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

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

1 Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire de Chimie Théorique (UMR7616), 4 place Jussieu F-75252 Paris, France
2 CNR-ION DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy
‡‡MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

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 electrons of the system, the scaling of \( O(N^3) \) per derivative we have recently reported is here lowered to \( O(N^2) \) for the entire set of derivatives. As a function of the number of electrons \( N \), the scaling of \( O(N^3) \) per derivative is lowered to \( O(N^2) + O(NN_e) \), the same as computing the energy alone in the sampling process. The scaling is demonstrated on linear polyenes up to \( C_{10}H_{18} \) and the efficiency of the method is illustrated with the structural optimization of butadiene and octatetraene with Jastrow—Slater wave functions comprising as many as 200 000 determinants and 60 000 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 \( \Psi(R) \), where \( R = (r_1...r_N) \) represents the coordinates of the \( N \) electrons of the system. Here, the vector \( r_i = (x_i,y_i,z_i) \) indicates the three spatial coordinates of the electron \( i \), and its spin component \( \sigma_i \) \( (\sigma_i = \pm \frac{1}{2}) \). This flexibility stems from the fact that integrals are not computed analytically but from a stochastic sampling. For example, the variational energy is

\[
E = \int dR \Psi^2(R) \frac{\hat{H} \Psi}{\Psi}(R)
\]

where \( \hat{H} \) is the Hamiltonian and \( \Psi \) is normalized. Equation 1 can be interpreted as the expectation value of a random variable, the so-called local energy \( E_i = \int \Psi^2 \hat{H} \Psi \), on the probability density \( \Psi^2(R) \). 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 \( \Psi \), \( E \) is typically computed with a scaling \( O(N^2) \) in memory requirements and \( O(N) \) in CPU per Monte Carlo step. In practice, one needs to optimize the parameters of \( \Psi \) and the geometry of a molecular system. Despite the availability of stable wave function optimization methods, such techniques remain costly, and one of the main reasons is that a large number of derivatives of \( E \) (typically \( O(N^3) \)) must 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 \( 3N_{\text{dms}} = O(N) \) intermolecular forces can be calculated with a scaling \( O(N^2) \) instead of \( O(N^3) \) with the use of algorithmic differentiation techniques. We have recently recovered the same reduction using transparent matrical formulas and extended it to the \( O(N^2) \) orbital coefficients.

For multireference expansions, the Jastrow—Slater wave function takes the form

\[
\Psi(R) = J(R) \sum_{i=0}^{N_e} c_i D_i
\]

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

\[
\Phi = \sum_{i=0}^{N_e} c_i \text{det}(A_i)
\]

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

\[
\Lambda_{ij} = \phi_j(r_i)
\]

Received: June 23, 2017
Published: September 5, 2017
For this type of wave functions, Clark et al.\(^4\) have proposed a method to compute \(\Psi\) with a scaling \(O(N_e)\) and \(E_t\) with a scaling \(O(\text{NNN})\). We have further reduced this scaling to \(O(N_e)\) and extended it to any derivative of \(E_t\).\(^3\) The derivatives of \(E_t\) 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\(^5\) that for any one-body operator \(\tilde{O}\), such as the Laplacian or the nonlocal pseudopotential contributions to \(E_t\), the expression \((\tilde{O} \partial_{\tilde{A}}) / \partial_{\tilde{A}}\) is a first-order derivative with respect to a suitable parameter \(\lambda\). For instance, while being a sum of second-order derivatives with respect to the electron coordinates, the Laplacian acting on any determinant \(A_t\) of the expansion can be recast in the form

\[
\frac{\nabla^2 D_t}{D_t} = \frac{\nabla^2 \det A_t}{\det A_t} = \text{tr}(A_t^{-1} \partial_i \partial_j A_t) = \frac{d}{d\lambda} \ln \det(A_t + \lambda B_t)|_{\lambda=0}
\]

