



HAL
open science

Optimization of large determinant expansions in quantum Monte Carlo

Abdallah Ammar, Emmanuel Giner, Anthony Scemama

► **To cite this version:**

Abdallah Ammar, Emmanuel Giner, Anthony Scemama. Optimization of large determinant expansions in quantum Monte Carlo. *Journal of Chemical Theory and Computation*, 2022, 18 (9), pp.5325-5336. 10.1021/acs.jctc.2c00556 . hal-03716808

HAL Id: hal-03716808

<https://hal.science/hal-03716808v1>

Submitted on 7 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimization of large determinant expansions in quantum Monte Carlo

Abdallah Ammar,¹ Emmanuel Giner,² and Anthony Scemama^{1,*}

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

²Laboratoire de Chimie Théorique (UMR 7616), Université Paris Sorbonne, CNRS, France

We present a new method for the optimization of large configuration interaction (CI) expansions in the quantum Monte Carlo (QMC) framework. The central idea here is to replace the non-orthogonal variational optimization of CI coefficients performed in usual QMC calculations by an orthogonal non-Hermitian optimization thanks to the so-called transcorrelated (TC) framework, the two methods yielding the same results in the limit of a complete basis set. By rewriting the TC equations as an effective self-consistent Hermitian problem, our approach requires the sampling of a single quantity per Slater determinant, leading to minimal memory requirements in the QMC code. Using analytical quantities obtained from both the TC framework and the usual CI-type calculations, we also propose improved estimators which reduce the statistical fluctuations of the sampled quantities by more than an order of magnitude. We demonstrate the efficiency of this method on wave functions containing $10^5 - 10^6$ Slater determinants, using effective core potentials or all-electron calculations. In all the cases, a sub-milliHartree convergence is reached within only two or three iterations of optimization.

I. INTRODUCTION

The accurate description of highly correlated systems such as transition states, magnetic systems and excited states requires a multi-configurational form of the wave function. For most problems, the complete active space (CAS) picture already contains enough information to give a qualitatively correct description of the physics governing the electronic structure. There exists nevertheless some systems in which the dynamical correlation is strongly coupled to the static correlation,¹⁻⁵ such that the dominant part of the wave function (typically its projection onto a valence-like CAS) is severely impacted by the full treatment of dynamical correlation in the optimization process. Multiple methods were designed to take into account the feedback of the dynamical correlation on the reference: one can cite for instance internally decontracted configuration interaction (CI)^{6,7} or coupled cluster (CC)^{8,9} methods, f_{12} methods combined with multi-reference CI,¹⁰ range-separated density functional theory (DFT) combined with CAS,¹¹⁻¹³ or the shifted- B_k method^{14,15}.

In this paper, we focus on Quantum Monte Carlo (QMC) approaches, in which the N -electron wave function Ψ is expressed as

$$\Psi(\mathbf{r}) = \Phi(\mathbf{r}) e^{J(\mathbf{r})}, \quad (1)$$

where \mathbf{r} is the \mathbb{R}^{3N} electronic configuration space. In Eq. (1), Φ is a multi-determinant expansion and $\exp(J)$ is a Jastrow correlation factor, taking explicitly into account electron-electron distances. The Jastrow factor allows the wave function $\Psi(\mathbf{r})$ to fulfill the exact Kato's cusp conditions¹⁶, but also introduces a large amount of short-range dynamical correlation which are usually very costly to capture within a usual Slater determinant expansion. Therefore, the functional form of Eq. (1) is very flexible and can be used to treat complex correlation effects, provided that one manages to optimize the different parameters.

As opposed to f_{12} methods where the correlation factor is projected into a space orthogonal to the multi-determinant expansion, the Jastrow factor has a significant overlap with the

determinant expansion. This last statement particularly motivates the need for an optimization of the parameters of Φ in the presence of the Jastrow factor, which is not straightforward as it involves $3N$ -dimensional integrals that cannot be evaluated exactly in the general case. As a consequence, the necessary matrix elements are sampled in a variational Monte Carlo (VMC) simulation and they are subject to statistical noise. In the general context of VMC simulations, an important aspect is precisely the amplitude of the statistical fluctuations of the quantities needed to optimize the wave function, which eventually determines the actual applicability of a given computational algorithm within a reasonable CPU time.

One of the most common methods to perform such an optimization is the standard linear method (LM), where the Schrödinger equation is projected in the self-plus-tangent space, i.e., the space spanned by the current wave function and its first derivatives with respect to the N_p parameters.¹⁷⁻²⁰ The new parameters are obtained by solving a generalized eigenvalue problem (GEP) after building the Hamiltonian and overlap matrices, both of size $N_p \times N_p$, through a VMC calculation. In practice, the standard LM is limited to few thousands of parameters²¹ due to the large memory requirement to sample and store these matrices.

Another class of optimization methods constrained by the storage of large matrices is the stochastic reconfiguration (SR) method.²²⁻²⁴ In this method, the imaginary time evolution operator is expanded to first order and projected iteratively in the self-plus-tangent space. At each step of this procedure, one must build a $N_p \times N_p$ overlap-like matrix and solve a system of linear equations. The wave function is updated after each iteration until the energy has converged.

Several solutions have been proposed to address the memory bottleneck of the LM and SR methods. Neuscamman and co-workers^{25,26} have suggested to employ Krylov subspace solvers to avoid building the matrices explicitly. Although this approach allows to enlarge significantly the number of variational parameters in the optimization, it requires additional matrix-vector multiplications as the Hamiltonian and overlap matrices are sampled, which increases the sampling effort. Moreover, ill-conditioned matrices constitute

a challenge in the LM.²⁷ A similar strategy relies on the Jacobi-Davidson method,^{28–30} which generalizes Davidson’s method.³¹ Recently, a variant of the LM, termed the blocked linear method (BLM), has been introduced to alleviate the difficulties of storing large matrices.²⁷

At the other end of the scale, the gradient descent (GD) approaches³² have a smaller memory footprint as the required storage scales linearly with the number of parameters. These approaches exploit the recent developments in the field of deep-learning algorithms of neural networks and make use of some adaptive GD flavours to perform the optimization directly.^{33–37} In their recent work³⁸, Otis and Neuscamman have presented a comparison between LM and GD techniques, and they have developed a hybrid scheme which combines BLM and GD methods.

In this work we focus on the optimization of the CI coefficients, in view of optimizing very large CI expansions for an arbitrary choice of Jastrow factor. For this purpose, we present an iterative scheme to optimize the wave function by combining VMC and the transcorrelated (TC) approach, motivated by i) the fact that the TC formalism doesn’t require the sampling of the overlap matrix and ii) that the VMC and TC approaches lead to the same wave function in the limit of a complete basis set.

In the TC framework the correlation effects brought by the Jastrow factor are incorporated directly in the original Hamiltonian through a similarity transformation $\hat{H}_{\text{TC}} \equiv e^{-\hat{J}} \hat{H} e^{\hat{J}}$. This approach, originally proposed by Hirschfelder to remove electron-electron poles from the original Hamiltonian,³⁹ was then further developed by Boys and Handy who referred to it as a *TC method* where the authors optimize both the Jastrow factor and the orbitals of a single Slater determinant.^{40–45} The TC method was then further developed in the beginning of the 2000’s by Ten-No^{46–48} and co-workers through the use of a universal correlation factor (*i.e.* the same correlation factor for all systems) with a relatively flexible form for the wave function expansion (such as perturbation theory or linearized coupled cluster). Developments using VMC to compute the variance of the TC Hamiltonian in order to remain within a variational framework were carried by Umezawa and co-workers^{49–54}. A recent renewal of the TC methods were brought by Alavi and coworkers where they used a flexible form for both the correlation factor and the wave function^{55–57} thanks to the use of a version of the full configuration interaction quantum Monte Carlo (FCIQMC)^{58,59} method adapted for a non-Hermitian Hamiltonian containing up to three-electron interactions.

The algorithm exposed here proposes to bypass the usual VMC optimization, and use instead the TC method to optimize the CI coefficients of Jastrow-Slater wave functions for an arbitrary Jastrow factor. To do so, we reformulate the non-Hermitian TC approach in terms of an Hermitian self-consistent dressing of the standard Hamiltonian \hat{H} which accounts for the effect of the Jastrow factor. These equations are then projected in a basis of Slater determinants, and as the scheme is Hermitian, the standard Davidson algorithm can be employed to optimize the wave function. The advantage of the present scheme is that one samples the action of the dressed

Hamiltonian on the CI vector, which requires the sampling of a single quantity per CI coefficient, leading to a minimal memory footprint. The convergence to the chemical accuracy of the self-consistent procedure is reached with typically two or three iterations. Another interesting aspect of the present scheme is that the sampled dressing matrix has a zero-variance property which leads to relatively small fluctuations as compared to the estimators used in the LM, and the fluctuations can be further reduced by introducing the deterministic computation of an auxiliary TC Hamiltonian.

The paper is organized as follows. In section II, we discuss the most common schemes used to optimize CI coefficients in the presence of a Jastrow factor. We then give in Sec. III the description of our new algorithm: in Sec. III A we provide a brief description of the TC formalism and its connection with VMC, we present the general ideas of our optimization scheme for large CI expansions in Sec. III B and present in Sec. III C the different numerical strategies employed to compute the dressing elements with minimal statistical fluctuations. In Sec. IV, we present numerical tests validating the present approach. In Sec. IV A we test the feasibility of the present approach on the Beryllium atom together with the actual impact of the incompleteness of the N -electron basis set. Finally, in Sec. IV B we optimize several CI expansions for small molecules made of a few hundred thousand Slater determinants.

