}\rangle}. \tag{A13}$$

Then, the contribution of T_A to the correlation energy of the super system $A\cdots B$ is simply

$$e^{(2)A+B}_{T_A} = \frac{\langle\psi^{(0)A+B}|H_A+H_B|\tilde{\psi}^{(1)A+B}_{T_A}\rangle}{\Delta E^{(0)A+B}_{T_A}}. \tag{A14}$$

One can then notice that as T_A only acts on the orbitals of A, one has

$$\langle\mathrm{I_A}|\otimes\langle\mathrm{J_B}|(H_A+H_B)\ T_A|\mathrm{J_B}\rangle\otimes|\mathrm{I_A}\rangle = \langle\mathrm{J_B}\,|\mathrm{J_B}\rangle\langle\mathrm{I_A}|H_A\ T_A|\mathrm{I_A}\rangle, \tag{A15}$$

and consequently the zeroth-order wave function of system B can be factorized in Eq. (A12)

$$\begin{aligned} |\tilde{\psi}^{(1)A+B}_{T_A}\rangle &= \sum_{\mathrm{I_B}} c_{\mathrm{I_B}}\,|\mathrm{I_B}\rangle\otimes\sum_{\mathrm{I_A}} c_{\mathrm{I_A}}\,\langle\mathrm{I_A}|H_A\ T_A|\mathrm{I_A}\rangle\ T_A\ \ |\mathrm{I_A}\rangle \\ &= |\psi^{(0)B}\rangle\otimes|\tilde{\psi}^{(1)A}_{T_A}\rangle. \end{aligned} \tag{A16}$$

This form for $|\tilde{\psi}^{(1)A+B}_{T_A}\rangle$ is crucial, as it has a product structure, implying that it will not suffer from any size consistency and separability issues. Indeed, the numerator of Eq. (A14) simply reduces to

$$\langle\psi^{(0)A+B}|H_A+H_B\,|\tilde{\psi}^{(1)A+B}_{T_A}\rangle = \langle\psi^{(0)A}|H_A|\tilde{\psi}^{(1)A}_{T_A}\rangle, \tag{A17}$$

and the denominator of the same Eq. (A14) is then

$$\begin{aligned} \Delta E^{(0)A+B}_{T_A} &= E^{(0)A+B} - \frac{\langle\tilde{\psi}^{(1)A}_{T_A}|H_A|\tilde{\psi}^{(1)A}_{T_A}\rangle}{\langle\tilde{\psi}^{(1)A}_{T_A}|\tilde{\psi}^{(1)A}_{T_A}\rangle} - E^{(0)B} \\ &= E^{(0)A} - \frac{\langle\tilde{\psi}^{(1)A}_{T_A}|H_A\,|\tilde{\psi}^{(1)A}_{T_A}\rangle}{\langle\tilde{\psi}^{(1)A}_{T_A}|\tilde{\psi}^{(1)A}_{T_A}\rangle} \\ &= \Delta E^{(0)A}_{T_A}, \end{aligned} \tag{A18}$$

and therefore,

$$e^{(2)A+B}_{T_A} = e^{(2)A}_{T_A}. \tag{A19}$$

Consequently, the JM-MRPT2 is strictly separable provided that a partition of the Hamiltonian in terms of H_A and H_B can be done, which supposes local orbitals.

2. Multi-partitioning of the Hamiltonian

In contrast with the CIPSI or shifted-B_k approaches, a given perturber determinant $|\mu\rangle$ has as much zeroth-order energies as reference determinants $|\mathrm{I}\rangle$ with which it interacts (i.e., $\langle\mu|H|\mathrm{I}\rangle\neq0$) within the JM-MRPT2 framework. This formally implies that the zeroth-order Hamiltonian depends on the reference determinant $|\mathrm{I}\rangle$, just as in the MUPA approach. The present paragraph proposes to briefly highlight the link existing between these two approaches.

Using the JM ansatz for the wave function [see Eq. (53)] and projecting the Schrödinger equation onto a given perturber $|\mu\rangle$ lead to

$$\sum_{\mathrm{I}} c_{\mathrm{I}}\left(\langle\mu|H|\mathrm{I}\rangle + \sum_{\mu'}\langle\mu|H|\mu'\rangle\, t_{\mathrm{I}\mu'}\right) + \mathcal{R} = E\sum_{\mathrm{I}} c_{\mathrm{I}}t_{\mathrm{I}\mu}, \tag{A20}$$

where $\mathcal{R}$ contains all terms in the coupled cluster equation containing higher or equal powers of T_{I} than $(T_{\mathrm{I}})^2$. Retaining all terms of first order in $t_{\mathrm{I}\mu'}$ in Eq. (A20) leads to the equations of linearized coupled cluster type

$$\sum_{\mathrm{I}} c_{\mathrm{I}}\left(\langle\mu|H|\mathrm{I}\rangle + \sum_{\mu'}\langle\mu|H|\mu'\rangle\, t_{\mathrm{I}\mu'}\right) = e^{(0)}\sum_{\mathrm{I}} c_{\mathrm{I}}t_{\mathrm{I}\mu}, \tag{A21}$$

which can be written as

$$\sum_{\mathrm{I}} c_{\mathrm{I}}\left(\langle\mu|H|\mathrm{I}\rangle + \sum_{\mu'}\langle\mu|H|\mu'\rangle\, t_{\mathrm{I}\mu'} - e^{(0)}t_{\mathrm{I}\mu}\right) = 0. \tag{A22}$$

Just as in the spirit of the UGA-SSMPRT2 of Mukherjee *et al.*, Eq. (A22) is solved independently for all references $|\mathrm{I}\rangle$, leading to

$$\langle\mu|H|\mathrm{I}\rangle + t_{\mathrm{I}\mu}\,\langle\mu|H|\mu\rangle + \sum_{\mu'\neq\mu}\langle\mu|H|\mu'\rangle\, t_{\mathrm{I}\mu'} = e^{(0)}t_{\mathrm{I}\mu}. \tag{A23}$$

As each equation is solved independently, one can use a different partitioning of the Hamiltonian according to the reference determinants $|\mathrm{I}\rangle$,

$$\begin{aligned} H &= H^{(0)}_{\mathrm{I}} + \lambda V_{\mathrm{I}}, \\ H^{(0)}_{\mathrm{I}} &= e^{(0)}\,|\mathrm{I}\rangle\langle\mathrm{I}| + \sum_{\mu} e^{(0)}_{\mathrm{I}\mu}\,|\mu\rangle\langle\mu|. \end{aligned} \tag{A24}$$

In Eq. (A24), retaining all terms at first order in λ leads to

$$\langle\mu\,|V_{\mathrm{I}}\,|\mathrm{I}\rangle + t^{(1)}_{\mathrm{I}\mu}\,\langle\mu\,|H^{(0)}_{\mathrm{I}}\,|\mu\rangle + \sum_{\mu'\neq\mu}\langle\mu\,|H^{(0)}_{\mathrm{I}}|\,\mu'\rangle\, t^{(1)}_{\mathrm{I}\mu'} = e^{(0)}t^{(1)}_{\mathrm{I}\mu}. \tag{A25}$$

By defining the zeroth-order energies $e_{\mathrm{I}\mu}^{(0)}$ as

$$e_{\mathrm{I}\mu}^{(0)} = \frac{\langle \tilde{\psi}_{T_{\mathrm{I}\mu}}^{(1)} | H^D | \tilde{\psi}_{T_{\mathrm{I}\mu}}^{(1)} \rangle}{\langle \tilde{\psi}_{T_{\mathrm{I}\mu}}^{(1)} | \tilde{\psi}_{T_{\mathrm{I}\mu}}^{(1)} \rangle}, \tag{A26}$$

where $|\tilde{\psi}_{T_{\mathrm{I}\mu}}^{(1)}\rangle$ is defined in Eq. (15) with the excitation operator $T_{\mathrm{I}\mu}$ which connects $|\mathrm{I}\rangle$ and $|\mu\rangle$ [see Eq. (6)], one can recover the expression of the amplitudes used in the JM-MRPT2 approach

$$t_{\mathrm{I}\mu}^{(1)} = \frac{\langle \mu | H | \mathrm{I} \rangle}{e^{(0)} - e_{\mathrm{I}\mu}^{(0)}}. \tag{A27}$$

Also, one can notice that the zeroth-order energies of the MUPA and JM-MRPT2 methods coincide for the 2h2p but also for the 1h2p and 2h1p. Indeed, the energy denominators appearing in the two latter classes imply the generalization of ionization potential [see Eq. (43)] and electronic affinities whose definition is identical in the MUPA and JM-MRPT2 methods. Therefore, in the case of the double excitations amplitudes, one can see the JM-MRPT2 method as the generalization of the MUPA method to all possible operations appearing in the active space for the definition of energy denominators.

