

Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow–Slater Expansions

Roland Assaraf, Saverio Moroni, Claudia Filippi

▶ **To cite this version:**

Roland Assaraf, Saverio Moroni, Claudia Filippi. Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow–Slater Expansions. Journal of Chemical Theory and Computation, 2017, 13 (11), pp.5273 - 5281. 10.1021/acs.jctc.7b00648 . hal-01635017

HAL Id: hal-01635017

https://hal.sorbonne-universite.fr/hal-01635017

Submitted on 14 Nov 2017

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.

JCTC Journal of Chemical Theory and Computation

UPMC SORBONNE UNIVERSITÉS

Bibliothèque Universitaire

Subscriber access provided by BUPMC - Bibliothèque Universitaire Pierre et Marie Curie

Article

Optimizing the energy with quantum Monte Carlo: A lower numerical scaling for Jastrow-Slater expansions

Roland Assaraf, Saverio Moroni, and Claudia Filippi

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.7b00648 • Publication Date (Web): 05 Sep 2017

Downloaded from http://pubs.acs.org on September 7, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

Journal of Chemical Theory and Computation is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Optimizing the energy with quantum Monte Carlo: A lower numerical scaling for Jastrow-Slater expansions

Roland Assaraf,[*,†] S. Moroni,[*,‡] and Claudia Filippi[*,R]

†*Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire de Chimie Théorique (UMR7616), 4 place Jussieu F-75252 Paris, France*

‡*CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy*

RMESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

E-mail: assaraf@lct.jussieu.fr; moroni@democritos.it; c.filippi@utwente.nl

Abstract

We present an improved formalism for quantum Monte Carlo calculations of energy derivatives and properties (e.g. the interatomic forces), with a multideterminant Jastrow-Slater function. As a function of the number N_e of Slater determinants, the numerical scaling of $O(N_e)$ per derivative we have recently reported is here lowered to $O(N_e)$ for the entire set of derivatives. As a function of the number of electrons N, the scaling to optimize the wave function and the geometry of a molecular system is lowered to $O(N^3) + O(NN_e)$, the same as computing the energy alone in the sampling process. The scaling is demonstrated on linear polyenes up to $C_{60}H_{62}$ and the efficiency of the method is illustrated with the structural optimization of butadiene and octate-

traene with Jastrow-Slater wave functions comprising as many as 200000 determinants and 60000 parameters.

1 Introduction

Quantum Monte Carlo methods (QMC) are first-principle methods which can efficiently solve the Schrödinger equation. For fermionic systems, they are powerful variational approaches because they can handle a large variety of variational wave functions $\Psi(\mathbf{R})$, where $\mathbf{R} = (\mathbf{r}_1 \dots \mathbf{r}_N)$ represents the coordinates of the N electrons of the system. Here, the vector $\mathbf{r}_i = (x_i, y_i, z_i, \sigma_i)$ indicates the 3 spatial coordinates of the electron i, (x_i, y_i, z_i) and its spin component σ_i ($\sigma_i = \pm\frac{1}{2}$). This flexibility stems from the fact that integrals are not computed analytically but from a stochastic sampling. For example, the variational energy is

$$E = \int d\mathbf{R}\Psi^2(\mathbf{R})\frac{\hat{H}\Psi}{\Psi}(\mathbf{R}) \tag{1}$$

where $\hat{H}$ is the Hamiltonian and Ψ is normalized. Eq. 1 can be interpreted as the expectation value of a random variable, the so-called local energy $E_L = \hat{H}\Psi/\Psi$, on the probability density $\Psi^2(\mathbf{R})$. QMC methods can be used as benchmark methods also for relatively large systems thanks to their favorable scaling with the number of particles N. For a given parametrization of Ψ, E is typically computed with a scaling $O(N^2)$ in memory requirements and $O(N^3)$ in CPU per Monte Carlo step. In practice, one needs to optimize the parameters of Ψ and the geometry of a molecular system. Despite the availability of stable wave function optimization methods,[1] such techniques remain costly and one of the main reasons is that a large number of derivatives of E (typically $O(N^2)$) has to be computed. Lowering the numerical scaling per derivative is therefore important. For single determinants, Sorella *et al.* have found that the low-variance estimators of the $3N_{\text{atoms}} = O(N)$ intermolecular forces can be calculated with a scaling $O(N^3)$ instead of $O(N^4)$ with the use of algorithmic differentiation techniques.[2] We have recently recovered the same reduction using transparent matricial formulas and

extended it to the $O(N^2)$ orbital coefficients.[3]

For multi-reference expansions, the Jastrow-Slater wave function takes the form

$$\Psi(\mathbf{R}) = J(\mathbf{R}) \sum_{I=0}^{N_e} c_I D_I \,, \tag{2}$$

where $J(R)$ is a positive Jastrow correlation factor, D_0 is a reference Slater determinant and D_I with $I > 0$ are N_e additional excited determinants. More explicitly, the determinantal component can be written as

$$\Phi = \sum_{I=0}^{N_e} c_I \det(A_I) \,. \tag{3}$$

where, for a system including N electrons, the matrix A_I is an $N \times N$ Slater matrix, built from N of the $N_{\rm orb}$ molecular spin-orbitals $\phi_i(\mathbf{r})$ $(1 \leq i \leq N_{\rm orb})$. Mathematically, A_I comprises N columns of the $N \times N_{\rm orb}$ matrix $\tilde{A}$ defined as follows

$$\tilde{A}_{ij} = \phi_j(\mathbf{r}_i) \,. \tag{4}$$

For this type of wave functions, Clark *et al.*[4] have proposed a method to compute Ψ with a scaling $O(N_e)$ and the corresponding E_L with a scaling $O(NN_e)$. We have further reduced this scaling to $O(N_e)$ and extended it to any derivative of E_L.[3] The derivatives of E_L are useful because they are involved in low-variance estimators for forces and observables.[5–7]

At the origin of this reduction is the observation[3] that for any one-body operator $\hat{O}$, such as the Laplacian or the non-local pseudopotential contributions to E_L, the expression $(\hat{O}D_I)/D_I$ is a first-order derivative with respect to a suitable parameter λ. For instance, while being a sum of second-order derivatives with respect to the electron coordinates, the Laplacian acting on any determinant A_I of the expansion can be recast in the form

$$\frac{\nabla^2 D_I}{D_I} = \frac{\nabla^2 \det A_I}{\det A_I} = \mathrm{tr}(A_I^{-1} B_I) = \frac{d}{d\lambda} \ln \det(A_I + \lambda B_I)|_{\lambda=0} \,. \tag{5}$$

where we have introduced an operator-specific matrix which, for the current example of the

Laplacian, is the $N \times N$ matrix $(B_I)_{ki} = \nabla^2 \phi_k(\mathbf{r}_i)$ constructed from the orbitals occupied in A_I.

In this paper, we show that the scaling $O(N_e)$ per derivative can be further improved to $O(N_e)$ for *any* set of derivatives of Ψ and E_L. The core observation is that the determinantal part Φ is a function of the matrix elements $\tilde{A}_{ij}$, and that any derivative of Φ can be computed using a simple trace formula involving the matrix Γ defined as the logarithmic gradient of Φ with respect to $\tilde{A}$. The first derivatives of the local energy $\partial_\mu E_L$ can then be expressed as traces involving Γ and *one* of its derivative $\partial_\lambda \Gamma$: many derivatives of Ψ and E_L are obtained efficiently because the matrices Γ and $\partial_\lambda \Gamma$ are computed only once for the whole set of parameters $\{\mu\}$. Consequently, the calculation of all derivatives of E with respect to all parameters of the wave function (Jastrow parameters, orbital coefficients, the coefficients of the expansion $\{c_I\}$, and all nuclear positions) has now the same scaling as the calculation of E alone, opening the path to full optimization of large multideterminant expansions.