(5)

where we have introduced an operator-specific matrix which, for the current expression of the Laplacian, is the \(N \times N\) matrix \((B_t)_{ij} = \nabla^2 \phi_i(r)\) constructed from the orbitals occupied in \(A_t\). 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 \(\Psi\) and \(E_t\). The core observation is that the determinantal part \(\Phi\) is a function of the matrix elements \(\gamma_{\mu\nu}\) and that any derivative of \(\Phi\) can be computed using a simple trace formula involving the matrix \(\Gamma\) defined as the logarithmic gradient of \(\Phi\) with respect to \(\tilde{A}\). The first derivatives of the local energy \(\partial_{\tilde{A}} E_t\) can then be expressed as traces involving \(\Gamma\) and one of its derivatives \(\partial_{\tilde{A}} \Gamma\). Many derivatives of \(\Psi\) and \(E_t\) are obtained efficiently because the matrices \(\Gamma\) and \(\partial_{\tilde{A}} \Gamma\) 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 \(\tilde{c}\), 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 \(\Gamma\). In section 3, we present a formula to compute \(\Gamma\) at a cost \(O(N^3) + O(N_e)\) and, in section 4, discuss the formulas for the second derivatives of \(\Phi\) and, specifically, the first derivatives of \(E_t\). In section 5, we demonstrate the scaling of the computation of interatomic forces with multideterminant wave functions on polyenes up to \(C_{20}H_{62}\) 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 \(\Phi\) 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 \(\Phi\) with respect to a given parameter \(\mu\) in \(\tilde{A}\) is obtained from the chain rule

\[
\frac{\partial}{\partial \tilde{A}_{ij}} \ln \Phi = \nabla \Phi \cdot \nabla \tilde{A}_{ij} = \text{tr}(\Gamma \nabla \tilde{A})
\]

(6)

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

\[
\Gamma_{ij} = \frac{\partial \ln \Phi}{\partial \tilde{A}_{ij}}
\]

(7)

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

Finally, also quantities such as 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.\(^5\)

3. EFFICIENT EVALUATION OF THE MATRIX \(\Gamma\)

3.1. Convenient Expression for \(\Phi\). The determinants of the Slater matrices \(A_t\) can be computed efficiently because \(A_t\) usually differs by a few columns from the reference Slater matrix \(A_0\). For example, let \(A_t\) be the \(4 \times 4\) Slater matrix built with the orbitals \(\phi_1, \phi_2, \phi_3, \phi_4\):

\[
A_t = (\tilde{A}_1 \quad \tilde{A}_2 \quad \tilde{A}_3 \quad \tilde{A}_4)
\]

(8)

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

\[
A_{t_1} = (\tilde{A}_1 \quad \tilde{A}_2 \quad \tilde{A}_3 \quad \tilde{A}_4)
\]

(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 \(A_t\) is

\[
\det(A_t) = \det(A) \det(A^{-1} A_t)
\]

and

\[
A^{-1} A_t = (A^{-1} \tilde{A}_1 \quad A^{-1} \tilde{A}_2 \quad A^{-1} \tilde{A}_3 \quad A^{-1} \tilde{A}_4)
\]

\[
\begin{pmatrix}
1 & 0 & (A^{-1} \tilde{A})_{15} & (A^{-1} \tilde{A})_{17} \\
0 & 1 & (A^{-1} \tilde{A})_{25} & (A^{-1} \tilde{A})_{27} \\
0 & 0 & (A^{-1} \tilde{A})_{35} & (A^{-1} \tilde{A})_{37} \\
0 & 0 & (A^{-1} \tilde{A})_{45} & (A^{-1} \tilde{A})_{47}
\end{pmatrix}
\]

(10)

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

\[
\det(A^{-1} A_t) = \det\begin{pmatrix}
(A^{-1} \tilde{A})_{15} & (A^{-1} \tilde{A})_{17} \\
(A^{-1} \tilde{A})_{25} & (A^{-1} \tilde{A})_{27} \\
(A^{-1} \tilde{A})_{35} & (A^{-1} \tilde{A})_{37} \\
(A^{-1} \tilde{A})_{45} & (A^{-1} \tilde{A})_{47}
\end{pmatrix}
\]

(11)