[1] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[2] J. Goldstone, Proc. R. Soc. London Ser. A **239**(1217), 267 (1957).
[3] F. Coester, Nucl. Phys. **7**, 421 (1958).
[4] F. Coester and H. Kummel, Nucl. Phys. **17**, 477 (1960).
[5] J. Cizek, J. Chem. Phys. **45**, 4256 (1966).
[6] R. J. Bartlett and G. Purvis, Phys. Scr. **21**, 255 (1980).
[7] R. J. Bartlett, J. D. Watts, and L. Noga, Chem. Phys. Lett. **165**, 513 (1990).
[8] R. J. Bartlett, J. D. Watts, and L. Noga, Chem. Phys. Lett. **167**, 609 (1990).
[9] B. H. Brandow, Rev. Mod. Phys. **39**, 771 (1967).
[10] D. Hegarty and M. Robb, Mol. Phys. **37**, 1445 (1979).
[11] Ph. Durand and J. P. Malrieu, *Advances in Chemical Physics* (Wiley, New York, 1987), Vol. 67, p. 321.
[12] S. Evangelisti, J. P. Daudey, and J. P. Malrieu, Phys. Rev. A **35**, 4930 (1987).
[13] M. R. Hoffmann, D. Datta, S. Das, D. Mukherjee, A. Szabados, Z. Rolik, and P. R. Surjan, J. Chem. Phys. **131**, 204104 (2009).
[14] D. I. Lyakh, M. Musial, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).
[15] B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).
[16] S. Evangelisti, J. P. Daudey, and J. P. Malrieu, Chem. Phys. **75**, 91 (1983).
[17] M. Caffarel, E. Giner, A. Scemama, and A. Ramírez-Solís, J. Chem. Theory Comput. **10**, 5286–5296 (2014).
[18] E. Giner and C. Angeli, J. Chem. Phys. **143**, 124305 (2015).
[19] A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, J. Chem. Phys. **141**, 244110 (2014).
[20] E. Giner, A. Scemama, and M. Caffarel, J. Chem. Phys. **142**, 044115 (2015).
[21] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, J. Chem. Phys. **144**, 151103 (2016).
[22] E. Giner, R. Assaraf, and J. Toulouse, Mol. Phys. **114**(7–8), 910–920 (2016).
[23] J. P. Malrieu, P. Durand, and J. P. Daudey, J. Phys. A: Math. Gen. **18**, 809 (1985).
[24] H. Nakano, J. Nakatani, and K. Hirao, J. Chem. Phys. **114**, 1133 (2001).
[25] B. Kirtman, J. Chem. Phys. **75**, 798 (1981).
[26] K. Andersson, P. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[27] K. Andersson, P. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**(2), 1218 (1992).
[28] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J. P. Malrieu, J. Chem. Phys. **114**, 10252 (2001).
[29] C. Angeli, R. Cimiraglia, and J. P. Malrieu, Chem. Phys. Lett. **350**, 297 (2001).
[30] C. Angeli, R. Cimiraglia, and J. P. Malrieu, J. Chem. Phys. **117**, 9138 (2002).
[31] K. G. Dyall, J. Chem. Phys. **102**, 4909 (1995).
[32] A. Sokolov and G. K.-L. Chan, J. Chem. Phys. **144**, 064102 (2016).
[33] J. Finley, P.-A. Malmqvist, B. O. Roos, and L. Serrano-Andrés, Chem. Phys. Lett. **288**, 299 (1998).
[34] C. Angeli, S. Borini, M. Cestari, and R. Cimiraglia, J. Chem. Phys. **121**(9), 4043–4049 (2004).
[35] J. L. Heully, J. P. Malrieu, and A. Zaitsevskii, J. Chem. Phys. **105**, 6887 (1996).
[36] P. Ghosh, S. Chattopadhyay, D. Jana, and D. Mukherjee, Int. J. Mol. Sci. **3**, 733 (2002).
[37] U. S. Mahapatra, B. Datta, and D. Mukherjee, Chem. Phys. Lett. **299**, 42 (1999).
[38] U. S. Mahapatra, B. Datta, and D. Mukherjee, J. Phys. Chem. A **103**, 1822 (1999).
[39] A. Sen, S. Sen, P. K. Samanta, and D. Mukherjee, J. Comput. Chem. **36**, 670–688 (2015).
[40] S. Sharma and A. Alavi, J. Chem. Phys. **143**, 102815 (2015).
[41] S. Sharma, G. Jeanmairet, and A. Alavi, J. Chem. Phys. **144**, 034103 (2016).
[42] G. Jeanmairet, S. Sharma, and A. Alavi, J. Chem. Phys. **146**, 044107 (2017).
[43] B. Jeziorski and H. Monkhorst, Phys. Rev. A **24**, 1668 (1981).
[44] A. Zaitsevskii and J. P. Malrieu, Chem. Phys. Lett. **233**, 597 (1995).
[45] A. Zaitsevskii and J. P. Malrieu, Chem. Phys. Lett. **250**, 366 (1996).
[46] A. Zaitsevskii and J. P. Malrieu, Theor. Chem. Acc. **96**, 269 (1997).
[47] G. Li Manni, D. Ma, F. Aquilante, J. Olsen, and L. Gagliardi, J. Chem. Theory Comput. **9**, 3375–3384 (2013).
[48] Z. Gershgorn and I. Shavitt, Int. J. Quantum Chem. **2**, 751 (1968).
[49] D. C. Rawlings and E. R. Davidson, Chem. Phys. Lett. **98**, 424–427 (1983).
[50] H. Nakano, J. Chem. Phys. **99**, 7983 (1993).
[51] S. Wilson and I. Hubac, *Brillouin-Wigner Methods for Many-Body Systems* (Springer Science & Business Media, 2009), ISBN: 978-90-481-3373-4.
[52] J. Miralles, O. C. Castell, R. Caballol, and J. P. Malrieu, Chem. Phys. **172**, 33–43 (1993).
[53] C. Daday, S. Smart, G. Booth, A. Alavi, and C. Filippi, J. Chem. Theory Comput. **8**, 4441–4451 (2012).
[54] A. Scemama, T. Applencourt, Y. Garniron, E. Giner, G. David, and M. Caffarel (2016). "Quantum package v1.0," Zenodo. http://dx.doi.org/10.5281/zenodo.200970, https://github.com/LCPQ/quantum_package.
[55] J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, "General atomic and molecular electronic-structure system," J. Comput. Chem. **14**, 1347–1363 (1993).
[56] F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitonak, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, "MOLCAS 7: The next generation," J. Comput. Chem. **31**, 224–247 (2010).
[57] P. O. Widmark, P. A. Malmqvist, and B. Roos, Theor. Chim. Acta **77**, 291 (1990).
[58] F. A. Evangelista and J. Gauss, J. Chem. Phys. **134**(11), 114102 (2011).
[59] M. Hanauer and A. Kohn, J. Chem. Phys. **134**(20), 204111 (2011).
[60] M. Hanauer and A. Kohn, J. Chem. Phys. **136**(20), 204107 (2012).
[61] M. Hanauer and A. Kohn, J. Chem. Phys. **137**(13), 131103 (2012).
[62] F. A. Evangelista, A. C. Simmonett, H. F. Schaefer III, D. Mukherjee, and W. D. Allen, Phys. Chem. Chem. Phys. **11**, 4728–4741 (2009).
[63] C. Angeli, J. Comput. Chem. **30**, 1319–1333 (2009).
[64] C. Angeli, Int. J. Quantum Chem. **110**(13), 2436–2447 (2010).
[65] R. Cimiraglia, J. P. Malrieu, M. Persico, and F. Spiegelmann, J. Phys. B: At. Mol. Phys. **18**, 3073–3084 (1985).
[66] E. Giner, G. David, A. Scemama, and J. P. Malrieu, J. Chem. Phys. **144**, 064101 (2016).

Appendix B

Résumé français

Contents

B.1 Introducion

La chimie quantique est un domaine qui nécessite d'effectuer des calculs de plus en plus coûteux. En ce qui concerne les méthodes de fonction d'onde, qui sont l'objet de cette thèse, le scaling varie entre $\mathcal{O}(N^5)$ et $\mathcal{O}(N^8)$, avec N le nombre d'électrons du système. Ce scaling très élevé nécessite à la fois la mise en place d'approximations permettant de le réduire, et la conception d'algorithmes efficaces capables de tirer avantage

des architectures informatiques modernes. C'est sur ce second aspect que les présents travaux portent plus particulièrement.

La plupart des codes de chimique quantique encore utilisées actuellement (Molpro[3], Molcas[4], or Gaussian[5]…), ont débuté leur développement dans les années 90. À cette époque, l'augmentation de la vitesse de calcul était liée à l'augmentation de fréquence des processeurs. Toutefois, depuis une douzaine d'années, celle-ci se heurte à des barrières physiques difficilement franchissables. En conséquence, les accès mémoire, voir disque, ont vu leurs coûts relatifs augmenter,[8] et sont devenu le goulot d'étranglement ; de plus, la réduction du temps d'exécution devant dorénavant passer par la multiplication du nombre d'unités de calcul, les algorithmes doivent etre repensés pour des architectures parallèles.[7] De multiples groupes travaillent actuellement à moderniser les codes traditionnels de chimie quantique, jusque-là pensés dans un mode séquentiel. Cette thèse s'inscrit dans cette démarche.

Certaines méthodes sont de part leur conception même adaptées aux architectures parallèles, en particulier les méthodes de type Monte-Carlo (stochastiques), qui sont par nature composées d'une multitudes de tâches indépendantes, ce qui en rend la parallélisation facile et efficace (*embarassingly parallel*). De plus, elles permettent généralement de déterminer de manière approchée des valeurs dont le calcul exacte est excessivement coûteux. Une partie du travail a consisté à intégrer un aspect Monte-Carlo aux méthodes traditionnelles d'*interaction de configuration* (IC).

Le Quantum Package [12] développé au LCPQ est une suite de codes de méthodes de fonction d'onde, dont l'objectif premier n'est pas d'être utilisé massivement en production, mais plutôt de permettre le développement et l'expérimentation de nouvelles méthodes de manière simple, y compris pour ce qui est de l'aspect parallèle. Pour cette raison, la base du code est de type *determinant-driven*, c'est à dire itérant sur des déterminants, contrairement à l'approche plus habituelle qui consiste à itérer sur des intégrales biélectroniques (on parle alors d'approche *integral-driven*). Bien que l'approche determinant-driven soit typiquement moins efficace - le nombre de déterminants étant généralement bien supérieur au nombre d'integrales - elle est également moins complexe et plus flexible,[13] ce qui rejoint les objectifs du Quantum Package . Si elle s'avère mieux adaptée aux architectures parallèles, elle pourrait connaître un regain de popularité.

La première étape a été l'accélération et la parallélisation de la diagonalisation de Davidson, qui est un point central de toute méthode d'IC.

Par la suite, il a fallu améliorer l'algorithme de séléction de déterminants utilisé par le Quantum Package pour bâtir des fonctions d'onde compactes. En résumé, cet algorithme de séléction appelé *Configuration Interaction using a Perturbative Selection (CIPSI)*,[11] consiste à intégrer progressivement à une fonction d'onde variationnelle les déterminants externes avec lesquelles elle interagit le plus.

Les améliorations importantes qui ont été apportées à cet algorithme sont en eux-

même le résultat le plus important de ce travail, mais ont également servi de base aux travaux subséquents. En effet, implémenter cette méthode de manière efficace soulève le problème fondamentale de connecter la fonction d'onde à l'espace externe, c'est à dire d'accéder aux informations qu'elle ne contient pas directement.

La problématique suivante a été de tirer parti au maximum de l'information fournie par l'algorithme CIPSI.

L'une de ces informations est E_{PT2} la contribution perturbative au second ordre, si liée que son calcul est parfois confondu avec l'algorithme CIPSI en tant que sélection. E_{PT2} fournit une approximation de l'énergie de corrélation "perdue" de part la troncature de la fonction d'onde ; de ce fait, si elle est ajoutée à l'énergie variationnelle, elle donne une approximation de l'énergie Full-CI, soit l'énergie exacte du système pour une base donnée. En effet, le critère utilisé par CIPSI pour sélectionner de nouveaux déterminants, est leurs contribution perturbative au second ordre, calculée pour chaque déterminant externe (approximations mises à part) ; la somme de ces contributions n'est nul autre que E_{PT2}, qui peut donc, en principe, être calculée au cours de l'algorithme de sélection. Toutefois, en pratique, la sélection peut subir des approximations bien plus drastiques que le calcul de E_{PT2}, car identifier les contributions les plus importantes peut se passer d'explorer des espaces de déterminants externes dans lesquelles les contributions seront prévisiblement petites, ou même de calculer les contributions avec précision. Calculer la somme des contributions, en revanche, peut difficilement se passer d'être précis et exhaustif, de part l'effet de masse.

Pour cette raison, le calcul de E_{PT2} - tel qu'implémenté, comme un "sous-produit" de la sélection - était bien plus coûteux que la sélection, et souvent trop coûteux, ce qui conduisait en pratique à tronquer le calcul. Ce problème à pu être réglé par l'intégration d'un aspect Monte-Carlo, afin d'en retirer les bénéfices habituels, à savoir un résultat d'une précision acceptable pour une fraction du coût, et une parallélisation relativement simple et efficace.

Nous avons ensuite pu déplorer que notre calcul de E_{PT2}, alors qu'il fournit des informations détaillées sur les interactions entre la fonction d'onde et l'espace externe, ne permette qu'une correction globale de l'énergie, et pas de la fonction d'onde elle-même. En se basant sur la méthode dite shifted-B_k et l'utilisation de matrices habillées,[67, 68, 69, 70, 71, 72, 73] la fonction d'onde a pu être corrigée en fonction des informations obtenues lors du calcul de E_{PT2} ; puis de manière plus générale, un système permettant de raffiner la fonction d'onde sous l'effet d'un espace externe estimé stochastiquement a été mis en place. Ce système a été testé avec l'espace externe impliqué par l'approche shifted-B_k (où les coefficients des déterminants externes sont estimés perturbativement) et par une approche de type MR-CCSD développée précédemment.

Les considérations techniques de ces implémentations n'ont bien sûr pas été abordées en détails dans les différents articles produits au cours de cette thèse. En ce qui

me concerne, mon travail a porté sur les implémentations au moins autant que sur la théorie sous-jacente, c'est pourquoi ce manuscrit est une opportunité d'aborder les questions algorithmiques plus en détail. Dans la mesure où ces questions pourraient être d'un interet particulier pour ceux cherchant à comprendre en profondeur l'implémentation du QUANTUM PACKAGE , j'ai choisi de le rédiger en anglais.

B.2 Calcul *determinant-driven* des éléments de matrice de $\widehat{H}$

Un déterminant de Slater (simplement appelé "déterminant" dans la suite du texte) peut être vu comme un ensemble d'opérateurs de création agissant sur le vide.

$$a_i^\dagger a_j^\dagger a_k^\dagger |\rangle = |I\rangle \tag{B.1}$$

De part la nature fermionique des électrons, permuter deux opérateurs conduit à un changement de signe.

$$a_j^\dagger a_i^\dagger = -a_i^\dagger a_j^\dagger \tag{B.2}$$

$$a_j^\dagger a_i^\dagger a_k^\dagger |\rangle = -|I\rangle \tag{B.3}$$

On peut voir qu'un déterminant peut se décomposer en deux informations :

- L'ensemble des spinorbitales occupées.
- Un signe, ou "facteur de phase".