In the next Section, we outline the main idea and introduce the matrix Γ. In Section 3, we present a formula to compute Γ at a cost $O(N^3) + O(N_e)$ and, in Section 4, discuss the formulas for the second derivative of Φ and, specifically, the first derivatives of E_L. In Section 5, we demonstrate the scaling of the computation of interatomic forces with multideterminant wave functions on polyenes up to $C_{60}H_{62}$ and, in the last Section, apply the scheme to the optimization of multideterminant wave functions and geometries of butadiene and octatetraene.

2 Derivatives of the determinantal expansion

In general, one needs to compute many derivatives of Φ with respect to different parameters of $\tilde{A}$. These parameters can be the electron coordinates, nuclei coordinates, orbital coefficients, basis-function parameters and so on. The derivative of Φ with respect to a given parameter

μ in $\tilde{A}$ is obtained from the chain rule

$$\partial_\mu \ln(\Phi) = \frac{\partial \ln(\Phi)}{\partial \tilde{A}_{ij}} \partial_\mu \tilde{A}_{ij} = \mathrm{tr}(\Gamma \partial_\mu \tilde{A}) \,, \tag{6}$$

where a summation on repeated indices is implied and we have introduced Γ, that is, the gradient of $\ln(\Phi)$ with respect to the matrix elements of $\tilde{A}$

$$\Gamma_{ji} = \frac{\partial \ln \Phi}{\partial \tilde{A}_{ij}} \,. \tag{7}$$

The trace formula (6) is at the core of greater efficiency in computing many derivatives of Φ because the $N \times N_{\mathrm{orb}}$ matrix Γ depends only on $\tilde{A}$ and not on $\partial_\mu \tilde{A}$. For a given configuration $\mathbf{R}$ in the Monte Carlo sample, Γ is computed only once for the entire set of derivatives. In addition, Γ can be evaluated efficiently, at a cost $O(N^3) + O(N_e)$ as we will see in the next Section. Once Γ is computed and stored, any new derivative $\partial_\mu \ln(\Phi)$ requires calculating besides $\partial_\mu \tilde{A}$ the trace (6) at a cost $O(N_{\mathrm{orb}} \times N)$. What is important here is that this scaling is independent of N_e and leads to vast improvements over previous methods[3,4] when N_e and the number of derivatives are large.

Finally, also quantities like the local energy or the value of the wave function after one electron move, can be computed using this trace formula (6). This is because one-body operators can be also expressed as first order derivatives of $\ln \Phi$ when applied to a Jastrow-Slater expansion.[3]

3 Efficient evaluation of the matrix Γ

3.1 Convenient expression for Φ

The determinants of the Slater matrices A_I can be computed efficiently because A_I usually differs by a few columns from the reference Slater matrix A_0. For example, let A_0 be the

4×4 Slater matrix built with the orbitals $\phi_1, \phi_2, \phi_3, \phi_4$:

$$A_0 = \left(\tilde{A}_1 \quad \tilde{A}_2 \quad \tilde{A}_3 \quad \tilde{A}_4 \right), \tag{8}$$

where the notation $\tilde{A}_i$ stands for the i^{th} column of $\tilde{A}$. The Slater matrix of a double excitation $(3,4) \rightarrow (5,7)$ is

$$A_I = \left(\tilde{A}_1 \quad \tilde{A}_2 \quad \tilde{A}_5 \quad \tilde{A}_7 \right), \tag{9}$$

which differs from A_0 only in the 2 last columns. Hereafter we will drop the subscript "0" of the reference Slater matrix to simplify the notation. The determinant of A_I is

$$\det(A_I) = \det(A)\det(A^{-1}A_I)$$

and

$$A^{-1}A_I = \left(A^{-1}\tilde{A}_1 \quad A^{-1}\tilde{A}_2 \quad A^{-1}\tilde{A}_5 \quad \tilde{A}^{-1}\tilde{A}_7 \right) = \begin{pmatrix} 1 & 0 & (A^{-1}\tilde{A})_{15} & (A^{-1}\tilde{A})_{17} \\ 0 & 1 & (A^{-1}\tilde{A})_{25} & (A^{-1}\tilde{A})_{27} \\ 0 & 0 & (A^{-1}\tilde{A})_{35} & (A^{-1}\tilde{A})_{37} \\ 0 & 0 & (A^{-1}\tilde{A})_{45} & (A^{-1}\tilde{A})_{47} \end{pmatrix} \tag{10}$$

where a column of the identity matrix arises whenever A_I and A share the same column. The determinant of $A^{-1}A_I$ is readily evaluated:

$$\det(A^{-1}A_I) = \det \begin{pmatrix} (A^{-1}\tilde{A})_{35} & (A^{-1}\tilde{A})_{37} \\ (A^{-1}\tilde{A})_{45} & (A^{-1}\tilde{A})_{47} \end{pmatrix}. \tag{11}$$

More generally, the determinant of $A^{-1}A_I$ for a k_I^{th}-order excitation is the determinant of a $k_I \times k_I$ submatrix. Such a submatrix can always be written as follows

$$\alpha_I = P_I^T A^{-1} \tilde{A} Q_I \,, \tag{12}$$

where, in our example,

$$P_I = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{13}$$

and

$$Q_I = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix} . \tag{14}$$

In general, P_I is such that AP_I are the columns of A which differ from those of A_I, and Q_I is such that $\tilde{A}Q_I = A_I P_I$. In other words P_I (applied on the right of A) selects the columns of A from which excitations are built, and Q_I (applied on the right of $\tilde{A}$) selects the columns of $\tilde{A}$ to which excitations are built. To summarize, the expression

$$\det(A_I) = \det(A)\det(P_I^T A^{-1} \tilde{A} Q_I) \tag{15}$$

enables us to compute the determinant of a large $N \times N$ matrix as the determinant of a small $k_I \times k_I$ submatrix of $A^{-1}\tilde{A}$. This expression can also be proven using the determinant lemma.[3,4] Finally, the convenient expression for Φ to efficiently compute Γ is:

$$\Phi = \det(A) \times \sum_I c_I \det(P_I^T A^{-1} \tilde{A} Q_I) \,. \tag{16}$$

3.2 Convenient expression for Γ

Introducing the matrix R such that $A = \tilde{A}R$, the expression for Φ in Eq. 16 is explicitly a function of $\tilde{A}$. In particular, the summation on the r.h.s. of Eq. 16

$$\chi \equiv \frac{\Phi}{\det(A)} = \sum_I c_I \det(P_I^T A^{-1} \tilde{A} Q_I) \tag{17}$$

is a polynomial function depending on the matrix elements of

$$T \equiv A^{-1}\tilde{A} = (\tilde{A}R)^{-1}\tilde{A} \,. \tag{18}$$

The order of this polynomial is the order of the highest-order excitation. It is usually low (typically $k_I < 4$). Applying the chain rule and using the convention of summation over repeated indices, we obtain

$$\begin{aligned} \partial_\mu \ln(\Phi) &= \partial_\mu \ln \det(A) + \partial_\mu \ln \chi \\ &= \mathrm{tr}(A^{-1}\partial_\mu A) + \frac{\partial \ln \chi}{\partial T_{ij}} \partial_\mu T_{ij} \\ &= \mathrm{tr}(A^{-1}\partial_\mu A) + \mathrm{tr}(Y \partial_\mu T) \,, \end{aligned} \tag{19}$$

where

$$Y_{ji} \equiv \frac{\partial \ln \chi}{\partial T_{ij}} = \frac{1}{\chi} \frac{\partial \chi}{\partial T_{ij}} \,. \tag{20}$$

It is simple to show that

$$\frac{\partial \chi}{\partial T_{ij}} = \sum_{I>0}^{N_e} c_I \det(\alpha_I)(Q_I \alpha_I^{-1} P_I^T)_{ji} \,. \tag{21}$$

The derivative of T is given by

$$\partial_\mu T = -A^{-1}\partial_\mu A \; A^{-1}\tilde{A} + A^{-1}\partial_\mu \tilde{A} \,. \tag{22}$$

Finally, writing $A = \tilde{A}R$ and using the cyclic property of the trace, we obtain

$$\partial_\mu \ln(\Phi) = \mathrm{tr}(\Gamma \partial_\mu \tilde{A}) \,, \tag{23}$$

where

$$\begin{aligned} \Gamma &= RA^{-1} + \; (1 - RA^{-1}\tilde{A})YA^{-1} \\ &= \left[R(1 - A^{-1}\tilde{A}Y) + Y\right] A^{-1} \,. \end{aligned} \tag{24}$$

For example, if the occupied orbitals are the N first ones, the matrix Γ is

$$\Gamma = \begin{pmatrix} A^{-1} - A^{-1}\tilde{A}YA^{-1} \\ Y_{\mathrm{virt}} A^{-1} \end{pmatrix} , \tag{25}$$

where the first line is a $N \times N$ matrix. The second line is a $(N_{\mathrm{orb}} - N) \times N$ matrix where Y_{virt} represents the non-zero lines of Y, i.e. the last $N_{\mathrm{virt}} \equiv N_{\mathrm{orb}} - N$ lines.