II. WAVE FUNCTION OPTIMIZATION IN THE PRESENCE OF A JASTROW FACTOR

A. General context

Consider a ground-state N -electron wave function Φ expressed in a basis of Slater determinants $\mathcal{B} = \{D_I, 1 \leq I \leq N_{\text{det}}\}$ obtained with orthonormal spin-orbitals

$$\Phi(\mathbf{r}) = \sum_{I=1}^{N_{\text{det}}} c_I D_I(\mathbf{r}), \quad (2)$$

where N_{det} is the number of determinants. In the QMC framework, a relatively cheap and efficient way of increasing the amount of correlation energy described by the wave function is to introduce a Jastrow factor $J(\mathbf{r})$, which captures short-range effects that cannot be easily described by the finite determinant basis set

$$\Psi(\mathbf{r}) = \Phi(\mathbf{r}) e^{J(\mathbf{r})} = \sum_{I=1}^{N_{\text{det}}} c_I D_I(\mathbf{r}) e^{J(\mathbf{r})}, \quad (3)$$

where, generally, $J(\mathbf{r})$ is a function of electron-nuclear, electron-electron, and electron-electron-nuclear distances. Because of the large overlap between the Jastrow factor and the determinant basis \mathcal{B} , the optimal CI coefficients are not those obtained by simply minimizing the variational energy of the wave function of Eq. (2). Our goal in this work is to implement an efficient scheme to optimize the CI coefficients for large N_{det} in the presence of a general Jastrow factor $J(\mathbf{r})$.

Two methods widely used to perform such an optimization, the LM and SR, are briefly discussed in this section.

B. Linear method

The most natural way to optimize the CI coefficients is to express the CI problem in the basis of determinants augmented by a Jastrow factor $\mathcal{B}_J = \{D_I e^J, 1 \leq I \leq N_{\text{det}}\}$. This basis is not orthonormal, so in addition to the Hamiltonian matrix elements

$$H_{IK} = \langle D_I e^J | \hat{H} | D_K e^J \rangle, \quad (4)$$

one needs also to compute the overlap matrix elements

$$S_{IK} = \langle D_I e^J | D_K e^J \rangle. \quad (5)$$

As the forms commonly used for the Jastrow factor are too complicated to integrate analytically, these $3N$ -dimensional integrals need to be sampled using a VMC sampling:

$$H_{IK} = \left\langle \frac{D_I e^J}{\Psi} \frac{\hat{H} (D_K e^J)}{\Psi} \right\rangle_{\Psi^2}, \quad (6)$$

$$S_{IK} = \left\langle \frac{D_I e^J}{\Psi} \frac{D_K e^J}{\Psi} \right\rangle_{\Psi^2}, \quad (7)$$

where $\langle \dots \rangle_{\Psi^2}$ denotes the stochastic average over the Monte Carlo samples drawn with the $3N$ -dimensional density Ψ^2 .

The CI problem can now be solved as a GEP

$$\mathbf{H} \mathbf{C} = E \mathbf{S} \mathbf{C}. \quad (8)$$

Provided that the Jastrow factor is kept constant, solving this GEP provides directly the best variational coefficients in the basis \mathcal{B}_J . However, several difficulties are encountered in this approach. The statistical errors obtained from the VMC calculation on H_{IK} and S_{IK} can be important, and frequently larger than the matrix element itself, leading to an effective optimization of only a reduced set of the parameters of the wave function. Therefore, this approach is in practice limited by the expensive memory requirements for the storage of the sampled matrices \mathbf{H} and \mathbf{S} .

The memory cost scales as $O(N_{\text{det}}^2)$, so when the number of parameters becomes as large as a few thousands this simple approach becomes prohibitive. A first solution to this memory bottleneck is to employ a Krylov subspace solver to eschew building \mathbf{H} and \mathbf{S} explicitly.²⁵ This improvement has led to the optimization of up to 5×10^5 variational parameters. The BLM is an alternative approach²⁷ in which the space of determinants \mathcal{B}_J is divided into N_b blocks. A GEP is solved in each block to generate a set of N_k eigenvectors. Those eigenvectors are then fixed and employed as directions to find a new direction for the full space \mathcal{B}_J . The memory cost of this approach scales approximately as $O(N_{\text{det}}^2/N_b)$. The BLM was recently applied in combination with GD methods.³⁸

C. Stochastic reconfiguration

The SR is an alternative method where the wave function is iteratively improved by applying the first order expansion of the imaginary time evolution operator $\exp(-\tau \hat{H}) \approx (1 - \tau \hat{H})$.²²⁻²⁴ Instead of solving a GEP one has to solve a set of N_{det} linear equations

$$\bar{\mathbf{S}} \delta \mathbf{c} = -\frac{\tau}{2} \mathbf{g}, \quad (9)$$

where

$$\bar{S}_{IK} = S_{IK} - \left\langle \frac{D_I e^J}{\Psi} \right\rangle_{\Psi^2} \left\langle \frac{D_K e^J}{\Psi} \right\rangle_{\Psi^2} \quad (10)$$

$$g_I = 2 \left[\left\langle \frac{D_I e^J}{\Psi} \frac{\hat{H} \Psi}{\Psi} \right\rangle_{\Psi^2} - \left\langle \frac{\hat{H} \Psi}{\Psi} \right\rangle_{\Psi^2} \left\langle \frac{D_I e^J}{\Psi} \right\rangle_{\Psi^2} \right]. \quad (11)$$

In its original formulation, the SR required storing an overlap-like matrix of $N_{\text{det}} \times N_{\text{det}}$ dimension which restricts the optimization to a few thousand coefficients. However, this memory bottleneck can be bypassed by using a conjugate gradient iterative solver to solve Eq. (9).²⁵ In this scheme, the explicit matrix $\bar{\mathbf{S}}$ is not required, but one needs to store instead Monte Carlo samples over M steps making the storage become $O(M \times N_p)$. This improvement allowed to optimize CI expansions of few hundred thousand parameters.^{60,61}

III. USING THE TRANSCORRELATED FORMALISM TO APPROXIMATE VMC

In this section, we present a new iterative scheme for the optimization of large CI expansions in the presence of an arbitrary Jastrow factor. This approach lies in the framework of Krylov subspace solvers. The memory requirement of the algorithm is minimal and scales as $O(N_{\text{det}})$, the convergence within chemical accuracy is reached typically in two or three iterations, and the method takes advantage of improved estimators which drastically reduce statistical fluctuations. In Sec. III A we present the main idea of our approach, in Sec. III B we expose the general iterative scheme used to optimize the wave functions, and in Sec. III C we detail the different strategies to reduce the statistical fluctuations.

A. General idea

In the TC formalism, the effect of the Jastrow factor is incorporated in the Hamiltonian through a similarity transformation³⁹⁻⁴⁵

$$\hat{H}_J \equiv e^{-J} \hat{H} e^J. \quad (12)$$

Therefore, solving exactly the Schrödinger equation

$$\hat{H} \Psi = E \Psi, \quad (13)$$

with a wave function $\Psi(\mathbf{r}) = \Phi(\mathbf{r}) e^{J(\mathbf{r})}$ as defined in Eq. (3) is equivalent to solve the TC eigenvalue equation

$$\hat{H}_J \Phi = E \Phi, \quad (14)$$

in the limiting case where Φ is expanded in a complete N -electron basis set. The effective TC Hamiltonian can be written as

$$\hat{H}_J = \hat{H} + \hat{\Delta}_J, \quad (15)$$

with

$$\hat{\Delta}_J = -\frac{1}{2} \sum_{i=1}^N \left[\nabla_i^2 J + \nabla_i J \cdot \nabla_i J + 2 \nabla_i J \cdot \nabla_i \right]. \quad (16)$$

Note that the $\hat{\Delta}_J$ operator is non-Hermitian due to the gradient operator in the last term of Eq. (16). As a consequence each eigenvalue is associated with a couple of left and right eigenvectors, and the variational principle does not apply anymore in the sense that there is no functional $F[\Phi]$ of a N -electron $|\Phi\rangle$ satisfying $F[\Phi] \geq E_0$ where E_0 is the exact ground state energy.

The central idea of this paper is to use the TC framework to optimize a CI wave function rather than the usual VMC framework. The main technical differences between the TC and the VMC approaches are first that in the TC framework the CI problem is expressed in an orthonormal basis, and secondly that the TC Hamiltonian involves *at most* three-electron interactions. Indeed, in the usual VMC approach the CI problem is solved in the non-orthonormal basis of determinants multiplied by the Jastrow factor, and the matrix elements of the Hamiltonian involve up to N -electron interactions.

These two approaches are nevertheless not equivalent when working with an incomplete basis set. Provided a given Jastrow factor J , if one uses an incomplete set of Slater determinants \mathcal{B} and the associated projector $\hat{P}_{\mathcal{B}}$ to project the TC equation and the VMC equation, one obtains the following projected operators

$$\hat{H}_{\text{TC}}^{\mathcal{B}} = \hat{P}_{\mathcal{B}} e^{-J} \hat{H} e^J \hat{P}_{\mathcal{B}}, \quad (17)$$

$$\hat{H}_{\text{VMC}}^{\mathcal{B}} = \hat{P}_{\mathcal{B}} e^{+J} \hat{H} e^J \hat{P}_{\mathcal{B}}, \quad (18)$$

which differ only by the sign in the first exponential. In the limit of a complete N -electron basis, the TC equation and the Hermitian non-orthogonal GEP of Eq. (8) lead to the same eigenvectors and eigenvalues. However, as shown in the appendix (see Sec. VI) the right-eigenvectors of $\hat{H}_{\text{TC}}^{\mathcal{B}}$ do not coincide with the eigenvectors of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$, although being usually very close. This means that the former is not optimal in the sense of the variational energy. Nonetheless, the statistical fluctuations induced by the finite sampling of the matrix elements of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ are such that in practice it is very hard to obtain either the exact optimum in VMC, and as we shall see in the Sec. IV A, the numerical solutions of the TC and VMC problems are close enough to be indistinguishable within reasonable statistical fluctuations.