L'approche determinant-driven implique d'itérer sur des déterminants, et par conséquent, nécessite de calculer explicitement et de manière intensive des éléments de matrice de $\widehat{H}$ l'hamiltonien électronique non-relativiste. C'est ce que permettent les règles de Slater-Condon. Avec $|D\rangle$ un déterminant de Slater et $\left|D_{pq}^{rs}\right\rangle$ le déterminant obtenu à partir de $|D\rangle$ par la substitution des spinorbitales p et q par les spinoribtales r et s ; en d'autres termes par l'action de l'opérateur d'excitation $\hat{T}_{pq}^{rs}$:

$$\langle D|\hat{H}|D\rangle = \sum_{i\in|D\rangle} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i\in|D\rangle}\sum_{j\in|D\rangle}\Big[(ii|jj) - (ij|ij)\Big] \tag{B.4}$$

$$\left\langle D\middle|\hat{H}\middle|D_p^r\right\rangle = \langle p|\hat{h}|r\rangle + \sum_{i\in|D\rangle}\Big[(pr|ii) - (pi|ri)\Big] \tag{B.5}$$

$$\left\langle D\middle|\hat{H}\middle|D_{pq}^{rs}\right\rangle = (pr|qs) - (ps|qr) \tag{B.6}$$

$$\left\langle D\middle|\hat{H}\middle|D_{pqt\ldots}^{rsu\ldots}\right\rangle = 0 \tag{B.7}$$

avec $\hat{h}$ la partie mono-électronique (énergie cinétique et potentiel électron-noyau),

$$\langle p|\hat{h}|r\rangle = \int d\mathbf{x}\ \phi_p^*(\mathbf{x})\left(-\frac{1}{2}\nabla + V_1(\mathbf{x})\right)\phi_h(\mathbf{x}), \tag{B.8}$$

$i \in |D\rangle$ signifiant que la spinorbitale i est occupée dans le déterminant $|D\rangle$, et

$$(ij|kl) = \int d\mathbf{x}_1 \int d\mathbf{x}_2\ \phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\phi_k^*(\mathbf{x}_2)\phi_l(\mathbf{x}_2) \tag{B.9}$$

une intégrale biélectronique. Ainsi qu'on le voit, ces calculs impliquent :

- D'être capable de déterminer l'excitation liant deux déterminants.
- De pouvoir accéder rapidement aux intégrales biélectroniques correspondantes.

Les intégrales biélectroniques étant potentiellement trop nombreuses pour être directement indicées à partir des indices d'orbitale qui la définissent - cela nécessiterait un stockage de l'ordre de ${N_{\text{orb}}}^4$ avec N_{orb} le nombre d'orbitales - une table de hash est un choix naturel pour stocker et accéder aux éléments non-nuls en temps constant. Une table de hash "maison" spécifiquement conçue pour les intégrales est implémentée dans le Quantum Package . Dans le cas où les intégrales recherchées sont aléatoires, elle permet un accès seulement deux fois plus lent qu'un accès direct dans un tableau à 4 dimensions. Cette différence s'accentue toutefois dans le cas où l'accès suit un certain schéma ; en pratique, un calcul CIPSI complet est de l'ordre de deux fois plus long si il utilise une table de hash plutôt qu'un accès direct. Pour atténuer ce problème, les intégrales impliquant uniquement les 128 orbitales les plus proches du niveau de Fermi sont stockées dans un tableau à 4 indices (2 Gio).

Les déterminants sont stockés avec la représentation dite *bitstring*. Cette approche est fondamentale dans l'implémentation du Quantum Package . Chaque spinorbitale est associée à un bit (contenu dans une variable de type integer 64 bits), qui prend la valeur de son nombre d'occupation. En d'autres termes 0 est associé au statut "inoccupé" et 1 au statut "occupé". Une variable integer de 64 bits peut donc stocker l'occupation de 64 spinorbitales. Le nombre de variables integer nécessaire pour stocker N_{orb} spinorbitales est

$$N_{\text{int}} = \left\lfloor \frac{N_{\text{orb}} - 1}{64} \right\rfloor + 1. \tag{B.10}$$

Pour des raisons pratiques, les spinorbitales $\uparrow$ et $\downarrow$ sont stockées sur des variables séparées. Par conséquent, la représentation interne d'un déterminant est un tableau à 2 dimensions, la dimension externe de taille 2 (spin $\uparrow$ et $\downarrow$), et l'interne de taille N_{int}.

Cette représentation ne permet toutefois pas de stocker le facteur de phase, par conséquent cette elle ne peut stocker qu'une "valeur absolue" de déterminant.

La représentation bitstring est une manière compacte de stocker des déterminants, mais c'est plus qu'une méthode de stockage ; en effet, elle permet de tirer parti de la capacité des processeurs à effectuer des opérations bit à bit de manière efficace. Plutôt que d'interpréter cette représentation comme une liste de nombres d'occupations, on peut l'interpréter comme la définition d'un ensemble.

On appel bitstring un tableau d'integers 64 bits de taille N_{int}, et d'une manière générale, il peut définir un ensemble d'orbitales ; les orbitales contenues dans l'ensemble sont celles dont le bit associé est non nul. La représentation d'un déterminant peut alors être vue comme une paire de bitstrings associés aux spinorbitales $\uparrow$ et $\downarrow$, respectivement, et donc définissant un ensemble de spinorbitales (en l'occurrence, les spinorbitales occupées). On appel de tels objet des $\uparrow\downarrow$-bitstrings.

```
! I est un updown-bitstring
! I_up et I_down sont des bitstrings

integer*8 :: I(N_int, 2)
integer*8 :: I_up(N_int), I_down(N_int)
...! On charge un determinant dans I
I_up   (:) = I(:,1)
I_down (:) = I(:,2)
```

Certaines instructions disponibles sur les processeurs modernes peuvent être ramenées à des opérations sur des ensembles. L'instruction AND par exemple (*bitwise AND*, fonction `IAND` en Fortran), correspond à l'intersection.

$$A = \texttt{IAND}(B, C) \tag{B.11}$$

Si on interpréte A, B et C comme des bitstrings, A définit l'intersection entre B et C. A titre d'exemple, en utilisant les instructions suivantes:

- $\text{POPCNT}(I)$: Retourne le nombre de bits non nuls dans un integer I.
  $\text{POPCNT}(00011000_2) = 2$.
- $\texttt{IEOR}(I, J)$: "ou exclusif" bit à bit.
  $\texttt{IEOR}(1100_2, 1010_2) = 0110_2$.
- $\texttt{IAND}(I, J)$: "et" bit à bit.
  $\texttt{IAND}(1100_2, 1010_2) = 1000_2$.

On peut déterminer le degré ainsi que les les trous et particules impliqués dans une excitation. Avec B et C les $\uparrow\downarrow$-bitstrings définissants deux déterminants $|B\rangle$ et $|C\rangle$ (par soucis de simplification on oublie leur caractère de tableau) :

$$d = \text{POPCNT}(\texttt{IEOR}(B, C)) \tag{B.12}$$

En toutes lettres, d est le nombre de spinorbitales qui sont occupées dans exactement l'un de $|B\rangle$ ou de $|C\rangle$, autrement dit celles dont l'occupation diffère ; une excitation impliquant un changement d'occupation dans 2 spinorbitales, le degré d'excitation entre $|B\rangle$ et $|C\rangle$ est de $d/2$.

Ensuite :

$$A = \texttt{IAND}(C, \texttt{IEOR}(B, C)) \tag{B.13}$$

Ici A est un $\uparrow\downarrow$-bitstring qui, en toutes lettres, contient les spinorbitales présentes dans $|C\rangle$ et dont l'occupation diffère entre $|B\rangle$ et $|C\rangle$. Autrement dit, les particules impliquées dans l'excitation $\hat{T}$ telle que $|C\rangle = \hat{T}\,|B\rangle$. De la même manière les trous impliqués dans $\hat{T}$ sont déterminés par $\texttt{IAND}(B, \texttt{IEOR}(B, C))$

Puisque notre représentation en $\uparrow\downarrow$-bitstring ne stock pas le signe d'un déterminant, lorsqu'une excitation est appliquée, il faut calculer un facteur de phase, qui peut être de 1 ou -1. Pour une excitation $\hat{T}_p^q$, ce facteur est lié à la parité du nombre d'électrons entre les spinorbitales p et q. Dans la mesure où le facteur de phase doit être calculé de nombreuses fois sur un même déterminant, une manière efficace de le faire consiste à stocker pour chaque spinorbitale, le nombre d'électrons dans les spinorbitales inférieures ; il suffit ensuite d'une soustraction pour connaître le nombre d'électrons entre deux spinorbitales. En pratique, il n'est pas nécessaire de stocker le nombre, mais seulement sa parité (connaissant les parités de A et B il est trivial de déterminer la parité de $A - B$) ; ce type de stockage est appelé *phase mask*.

B.3 Diagonalisation de Davidson

D'un point de vu algorithmique, la question posée se résume au calcul (mais pas au stockage) de la matrice hamiltonienne. Le calcul d'un élément de matrice peut être fait efficacement, mais le calcul de chacun des ${N_{\text{det}}}^2$ éléments (avec N_{det} le nombre de déterminants dans la fonction d'onde variationnelle) reste extrêmement coûteux.

Afin de réduire le nombre d'éléments à considérer, on peut créer des sous-ensembles de déterminants identifiables comme entièrement déconnectés de certains autres sous-ensembles. Ainsi, les déterminants peuvent par exemple être regroupés en fonction de leur partie de spin $\uparrow$. Si la partie $\uparrow$ qui définie un groupe, est plus que doublement excitée par rapport à la partie $\uparrow$ qui définie un autre groupe, on peut immédiatement en déduire qu'aucune connexion ne pourra être trouvée entre les déterminants de l'un et de l'autre groupe.

La version implémentée, pour une meilleur efficacité, disjoint les différents types d'excitation.

- Excitations $\uparrow\uparrow$ et $\uparrow$: les déterminants liés par une telle excitation partagent par définition la même partie de spin $\downarrow$. Les déterminants sont donc regroupés en fonction leur partie de spin $\downarrow$, et chaque déterminant n'a besoin d'être comparé

qu'aux déterminants du même groupe. Du fait le la faible taille des groupes, cette recherche est très peu coûteuse.

- Excitations $\downarrow\downarrow$ et $\downarrow$: situation symétrique avec la précedente, les déterminants sont donc groupés en fonction de leur partie de spin $\uparrow$.
- Excitations $\uparrow\downarrow$: la vaste majorité du temps de calcul est consacré à celles-ci. Les déterminants sont regroupés selon leur partie $\uparrow$ (arbitrairement). Une excitation $\uparrow\downarrow$ ne pourra être trouvée qu'entre deux groupes dont les parties $\uparrow$ qui les définissent sont simplement excitées l'une par rapport à l'autre.

B.4 Sélection avec le critère CIPSI

L'algorithme CIPSI consiste à construire itérativement la fonction d'onde variationnelle, en lui ajoutant à chaque itération les déterminants externes qui interagissent le plus avec elle.

L'itération n du CIPSI peut être décrite comme suit:

1. La fonction variationnelle $\left|\Psi^{(n)}\right\rangle$ est définie sur un ensemble de déterminants $\{|D_I\rangle\}^{(n)}$ dans lequel on diagonalise $\widehat{H}$

$$\left|\Psi^{(n)}\right\rangle = \sum_I c_I^{(n)} |D_I\rangle \tag{B.14}$$

2. Pour chaque déterminant $|\alpha\rangle$ dit *externe*, $|\alpha\rangle \notin \{|D_I\rangle\}^{(n)}$, on calcul la contribution perturbative

$$e_\alpha = \frac{\left\langle \Psi^{(n)} \middle| \hat{H} \middle| \alpha \right\rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}. \tag{B.15}$$

   $E^{(n)}$ dépend de la théorie de perturbation utilisée (dans notre cas, Epstein-Nesbet, $E^{(n)}$ correspond à l'énergie variationnelle de $\left|\Psi^{(n)}\right\rangle$. Toutefois une autre théorie pourrait être utilisée).