3.3 One-body operators and first-order derivatives of Φ

First-order derivatives of Φ can be computed with the trace formula (6) which involves the Γ matrix. One-body operators acting on the wave function can be also expressed as first-order derivatives of $\ln \Phi$ when applied to a Jastrow-Slater expansion as we have shown in Ref.[3]

and briefly recalled for the Laplacian in Eq. 5. The local energy can also be written as a first-order logarithmic derivative of the determinantal part where $\tilde{A}$ has been replaced by

$$\tilde{A}_\lambda = \tilde{A} + \lambda \tilde{B} \tag{26}$$

and $\tilde{B}$ is an appropriate matrix depending on the orbitals, the Jastrow factor, and their derivatives. In particular, the reference Slater determinant A has been replaced by $A_\lambda = A + \lambda B$. The determinantal part of the wave function is now

$$\Phi = \det(A_\lambda) \left[\sum_I c_I \det(P_I^T A_\lambda^{-1} \tilde{A}_\lambda Q_I) \right] . \tag{27}$$

From this expression, one can compute the local energy

$$E_L = \partial_\lambda (\ln \Phi) = \mathrm{tr}(\Gamma \tilde{B}) . \tag{28}$$

In the presence of the Jastrow factor, one recovers the same trace expression for the local energy of Ψ but with a matrix $\tilde{B}$ also depending on $J(\mathbf{R})$ and its derivatives.[3]

4 Second-order derivatives

The second derivative of Φ can be written in terms of Γ and its derivative as

$$\begin{aligned} \partial_\lambda \partial_\mu \ln(\Phi) &= \partial_\lambda \mathrm{tr}(\Gamma \partial_\mu \tilde{A}) \\ &= \mathrm{tr}(\Gamma\, \partial_{\lambda\mu} \tilde{A}) + \mathrm{tr}(\partial_\lambda \Gamma\ \partial_\mu \tilde{A}) . \end{aligned} \tag{29}$$

Example of the derivative of the local energy

When computing improved estimators of derivatives of the energy E, we need also the derivatives of the local energy E_L. It follows from Eq. 28 that the derivative of the local

energy with respect to a given parameter μ is

$$\begin{aligned} \partial_\mu E_L &= \partial_\lambda \partial_\mu \ln(\Phi) \\ &= \mathrm{tr}(\Gamma\, \partial_\mu \tilde{B}) + \mathrm{tr}(\partial_\lambda \Gamma\ \partial_\mu \tilde{A})\,. \end{aligned} \tag{30}$$

The order of the derivatives has been chosen so that $\tilde{A}$ and not Γ is differentiated with respect to μ. Consequently, the matrix $\partial_\lambda \Gamma$ does not depend on the parameter μ and has to be computed only once, whatever the number of second derivatives we need. Once $\partial_\lambda \Gamma$ has been computed, the calculation of $\partial_\mu E_L$ involves (besides $\partial_\mu \tilde{A}$ and $\partial_\mu \tilde{B}$) two traces which can be computed at a cost $O(N N_{\mathrm{orb}})$. Importantly, such a calculation does not depend on N_e in contrast to what was presented in Ref.[3]

Efficient calculation of $\partial_\lambda \Gamma$

The derivative of Γ is

$$\partial_\lambda \Gamma = \left[-\Gamma B + \partial_\lambda Y + R(\partial_\lambda T\ Y + T \partial_\lambda Y)\right] A^{-1}\,, \tag{31}$$

where

$$\partial_\lambda T = A^{-1}(\tilde{B} - BT) \equiv \tilde{M}\,. \tag{32}$$

Applying the chain rule, we obtain

$$\partial_\lambda Y_{ij} = Z_{ijkl} \tilde{M}_{kl}\,, \tag{33}$$

where

$$Z_{ijkl} \equiv \frac{\partial^2 \ln \chi}{\partial T_{ij} \partial T_{kl}} \quad (34)$$

$$= \frac{1}{\chi} \frac{\partial^2 \chi}{\partial T_{ij} \partial T_{kl}} - Y_{ij} Y_{kl} \,. \quad (35)$$

It follows from Eq. 21 that

$$\frac{\partial^2 \chi}{\partial T_{ij} \partial T_{kl}} = \sum_{I>0}^{N_e} c_I \det(\alpha_I) \left[(Q_I \alpha_I^{-1} P_I^T)_{ji} (Q_I \alpha_I^{-1} P_I^T)_{lk} - (Q_I \alpha_I^{-1} P_I^T)_{jk} (Q_I \alpha_I^{-1} P_I^T)_{li} \right] . \quad (36)$$

We can compute the derivatives of χ avoiding the evaluation of inverse matrices. That will be presented in the appendix.

Derivatives with respect to the linear coefficients

The derivatives of a local quantity with respect to the expansion coefficients require instead to evaluate the action of the one-body operator on each excited determinant A_I separately (Eq. 15). For instance, as we have shown in Ref.,[3] the derivative of the local energy with respect to c_I is given by

$$\begin{aligned} \partial_{c_I} E_L &= \partial_\lambda \partial_{c_I} (\ln \Phi) \\ &= \frac{\det(A)}{\Phi} \partial_\lambda \det(P_I^T T Q_I) \\ &= \frac{\det(A)}{\Phi} \mathrm{tr}(\alpha_I^{-1} P_I^T \partial_\lambda T Q_I) \\ &= \frac{\det(A)}{\Phi} \mathrm{tr}(\alpha_I^{-1} P_I^T \tilde{M} Q_I) \,. \end{aligned} \quad (37)$$

These quantities are needed in the optimization of the energy with respect to the linear coefficients and can be computed at a cost $O(N_e)$.

5 Numerical scaling

In practice, for each step of the Monte Carlo algorithm, we need to compute $\tilde{A}$, A^{-1}, and $T = A^{-1}\tilde{A}$ at a cost of at most $O(N^3)$ (products and inversions of matrices). Then, we need to calculate the first and second derivatives of χ with respect to T (Eqs. 21 and 36) at a cost $O(N_e)$ (a few sums and products for each excitation). The related tensors Y and Z are also computed at a cost $O(N_e)$. $\partial_\lambda Y$ is computed at a cost $O(N_d)$ where N_d is the total number of double excitations involved in any k_I^{th} order excitation ($k_I \geq 2$), where of course $N_d < N_e$. Finally, Γ and $\partial_\lambda \Gamma$ are computed at a cost $O(N^3)$ (product of matrices).

In particular, computing the $3N_{\text{atoms}}$ components of the inter-atomic forces with improved estimators has a scaling

$$O(N^3) + O(N^2 N_{\text{atoms}}) + O(N_{\text{atoms}} N N_{\text{virt}}) + O(N_e)\,. \tag{38}$$

Assuming that $N_{\text{virt}} = O(N) = O(N_{\text{atoms}})$, this scaling simplifies

$$O(N^3) + O(N_e)\,. \tag{39}$$

This is significantly more efficient than the scaling [1] $O(N^3)+O(N^2 N_{\text{virt}} N_{\text{atoms}})+O(N_e N_{\text{atoms}})$ presented in our previous work,[3] in the large N_e, N_{atoms} or N, N_{atoms}, N_{virt} regimes. The term $O(N^2 N_{\text{act}} N_{\text{atoms}})$ is no more present because here we avoid computing $\partial_\mu T$. Regarding the sampling process, when one-electron moves are used (see appendix), the total numerical cost for a full sweep (all the electrons are moved once) is $\sim O(N^3) + O(N N_e)$.