B. Obtaining right-eigenvectors with a Hermitian dressing of the Hamiltonian

Working in a TC framework implies that one needs to rely on non-Hermitian variants of the Davidson algorithm⁶² to obtain the ground state eigenvector. In the present section we describe an alternative procedure to obtain a given eigenstate of the TC Hamiltonian involving an effective non-linear Hermitian Hamiltonian, which can then be easily used with a standard Hermitian eigensolver. This idea was initially proposed in the context of single-reference coupled cluster,⁶³ and further extended to multi-reference coupled cluster⁶⁴ and to the application of the shifted- B_k method to selected CI.¹⁵

We denote here the projection of the TC Hamiltonian in the basis of Slater determinants \mathcal{B} by the matrix $\bar{\mathbf{H}}$, and we use an iterative scheme to obtain a specific right-eigenstate of $\bar{\mathbf{H}}$. The main idea is to iteratively build a state-specific Hermitian effective Hamiltonian $\tilde{\mathbf{H}}^{(\Phi)}$ which has, at convergence, the same eigenvalue and eigenvector as $\bar{\mathbf{H}}$ for the desired state. Remark that as the present scheme is state-specific, all the other eigenvectors and eigenvalues of $\tilde{\mathbf{H}}^{(\Phi)}$ are not considered, and hence they are allowed to be different from those of $\bar{\mathbf{H}}$.

We search for a symmetric dressing matrix $\Delta^{(\Phi)}$ parameterized by Φ such that

$$\tilde{\mathbf{H}}^{(\Phi)} \Phi = (\bar{\mathbf{H}} + \Delta^{(\Phi)}) \Phi = \bar{\mathbf{H}} \Phi. \quad (19)$$

The simplest solution is a diagonal dressing matrix:

$$\Delta_{II}^{(\Phi)} = \frac{V_I}{c_I}; \Delta_{I,J \neq I}^{(\Phi)} = 0 \quad (20)$$

obtained from the *dressing vector* $V^{(\Phi)}$ defined as

$$V_I^{(\Phi)} = \sum_{K=1}^{N_{\text{det}}} (\bar{\mathbf{H}}_{IK} - \mathbf{H}_{IK}) c_K. \quad (21)$$

Iteratively dressing the Hamiltonian using the previously obtained eigenvector converges to the desired solution.

Nevertheless, using Eq. (20) is numerically unstable since the coefficients c_I can be zero, or extremely small. Instead, we propose to use a column dressing in column L , chosen because $|c_L|$ has the largest magnitude among all coefficients:

$$\Gamma_{IL}^{(\Phi)} = \frac{V_I^{(\Phi)}}{c_L}; \Delta_{I,J \neq L}^{(\Phi)} = 0 \quad (22)$$

Then, we build the symmetric matrix $\Lambda^{(\Phi)}$ such that $\Delta_{LI}^{(\Phi)} = \Delta_{IL}^{(\Phi)} = \Gamma_{IL}^{(\Phi)}$. Doing this, the effect of the Jastrow factor is counted twice, and the double-counting can be removed by introducing an extra term in the diagonal element cancelling out the double counting:

$$\begin{cases} \Delta_{LL}^{(\Phi)} = 2\Gamma_{LL}^{(\Phi)} - \sum_{K=1}^{N_{\text{det}}} \Gamma_{KL} c_K / c_L \\ \Delta_{LI}^{(\Phi)} = \Delta_{IL}^{(\Phi)} = \Gamma_{IL}^{(\Phi)} & I \neq L \\ \Delta_{IK}^{(\Phi)} = \Delta_{KI}^{(\Phi)} = 0 & I \neq L, K \neq L \end{cases} \quad (23)$$

As a result, the dressing matrix $\Lambda^{(\Phi)}$ and thus, the matrix $\tilde{\mathbf{H}}^{(\Phi)}$ to be diagonalized is symmetric.

To summarize, given a Jastrow factor and a CI wave function Φ ,

1. within a VMC calculation, we sample the dressing vector V^Φ and the variational energy,
2. we build the matrix elements of $\tilde{\mathbf{H}}^{(\Phi)}$ by combining the matrix elements of \mathbf{H} with the sampled quantities
3. we employ Davidson's algorithm to extract the ground state right eigenvector of the matrix $\tilde{\mathbf{H}}^{(\Phi)}$, and we update the wave function Φ .

These steps are iterated until convergence. Note that only the N_{det} elements Γ_{IL} need to be sampled in QMC. All the other required quantities (\mathbf{H}_{JK}) can be obtained with standard deterministic computational chemistry codes.

C. Calculation of the dressing elements

1. Naive estimators

The integrals required to build the dressing elements $\Lambda^{(\Phi)}$ are

$$\gamma_I^{(\Psi)} = \langle D_I | \hat{H}_J | \Phi \rangle - \langle D_I | \hat{H} | \Phi \rangle, \quad (24)$$

and the second term of the right-hand side of Eq. (24) can be evaluated analytically for Gaussian-type one-electron orbitals. The calculation of the first term of the right-hand side of Eq. (24) is however not trivial since it cannot be calculated in closed form for a general functional form of the Jastrow factor. It involves three-electron integrals which can be rather expensive and it requires the storage of six-index quantities. We will show in this section how these integrals can be effectively evaluated in a VMC framework.

The most basic estimator to evaluate the integrals

$$\alpha_I^{(\Psi)} = \langle D_I | \hat{H}_J | \Phi \rangle \quad (25)$$

in a VMC sampling is

$$a_I^{(\Psi)}(\mathbf{r}) = \frac{D_I(\mathbf{r})}{\Phi(\mathbf{r})} E_{\text{loc}}^{(J)}(\mathbf{r}), \quad (26)$$

where the local energy is defined as

$$E_{\text{loc}}^{(J)}(\mathbf{r}) = \frac{\hat{H}_J \Phi(\mathbf{r})}{\Phi(\mathbf{r})} = \frac{\hat{H}(\Phi(\mathbf{r}) e^{J(\mathbf{r})})}{\Phi(\mathbf{r}) e^{J(\mathbf{r})}}, \quad (27)$$

and Φ^2 is the density used to draw the samples of M configurations:

$$\langle a_I^{(\Psi)} \rangle_{\Phi^2} \equiv \frac{1}{M} \sum_{i=1}^M a_I^{(\Psi)}(\mathbf{r}_i) \xrightarrow{M \rightarrow \infty} \alpha_I^{(\Psi)}. \quad (28)$$

In order to reduce the fluctuations on the dressing elements, we will present two improved estimators based on the so-called *control variates* technique^{65,66}. The general idea is to combine the estimator of Eq (26) with a correlated function in the $3N$ -dimensional space of which the integral can be evaluated in a deterministic way.

2. Improved estimator

Consider the modified electronic Hamiltonian

$$\hat{H}_V = \hat{T} + \hat{V}_{\text{n-e}} + \sum_{i < j} V(r_{ij}) \quad (29)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, \hat{T} is the usual kinetic operator, $\hat{V}_{\text{n-e}}$ is the Coulomb interaction between electrons and nuclei, and $V(r_{ij})$ is a model potential for electron-electron interaction which allows an efficient and deterministic evaluation of integrals in a Slater determinant basis $\langle D_I | V | D_J \rangle$. The local energy associated with this potential reads

$$E_{\text{loc}}^{(V)}(\mathbf{r}) = \frac{\hat{H}_V \Phi(\mathbf{r})}{\Phi(\mathbf{r})}. \quad (30)$$

The first improved estimator we propose is

$$\bar{a}_I^{(V)}(\mathbf{r}) = \frac{D_I(\mathbf{r})}{\Phi(\mathbf{r})} (E_{\text{loc}}^{(J)}(\mathbf{r}) - E_{\text{loc}}^{(V)}(\mathbf{r})) + \bar{\beta}_I^{(V)}, \quad (31)$$

where the control variate integrals

$$\bar{\beta}_I^{(V)} = \langle D_I | \hat{H}_V | \Phi \rangle \quad (32)$$

are analytically known, *i.e.* with zero statistical fluctuations.

We have first used the pure Coulomb potential for the dressing:

$$V_1(r_{ij}) = \frac{1}{r_{ij}}. \quad (33)$$

This boils down to using the usual Coulomb integrals as a reference. If no Jastrow factor is used, $\hat{H}_J = \hat{H}$ and $\bar{a}_I^{(V)}(\mathbf{r}) = \bar{\beta}_I^{(V)}$ with zero variance. Therefore, we expect the magnitude of the fluctuations of $\bar{a}_I^{(V)}$ to increase with the complexity of the Jastrow factor, even though the fluctuations are expected to remain small.

Although the estimator of Eq. (31) has lower fluctuations than the bare estimator (Eq. (26)), one can notice that, as long as the Jastrow factor $J(\mathbf{r})$ satisfies the cusp conditions, the local energy $E_{\text{loc}}^{(J)}(\mathbf{r})$ does not diverge when $r_{12} \rightarrow 0$. Therefore the introduction of the usual local energy $E_{\text{loc}}^{(V)}(\mathbf{r})$ in $\bar{a}_I^{(V)}(\mathbf{r})$ introduces a divergence when $r_{12} \rightarrow 0$ due to the bare Coulomb potential. To eliminate this problem, one can simply replace the bare Coulomb potential by a repulsive non-divergent potential and we propose here to use the long-range component of the potential commonly used in range-separated DFT:⁶⁷

$$V_2(r_{ij}) = \frac{\text{erf}(\mu r_{ij})}{r_{ij}}. \quad (34)$$

As such a potential is used here only to reduce the fluctuations of $\bar{a}_I^{(V)}(\mathbf{r})$, the choice of the parameter μ is arbitrary and does not introduce any bias, so it can be optimized to minimize the fluctuations of $\bar{a}_I^{(V)}$. As opposed to range-separated DFT where the common choice for μ is 1/2, here we use large values of μ (typically larger than 10), such that the model potential becomes close to the Coulomb potential, but without the divergence. As expected, we have observed that V_2 further reduces the fluctuations of the sampled quantities.