3. On extrait $\{|\alpha_\star\rangle\}^{(n)}$ le sous-ensemble des déterminants $|\alpha\rangle$ de plus grande contribution e_α, et on les ajoute à la fonction variationnelle.

$$\{|D_I\rangle\}^{(n+1)} = \{|D_I\rangle\}^{(n)} \cup \{|\alpha_\star\rangle\}^{(n)} \tag{B.16}$$

4. On passe à l'itération $n+1$ si un critère d'arrêt n'est pas atteint.

L'ancienne implémentation réalisait cet algorithme de manière relativement naïve, chaque déterminant externe étant généré individuellement puis comparé à chaque déterminant de la fonction d'onde variationnelle. Le très important gain de performance repose sur deux améliorations:

- Le filtrage des déterminants internes. Les déterminants externes sont créés en prenant un déterminant interne, qualifié de *générateur*, et en lui appliquant toutes des simples et doubles excitations possibles, en d'autres termes en créant tous les déterminants qui lui sont connectés. Chacun des déterminant externe ainsi généré, doit être comparé à chaque déterminant interne pour le calcul de e_α. Ainsi, pour qu'un déterminant interne soit connecté à un déterminant externe, il ne peut pas être "éloigné" du générateur par plus de 4 excitations. Par conséquent, tous les déterminants plus que quadruplement excités par rapport au générateur courant, peuvent être ignorés dans le calcul des e_α. Ce filtrage peut être raffiné quand le générateur devient un générateur doublement ionisé, ainsi que sera discuté dans le point suivant.

- La détermination "par batch" des connexions. Plutôt que de considérer les déterminants externes "un par un", l'unité de base est un générateur doublement ionisé $\left|G_{pq}\right\rangle$, soit $\left|G\right\rangle$ ionisé dans les spinorbitales p et q. Il implique les déterminants $\hat{T}_{pq}^{rs}\left|G\right\rangle$ pour toutes les valeurs de r et s ; en effet, comparer $\left|G_{pq}\right\rangle$ à un déterminant interne, permet de déterminer systématiquement quelles sont les valeurs de r et s qui correspondent à un déterminant connecté au dit déterminant interne. De même qu'on a pu filtrer certains déterminants internes en fonction du générateur, on peut en filtrer en fonction des ionisations p puis q.

On peut tenter de résumer la différence entre l'ancienne et la nouvelle implémentation à l'aide des figures B.1 et B.2.

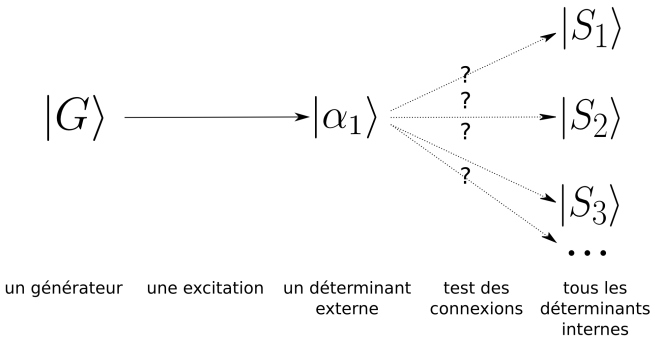


Figure B.1: Ancienne implémentation simple du CIPSI (représentation incomplète).

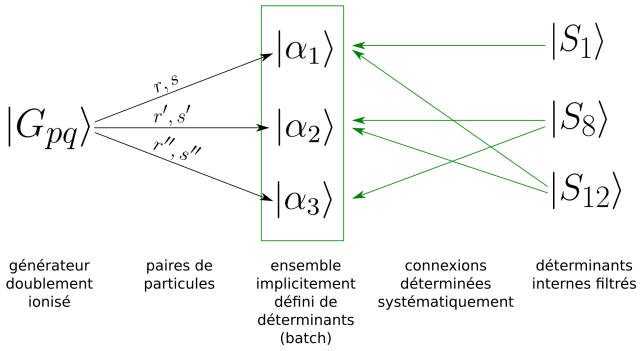


Figure B.2: Implémentation actuelle du CIPSI (représentation incomplète).

B.5 Calcul de la contribution perturbative au second ordre

La contribution perturbative au second ordre, qui donne accès à une valeur approximative de l'énergie Full-CI, est essentiellement la somme des contributions individuelles calculées au cours de l'algorithme CIPSI. Cependant son calcul est bien plus coûteux que celui du CIPSI, car moins d'approximations sont possibles. On peut également noter que la valeur exacte de E_{PT2} ne présente que peu d'intérêt, dans la mesure où elle ne sert que comme approximation de l'énergie Full-CI. De ce fait, un algorithme hybride stochastique/déterministe a été implémenté, donnant accès à E_{PT2} avec une précision suffisante pour une fraction du coût du calcul complet. La contribution élémentaire n'est pas la contribution d'un seul déterminant externe, mais la somme pour 1 déterminant interne des contributions de tous les déterminants externes pouvant être générés à partir de lui mais d'aucun déterminant de coefficient supérieur en valeur absolue. La raison est d'abord technique, la somme des contributions de déterminants "proches" (en l'occurrence car tous connectés à un générateur particulier) peut être calculée efficacement. Les conséquences de ce regroupement sont:

- Le nombre de contributions élémentaires est de N_{det}, et donc assez faible pour que chacune soit stockée en mémoire. Ainsi, contrairement au cas général dans un calcul Monte-Carlo, quand un élément est tiré, sa valeur peut être stockée et simplement ré-utilisée si cet élément est tiré à nouveau. On peut noter que la valeur exacte de E_{PT2} sera ainsi connue quand chaque élément aura été tiré, et donc pour un coût presque égale à celui du calcul exacte déterministe.
- Les valeurs absolues de ces contributions élémentaires décroissent rapidement avec le coefficient du générateur associé.

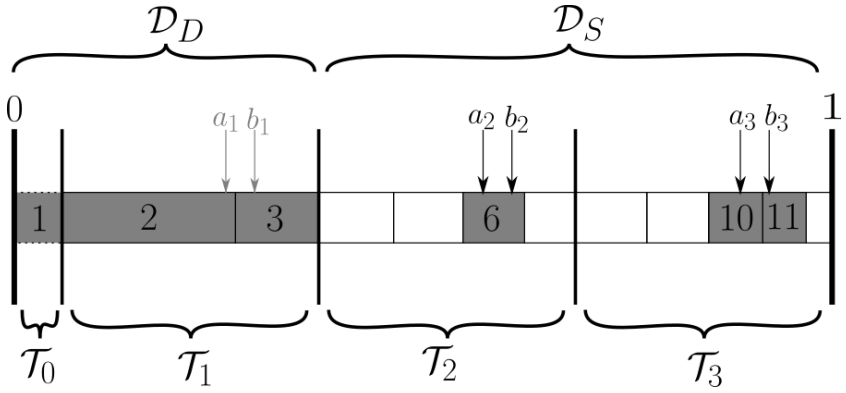
Ce deuxième point guide la manière dont est conduit le calcul Monte-Carlo.

On tri les déterminants internes par coefficients absolus décroissants, par conséquent la majeure partie de la contribution est contenue dans le “début” de la fonction d’onde ainsi trié. Dans un premier temps on partage les déterminants en un intervalle déterministe $\mathcal{D}_D$ et un intervalle stochastique $\mathcal{D}_S$.

La partie déterministe, qui regroupe les contributions les plus importantes, est calculée entièrement. De ce fait la barre d’erreur ne porte que sur les plus petites contributions, ce qui permet de la réduire considérablement. La partie stochastique est séparée en intervalles appelés "dents" et notés $\mathcal{T}_1$, $\mathcal{T}_2$ ... chacune contenant des contributions globalement plus faibles que celles de la dent précédente. Les valeurs qui serviront d’échantillon au sens statistique du terme, que l’on appel des *peignes*, seront des sommes d’un déterminant tiré dans chaque dent ; en bref, la somme de “un grand, un moyen et un petit” aura en toute logique une variance plus faible que la somme de “trois au hasard”. Lorsque toutes les contributions d’une dent ont été calculées, cette dent est déplacée dans la partie déterministe $\mathcal{D}_D$, ce qui permet de réduire encore la fraction des contributions sur laquelle porte la barre d’erreur.

La figure ci-dessous résume ce procédé. Chaque case correspond à un déterminant générateur, et sa largeur à la probabilité que ce déterminant soit tiré. La dent $\mathcal{T}_0$ est spéciale et fait toujours partie de $\mathcal{D}_D$, par conséquent aucun tirage n’est fait dedans. On a ici tiré deux peignes a et b, dont les valeurs sont la somme des contributions des générateurs désignés par les flèches a_1, a_2, a_3 et b_1, b_2, b_3 respectivement.

Comme toutes les contributions de $\mathcal{T}_1$ ont été calculées, elles sont déplacées dans $\mathcal{D}_D$, et les générateurs 2 et 3 ne sont plus utilisés dans l’estimation stochastique, qui se fait uniquement avec les générateurs 6, 10 et 11.



B.6 Habillage stochastique de matrice

Le calcul d’un habillage de matrice repose sur le même principe que le calcul de E_{PT2} ; la valeur recherchée est une somme dont chaque élément est associé à un déterminant externe.

De la même manière que pour le calcul de E_{PT2}, nous allons ici regrouper les contributions en N_{det} ensembles chacun associé à un déterminant interne, ce qui nous donne N_{det} contributions élémentaires. Le chapitre "Stochastic matrix dressing" porte essentiellement sur la manière dont l'estimation est calculée à partir des contributions élémentaires ; le calcul de ces contributions élémentaires elles-mêmes, dans un cadre général, est développé dans le chapitre suivant, "Application of stochastic matrix dressing to MR-CCSD".

Un changement majeur est que ces contributions élémentaires, qui sont des scalaires dans le cas du calcul de E_{PT2}, deviennent des vecteurs de taille N_{det} dans le cas de l'habillage de matrice. Cela soulève une difficulté supplémentaire: il est possible de stocker N_{det} scalaires, pas N_{det} vecteurs de taille N_{det}. Comment alors éviter de recalculer une contribution si elle est tirée de multiples fois?

Cela a pu être réalisé par l'introduction de "checkpoints" pré-déterminés (d'une à quelques dizaines), en dehors desquels un résultat ne peut pas être obtenu. L'idée est que dans un schéma Monte-Carlo, même "exotique" comme notre schéma hybride déterministe/stochastique, le résultat estimé est une combinaison linéaire des échantillons. Avec $\boldsymbol{\delta}^m$ l'estimation à un moment m du calcul Monte-Carlo et $\boldsymbol{\delta}_I$ la contribution élémentaire liée au déterminant interne d'indice I:

$$\boldsymbol{\delta}^m = \sum_{I=1}^{N_{\text{det}}} \mu_I^m \boldsymbol{\delta}_I \tag{B.17}$$

On peut pré-calculer les coefficients μ_I^m de cette combinaison linéaire sans avoir accès à aucune contribution élémentaire. $\boldsymbol{\delta}^m$ peut être initialisé au vecteur nul, puis construit incrémentalement au fur et à mesure que les contributions élémentaires sont calculées. Lorsque la contribution $\boldsymbol{\delta}_I$ est calculée, le checkpoint m et tous les autres sont mis à jour:

$$\boldsymbol{\delta}^m \leftarrow \boldsymbol{\delta}^m + \mu_I^m \boldsymbol{\delta}_I \tag{B.18}$$

De cette manière la valeur $\boldsymbol{\delta}_I$ n'a pas besoin d'être stockée.