In Fig. 1, we demonstrate this favorable scaling in the variational Monte Carlo (VMC) computation of the interatomic forces for multi-determinant Jastrow-Slater wave functions using the sequence of molecules C_nH_{n+2} with n between 4 and 60. For each system, the ratio of the CPU time of computing all interatomic forces to the time of evaluating only the

[1] or $O(N^3) + O(N^2 N_{\text{act}} N_{\text{atoms}}) + O(N_e N_{\text{atoms}})$

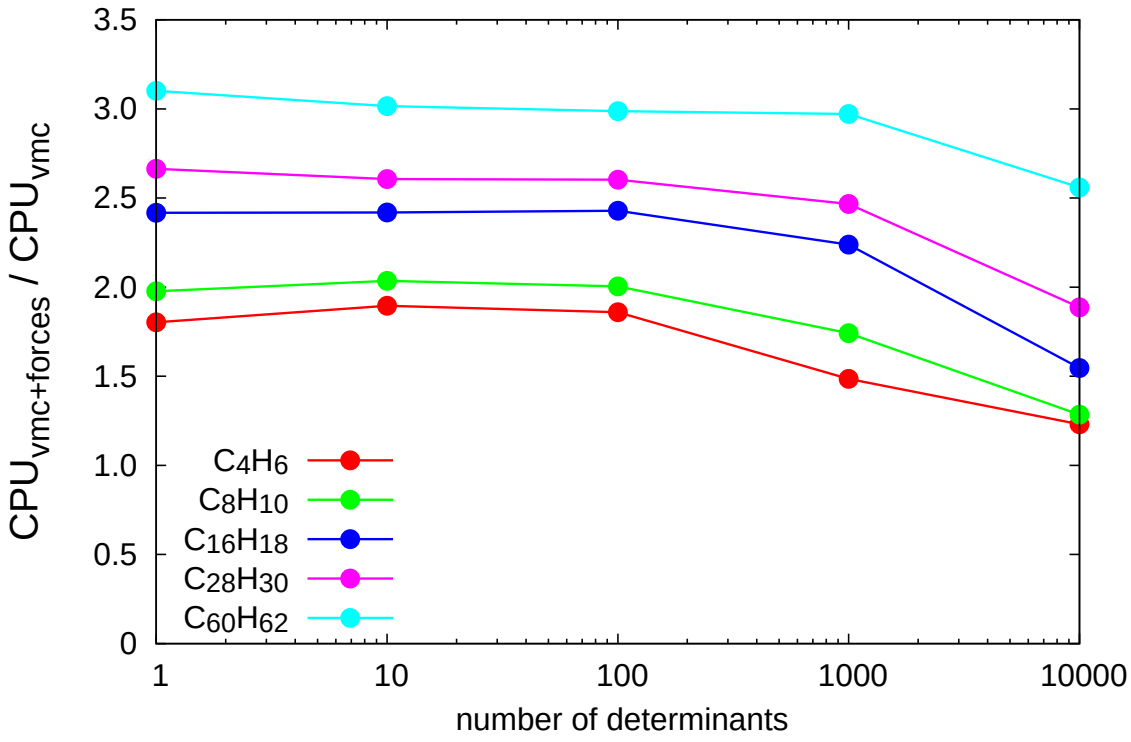


Figure 1: Ratio of the CPU time for a VMC calculation of the forces to the CPU time for the same simulation of the energy alone for the sequence of molecules C_nH_{n+2} with n between 4 and 60 and an increasing number of determinants in the Jastrow-Slater wave function. The forces are calculated after moving all the electrons once.

energy is initially constant and then decreases when the number of determinants exceeds about 100. For the largest $C_{60}H_{62}$, computing all interatomic gradients costs less than about 3 times a VMC simulation where one only evaluates the total energy. Finally, as shown in the Appendix, if we move one electron, many quantities can be updated so that, for each Monte Carlo step, the scaling is reduced to $O(N^2) + O(N_e)$. This leads to an overall scaling $O(N^3) + O(N_eN)$ when all the electrons have been moved. For an all-electron-move algorithm, the scaling is $O(N^3) + O(N_e)$ which could be more efficient when N_e is large.

6 Numerical results

We demonstrate the formulas above on the ground-state structural optimization in VMC of butadiene (C_4H_6) and octatetraene (C_8H_{10}) using large expansions in the determinantal component of the Jastrow-Slater wave function. All expansion coefficients, orbital and Jastrow parameters in the wave function are optimized together with the geometry. Given the large number of variational parameters (up to 58652) we employ the stochastic reconfiguration optimization method[8] in a conjugate gradient implementation[9] which avoids building and storing large matrices. We find that the simple conditioning of the overlap matrix introduced in Ref.[8] gives sufficient stability for all the cases considered here. In most of our calculations, to remove occasional spikes in the forces, we use an improved estimator of the forces obtained by sampling the square of a modified wave function close to the nodes.[10] To optimize the geometry, we simply follow the direction of steepest descent and appropriately rescale the interatomic forces. We employ the CHAMP code[11] with scalar-relativistic energy-consistent Hartree-Fock pseudopotentials and the corresponding cc-pVXZ[12,13] and aug-cc-pVXZ[14] basis sets with X=D,T, and Q. The Jastrow factor includes two-body electron-electron and electron-nucleus correlation terms.[15] The starting determinantal component of the Jastrow-Slater wave functions before optimization is obtained in multiconfiguration-self-consistent-field calculations performed with the program GAMESS(US).[16,17]

Table 1: Optimal bond lengths and BLA values (Å) of butadiene computed in VMC with the cc-pVQZ basis set and various choices of Jastrow-Slater expansions. The numbers of determinants and optimized parameters in the wave function are listed.

Expansion	No. det	No. param.	C-C	C=C	BLA (Å)
1 det	1	1404	1.45513(12)	1.32482(05)	0.13031(16)
CAS(4,4)	20	1547	1.45211(10)	1.33347(07)	0.11865(15)
CAS(4,16)	7232	4995	1.45160(15)	1.33422(13)	0.11738(16)
CAS(4,20)	18100	9147	1.45143(16)	1.33409(07)	0.11734(24)
CAS(10,10)	15912	6890	1.45858(09)	1.33694(06)	0.12163(13)
RAS(10,22)	45644	11094	1.45705(17)	1.33760(15)	0.11945(29)
CCSD(T)/CBS[a]			1.4548	1.3377	0.1171
CCSD(T)/CBS-corr[b]			1.4549	1.3389	0.1160

[a] Ref.;[18] [b] Ref.,[18] including a CCSDT(Q)(FC)/cc-pVDZ correction.

We first focus on the VMC geometrical optimization of butadiene. Despite its small size and apparent simplicity, predicting the bond length alternation (BLA) of butadiene remains a challenging task for quantum chemical approaches which lead to a spread of BLA values, mainly clustered around either 0.115 or 0.125 Å (see Table 2 in Ref.[19] for a recent compilation of theoretical predictions). In particular, Barborini and Guidoni[19] using VMC in combination with Jastrow-antisymmetrized geminal power (JAGP) wave functions find a best BLA value of 0.1244(6) Å, rather close to the BLA of 0.1251(7) Å they obtain using a single-determinant Jastrow-Slater wave function and clearly distinct from the CCSD(T) prediction of 0.116 Å computed in the complete basis set (CBS) limit and corrected for core-valence correlation, scalar-relativistic effects, and inclusion of quadruples.[18]

Here we explicitly address the relevance of the multireference character of the system, considering various expansions which correlate the π and σ electrons: a) a single determinant; b) the complete-active-space CAS(4,4), CAS(4,16), and CAS(4,20) expansions (20, 7232, and 18100 determinants, respectively) of the four π electrons in the bonding and antibonding π orbitals constructed from the $2p_z$, $3p_z$, $3d_{xz}$, $3d_{yz}$, and $4p_z$ atomic orbitals; c) a CAS(10,10) correlating the six σ and four π electrons of the carbon atoms in the corresponding bonding and antibonding π and σ orbitals (15912 determinants); d) the same CAS(10,10) expansion augmented with single and double excitations in the external space of 12 π orbitals and truncated with a threshold of 2×10^{-4} on the coefficients of the spin-adapted configuration state functions. This last choice results in a total of 45644 determinants and is denoted as a restricted-active-space RAS(10,22) expansion.