3. Further improved estimator

In order to propose a further improved estimator we introduce an auxiliary Jastrow factor \mathcal{J} which depends on a set of parameters $\{\mathbf{p}\}$, such that the corresponding TC Hamiltonian

$$\hat{H}_{\mathcal{J}} \equiv e^{-\mathcal{J}(\mathbf{p})} \hat{H} e^{\mathcal{J}(\mathbf{p})}, \quad (35)$$

allows a deterministic calculation for the integrals in the basis of determinants. With such a Jastrow factor, we can define the zero-variance (ZV) estimator

$$\bar{a}_I(\mathbf{r}) = \frac{D_I(\mathbf{r})}{\Phi(\mathbf{r})} \left(E_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) - E_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) \right) + \bar{\beta}_I, \quad (36)$$

with

$$E_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) = \frac{\hat{H}_{\mathcal{J}} \Phi(\mathbf{r})}{\Phi(\mathbf{r})}, \quad (37)$$

and

$$\bar{\beta}_I = \langle D_I | e^{-\mathcal{J}(\mathbf{p})} \hat{H} e^{\mathcal{J}(\mathbf{p})} | \Phi \rangle. \quad (38)$$

While the Jastrow factor \mathcal{J} is selected to mimic J , by optimizing the parameters \mathbf{p} , the statistical fluctuations of the estimators of interest can be significantly reduced.⁶⁸

The control variate quantities $\bar{\beta}_I$ can be prohibitively expensive to calculate, even in close form, due to the three-electrons terms (which generate a six-index tensor) inherent to any TC Hamiltonian. To avoid this complexity, we introduce the modified local energy

$$\epsilon_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) = E_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) - E_{3-e}^{(\mathcal{J})}(\mathbf{r}) \quad (39)$$

where we have dropped the three-electron contributions involved in the computation of the local energy $E_{3-e}^{(\mathcal{J})}$. Therefore, the improved estimator becomes

$$\tilde{a}_I(\mathbf{r}) = \frac{D_I(\mathbf{r})}{\Phi(\mathbf{r})} \left(E_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) - \epsilon_{\text{loc}}^{(\mathcal{J})}(\mathbf{r}) \right) + \tilde{\beta}_I, \quad (40)$$

with

$$\tilde{\beta}_I = \langle D_I | e^{-\mathcal{J}(\mathbf{p})} \hat{H} e^{\mathcal{J}(\mathbf{p})} - E_{3-e}^{(\mathcal{J})} | \Phi \rangle. \quad (41)$$

Note that the estimator of Eq. (40) is not biased since the three-electron terms arising from the Jastrow factor $J(\mathbf{r})$ are taken into account in $\tilde{a}_I(\mathbf{r})$, and the use of $\epsilon_{\text{loc}}^{(\mathcal{J})}(\mathbf{r})$ in Eq. (40) is only here to reduce the fluctuations. Therefore, the set of parameters $\{\mathbf{p}\}$ can still be optimized to minimize the variance of the dressing elements. One way to do that is, for instance, through minimizing the sum of the variance of integrals weighted with the squared CI coefficients:

$$\sum_{I=1}^{N_{\text{det}}} c_I^2 \text{Var}(\tilde{a}_I) = \text{Var} \left(E_{\text{loc}}^{(\mathcal{J})} - \epsilon_{\text{loc}}^{(\mathcal{J})} \right). \quad (42)$$

4. Choice of \mathcal{J} for further improved estimators

Several correlation factors that allow a deterministic calculation of integrals have been proposed in the literature of TC methods.^{46,47,69,70} We chose here to consider a two-electron Jastrow factor accounting for the cusp condition and the short-range part of the Coulomb hole, together with a one-body Jastrow factor which allows to compensate for the effect of the two-body Jastrow on the one-body density. Therefore, the general form of such a Jastrow factor reads

$$\mathcal{J} = \sum_{i<j}^N u(r_{ij}; \mu) - \sum_{i=1}^N \sum_{A=1}^M v(r_{iA}; \beta_A), \quad (43)$$

where, M is the number of nuclei, r_{iA} is the distance between the i -th electron and the A -th nucleus, and $\{\beta_A\}$ are some positive parameters.

Regarding the two-electron Jastrow factor, we used the recently proposed correlation factor tuned by a single-parameter⁷⁰

$$u(r_{ij}; \mu) = \frac{r_{ij}}{2} \left[1 - \text{erf}(\mu r_{ij}) \right] - \frac{\exp\left[-(\mu r_{ij})^2\right]}{2\sqrt{\pi}\mu}, \quad (44)$$

which imposes the electron-electron cusp conditions and whose corresponding TC Hamiltonian reproduces the effective Hamiltonian of RSDFT at leading-order in $1/r_{ij}$. The explicit form of the TC Hamiltonian obtained with such a Jastrow factor was given in Ref. 70.

Turning now to the one-body Jastrow factor chosen in the present work, its functional form reads

$$v(r_{iA}; \beta_A) = 1 - \exp(-\beta_A r_{iA}^2). \quad (45)$$

Within these definitions, the TC Hamiltonian corresponding to \mathcal{J} is given by

$$\begin{aligned} \hat{H}_{\mathcal{J}} = \hat{H} &- \sum_i \hat{v}[\{\beta_A\}](\mathbf{r}_i) - \sum_{i<j} (\hat{K}[\mu](\mathbf{r}_i, \mathbf{r}_j) + \hat{V}[\{\beta_A\}](\mathbf{r}_i, \mathbf{r}_j)) \\ &- \sum_{i<j<k} \hat{L}[\mu](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k), \end{aligned} \quad (46)$$

where the analytical expressions of the additional terms with respect to \hat{H} are

$$\begin{aligned} \hat{v}[\{\beta_A\}](\mathbf{r}_i) = &- \sum_A \beta_A e^{-\beta_A r_{iA}^2} \left[3 - 2\beta_A r_{iA}^2 + 2(\mathbf{r}_i - \mathbf{R}_A) \cdot \nabla_i \right] \\ &+ 2 \left[\sum_A \beta_A e^{-\beta_A r_{iA}^2} (\mathbf{r}_i - \mathbf{R}_A) \right]^2, \end{aligned} \quad (47)$$

$$\begin{aligned} \hat{K}[\mu](\mathbf{r}_i, \mathbf{r}_j) = &\frac{1 - \text{erf}(\mu r_{ij})}{r_{ij}} - \frac{\mu}{\sqrt{\pi}} e^{-(\mu r_{ij})^2} \\ &+ \frac{(1 - \text{erf}(\mu r_{ij}))^2}{4} - \left(\text{erf}(\mu r_{ij}) - 1 \right) \frac{\partial}{\partial r_{ij}}, \end{aligned} \quad (48)$$

$$\hat{V}[\{\beta_A\}](\mathbf{r}_i, \mathbf{r}_j) = - \left(\frac{1 - \operatorname{erf}(\mu r_{ij})}{r_{ij}} \right) \sum_A \beta_A \quad (49)$$

$$\times (\mathbf{r}_i - \mathbf{r}_j) \cdot \left[(\mathbf{r}_i - \mathbf{R}_A) e^{-\beta_A r_{iA}^2} - (\mathbf{r}_j - \mathbf{R}_A) e^{-\beta_A r_{jA}^2} \right],$$

and

$$\hat{L}[\mu](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = \frac{1 - \operatorname{erf}(\mu r_{ij})}{2r_{ij}} \mathbf{r}_{ij} \cdot \frac{1 - \operatorname{erf}(\mu r_{ik})}{2r_{ik}} \mathbf{r}_{ik} \quad (50)$$

$$+ \frac{1 - \operatorname{erf}(\mu r_{ji})}{2r_{ij}} \mathbf{r}_{ji} \cdot \frac{1 - \operatorname{erf}(\mu r_{jk})}{2r_{jk}} \mathbf{r}_{jk}$$

$$+ \frac{1 - \operatorname{erf}(\mu r_{ki})}{2r_{ik}} \mathbf{r}_{ki} \cdot \frac{1 - \operatorname{erf}(\mu r_{kj})}{2r_{kj}} \mathbf{r}_{kj}.$$

The computation of analytical or mixed analytical integrals on a Gaussian basis for $\hat{K}[\mu](\mathbf{r}_i, \mathbf{r}_j)$ and $\hat{L}[\mu](\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)$ have been given in Ref. 70, and similar schemes have been used for the integrals involving $\hat{v}[\{\beta_A\}](\mathbf{r}_i)$ and $\hat{V}[\{\beta_A\}](\mathbf{r}_i, \mathbf{r}_j)$.

IV. NUMERICAL RESULTS

In the present section we investigate the efficiency of the TC approach by performing a series of test calculations on both atomic and molecular systems. The wave function calculations were performed with the QUANTUM PACKAGE⁷¹ program, and the QMC calculations were made with QMC=CHEM.^{72,73} A plug-in was developed in QUANTUM PACKAGE to read the elements sampled with QMC=CHEM and add the deterministic control variate quantity for the dressing of the Hamiltonian to be diagonalized. Throughout this work, the initial wave function used in the iterative dressing is the ground state eigenvector of the usual Hamiltonian and the basis of Slater determinants is kept fixed.

In Sec. IV A we report a detailed study on the Be atom in order to compare the present optimization scheme with the usual linear method and stochastic reconfiguration. Then, in Sec. IV B we perform more realistic calculations with wave functions containing several hundreds of thousands of Slater determinants on atomic and molecular systems.

A. The Beryllium atom

In the present section we use a small system, the Be atom, in order to investigate several numerical aspects of the present work: i) the convergence of the iterative scheme described in Sec. III B used to obtain the right-eigenvector of a given non-Hermitian matrix, and ii) the effect of the incompleteness of the basis set on the discrepancy between the true variational minimum of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ and $\hat{H}_{\text{TC}}^{\mathcal{B}}$ as discussed in Sec. III. The numerical study of the reduction of the statistical fluctuations of the present scheme through improved estimators will be presented in Sec. IV B. Throughout Sec. IV A, we project both the TC and VMC Hamiltonians on the N -electron basis set \mathcal{B} made of the Hartree-Fock determinant in the cc-pCVDZ atomic basis set and all singly and doubly excited determinants, which results in a set of about 350 determinants. Also,

we use the following Jastrow factor $J(\mathbf{r})$ to define both the TC and VMC Hamiltonians

$$J(\mathbf{r}) = \sum_{i < j} u(r_{ij}; \mu), \quad (51)$$

with $\mu = 1.0$. In the case of the TC Hamiltonian, thanks to the simple analytical form of $J(\mathbf{r})$, the two- and three-body integrals involved in the TC Hamiltonian can be computed exactly, and we therefore avoid any problems related to the stochastic sampling inherent to VMC.