More generally, the determinant of \(A^{-1} A_t\) for a \(k^\text{th}\)-order excitation is the determinant of a \(k_i \times k_i\) submatrix. Such a submatrix can always be written as follows

\[
\alpha_t = P_i A^{-1} \tilde{A} Q_i
\]

(12)

where, in our example,

\[
P_i = \begin{pmatrix}
0 & 0 \\
0 & 0 \\
1 & 0 \\
0 & 1
\end{pmatrix}
\]

(13)
To summarize, the expression for the determinant of a small submatrix of 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.\textsuperscript{3,4} Finally, the convenient expression for $\Phi$ to efficiently compute $\Gamma$ is

$$\Phi = \text{det}(A) \times \sum_i c_i \text{det}(P_i^T A^{-1} \tilde{A} Q_i)$$

### 3.2. Convenient Expression for $\Phi$

Introducing the matrix $R$ such that $A = AR$, the expression for $\Phi$ in eq 16 is explicitly a function of $\tilde{A}$. In particular, the summation on the right hand side of eq 16

$$\chi \equiv \frac{\Phi}{\det(A)} = \sum_i c_i \text{det}(P_i^T A^{-1} \tilde{A} Q_i)$$

is a polynomial function depending on the matrix elements of $T \equiv A^{-1} \tilde{A} = (AR)^{-1} \tilde{A}$

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

$$\frac{\partial}{\partial \nu} \ln(\Phi) = \frac{\partial}{\partial \nu} \ln(A) + \frac{\partial}{\partial \nu} \ln(\chi) = \text{tr}(A^{-1} \partial \nu A) + \frac{\partial}{\partial T_{ij}} \ln(\chi) \frac{\partial T_{ij}}{\partial \nu}$$

$$= \text{tr}(A^{-1} \partial \nu A) + \text{tr}(Y \partial \nu T)$$

where

$$Y_{ij} \equiv \frac{\partial \ln(\chi)}{\partial T_{ij}} = \frac{1}{\chi} \frac{\partial \chi}{\partial T_{ij}}$$

It is simple to show that

$$\frac{\partial \chi}{\partial T_{ij}} = \sum_{i,j=0}^{N} c_i \text{det}(a_i)(Q_i a_i^{-1} P_i^T)_{ij}$$

The derivative of $T$ is given by

$$\partial \nu T = -A^{-1} \partial \nu AA^{-1} \tilde{A} + A^{-1} \partial \nu \tilde{A}$$

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

$$\partial \nu \ln(\Phi) = \text{tr}(\Gamma \partial \nu \tilde{A})$$

where

$$\Gamma = RA^{-1} + (1 - RA^{-1})Y A^{-1} = [R(1 - A^{-1} \tilde{A} Y) + Y]A^{-1}$$

For example, if the occupied orbitals are the $N$ first ones, the matrix $\Gamma$ is

$$\Gamma = \left( A^{-1} - A^{-1} \tilde{A} Y A^{-1} \right) Y_{\text{virt}} A^{-1}$$

where the first line is a $N \times N$ matrix. The second line is a $(N_{\text{eb}} - N) \times N$ matrix where $Y_{\text{virt}}$ represents the nonzero lines of $Y$, that is, the last $N_{\text{eb}} \equiv N_{\text{eb}} - N$ lines.

3.3. One-Body Operators and First-Order Derivatives of $\Phi$

First-order derivatives of $\Phi$ can be computed with the trace formula \textit{6} which involves the $\Gamma$ 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 $\tilde{A} = \tilde{A} + \lambda \tilde{B}$

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

$$\Phi = \det(A_j)[\sum_i c_i \text{det}(P_i^T A_j^{-1} \tilde{A} Q_i)]$$

From this expression, one can compute the local energy

$$E_L = \partial \nu (\ln \Phi) = \text{tr}(\Gamma \tilde{B})$$

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

4. SECOND-ORDER DERIVATIVES

The second derivative of $\Phi$ can be written in terms of $\Gamma$ and its derivative as

$$\partial \nu \partial \nu \ln(\Phi) = \partial \nu \text{tr}(\Gamma \partial \nu \tilde{A}) = \text{tr}(\Gamma \partial \nu \tilde{A}) + \text{tr}(\partial \nu \Gamma \partial \nu \tilde{A})$$