We start all runs from the same geometry and, after convergence, average the geometries over an additional 30-40 iterations. The results of these structural optimizations are summarized in Fig. 2. We find that the basis sets of triple- and quadruple-ζ quality yield values of BLA which are compatible within 1-1.5 standard deviations, namely, to better than 5×10^{-4} Å. The further addition of augmentation does not change the BLA as shown in the one-determinant case. In the following, we therefore focus on the cc-pVQZ bond lengths and

BLA values of butadiene, which are summarized in Table 1.

With a one-determinant wave function (case a), we obtain a BLA of 0.1303(2) Å which is higher than the value of 0.1251(6) Å reported in Ref.,[19] possibly due to their use of a basis set of quality inferior to triple-ζ. Moving beyond a single determinant, we observe a strong dependence of the result on the choice of active space. The inclusion of π-π correlation within 4, 16, and 20 π orbitals (case b) significantly decreases the BLA with respect to the one-determinant case with the CAS(4,16) and CAS(4,20) expansions yielding a BLA of 0.117 Å in apparent agreement with the CCSD(T)/CBS estimate of 0.116 Å. Accounting also for σ-π and σ-σ correlations in a CAS(10,10) (case c) leads however to a more substantial lengthening of the single than the double bond and a consequent increase of BLA. Finally, allowing excitations out of the CAS(10,10) in 12 additional π orbitals (case d) brings the double bond in excellent agreement with the CCSD(T)/CBS value and somewhat shortens the single bond, lowering the BLA to a final value of 0.119 Å. In summary, all choices of multi-determinant expansion in the Jastrow-Slater wave function represent a clear improvement with respect to the use of a single determinant, significantly lowering the value of BLA.

Finally, in Fig. 3, we demonstrate the ability of our method to optimize the structure and the many wave function parameters for the larger molecule C_8H_{10} when using a very large determinantal expansion. For this purpose, we employ the simple cc-pVDZ basis set and consider all single, double, and triple excitations in an expansion denoted as SDT(22,22), correlating 22 electrons in the 22 σ and π orbitals obtained from the carbon valence atomic orbitals. The wave function comprises a total of 201924 determinants and 58652 parameters. To illustrate the dependence of the energy on the choice of wave function, we also display the energy of the last iterations of a structural optimization of the same molecule with the minimal CAS(8,8) expansion over the π orbitals. At each iteration, we update both the wave function parameters and the atomic positions, the former with one step of the stochastic reconfiguration method and the latter along the down-hill direction of the interatomic forces. The energy of the SDT(22,22) wave function is distinctly lower than the one obtained with

the smaller active space and converged to better than 2 mHartree within about 80 iterations. The structural parameters converge much faster and reach stable values within the first 30 iterations.

7 Conclusion

In the last few years, a series of methodological advances[3,4,20] has enabled the use of large multi-determinant wave functions in real-space quantum Monte Carlo. In this work, we take a further step forward, reducing the scaling of the computational cost to obtain *all* the derivatives of both the trial function and the local energy. These are for instance needed in the low-variance estimators entering the optimization of the wave function and the structural parameters. This progress, described in Sections 5 and 6, is here summarized in Table 2. We believe that the possibility of optimizing molecular geometries and all the variational parameters in very large expansions for medium-size molecules will stimulate more extensive applications of real-space quantum Monte Carlo methods.

Table 2: Dominant scaling in the number of determinants N_e to compute the wave function, local energy, and one or M of their derivatives. We also compare with the standard algorithm using the Sherman Morrison formula (SMF) for the matrix updates in the sweep over the electrons.

	Sweep	E_L	$\partial_\mu \Psi$		$\partial_\mu \partial_\nu \Psi$, $\partial_\mu E_L$	
			1	M	1	M
SMF < 2011 [21]	$N^3\, N_e$	$N^3\, N_e$				
Ref. [4]	$N\, N_e$	$N\, N_e$				
Ref. [3]	$N\, N_e$	N_e	N_e	$M\, N_e$	N_e	$M\, N_e$
This work	$N\, N_e$	N_e	N_e	N_e	N_e	N_e

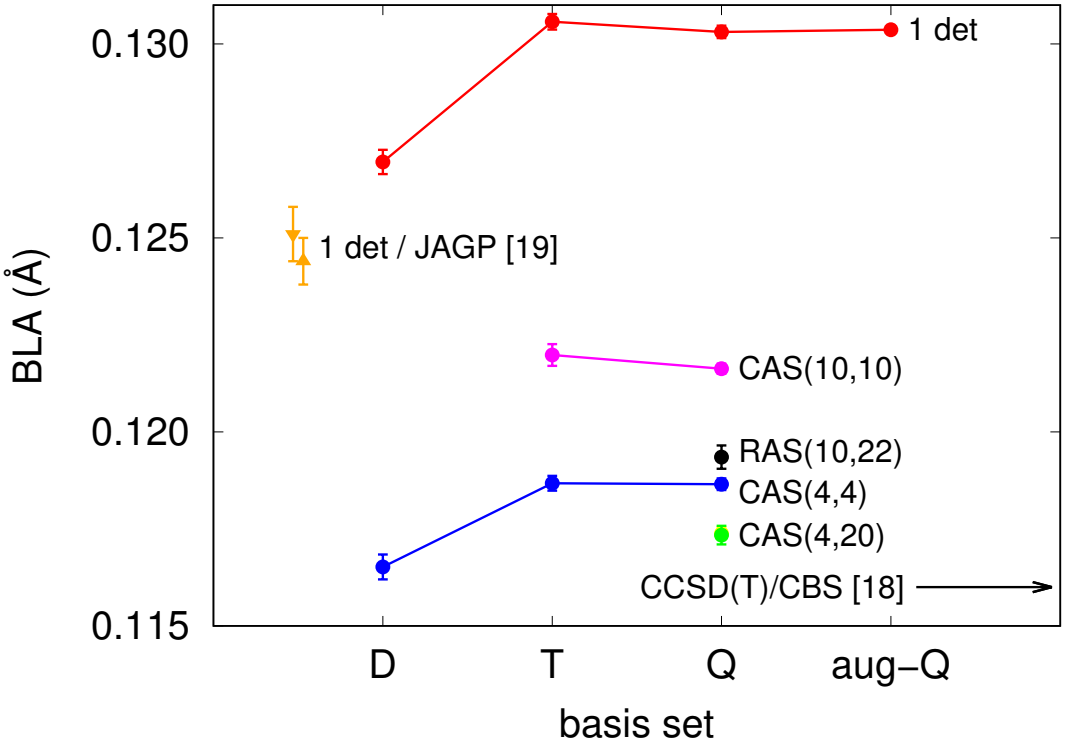


Figure 2: Bond length alternation (BLA) of C_4H_6 optimized in VMC for different basis sets and choices in the determinantal part of the Jastrow-Slater wave function. The atomic positions and all parameters of the wave function (expansion coefficients, orbital and Jastrow parameters) are simultaneously optimized. The CCSD(T) BLA in the CBS limit computed with various corrections[18] and the best value obtained with a Jastrow-antisymmetrized geminal power (JAGP)[19] are reported.