1. Iterative scheme to obtain right-eigenvector of the TC Hamiltonian

To analyze the iterative scheme leading to the lowest right-eigenvector of the TC Hamiltonian, we built explicitly the matrix of the TC Hamiltonian within \mathcal{B} and obtained as a reference the exact TC ground state eigenvalue and eigenvectors within \mathcal{B} by using a non-Hermitian eigensolver present in the LAPACK⁷⁴ library.

We report in Fig. 1 the convergence of the absolute error between the exact ground state eigenvalue and that obtained with the iterative scheme at a given iteration. From Fig. 1 one can notice that the iterative scheme converges, although not in a monotonic way for the first five iterations, towards the exact energy. A detailed analysis of the data shows that an error of about ~ 0.6 mH is reached in only two iterations. This experiment validates that without statistical fluctuations, this method returns the right eigenvector of the TC Hamiltonian within chemical accuracy in about 2 – 3 iterations. This observation appears to be true for more complex systems as it will be shown in the next sections.

One can now focus on the efficiency of the TC iterative eigensolver with respect to the optimization in the sense of the VMC energy (*i.e.* in the presence of the Jastrow factor of Eq. (51)). We reported in Fig. 1 the convergence of the VMC energy of the wave function at a given iteration of the iterative scheme. From Fig. 1 it appears that the first two iterations lower the VMC energy by about ~ 40 mH, the successive iterations having only a minor impact on the VMC energy are within two statistical deviations.

2. Effect of an incomplete basis set and comparison with other optimization methods

Having validated the approach of the iterative dressing technique to obtain the eigenvector of the TC Hamiltonian in a given Slater determinant basis set \mathcal{B} , we propose here to investigate the effect of the incompleteness of such a basis set. Indeed, as shown in the appendix (see Sec. VI), the ground state eigenvector of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ (*i.e.* the VMC Hamiltonian projected in \mathcal{B}) does not necessarily coincide with the ground state right-eigenvector of $\hat{H}_{\text{TC}}^{\mathcal{B}}$ (*i.e.* the TC Hamiltonian projected in \mathcal{B}) as long as \mathcal{B} is incomplete. Therefore the variational energy obtained with the latter is necessarily an upper

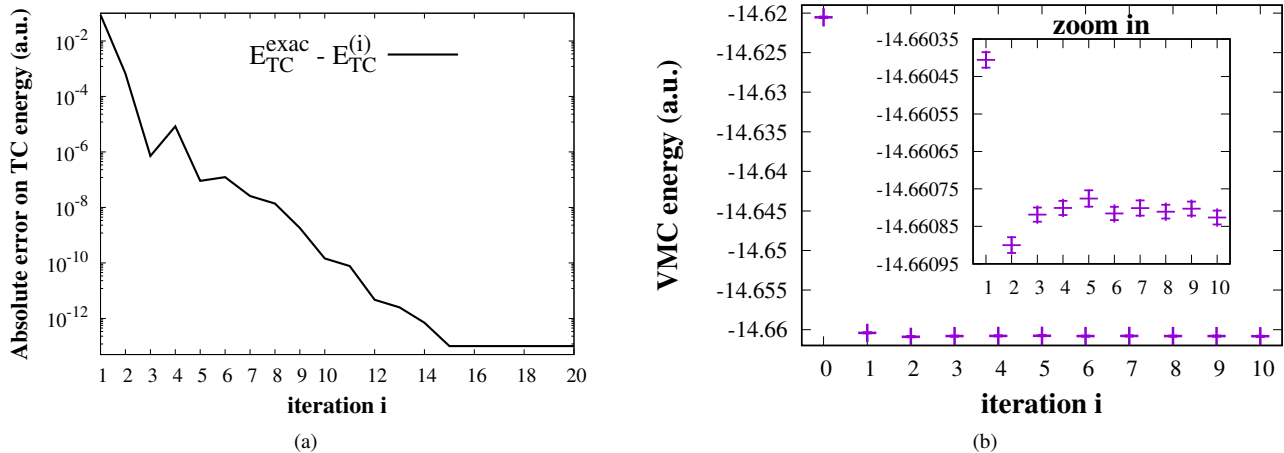


FIG. 1: Convergence of the TC and VMC energies along the iterative dressing. (a) Difference between the TC energy obtained by a direct diagonalization and that obtained by the iterative dressing technique. (b) VMC energy of the CI-J wavefunction before optimizing ($i = 0$) and during the iterative dressing.

bound to the ground state eigenvalue of $\hat{H}_{\text{VMC}}^{\text{B}}$, and we propose here to quantify this on a simple system such as the Be atom.

We sampled the matrix elements of $\hat{H}_{\text{VMC}}^{\text{B}}$ together with the corresponding overlap matrix and solved the GEP of the form of Eq. (8) to obtain the ground state eigenvector. The matrix elements are impacted by statistical fluctuations, so we have made simulations with increasingly large numbers of samples. To show the impact of these fluctuations on the energy of the obtained eigenvector, we report in Fig 2 the convergence of the variational energy of the eigenvector as a function of the wall-clock time used to sample the matrices with 72 CPU cores. We also report on the same figure the convergence of the SR approach applied to the same problem with a step size $\tau = 0.01$ (Eq.(9)). We compare these variational energies with those obtained with the two first iterations of the iterative dressing scheme. In order to make the comparison more fair, instead of using the exact analytical integrals involved in $\hat{H}_{\text{TC}}^{\text{B}}$, we sampled the dressing elements in VMC using the estimator $\bar{a}_i^{(V)}$ of Eq. (31) with the Coulomb potential of Eq. (33). At each iteration, the dressing vector was sampled using a run of 50 minutes on 72 cores. From Fig. 2 we can observe that the variational energies obtained with the two schemes converge essentially to the same energy as the one obtained with the TC optimization scheme, the latter being the lowest obtained within the CPU time spent. One can also notice the slow convergence and the erratic behaviour of the GEP approach, which is caused by the large fluctuations of the sampled matrix elements¹⁷. Regarding the SR approach, one can also see the slow rate of convergence. These calculations show that, even for a basis of Slater determinants far from being complete such as the CI with single and double substitutions (CISD) in a cc-pCVDZ basis set, the error with respect to the true minimum is negligible, and in that precise case the new scheme provides a lower variational energy than the two other schemes based on usual VMC approaches. This

demonstrates that the error induced by a finite sampling are more important than the finite basis approximations.

B. Large wave function optimizations on atomic and molecular systems

The optimization of CI coefficients in all-electron calculations is considered a difficult task because of the wide fluctuations of the sampled matrix elements. The large magnitude of the contributions to the local energy of the electrons close to the nucleus are responsible for these large fluctuations. As illustrated in Fig. 2, this strongly impacts the rate of convergence of the optimization algorithms. Here, we illustrate the efficiency of our improved estimators on all-electron calculations of C_2 , N_2 , O_2 , and F_2 with expansions made of a few hundred thousand parameters. It is worth mentioning that point-group symmetry was not exploited in the optimization.

Throughout this section, the following Jastrow factor was used:

$$J(\mathbf{r}) = \sum_{i < j}^N \frac{r_{ij}}{2(1 + b r_{ij})} - \sum_{i=1}^N \sum_{A=1}^M \left(\frac{\alpha_A r_{iA}}{1 + \alpha_A r_{iA}} \right)^2. \quad (52)$$

We performed all electrons calculations using Dunning's cc-pVTZ basis set.⁷⁵, and in order to investigate the behaviour of the present schemes within pseudo-potential calculations, we report calculations in the case of the N_2H_4 molecule using the Burkatzki-Filippi-Dolg effective core potentials (ECPs)⁷⁶ in the compatible double zeta basis set. The energy was computed in the determinant localization approximation.⁷⁷

For all diatomics, the bond lengths were taken to be the experimental ones given by Huber and Herzberg.⁷⁸ The geometry of N_2H_4 is the experimental geometry. The Jastrow factors were optimized at the Hartree-Fock (HF) level with a single determinant, and then frozen-core CIPSI calculations

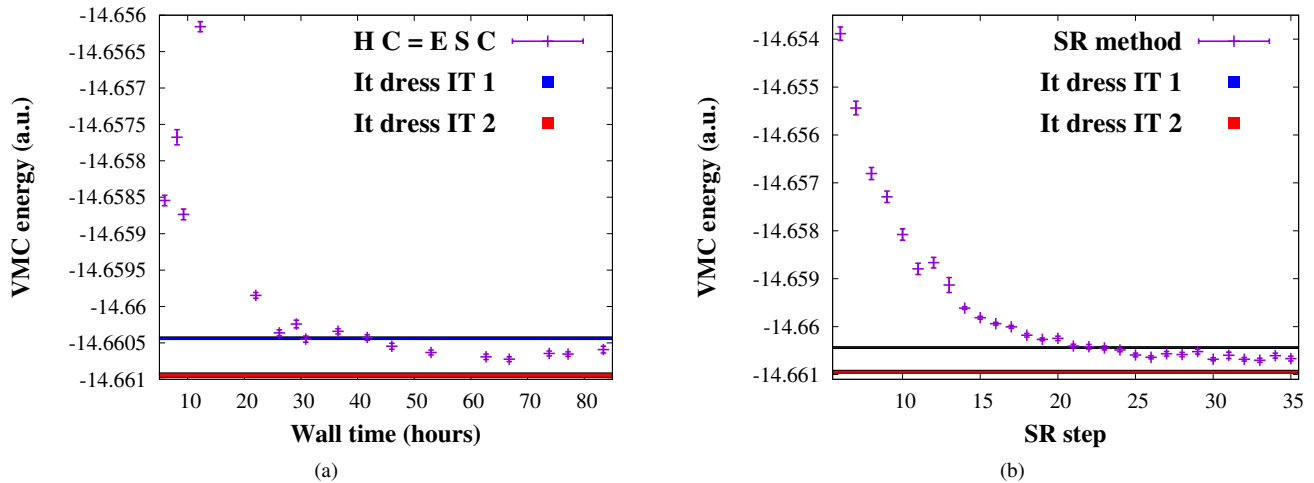


FIG. 2: Convergence of the variational energies of the ground state eigenvector of (a) the GEP method and (b) the SR method. On each figure we report the variational energies obtained with the two first iterations of the dressing scheme.