B.7 Application de l'habillage stochastique de matrice au MR-CCSD

L'habillage de matrice, de manière générale, permet de raffiner la fonction d'onde variationnelle sous l'effet d'un espace externe. Cet espace externe est défini par une fonction $Z(\alpha, \ldots)$, prenant en paramètre au minimum un déterminant externe, et retournant le coefficient à lui associer. Afin de rendre simple l'implémentation de n'importe quel espace externe, un framework a été crée, dans lequel il suffit de définir cette fonction ; l'espace externe correspondant est estimé stochastiquement selon la méthode développée au chapitre précédent, et la fonction d'onde variationnelle est modifiée sous

son effet. Dans le cas de la méthode shifted-B_k , dont il a été question dans le chapitre précédent, les coefficients externes sont estimés en perturbation. Dans ce chapitre, il s'agit des coefficients impliqués par la méthode MR-CCSD, qui avait précédemment été implémentée dans le cadre d'une publication.

De manière générale, la détermination d'un coefficient externe fait intervenir ses connexions avec l'espace variationnel. Pour des raisons de performance et de simplicité, il est vital que les déterminants internes auxquels se connecte le déterminant externe considéré, soient déterminés en amont de l'appel à la fonction Z - le contraire contraindrait à ré-explorer la fonction d'onde entière pour chaque déterminant externe.

Nous avons pu réutiliser la "machinerie" du calcul CIPSI. En effet, dans ce dernier, chaque déterminant externe est mis en rapport avec tous les déterminants internes auxquels il se connecte, chaque connexion étant un apport à sa contribution perturbative. Dans le cas présent, quand une connexion sera mise en évidence, plutôt que d'incrémenter la contribution perturbative associée du déterminant externe impliqué, on ajoutera le déterminant interne à une liste associée au déterminant externe. A terme cette liste contiendra donc tous les déterminants internes connectés (voir figure B.3).

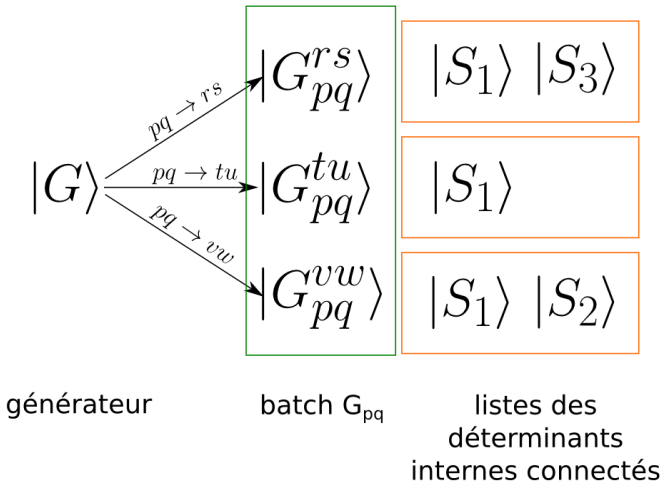
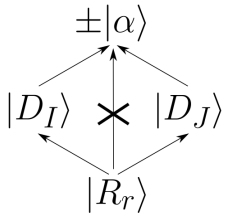


Figure B.3: Construction des listes de déterminants internes connectés pour tous les $|\alpha\rangle$ d'un batch G_{pq}.

Dans le cas particulier du MR-CCSD, le calcul de ces coefficients se fait en mettant en évidence, pour chaque déterminant externe, des structures en "losange".



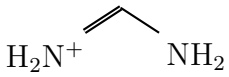
Avec $|\alpha\rangle$ le déterminant externe considéré, $|R_r\rangle$ un déterminant de la référence, $|D_I\rangle$ et $|D_J\rangle$ deux déterminants internes. Les flèches parallèles indiquent une connexion par la même excitation. La flèche verticale indique une absence de connexion (degré d'excitation supérieur ou égale à 3). On remarque que dans ce cas particulier, un losange peut être mis en évidence très simplement à l'aide de la fonction IEOR présentée précédemment, le critère étant:

$$\alpha \oplus D_I \oplus D_J \oplus R_r = 0 \tag{B.19}$$

Avec α, D_I ,D_J et R_r les $\uparrow\downarrow$-bitstrings définissants les déterminants correspondants, et $A \oplus B = \mathrm{IEOR}(A, B)$.

B.8 Mesures de performance

L'efficacité des implémentations est évaluée sur une cyanine,



dans l'état fondamental et le premier état excité, en base aug-cc-pVDZ avec les orbitales $1s$ des atomes C et N gelées.

La géometrie est celle de l'état fondamental, optimisée au niveau PBE0/cc-pVQZ. L'état fondamental est de type couche fermée, bien décrit en mono-référence. L'état excité est simplement excité, et requiert deux déterminants dans la réference $(1/\sqrt{2}(a\bar{b} + b\bar{a}))$.

L'espace Full-CI pour ce système est un CAS(18,111). L'énergie d'excitation de référence, obtenue au niveau CC3/ANO-L-VQZP, est de 7.18 eV.[84]

Les calculs ont été effectués sur le supercalculateur Olympe (CALMIP), chaque nœud est un dual-socket Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz avec 192Gio de RAM et 36 cœurs physiques. Dans la figure B.4, on trace la convergence des énergies de l'état fondamental et de l'état excité en fonction du nombre de déterminants, avec et sans la contribution perturbative au second ordre E_{PT2}. On observe que, bien que E_{PT2} soit relativement grand (~ 0.02 au), l'énergie d'excitation obtenue avec ou sans la correction perturbative est de 7.20 eV, ce qui est compatible avec l'énergie de référence obtenue dans une base plus grande.

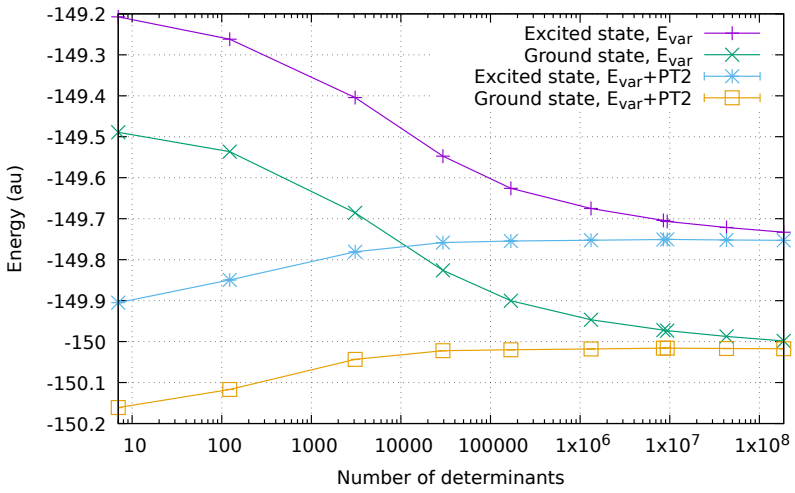


Figure B.4: Convergence de l'énergie de l'état fondamental et de l'état excité en fonction du nombre de déterminants dans l'espace variationnel.

B.8.1 Diagonalisation de Davidson

Nous mesurons le temps nécessaire à 1 itération de Davidson en fonction du nombre de déterminants dans la fonction d'onde variationnelle (figure B.5).

Le scaling obtenu correspond à ce qui est attendu, à savoir $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$.

Ensuite, en utilisant les deux plus grandes fonctions d'onde, nous mesurons le temps mural nécessaire au calcul en fonction du nombre de nœuds (figure B.6).

Dans la mesure où la communication croît en $\mathcal{O}(N_{\text{det}})$ alors que le calcul croît en $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$, l'efficacité parallèle augmente avec N_{det}.

Pour 50 nœuds avec la fonction d'onde à 42 959 496 déterminants, l'efficacité parallèle est de 76%

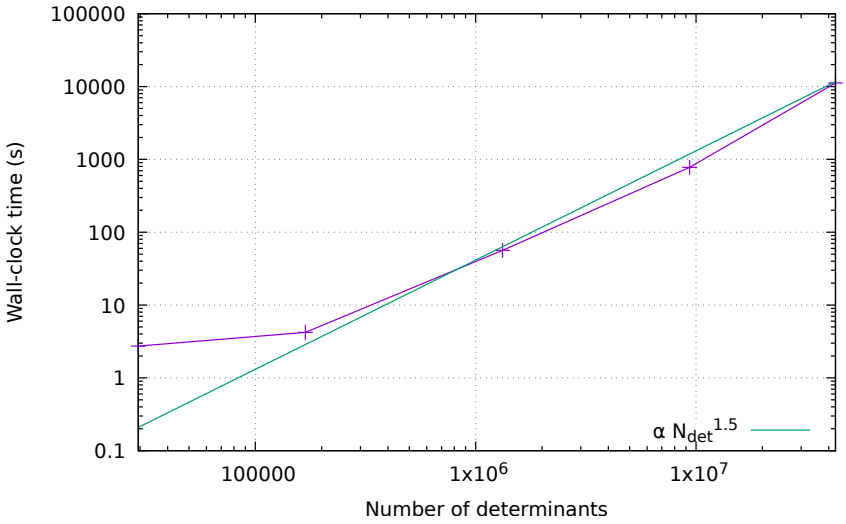


Figure B.5: Temps mural d'une itération de Davidson en fonction du nombre de détermiants dans la fonction d'onde.

B.8.2 CIPSI

Nous mesurons le temps nécessaire à une étape de sélection par CIPSI, en fonction du nombre de déterminants (figure B.7), puis du nombre de cœurs (figure B.8) avec la plus grande fonction d'onde (9 356 952 déterminants). L'accélération en fonction du nombre de déterminant est presque idéale ; toutefois, cela est dû à une approximation (le seuil n_g, non discuté dans ce résumé), qui fait que le nombre de déterminants générateurs tend à devenir constant. L'accélération en fonction du nombre de nœuds est également presque idéale, avec 95% d'efficacité parallèle pour 50 nœuds. Cela est en partie dû à la fragmentation qui permet des tâches équilibrées.

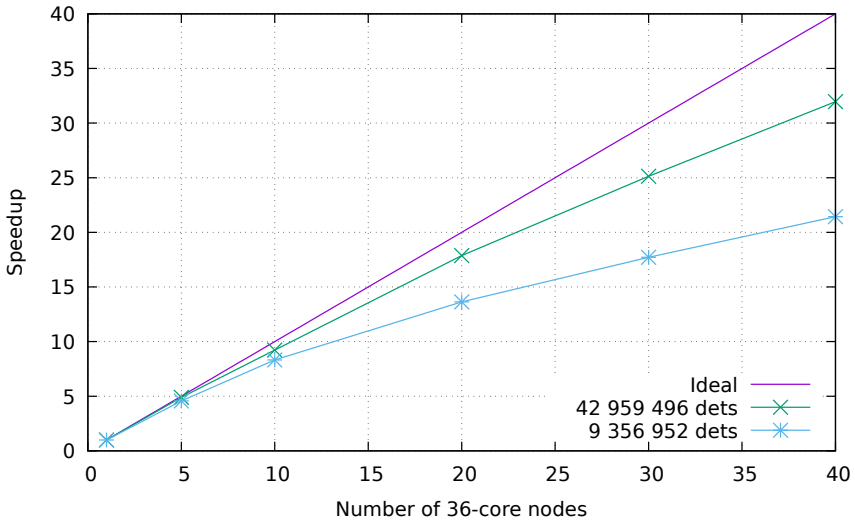


Figure B.6: Accélération pour une itération de Davidson en fonction du nombre de noeuds à 36 cœurs.