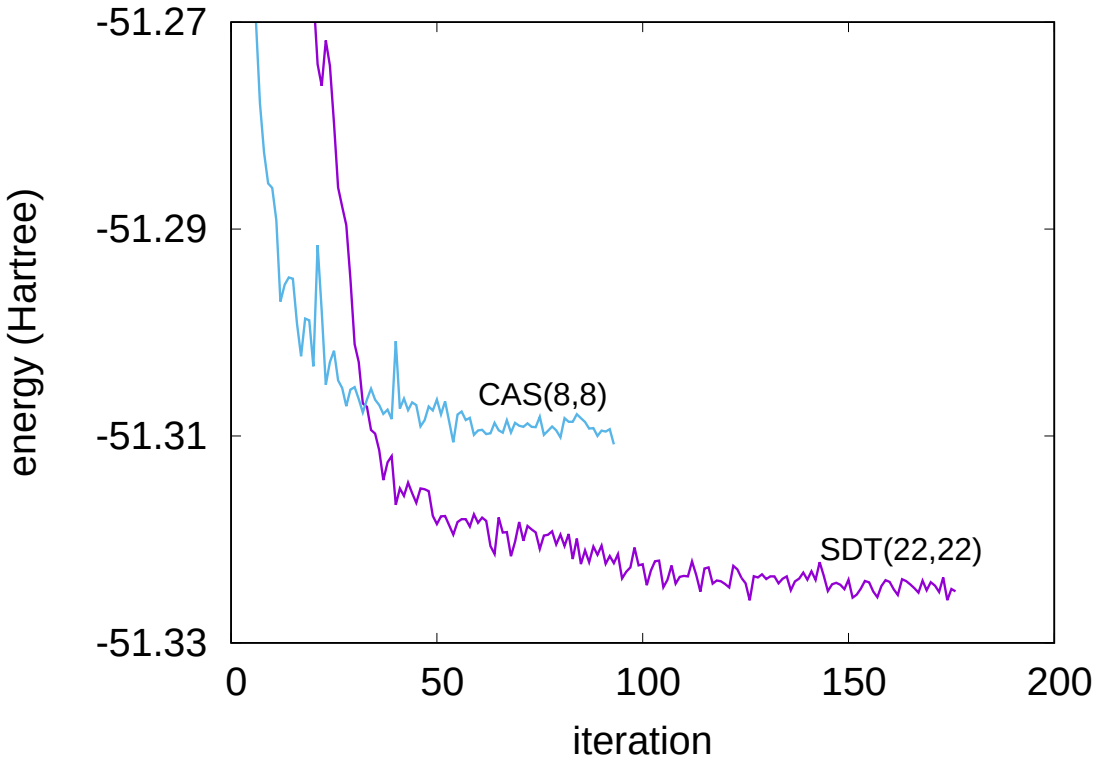


Figure 3: Total energy for a VMC geometry optimization of C_8H_{10} using two different expansions in the Jastrow-Slater wave function, that is, a full CAS(8,8) with 2468 determinants, and all single, double, and triple excitations in an expansion correlating 22 electrons in 22 orbitals for a total of 201924 determinants. The atomic positions and all parameters of the wave function (expansion coefficients, orbital and Jastrow parameters) are simultaneously optimized.

A Efficient calculation of Z, Y, χ

We demonstrate here that we do not need to compute explicitly the inverses of the submatrices α_I as in Eqs. (17, 22, and 36) or in Refs.[3,4] to obtain χ and its derivatives. These can be computed efficiently using recursion formulas.

Suppose that χ contains only third-order excitations (the generalization to an arbitrary order is straightforward). Let us rewrite the expression of χ (Eq. 17) as

$$\chi = \sum_{i_1<i_2<i_3, j_1<j_2<j_3} C_{i_1 i_2 i_3 j_1 j_2 j_3} \sum_p (-1)^p T_{i_1 p(j_1)} T_{i_2 p(j_2)} T_{i_3 p(j_3)} \,, \tag{40}$$

where p stands for a permutation of the indices (j_1, j_2, j_3), and $(-1)^p$ is the sign of the permutation. We note that this formula can also include first- and second-order excitations: a second-order excitation $(i_1 \to j_1, i_2 \to j_2)$ can be written as $(i_1, \to j_2, i_2 \to j_2, i_3 \to i_3)$, and a first-order excitation $(i_1 \to j_1)$ as $(i_1, \to j_1, i_2 \to i_2, i_3 \to i_3)$.

The starting point is that the tensor of second derivatives can be computed directly from the expression (40) as

$$\frac{\partial^2 \chi}{\partial T_{i_1 j_1} \partial T_{i_2 j_2}} = \sum_{i_3 j_3} (-1)^{p+q} C_{p(i_1) q(j_1) p(i_2) q(j_2) p(i_3) q(j_3)} T_{i_3 j_3} \,, \tag{41}$$

where p and q are the permutations ordering (i_1, i_2, i_3) and (j_1, j_2, j_3), respectively. Note that this tensor is antisymmetric with respect to the permutations of either the indices (i_1, i_2) or the indices (j_1, j_2), and we only need to compute and store the elements such that $i_1 < i_2$, and $j_1 < j_2$. The tensor of first order derivatives is

$$\frac{\partial \chi}{\partial T_{i_1 j_1}} = \frac{1}{2} \sum_{i_2 j_2} \frac{\partial^2 \chi}{\partial T_{i_1 j_1} \partial T_{i_2 j_2}} T_{i_2 j_2} \,, \tag{42}$$

and the value of χ is

$$\chi = \frac{1}{3} \sum_{ij} \frac{\partial \chi}{\partial T_{ij}} T_{ij} \,. \tag{43}$$

In practice, sparse representations of these tensors should be used. The formula (41) involves at most nine products and nine sums per excitation. The formulas (42) and (43) require less than $N^2 N_{\rm orb}^2$ and $N N_{\rm orb}$ operations (additions or multiplications), respectively. The method still scales like $O(N_e)$ but with a reduced prefactor because no divisions are involved and the number of operations is smaller. For example, expression (43) involves at most $N N_{\rm orb}$ multiplications and additions whereas (17) is a sum on N_e terms (N_e can be of order $N^3 N_{\rm orb}^3$ if third-order excitations are included).

B One-electron-move algorithms

To sample the density Ψ^2, we use the Metropolis-Hastings method[22,23] which is a stochastic dynamics in the space of configurations $\mathbf{R} = (\mathbf{r_1}, \mathbf{r}_2 \dots \mathbf{r}_N)$. For a given iteration, this method proposes a random move $\mathbf{R} \to \mathbf{R}'$ with a transition probability density $P(\mathbf{R} \to \mathbf{R}')$. The proposed move is accepted with the probability

$$\min\left(\frac{\Psi^2(\mathbf{R}')}{\Psi^2(\mathbf{R})}\frac{P(\mathbf{R}' \to \mathbf{R})}{P(\mathbf{R} \to \mathbf{R}')}, 1\right) . \tag{44}$$

If only one electron is moved (here the first, for example), the new configuration is $\mathbf{R}' = (\mathbf{r_1}', \mathbf{r}_2 \dots \mathbf{r}_N)$. The new extended Slater matrix $\tilde{A}'$ differs from $\tilde{A}$ only in the first line.

We introduce the matrix $\tilde{B}_e$ such that the first line of $\tilde{B}_e$ and $\tilde{A}'$ are the same but $\tilde{B}_e$ is zero elsewhere. Since Φ is a linear function of the modified line

$$\frac{\Phi(\mathbf{R}')}{\Phi(\mathbf{R})} \;=\; \partial_\lambda \ln \Phi(\tilde{A}) \,, \tag{45}$$

where we considered the following transformation $\tilde{A} \to \tilde{A} + \lambda \tilde{B}_e$. Using Eq. (19), we obtain

$$\frac{\Phi(\mathbf{R}')}{\Phi(\mathbf{R})} = \mathrm{tr}(A^{-1}\partial_\lambda A) + \mathrm{tr}(Y \partial_\lambda T) \,, \tag{46}$$

where we recall that $T = A^{-1}\tilde{A}$ and $\partial_\lambda T = A^{-1}\tilde{B}_e - A^{-1}B_e T$. The cost of this calculation is $O(NN_{\rm orb}) \sim O(N^2)$. When the first electron has been moved, T can be updated using the Sherman Morrison formula at a cost $O(NN_{\rm orb})$,[3] and Y which depends on T can be again computed at a cost $O(N_e)$. The total cost for a sweep (each electron has moved once) is $O(N^2 N_{\rm orb}) + O(NN_e)$. The matrix Γ and all derivatives are computed after each sweep.