### 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 $\nu$ is

$$\partial \nu E_L = \partial \nu \partial \nu \ln(\Phi) = \text{tr}(\Gamma \partial \nu \tilde{B}) + \text{tr}(\partial \nu \Gamma \partial \nu \tilde{A})$$

The order of the derivatives has been chosen so that $\tilde{A}$ and not $\Gamma$ is differentiated with respect to $\nu$. Consequently, the matrix $\partial \nu \Gamma$ does not depend on the parameter $\nu$ and has to be computed only once, whatever the number of second derivatives we need. Once $\partial \nu \Gamma$ has been computed, the calculation of $\partial \nu E_L$ involves (besides $\partial \nu \tilde{A}$ and $\partial \nu \tilde{B}$) two traces which can be computed at a cost $O(N_{\text{eb}} N_A)$. Importantly, such a calculation does not depend on $N_{\text{eb}}$ in contrast to what was presented in ref 3.
Efficient Calculation of $\partial \Gamma$. The derivative of $\Gamma$ is

$$\partial \Gamma = [-\Gamma B + \partial \gamma Y + R(\partial \gamma TY + T \partial \gamma Y)]A^{-1}$$

(31)

where

$$\partial \gamma T = A^{-1}(B - BT) \equiv \tilde{M}$$

(32)

Applying the chain rule, we obtain

$$\partial \gamma Y = Z_{ijkl} \tilde{M}_{kl}$$

(33)

where

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

(34)

$$= \frac{1}{\chi} \frac{\partial^2 \chi}{\partial T_{ij} \partial T_{kl}} - Y_{ij} Y_{lk}$$

(35)

It follows from eq 21 that

$$\frac{\partial^2 \chi}{\partial T_{ij} \partial T_{kl}} = \sum_{i,j=0}^{N} c_i \det(\alpha_j) \left[ (Q_i \alpha_i - P_i) (Q_j \alpha_j - P_j) \right]$$

$$- (Q_i \alpha_i - P_i) (Q_j \alpha_j - P_j)$$

(36)

We can compute the derivatives of $\chi$ 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-electron operator on each excited determinant $A_j$ 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

$$\partial \gamma E_i = \partial \gamma \partial (\ln \Phi)$$

$$= \frac{\det(\alpha)}{\Phi} \partial \gamma \det(\alpha_i^T \Phi)$$

$$= \frac{\det(\alpha)}{\Phi} \text{tr}(\alpha_i^{-1} P_i T Q_i)$$

$$= \frac{\det(\alpha)}{\Phi} \text{tr}(\alpha_i^{-1} P_i^T \tilde{M} Q_i)$$

(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 $A$, $A^{-1}$, and $T = A^{-1} \tilde{A}$ at a cost of at most $O(N^5)$ (products and inversions of matrices). Then, we need to calculate the first and second derivatives of $\chi$ 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 $O(N_e)$. $\partial \gamma Y$ is computed at a cost $O(N_e)$, where $N_e$ is the total number of double excitations involved in any $k$th order excitation ($k \geq 2$), where of course $N_i < N_e$. Finally, $\Gamma$ and $\partial \gamma \Gamma$ are computed at a cost $O(N^5)$ (product of matrices).

In particular, computing the $N_{\text{atom}}$ components of the interatomic forces with improved estimators has a scaling

$$O(N^3) + O(N^2 N_{\text{atom}}) + O(N_{\text{atom}} N N_{\text{virt}}) + O(N_e)$$

(38)

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

$$O(N^3) + O(N_e)$$

(39)

This is significantly more efficient than the scaling $O(N^5) + O(N^2 N_{\text{atom}}) + O(N_e)$ in the large $N_e$ regime. The term $O(N^2 N_{\text{atom}})$ is no more present because here we avoid computing $\partial \gamma T$. 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 Figure 1, we demonstrate this favorable scaling in the variational Monte Carlo (VMC) computation of the interatomic forces for multideterminant Jastrow–Slater wave functions using the sequence of molecules $\text{C}_n\text{H}_{n+2}$ 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.