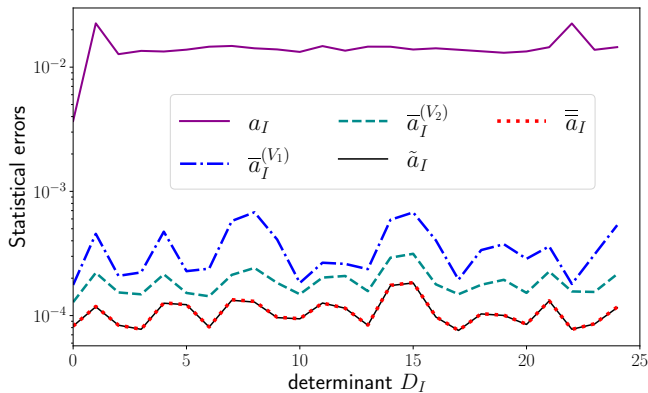


FIG. 3: Statistical errors of the estimators: a_I (Eq. (26)), $\bar{a}_I^{(V_1)}$ and $\bar{a}_I^{(V_2)}$ (Eq. (31)), $\bar{\bar{a}}_I$ (Eq. (36)), and \tilde{a}_I (Eq. (40)), obtained in a common VMC sampling.

were made in the full valence CI space using Hartree-Fock orbitals in the cc-pVTZ basis set to generate the initial CI expansions.

1. Reduction of statistical fluctuations through control variates

We begin this study by investigating the reduction of statistical fluctuations of the sampled quantities using the various control variates presented in Sec. III C. We report in Fig. 3 a comparison between the statistical errors obtained in a single VMC run for the 25 most important determinants of the N_2 molecule using different estimators: i) the standard estimator a_I given in Eq. (26); ii) the improved estimators $\bar{a}_I^{(V_1)}$ and $\bar{a}_I^{(V_2)}$ of Eq. (31), corresponding to the model potentials of Eqs. (33) and (34) with $\mu = 10$, respectively; iii) the TC-correlated estimators $\bar{\bar{a}}_I$ and \tilde{a}_I (Eqs. (36) and (40)) which involve the Jas-

trow factor \mathcal{J} with the parameters given in table I. For N_2 , as well as for other systems considered in the following, the parameters were optimized to reduce the fluctuations of the weighted variance given in Eq (42).

As apparent from Fig. 3, the fluctuations of the improved estimators are much smaller than those of the standard estimator. The singularity-free potential (estimator $\bar{a}_I^{(V_2)}$) yields smaller statistical errors than the Coulomb potential (estimator $\bar{a}_I^{(V_1)}$). The noise is further reduced using the TC-based estimators $\bar{\bar{a}}_I$ and \tilde{a}_I , and we can observe that neglecting the three-electron terms in the local energy does not increase the fluctuations. The latter is extremely important from a computational perspective as the TC calculations without any of the three-electron terms has essentially the same computational scaling as a standard calculation using the regular Hamiltonian (*i.e.* without any Jastrow factor).

To get an overview of the reduction of the statistical error on the complete set of coefficients, we compared for each estimator a the ratio of the sum of the standard deviations with the one obtained with the bare estimator α using the ratio

$$r(a) = \frac{\sum_{I=1}^{N_{\text{det}}} \sigma(a_I)}{\sum_{I=1}^{N_{\text{det}}} \sigma(\alpha_I)}. \quad (53)$$

We obtained $r(\bar{a}_I^{(V_1)}) = 0.027$, $r(\bar{a}_I^{(V_2)}) = 0.022$, $r(\bar{\bar{a}}_I) = 0.009$, $r(\tilde{a}_I) = 0.009$, showing that the reduction of statistical error is observed for all determinants and not only the most important ones.

2. Optimization of large wave functions for molecular systems

Having established the estimator with the lowest statistical fluctuations in Sec. IV B 1, we optimize the CI parameters of the CIPSI-Jastrow wave functions using the estimator $\tilde{a}_I(\mathbf{r})$

TABLE I: Parameters and energies (a.u.) associated with the different wave functions. The convergence of the energy with the iterations of optimization is shown in the bottom part of the table.

	C ₂	N ₂	O ₂	F ₂	N ₂ H ₄
Basis set	cc-pVTZ	cc-pVTZ	cc-pVTZ	cc-pVTZ	BFD/cc-pVDZ
N_{det}	287 998	298 250	123 711	191 538	1 000 000
r_0 (Å)	1.2425	1.0977	1.2075	1.4119	–
b	1.382	1.829	1.923	2.206	1.000
α	0.882	0.691	0.882	0.691	N:0.690, H:0.170
μ	0.738	2.161	2.263	1.619	–
β	0.635	4.288×10^{-3}	6.209×10^{-3}	0.410	–
HF energy	–75.40144	–108.98347	–149.65257	–198.75204	–21.73877
HF/J energy	–75.6587(1)	–109.2668(1)	–150.0060(1)	–199.0884(1)	–22.1008(1)
Selected CI energy	–75.77204	–109.35465	–150.09752	–199.26114	–22.11353
Energies along the optimization:					
CI/J IT 0	–75.8808(1)	–109.4785(1)	–150.2451(2)	–199.3766(3)	–22.1969(5)
CI/J IT 1	–75.8896(1)	–109.4986(1)	–150.2670(1)	–199.4297(2)	–22.2297(3)
CI/J IT 2	–75.8902(1)	–109.4986(1)	–150.2671(1)	–199.4300(2)	–22.2297(3)
CI/J IT 3	–75.8903(1)	–109.4989(1)	–150.2673(1)	–199.4306(2)	–22.2302(3)

given in Eq. (40). We begin our study with all electron calculations on the C₂, N₂, O₂ and F₂ molecules. The parameters of \mathcal{J} for each molecule and the corresponding VMC energies are reported in table I.

For all systems, we observe a sub-milliHartree convergence after two optimization iterations, although a single iteration could be sufficient. The gains in VMC energy are 10 mH for C₂, 20 mH for N₂, 22 mH for O₂ and 53 mH for F₂. It is remarkable that the optimization is still stable for the F₂ molecule which has the largest total energy. These results confirm that this optimization method can be used as a black box method in routine calculations.

We conclude our study by a calculation on the N₂H₄ molecule using a BFD ECP with the corresponding double zeta basis set. For this system, 1 000 000 CI coefficients are optimized. The results of the optimization are presented in table I and show that two iterations are required to converge to the minimum energy which is about $\sim 32 mE_h$ below the initial energy. This result confirms that the present scheme allows the optimization in an ECP framework of very large CI expansions even with the simplest estimator (Eqs. (31) and (33)).

V. SUMMARY

We have presented an iterative method to optimize large CI expansions in the presence of a general correlation factor. The main idea is to consider a similarity-transformation of the Hamiltonian by the Jastrow factor which results in an effective Hamiltonian, the so-called TC Hamiltonian, having the same right-eigenvectors than the usual ground state VMC eigenvector in the limit of a complete basis. The effect of the Jastrow factor in the TC approach can be written as an additive dressing of the standard Hamiltonian. The QMC simulations are then used only to sample the quantities required for such a dressing, and to compute the variational energy associated with the wave function. The main advantages of the

present approach are i) a large part of the quantities required to optimize the wave function are analytical (*i.e.* with zero fluctuations), ii) the number of quantities to sample is equal to the number of determinants, and iii) the fluctuations of the sampled quantities are small.

After having illustrated in Sec. IV A the robustness of the present approach on calculations on the Be atom even far from the complete basis set limit (*i.e.* in realistic cases), we have shown in Sec. IV B its efficiency on wave functions made of a few hundred thousand Slater determinants. We demonstrated in Sec. IV B 1 how one can significantly lower the statistical fluctuations of the computed quantities thanks to control variates, and in Sec. IV B 2 how it performs on molecular systems with wave functions with $10^5 - 10^6$ determinants. The efficiency of this approach comes from the mixing of the deterministic transcorrelated method, with the QMC method: a large fraction of the needed matrix elements can be computed in the standard framework of wave function methods, and only a small number of contributions come from the QMC simulations. Therefore, the quantities of interest have by nature a very low variance. It is possible to put more computational weight on the deterministic part of the calculation to reduce even more the statistical fluctuations, and the user has the flexibility to find the best compromise between the computational cost of the control variates and the QMC simulations.

We have also shown that this method has several advantages regarding computational considerations. First, the memory required in the QMC code is minimal: only N_{det} quantities need to be stored since the overlap matrix is not needed, and these quantities are computed using building blocks that are already needed for the computation of the energy. Hence, the extra computational cost is also minimal. Finally, it is important to mention that the fast convergence of the method (less than three iterations) is extremely important when considering massively parallel simulations. The only necessary blocking communications take place at the beginning and at the end of an iteration, so it is preferable to have an optimization algorithm with long computing phases and very few iterations than

the opposite.

In this paper, we have limited our examples to the optimization of CI expansions, but it is worth mentioning that this method can of course be applied in a super-CI framework for the optimization of the coefficients of the molecular orbitals. We plan to elaborate more on these aspects in a subsequent work.

ACKNOWLEDGMENTS

This work was performed using HPC resources from GENCI-TGCC (2021-gen1738) and from CALMIP (Toulouse) under allocation P22001, and was also supported by the European Centre of Excellence in Exascale Computing TREX — Targeting Real Chemical Accuracy at the Exascale. This project has received funding from the European Union’s Horizon 2020 — Research and Innovation program — under grant agreement no. 952165.

VI. APPENDIX

A. VMC, Transcorrelation and projection in an incomplete basis set

The aim of the present section is to make a proper link between the VMC optimization and the TC approach when they are both projected in the same basis set.