We note that, if one uses instead the expression involving Γ to update the wave function,

$$\frac{\Phi(\mathbf{R}')}{\Phi(\mathbf{R})} = \mathrm{tr}(\Gamma \tilde{B}_e)\,, \tag{47}$$

one would need to update Γ at each Monte Carlo step and incur the higher cost of $O(N^4) + O(NN_e)$ for a full sweep. This is because updating Γ requires the calculation of $\partial_\lambda \Gamma$ given in Eq. (31), where of course $\tilde{B}$ is replaced by $\tilde{B}_e$. In this equation, the product $(\partial_\lambda Y)A^{-1}$ scales like $O(N^3)$, unless Y is sparsely modified after one electron move (i.e. a few double excitations are involved).

Finally, also in the calculation of the drift of a single electron $\nabla_i \Phi/\Phi$ needed in the Monte Carlo sampling, it is better not to recompute Γ but to use formula (46) with $\partial_\lambda T = A^{-1}\tilde{B}_e^{\rm drift} - A^{-1}B_e^{\rm drift}T$, where the matrix $\tilde{B}_e^{\rm drift}$ is zero except the i^{th} row which equals $\nabla\phi_j(\mathbf{r}_i)$. However, if the sampling is modified to use a finite distribution at the nodes following Ref.,[10] the full drift has to be computed at each step. The resulting scaling is $O(N^4) + O(NN_e)$ per sweep, using Eq. (46) or (47) alike.

C Simple Expression of Γ for a Jastrow-Slater expansion

Here, we provide a simple (though not efficient) expression for Γ and some mathematical properties.

Simple expression for Γ

The determinantal contribution of the wave function written in Eq. (3) is

$$\Phi = \sum_{I=0}^{N_e} c_I \det(A_I)\,.$$

where A_I is a list of N columns of the $N \times N_{\rm orb}$ generalized Slater matrix $\tilde{A}$. We can then define a $N_{\rm orb} \times N$ matrix R_I such that

$$A_I = \tilde{A} R_I'\,, \tag{48}$$

which gives an explicit expression of Φ as a function of $\tilde{A}$

$$\Phi(\tilde{A}) = \sum_I c_I \det(\tilde{A} R_I)\,. \tag{49}$$

For example, given a 3×3 Slater matrix built on the orbitals (ϕ_1, ϕ_3, ϕ_4)

$$R_I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \end{pmatrix}\,.$$

The derivative of the determinantal expansion with respect to a parameter μ is

$$\begin{aligned} \partial_\mu \Phi &= \sum_I c_I \det(A_I) \mathrm{tr}(A_I^{-1} \partial_\mu A_I) \\ &= \sum_I c_I \det(A_I) \mathrm{tr}(A_I^{-1} \partial_\mu \tilde{A} R_I)\,. \end{aligned}$$

Using the linearity and the cyclic properties of the trace, we find

$$\frac{\partial_\mu \Phi}{\Phi} = \mathrm{tr}(\Gamma \partial_\mu \tilde{A}) \,, \tag{50}$$

where we can identify Γ

$$\Gamma = \frac{1}{\Phi} \sum_I c_I \det(A_I) R_I A_I^{-1} \,. \tag{51}$$

In the expression (51), the application of R_I on the left of A_I^{-1} dispatches the N lines of A_I^{-1} in a larger $N_{\mathrm{orb}} \times N$ matrix. Of course, a direct evaluation of (51) would be $O(N_e N^3)$ and would be too costly.

Properties of the matrix Γ

Γ is a right inverse of $\tilde{A}$, i.e.

$$\tilde{A}\Gamma = I_N \,, \tag{52}$$

where I_N is the identity matrix of order N. The proof is simple

$$\begin{aligned} \tilde{A}\Gamma &= \frac{1}{\Phi} \sum_I c_I \det(A_I) \tilde{A} R_I A_I^{-1} & (53) \\ &= \frac{1}{\Phi} \sum_I c_I \det(A_I) A_I A_I^{-1} = I_N \,. & (54) \end{aligned}$$

We now consider the $N_{\mathrm{orb}} \times N_{\mathrm{orb}}$ matrix $\Gamma\tilde{A}$ and resort to the transformation $\phi_i \to \phi_i + \mu_{ij}\phi_j$. The only non-zero column of the matrix $\partial\tilde{A}/\partial\mu_{ij}$ is the i^{th} column, which is the same as the j^{th} column of $\tilde{A}$. Therefore,

$$\frac{1}{\Phi}\frac{\partial \Phi}{\partial \mu_{ij}} = \mathrm{tr}\left(\Gamma \frac{\partial \tilde{A}}{\partial \mu_{ij}}\right) = (\Gamma\tilde{A})_{ij} \,, \tag{55}$$

meaning that $\Phi(\Gamma\tilde{A})_{ij}$ is the new value of the determinantal expansion when the orbital i has been replaced by the orbital j

$$\Phi(\Gamma\tilde{A})_{ij} = \sum_I c_I \det(A_I^{i\to j})\,. \tag{56}$$

In particular, if $i = j$,

$$\Phi(\Gamma\tilde{A})_{ii} = \sum_{I/\phi_i\in A_I} c_I \det(A_I)\,. \tag{57}$$

In other words, the main diagonal of $\Phi\Gamma\tilde{A}$ is made of restrictions of the summation in (3) to determinants containing a given orbital. As a by-product, if ϕ_i is common to all the determinants of the expansion, $(\Gamma\tilde{A})_{ii}$ is equal to 1. If $i \neq j$, $\Phi(\Gamma\tilde{A})_{ij}$ is the expansion (3) restricted to Slater determinants occupied by ϕ_i and not by ϕ_j

$$\Phi(\Gamma\tilde{A})_{ij} = \sum_{I/\phi_i\in A_I, \Phi_j\notin A_I} c_I \det(A_I^{i\to j})\,. \tag{58}$$

In particular, if the orbital j is common to all determinants, $(\Gamma\tilde{A})_{ij} = 0$. In conclusion, if there are $N_{\rm act}$ orbitals which can be excited (i.e. there are $N - N_{\rm act}$ orbitals common to all determinants), the following property holds: $\Gamma\tilde{A}$ contains a $N_{\rm orb} \times (N - N_{\rm act})$ block which is zero with the exception of a $(N - N_{\rm act}) \times (N - N_{\rm act})$ square sub-block which is the identity matrix.

D Calculation of Γ using the Sherman-Morrison-Woodbury formula

Here, we derive the expression (24) directly from the identity (51) using the Sherman-Morrison-Woodbury formula. The algebra is a bit more tedious. First, we remind some notations useful to explicit the matrix R_I and dependencies on $\tilde{A}$. A is the reference Slater

matrix and R is the matrix which selects the columns $\tilde{A}$ from which A is made

$$A = \tilde{A}R\,. \tag{59}$$

P_I is the matrix such that AP_I is the list of the k_I columns of A which differ from those of A_I (see for example Eq. 13). The $N \times N$ matrix $P_I P_I^T$ is a diagonal matrix: if i is the index of a column which differ in A and A_I, $(P_I P_I^T)_{ii} = 1$, while $(P_I P_I^T)_{ii} = 0$ otherwise. Consequently, the identity

$$A_I - A = (A_I - A)P_I P_I^T \tag{60}$$

holds. The list of excited orbitals are the columns of $A_I P_I$ and can be selected from $\tilde{A}$ with the aid of the $N_{\rm orb} \times k_I$ matrix Q_I such that

$$A_I P_I = \tilde{A} Q_I\,, \tag{61}$$

as in the example Eq. (14). With these definitions

$$A_I = A + (A_I - A)P_I P_I^T = \tilde{A}\left(R + (Q_I - RP_I)P_I^T\right)\,, \tag{62}$$