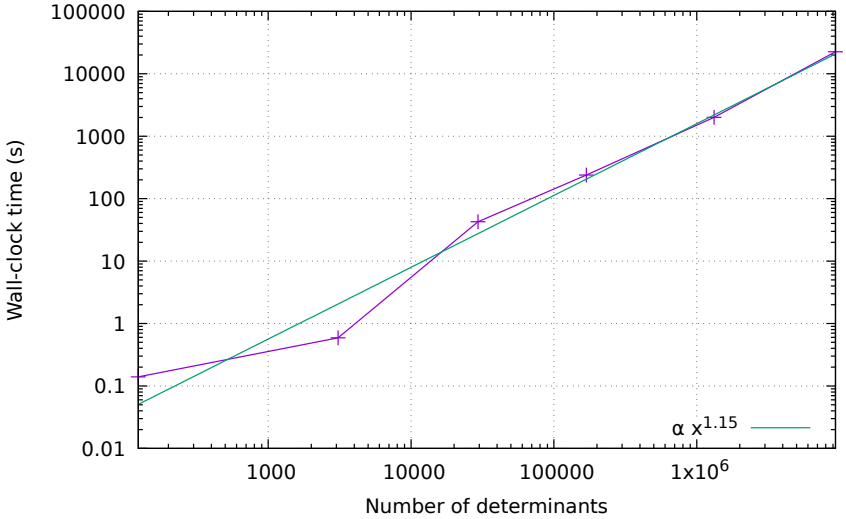


Figure B.7: Temps mural de la sélection CIPSI en fonction du nombre de déterminants dans la fonction d'onde.

B.8.3 Calcul de E_{PT2}

L'algorithme du calcul de E_{PT2} est très similaire à celui du CIPSI. On s'attend donc à un comportement similaire.

Le critère d'arrêt est une erreur relative de 1/1000. Puisque E_{PT2} diminu avec N_{det},

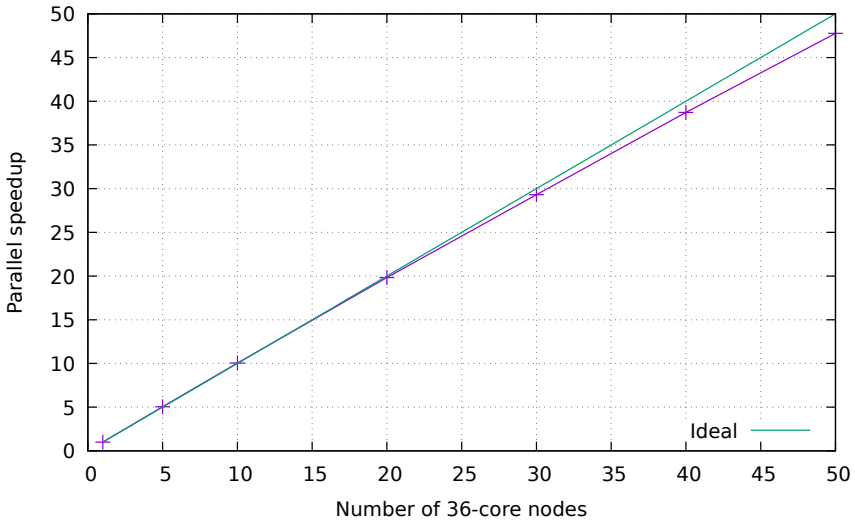


Figure B.8: Accélération parallèle pour la sélection CIPSI. La référence est un unique nœuds à 36 cœurs.

l'erreur acceptable diminue également avec N_{det}. Cependant, le coût du calcul stochastique par rapport au calcul complet reste relativement constant, autour de 5%.

Le scaling obtenu en fonction de N_{det} (figure B.9) est presque linéaire, $\mathcal{O}\left({N_{\text{det}}}^{1.15}\right)$, pour les plus grandes fonctions. Cela peut se comprendre, dans la mesure où pour un relativement petit nombre de déterminants, le nombre de déterminants externes produit est proportionnel à N_{det}, chacun potentiellement connecté à N_{det} déterminants internes, pour un scaling attendu de l'ordre de ${N_{\text{det}}}^2$. Toutefois, à mesure que la fonction variationnelle tend vers la fonction Full-CI, le nombre de déterminants internes auxquels un déterminant externe est connecté, est limité par le nombre de doubles excitations possibles, ce qui fait disparaître la seconde dépendance à N_{det}.

Le scaling obtenu en fonction du nombre de nœuds (figure B.10) est un peu moins satisfaisant que celui obtenu pour le CIPSI. L'efficacité parallèle avec 50 nœuds (1800 cœurs) est de 80%, contre 95% pour le CIPSI. On peut trouver deux raisons à cela:

- Le pré-calcul des peignes sur le processus maître délaye le début de la partie parallèle.
- Contrairement au CIPSI, le calcul est ici interrompu de manière imprévisible, lors que la barre d'erreur atteint le seuil requis. Par conséquent un certain nombre de tâches "superflues" peuvent être lancées, ce nombre étant d'autant plus grand que le nombre de cœurs est important ; dans la limite où N_{det} est égal au nombre de cœurs, c'est de fait toujours le calcul complet qui sera réalisé, alors que dans le mode mono-cœur aucune tâche superflue ne sera calculée.

Les tâches qualifiées de superflues pouvant néanmoins être utilisées pour réduire la barre d'erreur, ce calcul d'accélération n'est pas tout à fait pertinent car la barre d'erreur finale sera plus faible quand le nombre de nœuds est plus grand.

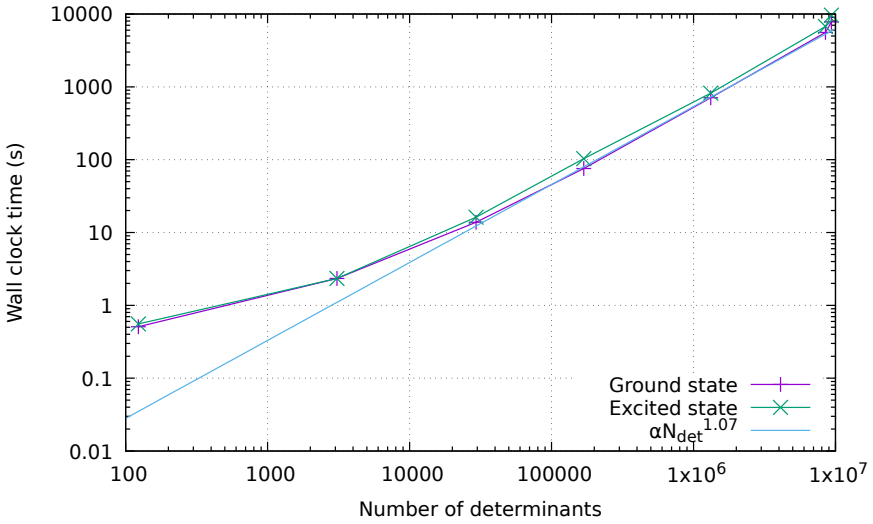


Figure B.9: Temps mural requis pour calculer la contribution perturbative E_{PT2} pour l'état fondamental et l'état excité, en fonction du nombre de déterminants dans la fonction d'onde.

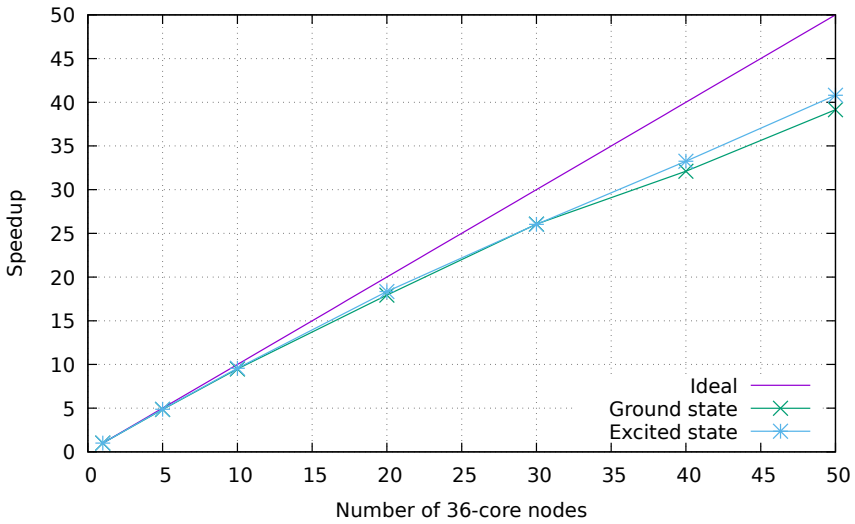


Figure B.10: Accélération parallèle pour le calcul de la contribution perturbative E_{PT2} pour l'état fondamental avec la plus grande fonction d'onde. Chaque nœud contient 36 cœurs physiques.

B.8.4 Habillage de matrice

L'algorithme d'habillage de matrice est similaire à celui du calcul de E_{PT2}, on s'attend donc à un comportement similaire.

Le critère d'arrêt est une erreur relative de 1/1000 sur l'énergie d'habillage $\langle \Psi | \hat{\Delta} | \Psi \rangle$, avec $\hat{\Delta}$ la matrice d'habillage.

Le scaling en fonction du nombre de déterminants (figure B.11) est de $\mathcal{O}\left({N_{\text{det}}}^{1.15}\right)$, ce qui est légèrement supérieur à celui trouvé pour E_{PT2}.

Ce coût additionnel est lié à la gestion des résultats par checkpoint, qui élimine le goulot d'étranglement lié aux communications.

L'accélération en fonction du nombre de nœuds (figure B.12) a été mesurée avec la fonction d'onde à 9 356 952 déterminants.

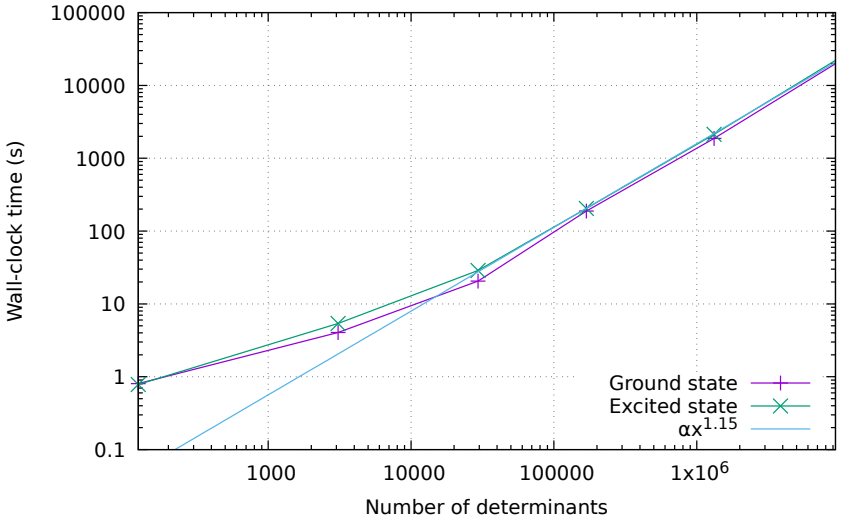


Figure B.11: Temps mural requis pour calculer l'habillage Shifted-B_k pour l'état fondamental et l'état excité, en fonction du nombre de déterminants dans la fonction d'onde.

B.9 Conclusion

Des améliorations aussi bien dans le mode séquentiel que parallèle ont été apportées au Quantum Package . Il est actuellement possible de réaliser des calculs sur $\sim$ 2000 cœurs avec des centaines de millions de déterminants dans l'espace variationnel, ce qui pourra conduire à la réalisation d'autres d'applications difficiles.

La diagonalisation de Davidson, au centre des méthodes variationnelles, souffre de l'impossibilité de stocker la matrice hamiltonienne, ce qui nous contraint à recalculer

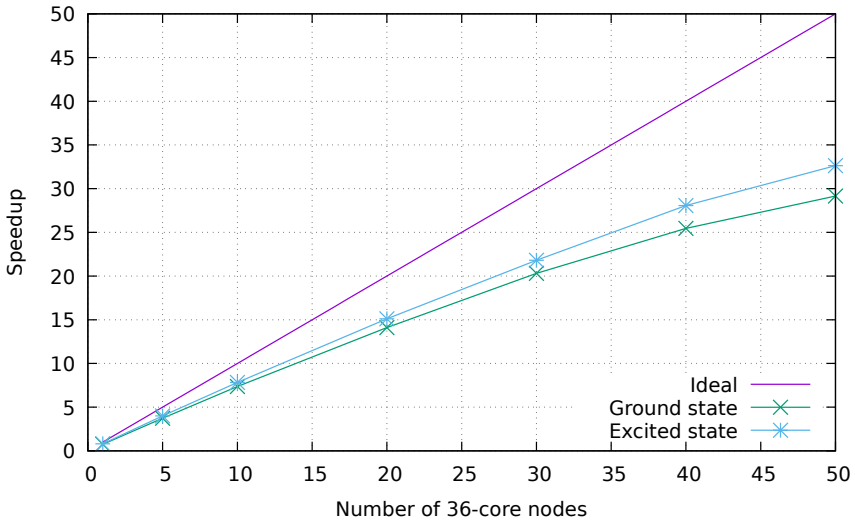


Figure B.12: Accélération parallèle pour le calcul de la matrice d'habillage de l'état fondamental avec la plus grande fonction d'onde. Chaque noeud contient 36 cœurs physiques.

les éléments de matrice *on the fly* à chaque itération. Malgré une méthode très efficace pour calculer les éléments de matrice,[45] et en particulier pour identifier un élément nul, l'implémentation initiale explorait chacun des $\sim {N_{\text{det}}}^2$ éléments de matrice.