1. Definition of the projector on a basis set

Let \mathcal{B} be a N -electron basis set, which can be for instance a set of N -electron Slater determinant

$$\mathcal{B} = \{|\mathbb{I}\rangle, \mathbb{I} = 1, N_{\mathcal{B}}\}, \quad (54)$$

which we will assume orthonormal for the sake of simplicity. Such a basis span a vector space $\mathcal{V}_{\mathcal{B}}$ which is made of any functions $|\Psi_{\mathcal{B}}\rangle$ being a linear combinations of the Slater determinants in \mathcal{B}

$$\mathcal{V}_{\mathcal{B}} = \left\{ |\Psi_{\mathcal{B}}\rangle = \sum_{\mathbb{I} \in \mathcal{B}} c_{\mathbb{I}} |\mathbb{I}\rangle, c_{\mathbb{I}} \in \mathbb{C} \right\}. \quad (55)$$

From \mathcal{B} one can build the projector $\hat{P}_{\mathcal{B}}$

$$\hat{P}_{\mathcal{B}} = \sum_{\mathbb{I}} |\mathbb{I}\rangle \langle \mathbb{I}|, \quad (56)$$

which verifies the projector property

$$\hat{P}_{\mathcal{B}} \hat{P}_{\mathcal{B}} = \hat{P}_{\mathcal{B}}. \quad (57)$$

Such a projector coincide with the identity for all wave functions in $\mathcal{V}_{\mathcal{B}}$

$$\hat{P}_{\mathcal{B}} |\Psi_{\mathcal{B}}\rangle = |\Psi_{\mathcal{B}}\rangle. \quad (58)$$

Therefore, in virtue of Eq.(57) one obtains that

$$\hat{P}_{\mathcal{B}} \hat{P}_{\mathcal{B}} |\Psi_{\mathcal{B}}\rangle = |\Psi_{\mathcal{B}}\rangle, \quad (59)$$

which implies that

$$(\hat{P}_{\mathcal{B}})^{-1} = \hat{P}_{\mathcal{B}}. \quad (60)$$

One can define the *complementary basis set* \mathcal{B}^{\perp} which is the set of Slater determinants such that it completes the basis set

$$\mathcal{B}^{\perp} = \{|\mathbb{L}\rangle \notin \mathcal{B}\}, \quad (61)$$

the corresponding vector space $\mathcal{V}_{\mathcal{B}}^{\perp}$

$$\mathcal{V}_{\mathcal{B}}^{\perp} = \left\{ |\Psi_{\mathcal{B}}^{\perp}\rangle = \sum_{\mathbb{L} \in \mathcal{B}^{\perp}} c_{\mathbb{L}} |\mathbb{L}\rangle \right\} \quad (62)$$

and the corresponding projector $\hat{P}_{\mathcal{B}}^{\perp}$ such that

$$\hat{P}_{\mathcal{B}} + \hat{P}_{\mathcal{B}}^{\perp} = \mathbf{1}, \quad (63)$$

where $\mathbf{1}$ is the identity operator defined on the complete basis made of the reunion of $\mathcal{V}_{\mathcal{B}}$ and $\mathcal{V}_{\mathcal{B}}^{\perp}$

$$\mathbf{1} |\Psi\rangle = |\Psi\rangle \quad \forall |\Psi\rangle \in \mathcal{V}_{\mathcal{B}} \cup \mathcal{V}_{\mathcal{B}}^{\perp}. \quad (64)$$

An important property is that the projector $\hat{P}_{\mathcal{B}}$ and the complementary projector $\hat{P}_{\mathcal{B}}^{\perp}$ are orthogonal

$$\hat{P}_{\mathcal{B}} \hat{P}_{\mathcal{B}}^{\perp} = \hat{P}_{\mathcal{B}}^{\perp} \hat{P}_{\mathcal{B}} = 0. \quad (65)$$

2. Link between the VMC and transcorrelated Hamiltonian within the same basis set

The purpose of the present section is to establish the formal link between the VMC Hamiltonian and the TC one projected in a basis set. Let J be a Jastrow factor, one can then define the VMC Hamiltonian as

$$H_{\text{VMC}} = e^J H e^J, \quad (66)$$

the corresponding TC Hamiltonian as

$$H_{\text{TC}} = e^{-J} H e^J, \quad (67)$$

which are of course related by

$$H_{\text{VMC}} = e^{2J} H_{\text{TC}}. \quad (68)$$

The corresponding operators projected onto a basis set \mathcal{B} are defined as

$$\hat{H}_{\text{VMC}}^{\mathcal{B}} = \hat{P}_{\mathcal{B}} H_{\text{VMC}} \hat{P}_{\mathcal{B}}, \quad (69)$$

$$\hat{H}_{\text{TC}}^{\mathcal{B}} = \hat{P}_{\mathcal{B}} H_{\text{TC}} \hat{P}_{\mathcal{B}}, \quad (70)$$

and we would like to express $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ in terms of $\hat{H}_{\text{TC}}^{\mathcal{B}}$. To do so we write

$$\begin{aligned}\hat{H}_{\text{VMC}}^{\mathcal{B}} &= \hat{P}_{\mathcal{B}} e^{2J} H_{\text{TC}} \hat{P}_{\mathcal{B}} \\ &= \hat{P}_{\mathcal{B}} e^{2J} \mathbf{1} H_{\text{TC}} \hat{P}_{\mathcal{B}} \\ &= \hat{P}_{\mathcal{B}} e^{2J} (\hat{P}_{\mathcal{B}} + \hat{P}_{\mathcal{B}}^{\perp}) (\hat{P}_{\mathcal{B}} + \hat{P}_{\mathcal{B}}^{\perp}) H_{\text{TC}} \hat{P}_{\mathcal{B}}\end{aligned}\quad (71)$$

which, in virtue of Eq. (65), reads

$$\hat{H}_{\text{VMC}}^{\mathcal{B}} = \hat{P}_{\mathcal{B}} e^{2J} \hat{P}_{\mathcal{B}} \hat{H}_{\text{TC}}^{\mathcal{B}} + \hat{P}_{\mathcal{B}} e^{2J} \hat{P}_{\mathcal{B}}^{\perp} H_{\text{TC}} \hat{P}_{\mathcal{B}}. \quad (72)$$

If we define $(e^{2J})_{\mathcal{B}} = \hat{P}_{\mathcal{B}} e^{2J} \hat{P}_{\mathcal{B}}$ and

$$\Delta_{\mathcal{B}}^{\text{TC-VMC}} = \hat{P}_{\mathcal{B}} e^{2J} \hat{P}_{\mathcal{B}}^{\perp} H_{\text{TC}} \hat{P}_{\mathcal{B}} \quad (73)$$

we can then write

$$\hat{H}_{\text{VMC}}^{\mathcal{B}} = (e^{2J})_{\mathcal{B}} \hat{H}_{\text{TC}}^{\mathcal{B}} + \Delta_{\mathcal{B}}^{\text{TC-VMC}}. \quad (74)$$

Of course, in the limit where \mathcal{B} is complete, one has that $\hat{P}_{\mathcal{B}}^{\perp} = 0$ and then $\Delta_{\mathcal{B}}^{\text{TC-VMC}} = 0$.

3. Link between the eigenvectors of the TC and VMC

When working on a complete basis set, we know that the eigenvectors of H_{VMC} coincides with the right-eigenvectors of H_{TC}

$$\begin{aligned}e^J H e^J |\Psi_i\rangle &= E_i e^{2J} |\Psi_i\rangle \\ \Leftrightarrow e^{-2J} e^J H e^J |\Psi_i\rangle &= E_i e^{-2J} e^{2J} |\Psi_i\rangle \\ \Leftrightarrow e^{-J} H e^J |\Psi^i\rangle &= E^i |\Psi^i\rangle.\end{aligned}\quad (75)$$

We want now to find the same kind of relationship when the operators are projected in a basis set \mathcal{B} . We start from the

eigenvalue equation for $\hat{H}_{\text{VMC}}^{\mathcal{B}}$

$$\hat{H}_{\text{VMC}}^{\mathcal{B}} |\Psi_{\mathcal{B}}^i\rangle = E_{\mathcal{B}}^i (e^{2J})_{\mathcal{B}} |\Psi_{\mathcal{B}}^i\rangle, \quad (76)$$

and inserting now the expression of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ in terms of $\hat{H}_{\text{TC}}^{\mathcal{B}}$ one obtains

$$\left((e^{2J})_{\mathcal{B}} \hat{H}_{\text{TC}}^{\mathcal{B}} + \Delta_{\mathcal{B}}^{\text{TC-VMC}} \right) |\Psi_{\mathcal{B}}^i\rangle = E_{\mathcal{B}}^i (e^{2J})_{\mathcal{B}} |\Psi_{\mathcal{B}}^i\rangle \quad (77)$$

and multiplying by the inverse of $(e^{2J})_{\mathcal{B}}$ (which is $(e^{-2J})_{\mathcal{B}}$) from the left it comes

$$\hat{H}_{\text{TC}}^{\mathcal{B}} |\Psi_{\mathcal{B}}^i\rangle + (e^{-2J})_{\mathcal{B}} \Delta_{\mathcal{B}}^{\text{TC-VMC}} |\Psi_{\mathcal{B}}^i\rangle = E_{\mathcal{B}}^i |\Psi_{\mathcal{B}}^i\rangle. \quad (78)$$

One can explicit the term $(e^{-2J})_{\mathcal{B}} \Delta_{\mathcal{B}}^{\text{TC-VMC}}$ which gives

$$(e^{-2J})_{\mathcal{B}} \Delta_{\mathcal{B}}^{\text{TC-VMC}} = \hat{P}_{\mathcal{B}} e^{-J} \left(e^{-J} \hat{P}_{\mathcal{B}} e^J - \mathbf{1} \right) H e^J \hat{P}_{\mathcal{B}}. \quad (79)$$

Defining

$$\delta_{\mathcal{B}} = -e^{-J} \hat{P}_{\mathcal{B}}^{\perp} e^J, \quad (80)$$

one obtains

$$(e^{-2J})_{\mathcal{B}} \Delta_{\mathcal{B}}^{\text{TC-VMC}} = \hat{P}_{\mathcal{B}} e^{-J} \delta_{\mathcal{B}} H e^J \hat{P}_{\mathcal{B}}. \quad (81)$$

Therefore, the fact that the eigenvectors of $\hat{H}_{\text{VMC}}^{\mathcal{B}}$ and $\hat{H}_{\text{TC}}^{\mathcal{B}}$ do not coincide comes from the fact that $\delta_{\mathcal{B}} \neq 0$ in an incomplete basis set.