and the matrix R_I which selects the columns $\tilde{A}$ from which A_I, is given by

$$R_I = R + (Q_I - RP_I)P_I^T\,. \tag{63}$$

Now, writing $A_I = A + (A_I - A)P_I P_I^T$ and applying the Sherman-Morrison-Woodbury formula, we obtain

$$\begin{aligned} A_I^{-1} &= A^{-1} - A^{-1}(A_I - A)P_I(1 + P_I^T A^{-1}(A_I - A)P_I)^{-1}\, P_I^T A^{-1} \\ &= A^{-1} - A^{-1}(A_I - A)P_I(P_I^T A^{-1} A_I P_I)^{-1}\; P_I^T A^{-1}\,, \end{aligned} \tag{64}$$

so that

$$A_I^{-1} = A^{-1} + P_I \alpha_I^{-1} P_I^T A^{-1} - A^{-1} \tilde{A} Q_I \alpha_I^{-1} P_I^T A^{-1} \,, \tag{65}$$

where we have introduced

$$\alpha_I \equiv P_I^T A^{-1} \tilde{A} Q_I \,. \tag{66}$$

Multiplying both sides of Eq. (65) by P_I^T gives the following identity

$$\begin{aligned} P_I^T A_I^{-1} &= P_I^T P_I \alpha_I^{-1} \; P_I^T A^{-1} \\ &= \alpha_I^{-1} \; P_I^T A^{-1} \,. \end{aligned} \tag{67}$$

Using this expression, we can simplify

$$\begin{aligned} R_I A_I^{-1} &= R A_I^{-1} + (Q_I - R R_I) \alpha_I^{-1} \; P_I^T A^{-1} \\ &= R A^{-1} - R A^{-1} \tilde{A} Q_I \alpha_I^{-1} P_I^T A^{-1} + Q_I \alpha_I^{-1} P_I^T A^{-1} \\ &= R A^{-1} + (1 - R A^{-1} \tilde{A}) Q_I \alpha_I^{-1} P_I^T A^{-1} \end{aligned} \tag{68}$$

From equations (68) and (51), we then obtain

$$\Gamma = R A^{-1} + \; (1 - R A^{-1} \tilde{A}) Y A^{-1} \,, \tag{69}$$

with

$$Y \equiv \frac{\det(A)}{\Phi} \sum_I c_I \det(\alpha_I) Q_I \alpha_I^{-1} P_I^T \tag{70}$$

and, of course,

$$\Phi \;=\; \det(A) \left(\sum_{I=1}^{N_e} c_I \det(\alpha_I) \right) \,. \tag{71}$$

Acknowledgement

C.F. acknowledges support from the Netherlands Organization for Scientific Research (NWO) for the use of the SURFsara supercomputer facilities.

References

(1) 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**, *98*, 110201.

(2) Sorella, S.; Capriotti, L. Algorithmic differentiation and the calculation of forces by quantum Monte Carlo. *J. Chem. Phys.* **2010**, *133*, 234111–234121.

(3) Filippi, C.; Assaraf, R.; Moroni, S. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 194105.

(4) Clark, B. K.; Morales, M. A.; McMinis, J.; Kim, J.; Scuseria, G. E. Computing the energy of a water molecule using multideterminants: A simple, efficient algorithm. *J. Chem. Phys.* **2011**, *135*, 244105.

(5) Assaraf, R.; Caffarel, M. Zero-Variance Principle for Monte Carlo Algorithms. *Phys. Rev. Lett.* **1999**, *83*, 4682–4685.

(6) Filippi, C.; Umrigar, C. J. Correlated sampling in quantum Monte Carlo: A route to forces. *Phys. Rev. B* **2000**, *61*, R16291–R16294.

(7) Assaraf, R.; Caffarel, M. Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces. *J. Chem. Phys.* **2003**, *119*, 10536–10552.

(8) Sorella, S.; Casula, M.; Rocca, D. Weak binding between two aromatic rings: feeling the van der Waals attraction by quantum Monte Carlo methods. *J. Chem. Phys.* **2007**, *127*, 014105.

(9) Neuscamman, E.; Umrigar, C. J.; Chan, G. K.-L. Optimizing large parameter sets in variational quantum Monte Carlo. *Phys. Rev. B* **2012**, *85*, 045103.

(10) C. Attaccalite, and S. Sorella, Stable Liquid Hydrogen at High Pressure by a Novel Ab Initio Molecular-Dynamics Calculation. *Phys. Rev. Lett.* **2008**, *100*, 114501.

(11) CHAMP is a quantum Monte Carlo program package written by C. J. Umrigar, C. Filippi, S. Moroni and collaborators.

(12) Burkatzki, M.; Filippi, C.; Dolg, M. Energy-consistent pseudopotentials for quantum Monte Carlo calculations. *J. Chem. Phys.* **2007**, *126*, 234105.

(13) For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. Dolg, M.; Filippi, C., private communication.

(14) We take the diffuse functions from the aug-cc-pVXZ basis sets in the EMSL Basis Set Library (http://bse.pnl.gov). We do not include the g functions in the cc-pVQZ and aug-cc-pVQZ basis sets.

(15) As Jastrow factor, we use the exponential of the sum of two fifth-order polynomials of the electron-nuclear and the electron-electron distances, respectively, and rescale the inter-particle distances as $R = (1 - \exp(-\kappa r))/\kappa$ with κ set to 0.6 a.u. We employ different electron-nucleus Jastrow factors to describe the correlation of an elecron with C and H. The total number of free parameters to be optimized in the Jastrow factor is 13 for all the systems considered here.

(16) 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.; Windus, T. L.; Dupuis, M.; Jr, J. A. M. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(17) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C., Frenking, G., Kim, K., Scuseria, G., Eds.; Elsevier: Amsterdam, 2011; Chapter 41, pp 1167–1190.

(18) Feller, D.; Craig, N. C. High Level ab Initio Energies and Structures for the Rotamers of 1,3-Butadiene. *J. Phys. Chem. A* **2009**, *113*, 1601–1607.

(19) Barborini, M.; Guidoni, L. -Conjugation in trans-1,3-Butadiene: Static and Dynamical Electronic Correlations Described through Quantum Monte Carlo. *J. Chem. Theory Comput.* **2015**, *11*, 508–517.

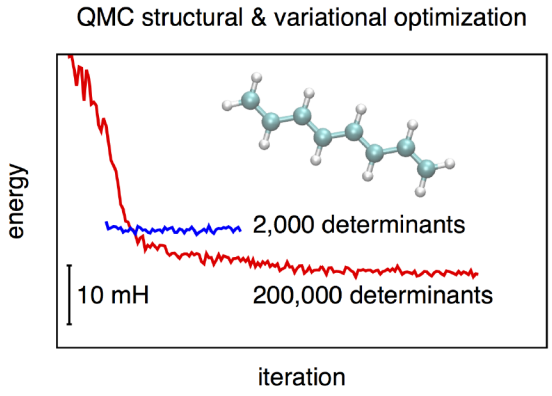
(20) Scemama, A.; Applencourt, T.; Giner, E.; Caffarel, M. Quantum Monte Carlo with very large multideterminant wavefunctions. *J Comp. Chem.* **2016**, *37*, 1866–1875.

(21) For an all-electron move, the second term in the $O(N^3) + O(N^2\,N_e) + O(N_e)$ scaling of the SMF algorithm is reduced to $O(N^2\,\sqrt{N_e})$ in Ref. 20. In our formulation, we have eliminated this term, so the cost of an all-electron move simply scales as $O(N_e)$.

(22) Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109.

(23) Toulouse, J.; Assaraf, R.; Umrigar, C. J. Introduction to the Variational and Diffusion Monte Carlo Methods. *Adv. Quantum Chem.* **2016**, *73*, 285 – 314.

TOC figure


QMC structural & variational optimization
energy
2,000 determinants
10 mH
200,000 determinants
iteration