![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 $\text{C}_n\text{H}_{n+2}$ 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.](https://example.com/figure1)

6. NUMERICAL RESULTS

We demonstrate the formulas above on the ground-state structural optimization in VMC of butadiene ($\text{C}_4\text{H}_6$) and octatetraene ($\text{C}_8\text{H}_{10}$) 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 58 652) we employ the stochastic reconfiguration optimization method in a conjugate gradient implementation 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. To optimize the

geometry, we simply follow the direction of steepest descent and appropriately rescale the interatomic forces. We employ the CHAMP code\textsuperscript{11} with scalar-relativistic energy-consistent Hartree–Fock pseudopotentials and the corresponding cc-pVXZ\textsuperscript{12,13} and aug-cc-pVXZ\textsuperscript{14} basis sets with $X = D$, $T$, and $Q$. The Jastrow factor includes two-body electron–electron and electron–nucleus correlation terms.\textsuperscript{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).\textsuperscript{16,17}

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\textsuperscript{19} for a recent compilation of theoretical predictions). In particular, Barborini and Guidoni\textsuperscript{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.\textsuperscript{18}

Here we explicitly address the relevance of the multireference character of the system, considering various expansions which correlate the $\pi$ and $\sigma$ 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 $\pi$ electrons in the bonding and antibonding $\pi$ orbitals constructed from the $2p_x$, $3p_x$, $3d_{\sigma\pi}$, and $4p_x$ atomic orbitals; (c) a CAS(10,10) correlating the six $\sigma$ and four $\pi$ electrons of the carbon atoms in the corresponding bonding and antibonding $\pi$ and $\sigma$ orbitals (15 912 determinants); (d) the same CAS(10,10) expansion augmented with single and double excitations in the external space of 12 $\pi$ orbitals and truncated with a threshold of $2 \times 10^{-6}$ on the coefficients of the spin-adapted configuration state functions. This last choice results in a total of 45 644 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 Figure 2. We find that the basis sets of triple- and quadruple-$\zeta$ quality yield values of BLA which are compatible within 1–1.5 standard deviations, namely, to better than 5 $\times 10^{-4}$ 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\textsuperscript{19}, possibly due to their use of a basis set of quality inferior to triple-$\zeta$. Moving beyond a single determinant, we observe a strong dependence of the result on the choice of active space. The inclusion of $\pi-\pi$ correlation within 4, 16, and 20 $\pi$ 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 $\sigma-\pi$ and $\sigma-\sigma$ correlations in a CAS(10,10) (case c) leads however to a more substantial

![Figure 2. Bond-length alternation (BLA) of C4H6 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\textsuperscript{19} and the best value obtained with a Jastrow-antisymmetrized geminal power (JAGP)\textsuperscript{19} are reported.](image)

7. CONCLUSION

In the past few years, a series of methodological advances\textsuperscript{3,4,20} has enabled the use of large multireference 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.

\section*{A. EFFICIENT CALCULATION OF Z, Y, AND χ}

We demonstrate here that we do not need to compute explicitly the inverses of the submatrices \( a_i \) as in eqs 17, 22, and 36 or in refs 3 and 4 to obtain \( \chi \) and its derivatives. These can be computed efficiently using recursion formulas.

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

\[
\chi = \sum_{i_1 < i_2 < i_3 < j_1 < j_2 < j_3} C_{i_1 j_1 i_2 j_2 i_3 j_3} \sum_P (-1)^P T_{\gamma_{i_1 j_1} \gamma_{i_2 j_2} \gamma_{i_3 j_3}} \tag{40}
\]

where \( P \) stands for a permutation of the indices \((i_1, j_1, i_2, j_2, i_3, j_3)\), 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 \((i_1 \rightarrow j_1, i_2 \rightarrow j_2)\) can be written as \((i_1 \rightarrow j_1, i_2 \rightarrow j_2, i_1 \rightarrow j_2, i_2 \rightarrow j_1)\), and a first-order excitation \((i_1 \rightarrow j_1)\) as \((i_1 \rightarrow j_1, i_2 \rightarrow j_2, i_1 \rightarrow j_2, i_2 \rightarrow j_1)\).