À présent, les déterminants sont séparés en plusieurs ensembles disjoints et souvent identifiables comme entièrement déconnectés les uns des autres ; de ce fait, la vaste majorité des éléments de matrice n'ont plus à être explicitement considérés, et le scaling est ramené à $\mathcal{O}\left({N_{\text{det}}}^{3/2}\right)$. Une variante de scaling linéaire est possible, mais n'a pas été implémentée du fait d'une empreinte mémoire trop importante. Cependant, l'implémentation pourrait être raffinée afin d'utiliser conjointement les deux méthodes.

Bien que la parallélisation de l'algorithme proposé soit difficile, l'implémentation distribuée qui en a été réalisée offre une accélération satisfaisante, de l'ordre de 35× pour 50 nœuds (1 800 cœurs).

L'algorithme de sélection CIPSI a été considérablement amélioré, ce qui a rendu possible des applications jusque là infaisables.[57, 89]

Les différentes optimisations réalisées portent à la fois sur la mise en évidence des connexions entre les espaces interne et externe (approche par batch de déterminants, filtrage), et le coût de calcul des éléments de matrice correspondants (phase mask, détermination implicite des excitations). L'implémentation distribuée offre une accélération presque idéale avec 50 nœuds.

D'un point de vue méthodologique, le CIPSI est toutefois d'une précision et donc d'un coût qui ne se justifie pas dans toutes les situations ; de meilleurs performances,

sans perte de précision, pourraient être obtenues en combinant le CIPSI avec un algorithme de sélection moins précis mais très peu coûteux, le *Heat-Bath Configuration Interaction (HCI)*. [38, 39]

Le calcul de la contribution perturbative au second ordre E_{PT2}, a également fait l'objet d'améliorations importantes. Bien qu'étant calculé par le même algorithme que CIPSI, obtenir E_{PT2} était plus coûteux car n'autorisant pas autant d'approximations. Dans la mesure où E_{PT2} est une somme d'une multitude de termes, une approche Monte-Carlo est pertinente, d'autant plus que E_{PT2} sert essentiellement à fournir une approximation de l'énergie Full-CI. Par conséquent, sa valeur exacte n'a que peu d'intérêt, l'important étant que la précision de son estimation soit supérieur à la précision typique avec laquelle $E_{\mathrm{var}} + E_{\mathrm{PT2}}$ approxime l'énergie Full-CI.

L'approche Monte-Carlo proposée, originale par son caractère hybride stochastique/déterministe, a été développée avec l'aide du groupe de Michel Caffarel, et permet d'obtenir E_{PT2} avec une précision satisfaisante pour quelques pourcents du coût du calcul exacte.

Toutefois, cette approche stochastique n'est utilisée que pour le calcul de E_{PT2} et pas pour la sélection. Une sélection stochastique aurait l'avantage qu'à chaque itération, n'importe quel déterminant externe pourrait potentiellement être généré. Ce n'est pas le cas avec l'approche actuelle, en raison du seuil n_g, les déterminants de faibles coefficients ne sont jamais utilisés comme générateurs, ce qui est la cause d'une légère erreur de size-consistance.

Afin de tirer le meilleur parti des données auxquelles nous pouvions déjà accéder, nous avons par la suite implémenté la méthode shifted-B_k , qui permet de raffiner la fonction d'onde en fonction des contributions énergétiques individuelles des déterminants externes. Cette méthode nécessite le calcul d'une matrice d'habillage, estimée stochastiquement de la même manière que E_{PT2}. Toutefois, de part le fait que l'estimation porte sur un vecteur et non plus sur un scalaire, des difficultés supplémentaires ont dû être surmontées. Cela a pu être fait au prix d'un compromis relativement modeste, consistant à pré-définir, au début du calcul, un certain nombre (de l'ordre d'une à quelques dizaines) de "checkpoints" au niveau desquels un résultat est obtenu ; entre deux checkpoints, aucun résultat n'est disponible.

Cet habillage stochastique de matrice a ensuite pu être généralisé dans un framework permettant de raffiner la fonction d'onde sous l'effet de n'importe quel espace externe (défini par les coefficients des déterminants externes). Une implémentation de la méthode *Multi-Reference Coupled Cluster Single and Double (MR-CCSD)* a ainsi pu être réalisée simplement,[91] et a permis d'explorer certaines possibilités de l'approche determinant-driven.

A l'heure actuelle ce framework est commode mais offre peu de possibilités. Certaines pistes sont explorées afin de le rendre plus flexible.

Durant les différentes étapes de l'évolution du Quantum Package , de plus en plus

d'applications ont été rendues possibles.[1, 57, 89, 91, 92, 93, 94]

Cela a donné au Quantum Package plus de visibilité, conduisant notamment à sa sélection en tant que benchmark dans le choix du nouveau supercalculateur du centre CALMIP. De plus, différents groupes se sont mis à l'utiliser pour des applications et le développement de nouvelles idées. Par exemple, le groupe de Claudia Filippi aux Pays-Bas utilise maintenant des fonctions d'onde CIPSI dans le cadre de calculs Monte-Carlo quantique,[95] et le groupe d'Argonne implémente actuellement des orbitales complexes afin d'adapter le Quantum Package à la chimie du solide.[96]

Bibliography

[1] E. Giner, C. Angeli, Y. Garniron, A. Scemama, and J.-P. Malrieu, "A jeziorski-monkhorst fully uncontracted multi-reference perturbative treatment. i. principles, second-order versions, and tests on ground state potential energy curves," *The Journal of Chemical Physics*, vol. 146, p. 224108, jun 2017.

[2] M. Gordon *Electronics*, vol. 38, no. 8, p. 114, 1965.

[3] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. M. III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, "Molpro, version 2018.1, a package of ab initio programs," 2018. see.

[4] G. Karlström, R. Lindh, P.-Å. Malmqvist, B. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, and L. Seijo, "Molcas: a program package for computational chemistry.," *Elsevier*, vol. 28, no. 2, pp. 222–239, 2003.

[5] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels,

Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09, revision b.01," 2009.

[6] "Lists | TOP500 Supercomputer Sites," Oct 2018. [Online; accessed 9. Oct. 2018].

[7] H. Sutter and J. Larus, "Software and the concurrency revolution," *Queue*, vol. 3, p. 54, sep 2005.

[8] Wm. A. Wulf and S. A. McKee, "Hitting the memory wall: implications of the obvious," *SIGARCH Comput. Archit. News*, vol. 23, pp. 20–24, Mar 1995.

[9] H. N. Khan, D. A. Hounshell, and E. R. H. Fuchs, "Science and research policy at the end of Moore's law," *Nature Electronics*, vol. 1, pp. 14–21, Jan 2018.

[10] G. H. Booth, A. J. W. Thom, and A. Alavi, "Fermion monte carlo without fixed nodes: A game of life, death, and annihilation in slater determinant space," *The Journal of Chemical Physics*, vol. 131, no. 5, p. 054106, 2009.

[11] B. Huron, J. P. Malrieu, and P. Rancurel, "Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions," *The Journal of Chemical Physics*, vol. 58, pp. 5745–5759, jun 1973.

[12] "Quantum package," Oct 2018. [Online; accessed 11. Sep. 2018].

[13] Àngels Povill and J. Rubio, "An efficient improvement of the string-based direct selected CI algorithm," *Theoretica Chimica Acta*, vol. 92, pp. 305–313, nov 1995.

[14] S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, "Convergence of an improved CIPSI algorithm," *Chemical Physics*, vol. 75, pp. 91–102, feb 1983.

[15] R. Pauncz, "The waller-hartree double determinant in quantum chemistry," *International Journal of Quantum Chemistry*, vol. 35, pp. 717–719, jun 1989.

[16] P. Löwdin, "P.-o. löwdin, adv. chem. phys. 2, 207 (1959).," *Adv. Chem. Phys.*, vol. 2, p. 207, 1959.

[17] C. C. J. Roothaan, "New developments in molecular orbital theory," *Reviews of Modern Physics*, vol. 23, pp. 69–89, apr 1951.

[18] S. Obara and A. Saika, "Efficient recursive computation of molecular integrals over cartesian gaussian functions," *The Journal of Chemical Physics*, vol. 84, pp. 3963–3974, apr 1986.

[19] M. Head-Gordon and J. A. Pople, "A method for two-electron gaussian integral and integral derivative evaluation using recurrence relations," *The Journal of Chemical Physics*, vol. 89, pp. 5777–5786, nov 1988.

[20] S. Ten-no, "An efficient algorithm for electron repulsion integrals over contracted gaussian-type functions," *Chemical Physics Letters*, vol. 211, pp. 259–264, aug 1993.

[21] P. M. W. Gill, M. Head-Gordon, and J. A. Pople, "An efficient algorithm for the generation of two-electron repulsion integrals over gaussian basis functions," *International Journal of Quantum Chemistry*, vol. 36, pp. 269–280, jun 1989.

[22] P. M. W. Gill and J. A. Pople, "The prism algorithm for two-electron integrals," *International Journal of Quantum Chemistry*, vol. 40, pp. 753–772, dec 1991.

[23] E. F. Valeev, "Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions." http://libint.valeyev.net/, 2018. version 2.5.0-beta.1.

[24] J. Zhang, "libreta: Computerized optimization and code synthesis for electron repulsion integral evaluation," *Journal of Chemical Theory and Computation*, vol. 14, pp. 572–587, jan 2018.

[25] S. Wilson, "Four-index transformations," in *Methods in Computational Chemistry*, pp. 251–309, Springer US, 1987.

[26] S. Rajbhandari, F. Rastello, K. Kowalski, S. Krishnamoorthy, and P. Sadayappan, "Optimizing the four-index integral transform using data movement lower bounds analysis," in *Proceedings of the 22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming - PPoPP '17*, ACM Press, 2017.

[27] A. C. Limaye and S. R. Gadre, "A general parallel solution to the integral transformation and second-order mo/ller–plesset energy evaluation on distributed memory parallel machines," *The Journal of Chemical Physics*, vol. 100, pp. 1303–1307, jan 1994.

[28] G. Fletcher, M. Schmidt, and M. Gordon, "Developments in parallel electronic structure theory," *Advances in chemical physics*, vol. 110, pp. 267–294, 1999.

[29] L. A. Covick and K. M. Sando, "Four-index transformation on distributed-memory parallel computers," *Journal of Computational Chemistry*, vol. 11, pp. 1151–1159, nov 1990.

[30] E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices," *Journal of Computational Physics*, vol. 17, pp. 87–94, jan 1975.

[31] L. Bytautas and K. Ruedenberg, "A priori identification of configurational deadwood," *Chemical Physics*, vol. 356, pp. 64–75, feb 2009.

[32] J. S. Anderson, F. Heidar-Zadeh, and P. W. Ayers, "Breaking the curse of dimension for the electronic schrodinger equation with functional analysis," *Computational and Theoretical Chemistry*, vol. 1142, pp. 66–77, oct 2018.

[33] C. F. Bender and E. R. Davidson, "Studies in configuration interaction: The first-row diatomic hydrides," *Phys. Rev.*, vol. 183, pp. 23–30, jul 1969.