* scemama@irsamc.ups-tlse.fr

¹ C. J. Calzado, C. Angeli, D. Taratiel, R. Caballol, and J.-P. Malrieu, *J Chem Phys* **131**, 044327 (2009).
² S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli, and M. Olivucci, *J Chem Theory Comput* **8**, 4069 (2012).
³ J. P. Malrieu, R. Caballol, C. J. Calzado, C. de Graaf, and N. Guihéry, *Chem Rev* **114**, 429 (2014).
⁴ W. Liu and M. R. Hoffmann, *Theor. Chem. Acc.* **133**, 1481 (2014).
⁵ M. Hapka, K. Pernal, and O. V. Gritsenko, *J. Phys. Chem. Lett.* **11**, 5883 (2020).
⁶ C. Angeli, C. J. Calzado, R. Cimiraglia, and J.-P. Malrieu, *J. Chem. Phys.* **124**, 234109 (2006).
⁷ G. Li Manni, D. Ma, F. Aquilante, J. Olsen, and L. Gagliardi, *J. Chem. Theory Comput.* **9**, 3375 (2013).
⁸ J. Meller, J. P. Malrieu, and R. Caballol, *J. Chem. Phys.* **104**, 4068 (1996).
⁹ U. S. Mahapatra, B. Datta, and D. Mukherjee, *J. Chem. Phys.* **110**, 6171 (1999).
¹⁰ T. Shiozaki, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **134**,

034113 (2011).

¹¹ A. Savin, in *Theoretical and Computational Chemistry*, Vol. 4 (Elsevier, Waltham, MA, USA, 1996) pp. 327–357.
¹² K. Sharkas, A. Savin, H. J. Aa. Jensen, and J. Toulouse, *J. Chem. Phys.* **137**, 044104 (2012).
¹³ K. Pernal and M. Hapka, *WIREs Comput. Mol. Sci.* **12**, e1566 (2022).
¹⁴ L. E. Nitzsche and E. R. Davidson, *J. Chem. Phys.* **68**, 3103 (1978).
¹⁵ Y. Garniron, A. Scemama, E. Giner, M. Caffarel, and P.-F. Loos, *J. Chem. Phys.* **149**, 064103 (2018).
¹⁶ T. Kato, *Communications on Pure and Applied Mathematics* **10**, 151 (1957).
¹⁷ M. Nightingale and V. Melik-Alaverdian, *Physical review letters* **87**, 043401 (2001).
¹⁸ C. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, *Physical review letters* **98**, 110201 (2007).
¹⁹ J. Toulouse and C. J. Umrigar, *The Journal of chemical physics* **126**, 084102 (2007).
²⁰ J. Toulouse and C. Umrigar, *The Journal of chemical physics* **128**,

- 174101 (2008).
- 21 B. K. Clark, M. A. Morales, J. McMinis, J. Kim, and G. E. Scuseria, *The Journal of chemical physics* **135**, 244105 (2011).
 - 22 S. Sorella, *Physical Review B* **64**, 024512 (2001).
 - 23 M. Casula, C. Attaccalite, and S. Sorella, *The Journal of Chemical Physics* **121**, 7110 (2004), <https://doi.org/10.1063/1.1794632>.
 - 24 S. Sorella, M. Casula, and D. Rocca, *The Journal of chemical physics* **127**, 014105 (2007).
 - 25 E. Neuscamman, C. Umrigar, and G. K.-L. Chan, *Physical Review B* **85**, 045103 (2012).
 - 26 E. Neuscamman, *Journal of chemical theory and computation* **12**, 3149 (2016).
 - 27 L. Zhao and E. Neuscamman, *Journal of chemical theory and computation* **13**, 2604 (2017).
 - 28 G. L. Sleijpen, A. G. Booten, D. R. Fokkema, and H. A. Van der Vorst, *BIT Numerical Mathematics* **36**, 595 (1996).
 - 29 G. L. Sleijpen and H. A. Van der Vorst, *SIAM review* **42**, 267 (2000).
 - 30 I. Sabzevari, A. Mahajan, and S. Sharma, *The Journal of chemical physics* **152**, 024111 (2020).
 - 31 E. R. Davidson, *J. Comput. Phys.* **17**, 87 (1975).
 - 32 A. Harju, B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, *Phys. Rev. Lett.* **79**, 1173 (1997).
 - 33 L. R. Schwarz, A. Alavi, and G. H. Booth, *Physical review letters* **118**, 176403 (2017).
 - 34 I. Sabzevari and S. Sharma, *Journal of chemical theory and computation* **14**, 6276 (2018).
 - 35 D. Luo and B. K. Clark, *Physical review letters* **122**, 226401 (2019).
 - 36 A. Mahajan and S. Sharma, *The Journal of Physical Chemistry A* **123**, 3911 (2019).
 - 37 J. McFarland and E. Manousakis, *Phys. Rev. A* **105**, 032815 (2022).
 - 38 L. Otis and E. Neuscamman, *Physical Chemistry Chemical Physics* **21**, 14491 (2019).
 - 39 J. O. Hirschfelder, *J. Chem. Phys.* **39**, 3145 (1963).
 - 40 S. F. Boys and N. C. Handy, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **309**, 209 (1969).
 - 41 S. F. Boys and N. C. Handy, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **310**, 63 (1969).
 - 42 S. F. Boys and N. C. Handy, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **310**, 43 (1969).
 - 43 S. F. Boys and N. C. Handy, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **311**, 309 (1969).
 - 44 N. Handy, *Molecular Physics* **21**, 817 (1971).
 - 45 N. Handy, *Molecular Physics* **23**, 1 (1972).
 - 46 S. Ten-no, *Chemical Physics Letters* **330**, 169 (2000).
 - 47 S. Ten-no, *Chemical Physics Letters* **330**, 175 (2000).
 - 48 O. Hino, Y. Tanimura, and S. Ten-no, *The Journal of Chemical Physics* **115**, 7865 (2001).
 - 49 N. Umezawa and S. Tsuneyuki, *J. Chem. Phys.* **119**, 10015 (2003).
 - 50 N. Umezawa and S. Tsuneyuki, *International journal of quantum chemistry* **91**, 184 (2003).
 - 51 N. Umezawa and S. Tsuneyuki, *The Journal of chemical physics* **119**, 10015 (2003).
 - 52 N. Umezawa and S. Tsuneyuki, *The Journal of chemical physics* **121**, 7070 (2004).
 - 53 N. Umezawa, S. Tsuneyuki, T. Ohno, K. Shirashi, and T. Chikyow, *The Journal of chemical physics* **122**, 224101 (2005).
 - 54 R. Prasad, N. Umezawa, D. Domin, R. Salomon-Ferrer, and W. A. Lester Jr, *The Journal of chemical physics* **126**, 164109 (2007).
 - 55 H. Luo and A. Alavi, *Journal of chemical theory and computation* **14**, 1403 (2018).
 - 56 K. Liao, T. Schraivogel, H. Luo, D. Kats, and A. Alavi, *arXiv preprint arXiv:2103.03176* (2021).
 - 57 T. Schraivogel, A. J. Cohen, A. Alavi, and D. Kats, *The Journal of Chemical Physics* **155**, 191101 (2021).
 - 58 G. H. Booth, A. J. Thom, and A. Alavi, *The Journal of chemical physics* **131**, 054106 (2009).
 - 59 D. Cleland, G. H. Booth, and A. Alavi, *The Journal of chemical physics* **132**, 041103 (2010).
 - 60 R. Assaraf, S. Moroni, and C. Filippi, *Journal of chemical theory and computation* **13**, 5273 (2017).
 - 61 M. Dash, S. Moroni, A. Scemama, and C. Filippi, *Journal of chemical theory and computation* **14**, 4176 (2018).
 - 62 K. Hirao and H. Nakatsuji, *J. Comput. Phys.* **45**, 246 (1982).
 - 63 I. Nebot-Gil, J. Sánchez-Marín, J. P. Malrieu, J. L. Heully, and D. Maynau, *J Chem Phys* **103**, 2576 (1995).
 - 64 E. Giner, G. David, A. Scemama, and J. P. Malrieu, *J Chem Phys* **144**, 064101 (2016).
 - 65 M. H. Kalos and P. A. Whitlock, in *Monte Carlo Methods* (John Wiley & Sons, Ltd, 2008) pp. 77–106.
 - 66 M. Caffarel, *The Journal of Chemical Physics* **151**, 064101 (2019).
 - 67 J. Toulouse, F. Colonna, and A. Savin, *Physical Review A* **70**, 062505 (2004).
 - 68 R. Assaraf and M. Caffarel, *Physical review letters* **83**, 4682 (1999).
 - 69 M. Nooijen and R. J. Bartlett, *The Journal of chemical physics* **109**, 8232 (1998).
 - 70 E. Giner, *The Journal of Chemical Physics* **154**, 084119 (2021).
 - 71 Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, *et al.*, *Journal of chemical theory and computation* **15**, 3591 (2019).
 - 72 A. Scemama, M. Caffarel, E. Oseret, and W. Jalby, in *High Performance Computing for Computational Science - VECPAR 2012* (Springer, Berlin, Germany, 2013) pp. 118–127.
 - 73 A. Scemama, M. Caffarel, E. Oseret, and W. Jalby, *J. Comput. Chem.* **34**, 938 (2013).
 - 74 E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).
 - 75 T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
 - 76 M. Burkatzki, C. Filippi, and M. Dolg, *J. Chem. Phys.* **126**, 234105 (2007).
 - 77 A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, *J. Chem. Phys.* **151**, 134105 (2019).
 - 78 G. H. K. P. Huber, "Molecular spectra and molecular structure: Constants of diatomic," in *Monte Carlo Methods* (Van Nostrand Reinhold, New York, 1979) pp. 112–490.