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

\[
\frac{\partial^2 \chi}{\partial T_{i_1 j_1} \partial T_{i_2 j_2}} = \sum_{i_3 j_3} (-1)^P P_{\gamma_{i_1 j_1} \gamma_{i_2 j_2} \gamma_{i_3 j_3}} \sum_P \sum_{i_1 j_1} C_{i_1 j_1 i_2 j_2 i_3 j_3} T_{\gamma_{i_3 j_3}} \tag{41}
\]

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

\[
\frac{\partial \chi}{\partial T_{i_1 j_1}} = \frac{1}{2} \sum_{i_2 j_2} \frac{\partial^2 \chi}{\partial T_{i_2 j_2} \partial T_{i_1 j_1}} = \frac{1}{2} \sum_{i_3 j_3} \sum_P \sum_{i_1 j_1} (-1)^P P_{\gamma_{i_1 j_1} \gamma_{i_2 j_2} \gamma_{i_3 j_3}} C_{i_1 j_1 i_2 j_2 i_3 j_3} T_{\gamma_{i_3 j_3}} \tag{42}
\]

and the value of \( \chi \) is

\[
\chi = \frac{1}{3} \sum \frac{\partial \chi}{\partial T_{i_1 j_1}} \tag{43}
\]

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 \( N^2 N_{\text{orb}} \) and \( N N_{\text{orb}} \) operations (additions or multiplications), respectively. The method still scales like \( O(N) \) but with a reduced prefactor because no divisions are involved and the number of operations is smaller. For example, expression 43 involves at most \( NN_{\text{orb}} \) multiplications and additions whereas (17) is a sum on \( N_{\text{orb}} \) terms (\( N_{\text{orb}} \) can be of order \( N^2 N_{\text{orb}} \) if third-order excitations are included).

\section*{B. ONE-ELECTRON-MOVE ALGORITHMS}

To sample the density \( \Psi'^2 \), we use the Metropolis–Hastings method\(^{22,23}\), which is a stochastic dynamics in the space of configurations \( R = (r_1, r_2, \ldots, r_N) \). For a given iteration, this method proposes a random move \( R \rightarrow R' \) with a transition probability density \( P(R \rightarrow R') \). The proposed move is accepted with the probability

\[
\min \left\{ \frac{\Psi'^2(R')}{\Psi^2(R)}, \frac{P(R' \rightarrow R)}{P(R \rightarrow R')} \right\} \tag{44}
\]
If only one electron is moved (here the first, for example), the new configuration is $R' = (r_1', r_2, ..., r_N)$. The new extended Slater matrix $\bar{A}'$ differs from $\bar{A}$ only in the first line.

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

$$\frac{\Phi(R')}{\Phi(R)} = \partial \ln \Phi(\bar{A})$$

where we considered the following transformation $\bar{A} \rightarrow \bar{A} + \lambda \bar{B}_e$. Using eq 19, we obtain

$$\frac{\Phi(R')}{\Phi(R)} = \text{tr}(\bar{A}^{-1} \partial \bar{A} + \text{tr}(\partial Y))$$

(46)

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

We note that, if one uses instead the expression involving $\Gamma$ to update the wave function,

$$\frac{\Phi(R')}{\Phi(R)} = \text{tr}(\Gamma \bar{B}_e)$$

(47)

one would need to update $\Gamma$ at each Monte Carlo step and incur the higher cost of $O(N^3) + O(NN)$ for a full sweep. This is because updating $\Gamma$ requires the calculation of $\partial \Gamma$, given in eq 31, where of course $\bar{B}$ is replaced by $\bar{B}_e$. In this equation, the product $(\partial \gamma Y)A^{-1}$ 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 $\Gamma_{\partial \Phi/\partial \Phi}$ needed in the Monte Carlo sampling, it is better not to recompute $\Gamma$ but to use formula 46 with $\partial T = A^{-1} \bar{B}_{\text{drift}} - A^{-1} \bar{B}_e \bar{B}_e T$, where the matrix $\bar{B}_{\text{drift}}$ is zero except the $i$th row which equals $V \partial \phi_i(r)$. 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 $O(N^3) + O(NN)$ per sweep, using eq 46 or 47 alike.