[34] J. L. Whitten and M. Hackmeyer, "Configuration interaction studies of ground and excited states of polyatomic molecules. i. the CI formulation and studies of formaldehyde," *The Journal of Chemical Physics*, vol. 51, pp. 5584–5596, dec 1969.

[35] Y. Ohtsuka and J. ya Hasegawa, "Selected configuration interaction method using sampled first-order corrections to wave functions," *The Journal of Chemical Physics*, vol. 147, p. 034102, jul 2017.

[36] J. P. Coe, "Machine learning configuration interaction," 2018.

[37] P. Knowles and N. Handy, "A new determinant-based full configuration interaction method," *Chemical Physics Letters*, vol. 111, pp. 315–321, nov 1984.

[38] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, "Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3674–3680, aug 2016.

[39] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, "Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory," *Journal of Chemical Theory and Computation*, vol. 13, pp. 1595–1604, mar 2017.

[40] G. H. Booth and A. Alavi, "Approaching chemical accuracy using full configuration-interaction quantum monte carlo: A study of ionization potentials," *The Journal of Chemical Physics*, vol. 132, p. 174104, may 2010.

[41] D. Cleland, G. H. Booth, and A. Alavi, "Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum monte carlo," *The Journal of Chemical Physics*, vol. 132, p. 041103, jan 2010.

[42] J. C. Greer, "Estimating full configuration interaction limits from a monte carlo selection of the expansion space," *The Journal of Chemical Physics*, vol. 103, pp. 1821–1828, aug 1995.

[43] J. Greer, "Monte carlo configuration interaction," *Journal of Computational Physics*, vol. 146, pp. 181–202, oct 1998.

[44] W. D. Maurer and T. G. Lewis, "Hash Table Methods," *ACM Comput. Surv.*, vol. 7, pp. 5–19, Mar 1975.

[45] A. Scemama and E. Giner, "An efficient implementation of Slater-Condon rules," *ArXiv [physics.comp-ph]*, p. 1311.6244, Nov. 2013.

[46] B. Liu, "The simultaneous expansion for the solution of several of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices," *Numerical Algorithms in Chemistry: Algebraic Method, Lawrence Berkeley Laboratory, University of California, California*, pp. 49–53, 1978.

[47] J. Olsen, P. Jørgensen, and J. Simons, "Passing the one-billion limit in full configuration-interaction (FCI) calculations," *Chem. Phys. Lett.*, vol. 169, pp. 463–472, Jun 1990.

[48] F. X. Gadea, "Large matrix diagonalization, comparison of various algorithms and a new proposal," *Chemical Physics Letters*, vol. 227, pp. 201–210, sep 1994.

[49] M. Crouzeix, B. Philippe, and M. Sadkane, "The davidson method," *SIAM Journal on Scientific Computing*, vol. 15, pp. 62–76, jan 1994.

[50] J. Choi, J. Dongarra, S. Ostrouchov, A. Petitet, D. Walker, and R. C. Whaley, "A proposal for a set of parallel basic linear algebra subprograms," LAPACK Working Note 100, Department of Computer Science, University of Tennessee, Knoxville, Knoxville, TN 37996, USA, May 1995. LAPACK Working Note #100. UT-CS-95-292, May 1995.

[51] I. J. Davis, "A Fast Radix Sort," *Comput. J.*, vol. 35, pp. 636–642, Dec 1992.

[52] M. P. Forum, "MPI: A Message-Passing Interface Standard," *University of Tennessee*, Apr 1994.

[53] L. Dagum and R. Menon, "OpenMP: An Industry-Standard API for Shared-Memory Programming," *IEEE Comput. Sci. Eng.*, vol. 5, pp. 46–55, Jan 1998.

[54] W. J. Bolosky and M. L. Scott, "False sharing and its effect on shared memory performance," *USENIX Association*, p. 3, Sep 1993.

[55] B. S. Fales, E. G. Hohenstein, and B. G. Levine, "Robust and Efficient Spin Purification for Determinantal Configuration Interaction," *J. Chem. Theory Comput.*, vol. 13, pp. 4162–4172, Sep 2017.

[56] E. Giner, *Coupling Configuration Interaction and quantum Monte Carlo methods: The best of both worlds.* Theses, Université de Toulouse, Oct. 2014.

[57] E. Giner, D. Tew, Y. Garniron, and A. Alavi, "Interplay between electronic correlation and metal-ligand delocalization in the spectroscopy of transition metal compounds: case study on a series of planar cu^{2+} complexes," 2018.

[58] R. Cimiraglia, "Many-body multireference mï¿½ller?plesset and epstein?nesbet perturbation theory: Fast evaluation of second-order energy contributions," *International Journal of Quantum Chemistry*, vol. 60, pp. 167–171, oct 1996.

[59] C. Angeli, R. Cimiraglia, M. Persico, and A. Toniolo, "Multireference perturbation CI i. extrapolation procedures with CAS or selected zero-order spaces," *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, vol. 98, pp. 57–63, Oct. 1997.

[60] L. Devroye, *Non-Uniform Random Variate Generation*. Springer Science & Business Media, Nov 2013.

[61] D. Michie, ""Memo" Functions and Machine Learning," *Nature*, vol. 218, p. 19, Apr 1968.

[62] R. A. Frost, "Using Memoization to Achieve Polynomial Complexity of Purely Functional Executable Specifications of Non-Deterministic Top-Down Parsers," *undefined*, 1994.

[63] R. V. Hogg, E. Tanis, and D. Zimmerman, *Probability and Statistical Inference*. Pearson Education, Jan 2014.

[64] Z. Gershgorn and I. Shavitt, "An application of perturbation theory ideas in configuration interaction calculations," *International Journal of Quantum Chemistry*, vol. 2, pp. 751–759, nov 1968.

[65] J. P. Malrieu, P. Durand, and J. P. Daudey, "Intermediate hamiltonians as a new class of effective hamiltonians," *Journal of Physics A: Mathematical and General*, vol. 18, no. 5, p. 809, 1985.

[66] I. Lindgren and J. Morrison, *Atomic Many-Body Theory, Vol. 13 of Springer Series in Chemical Physics*. Springer, Berlin, 1982.

[67] L. E. Nitzsche and E. R. Davidson, "A perturbation theory calculation on the 1pipi* state of formamide," *The Journal of Chemical Physics*, vol. 68, pp. 3103–3109, apr 1978.

[68] L. E. Nitzsche and E. R. Davidson, "Ab initio calculation of some vertical excitation energies of n-methylacetamide," *Journal of the American Chemical Society*, vol. 100, pp. 7201–7204, nov 1978.

[69] D. C. Rawlings and E. R. Davidson, "The rayleigh—schrodinger BK method applied to the lower electronic states of pyrrole," *Chemical Physics Letters*, vol. 98, pp. 424–427, jul 1983.

[70] P. M. Kozlowski, M. Dupuis, and E. R. Davidson, "The cope rearrangement revisited with multireference perturbation theory," *Journal of the American Chemical Society*, vol. 117, pp. 774–778, jan 1995.

[71] P. Kozlowski and E. Davidson, "Test of a new multi-reference møller-plesset perturbation theory," *Chemical Physics Letters*, vol. 222, pp. 615–620, jun 1994.

[72] P. M. Kozlowski and E. R. Davidson, "Considerations in constructing a multireference second-order perturbation theory," *The Journal of Chemical Physics*, vol. 100, pp. 3672–3682, mar 1994.

[73] P. Kozlowski and E. Davidson, "Construction of open shell perturbation theory invariant with respect to orbital degeneracy," *Chemical Physics Letters*, vol. 226, pp. 440–446, aug 1994.

[74] J. Goldstone, "Derivation of the brueckner many-body theory," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 239, pp. 267–279, feb 1957.

[75] S. R. Langhoff and E. R. Davidson, "Configuration interaction calculations on the nitrogen molecule," *International Journal of Quantum Chemistry*, vol. 8, pp. 61–72, jan 1974.

[76] L. Meissner, "Size-consistency corrections for configuration interaction calculations," *Chemical Physics Letters*, vol. 146, pp. 204–210, may 1988.

[77] H. P. Kelly and A. M. Sessler, "Correlation effects in many-fermion systems: Multiple-particle excitation expansion," *Physical Review*, vol. 132, pp. 2091–2095, dec 1963.

[78] H. P. Kelly, "Correlation effects in many fermion systems. II. linked clusters," *Physical Review*, vol. 134, pp. A1450–A1453, jun 1964.

[79] W. Meyer, "Ionization energies of water from PNO-CI calculations," *International Journal of Quantum Chemistry*, vol. 5, pp. 341–348, jun 1971.

[80] W. Meyer, "PNO–CI studies of electron correlation effects. i. configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane," *The Journal of Chemical Physics*, vol. 58, pp. 1017–1035, feb 1973.

[81] W. Meyer, "PNO-CI and CEPA studies of electron correlation effects," *Theoretica Chimica Acta*, vol. 35, no. 4, pp. 277–292, 1974.

[82] R. Ahlrichs, H. Lischka, V. Staemmler, and W. Kutzelnigg, "PNO–CI (pair natural orbital configuration interaction) and CEPA–PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. i. outline of the method for closed-shell states," *The Journal of Chemical Physics*, vol. 62, pp. 1225–1234, feb 1975.

[83] S. Koch and W. Kutzelnigg, "Comparison of CEPA and CP-MET methods," *Theoretica Chimica Acta*, vol. 59, pp. 387–411, jul 1981.

[84] R. Send, O. Valsson, and C. Filippi, "Electronic excitations of simple cyanine dyes: Reconciling density functional and wave function methods," *J. Chem. Theory Comput.*, vol. 7, pp. 444–455, 2011.

[85] A. A. Holmes, C. J. Umrigar, and S. Sharma, "Excited states using semistochastic heat-bath configuration interaction," *The Journal of Chemical Physics*, vol. 147, p. 164111, oct 2017.

[86] B. L. Hammond, W. A. Lester, and P. J. Reynolds, *Monte Carlo methods in ab initio quantum chemistry*, vol. 1. World Scientific, 1994.

[87] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Communication: Toward an improved control of the fixed-node error in quantum monte carlo: The case of the water molecule," *The Journal of Chemical Physics*, vol. 144, p. 151103, apr 2016.

[88] K. H. Mood and A. Lüchow, "Full wave function optimization with quantum monte carlo and its effect on the dissociation energy of FeS," *The Journal of Physical Chemistry A*, vol. 121, pp. 6165–6171, aug 2017.

[89] A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P.-F. Loos, "Excitation energies from diffusion monte carlo using selected configuration interaction nodes," *The Journal of Chemical Physics*, vol. 149, p. 034108, jul 2018.

[90] A. A. Holmes, H. J. Changlani, and C. J. Umrigar, "Efficient heat-bath sampling in fock space," *Journal of Chemical Theory and Computation*, vol. 12, pp. 1561–1571, mar 2016.

[91] Y. Garniron, E. Giner, J.-P. Malrieu, and A. Scemama, "Alternative definition of excitation amplitudes in multi-reference state-specific coupled cluster," *The Journal of Chemical Physics*, vol. 146, p. 154107, apr 2017.

[92] P.-F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, "A mountaineering strategy to excited states: Highly accurate reference energies and benchmarks," *Journal of Chemical Theory and Computation*, vol. 14, pp. 4360–4379, jul 2018.

[93] Y. Garniron, A. Scemama, E. Giner, M. Caffarel, and P.-F. Loos, "Selected configuration interaction dressed by perturbation," *The Journal of Chemical Physics*, vol. 149, p. 064103, aug 2018.

[94] Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, "Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory," *The Journal of Chemical Physics*, vol. 147, p. 034101, jul 2017.

[95] M. Dash, S. Moroni, A. Scemama, and C. Filippi, "Perturbatively selected configuration-interaction wave functions for efficient geometry optimization in quantum monte carlo," *Journal of Chemical Theory and Computation*, vol. 14, pp. 4176–4182, jun 2018.

[96] A. Benali, K. Gasperich, and T. Applencourt, "Cipsi for solid state.".