**3. SIMPLE EXPRESSION OF $\Gamma$ FOR A JASTROW–SLATER EXPANSION**

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

**Simple Expression for $\Gamma$**

The determinantal contribution of the wave function written in eq 3 is

$$\Phi = \sum_{i=0}^N c_i \det(A_i)$$

where $A_i$ is a list of $N$ columns of the $N \times N_{\text{orb}}$ generalized Slater matrix $\bar{A}$. We can then define a $N_{\text{orb}} \times N$ matrix $R_i$ such that

$$A_i = \bar{A} R_i$$

(48)

which gives an explicit expression of $\Phi$ as a function of $\bar{A}$

$$\Phi(\bar{A}) = \sum_j c_i \det(\bar{A} R_i)$$

(49)

For example, given a $3 \times 3$ Slater matrix built on the orbitals $(\phi_1, \phi_2, \phi_3)$

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{pmatrix}$$

The derivative of the determinantal expansion with respect to a parameter $\mu$ is

$$\frac{\partial \Phi}{\partial \mu} = \sum_j c_i \det(A_i) \text{tr}(A_i^{-1} \partial \mu \phi_i)$$

(50)

where we can identify $\Gamma$

$$\Gamma = \frac{1}{\Phi} \sum_j c_i \det(A_i) R_i A_i^{-1}$$

(51)

In 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 $N_{\text{orb}} \times N$ matrix. Of course, a direct evaluation of expression 51 would be $O(N_{\text{orb}} N^3)$ and would be too costly.

**Properties of the matrix $\Gamma$**

$\Gamma$ is a right inverse of $\bar{A}$, that is,

$$\bar{A} \Gamma = I_N$$

(52)

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

$$\bar{A} \Gamma = \frac{1}{\Phi} \sum_j c_i \det(A_i) A_i R_i A_i^{-1}$$

(53)

$$= \frac{1}{\Phi} \sum_j c_i \det(A_i) A_i A_i^{-1} = I_N$$

(54)

We now consider the $N_{\text{orb}} \times N_{\text{orb}}$ matrix $\Gamma \bar{A}$ and resort to the transformation $\phi_i \rightarrow \phi_i + \mu \phi_i$, the only non-zero column of the matrix $\partial \bar{A}/\partial \mu_i$ is the $i$th column, which is the same as the $j$th column of $\bar{A}$. Therefore,

$$\frac{1}{\Phi} \frac{\partial \Phi}{\partial \mu_i} = \text{tr} \left( \Gamma \frac{\partial \bar{A}}{\partial \mu} \right) = (\Gamma \bar{A})_{ij}$$

(55)

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

$$\Phi(\Gamma \bar{A})_{ij} = \sum_j c_i \det(A_i^{(\mu)})$$

(56)

In particular, if $i = j$

$$\Phi(\Gamma \bar{A})_{ii} = \sum_{i \neq \phi \in A_i} c_i \det(A_i)$$

(57)

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

\[
\Phi(\Gamma \vec{A})_{ij} = \sum_{l \, | \phi_l \in \Omega_i, \phi_l \notin \Omega_j} c_l \det(A_l^{i-j+1})
\]

(58)

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

**D. CALCULATION OF \( \Gamma \) USING THE SHERMAN–MORRISON–WOODBURY FORMULA**

Here, we derive expression 24 directly from the identity (eq 51) using the Sherman–Morrison–Woodbury formula. The algebra is a bit more tedious. First, we recall some notations useful to explicit the matrix \( R_i \) and dependencies on \( \vec{A} \). A is the reference Slater matrix and \( R \) is the matrix which selects the columns \( \vec{A} \) from which \( A \) is made

\[
A = \vec{A} R
\]

(59)

\( P_i \) is the matrix such that \( A P_i \) 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 differs in \( A \) and \( A_i \) \( (P_i P_i^T)_{ii} = 1 \), while \( (P_i P_i^T)_{ij} = 0 \) otherwise. Consequently, the identity

\[
A_i - A = (A_i - A) P_i P_i^T
\]

(60)

holds. The list of excited orbitals are the columns of \( A P_i \) and can be selected from \( \vec{A} \) with the aid of the \( N_{\text{orb}} \times k_i \) matrix \( Q_i \) such that

\[
A_i P_i = \vec{A} Q_i
\]

(61)

as in the example eq 14. With these definitions

\[
A_i = A + (A_i - A) P_i P_i^T = \vec{A} (R + (Q_i - RP_i) P_i^T)
\]

(62)

and the matrix \( R_i \) which selects the columns \( \vec{A} \) from which \( A_i \) is given by

\[
R_i = R + (Q_i - RP_i) P_i^T
\]

(63)

Now, writing \( A_i = A + (A_i - A) P_i P_i^T \) and applying the Sherman–Morrison–Woodbury formula, we obtain

\[
A_i^{-1} = A^{-1} - A^{-1}(A_i - A)(P_i (1 + P_i^T A^{-1} (A_i - A) P_i)^{-1}) P_i^T A^{-1}
\]

(64)

so that

\[
A_i^{-1} = A^{-1} + P_i \alpha_i^{-1} P_i^T A^{-1} - A^{-1} \vec{A} Q_i \alpha_i^{-1} P_i^T A^{-1}
\]

(65)

where we have introduced

\[
\alpha_i \equiv P_i^T A^{-1} \vec{A} Q_i
\]

(66)

Multiplying both sides of eq 65 by \( P_i^T \) gives the following identity

\[
P_i^T A_i^{-1} = P_i^T P_i \alpha_i^{-1} P_i^T A^{-1} = \alpha_i^{-1} P_i^T A^{-1}
\]

(67)

Using this expression, we can simplify

\[
R_i A_i^{-1} = RA_i^{-1} + (Q_i - RP_i) \alpha_i^{-1} P_i^T A^{-1}
\]

\[
= RA_i^{-1} - RA_i^{-1} \vec{A} Q_i \alpha_i^{-1} P_i^T A^{-1} + Q_i \rho_i^{-1} P_i^T A^{-1}
\]

\[
= RA_i^{-1} + (1 - RA_i^{-1} \vec{A}) Q_i \rho_i^{-1} P_i^T A^{-1}
\]

(68)

From eqs 68 and 51, we then obtain

\[
\Gamma = RA_i^{-1} + (1 - RA_i^{-1} \vec{A}) Y A_i^{-1}
\]

(69)

with

\[
Y \equiv \det(A) \sum_i c_i \det(\alpha_i) Q_i \rho_i^{-1} P_i^T
\]

(70)

and, of course,

\[
\Phi = \det(A) \sum_i c_i \det(\alpha_i)
\]

(71)

**AUTHOR INFORMATION**

**Corresponding Authors**

*E-mail: assaraf@lct.jussieu.fr.*

*E-mail: moroni@democritos.it.*

*E-mail: c.filippi@utwente.nl.*

**ORCID**

Claudia Filippi: 0000-0002-2425-6735

**Funding**

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

**Notes**

The authors declare no competing financial interest.

**REFERENCES**


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


(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 interparticle distances as \( R = (1 - \exp(-\kappa r))/\kappa \) with \( \kappa \) set to 0.6 au. We employ different electron–nucleus Jastrow factors to describe the correlation of an electron 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.


(21) For an all-electron move, the second term in the \( O(N^6) + O(N^5)N + O(N) \) scaling of the SMF algorithm is reduced to \( O(N^2\sqrt{N}) \) in ref 20. In our formulation, we have eliminated this term, so the cost of an all-electron move simply scales as \( O(N) \).
