HAL
open science

Perturbatively Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo

Monika Dash, Saverio Moroni, Anthony Scemama, Claudia Filippi

To cite this version:

Monika Dash, Saverio Moroni, Anthony Scemama, Claudia Filippi. Perturbatively Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo. Journal of Chemical Theory and Computation, 2018, 14 (8), pp.4176-4182. 10.1021/acs.jctc.8b00393 . hal-01873356

HAL Id: hal-01873356
https://hal.science/hal-01873356v1

Submitted on 29 Jan 2020

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.

Perturbatively selected configuration-interaction wave functions for efficient geometry optimization in quantum Monte Carlo

Monika Dash,[1] Saverio Moroni,[2, a)] Anthony Scemama,[3, b)] and Claudia Filippi[1, c)]
[1)] *MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*
[2)] *CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy*
[3)] *Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, France*

arXiv:1804.09610v1 [physics.chem-ph] 25 Apr 2018

We investigate the performance of a class of compact and systematically improvable Jastrow-Slater wave functions for the efficient and accurate computation of structural properties, where the determinantal component is expanded with a perturbatively selected configuration interaction scheme (CIPSI). We concurrently optimize the molecular ground-state geometry and full wave function – Jastrow factor, orbitals, and configuration interaction coefficients– in variational Monte Carlo (VMC) for the prototypical case of 1,3-trans-butadiene, a small yet theoretically challenging π-conjugated system. We find that the CIPSI selection outperforms the conventional scheme of correlating orbitals within active spaces chosen by chemical intuition: it gives significantly better variational and diffusion Monte Carlo energies for all but the smallest expansions, and much smoother convergence of the geometry with the number of determinants. In particular, the optimal bond lengths and bond-length alternation of butadiene are converged to better than one mÅ with just a few thousand determinants, to values very close to the corresponding CCSD(T) results. The combination of CIPSI expansion and VMC optimization represents an affordable tool for the determination of accurate ground-state geometries in quantum Monte Carlo.

I. INTRODUCTION

Quantum Monte Carlo methods are a class of *ab initio* approaches which solve the interacting Schrödinger equation stochastically. The most widely used variants of QMC are the variational (VMC) and diffusion Monte Carlo (DMC). Thanks to their favorable scaling with the number of particles and the ease of parallelization, they have often been employed to benchmark electronic properties, in particular, total energies of relatively large molecules as well as solids. Recently, it has been shown[1,2] that it is possible to compute derivatives of the energy at the same computational cost per Monte Carlo step as evaluating the energy itself, also when employing large determinantal expansions in the commonly used Jastrow-Slater QMC wave functions. Consequently, one can compute all derivatives necessary for the optimization of the structure of a system very efficiently and, simultaneously, of the variational parameters in the wave function, as was demonstrated on the structural optimization of short polyenes with expansions comprising over 200,000 determinants[2]. These developments also allow us to thoroughly explore the sensitivity of QMC calculations to the choice of the Slater expansion, namely, the set of orbitals which one must correlate and the truncation of the active space. It was shown that an instructed guess of the orbitals based on chemical intuition can lead to significant variations in VMC energies and structures of a molecule as small as butadiene even when employing large expansions in the presence of a Jastrow correlation factor[2].

To overcome the limitations of an *a priori* approach in the choice of the orbital set and achieve a compact description of the determinantal component in QMC, a promising alternative is to employ a selected CI algorithm such as the CIPSI (configuration interaction using a perturbative selection done iteratively) method. CIPSI was originally introduced by Huron *et al.* in 1973[3] and continually tested[4–10], improved[10–13], and assessed in comparison to full-CI (FCI) expansions[10,14,15] in combination with a variety of orbital descriptions[16]. In the last few years, there has been renewed interest in the development of selected CI approaches[17–21] to accurately calculate the ground- and excited-state energies of small molecules, establishing the competence of these approaches for benchmarking applications. The use of these expansions has however only been marginally explored within the QMC framework[22–27], where it was shown to yield very good DMC energies at the price of employing large CIPSI wave functions.

Here, we complement the perturbatively selected CIPSI determinants with a Jastrow factor and VMC optimization of the full wave function (Jastrow parameters, orbitals and CI coefficients), and investigate the ability of the resultant Jastrow-CIPSI wave functions to obtain accurate molecular geometries in VMC, as well as corresponding VMC and DMC energies, with relatively compact expansions. We focus on the butadiene molecule, where an accurate estimation of the bond length alternation (BLA) is theoretically quite challenging, demanding a proper description of correlation effects, both static and dynamic. As mentioned above, significant variations in energies and structural parameters were observed in VMC when correlating different sets of σ and π orbitals in a fully optimized Jastrow-Slater wave function[2]: with a large expansion comprising over 45,000 determinants in a restricted active space, good agreement with coupled cluster with singles, doubles, and perturbative triples in the complete basis set limit (CCSD(T)/CBS) was obtained in the bond lengths. This number of determinants is however surprisingly large for

[a)] Electronic mail: moroni@democritos.it
[b)] Electronic mail: scemama@irsamc.ups-tlse.fr
[c)] Electronic mail: c.filippi@utwente.nl

the description of the ground state of such a small system. Indeed, we find here that a determinantal description with CIPSI yields a much smoother convergence in the structural parameters and QMC energies when the size of the expansion is systematically incremented. Correspondingly, very accurate values for these physical properties are already obtained with only a few thousand determinants.

To construct the CIPSI component of the QMC trial wave functions, we follow two different schemes: one where we systematically "expand" the wave function by adding important determinants at every step, and the other where we first generate an extremely large wave function and subsequently "truncate" it to obtain requisite sizes. For a fixed size of the CI expansion, the "truncation" scheme is expected to be more accurate than the "expansion" one since the overlap of the wave function with the FCI wave function is larger. The difference can be important in systems where the CI coefficients change dramatically with the number of determinants in the wave function[28], but generation of the large wave function to initiate "truncation" might not be entirely feasible for larger systems where the "expansion" scheme would therefore represent a computationally less expensive route. For small systems such as butadiene, however, we can employ both strategies and, consequently, draw an assessment of their relative performance. Here, we find that the truncation scheme provides a faster and smoother convergence of the bond lengths and VMC energies of butadiene. We note that we are looking at extremely small variations in the structural parameters (well below mÅ) when comparing and establishing the convergence. Finally, while we primarily employ multi-configurational self consistent field (MCSCF) canonical orbitals in the CIPSI algorithm and as starting orbitals in QMC, we additionally test the use of natural orbitals. Rather peculiarly, expansions with natural orbitals commensurate in size with those with canonical orbitals consistently converge to slightly higher VMC energies.

The paper is organized as follows. In section II, we describe the CIPSI algorithm and the functional form of the Jastrow-Slater wave functions and, in Section III, we report the computational details. The numerical results obtained for the structural optimization of butadiene and a comparison with our previous Jastrow-CAS results[2] are given in Section IV.

II. METHODS

CIPSI is an iterative CI selection algorithm that allows us to perturbatively select determinants from the FCI space. Starting with an initial reference wave function, additional determinants are added to the expansion based on their effective second-order energy contribution which is required to be greater than a fixed or an iteratively modifiable threshold. Selection iterations can be performed until a target number of determinants is reached or until some other selection criterion is met[29]. In the process, the quality of the wave function is systematically improved and converges towards the FCI solution.

A step-by-step description of the CIPSI algorithm is provided in Refs. 22, 24, 25, and 30. The scheme is briefly iterated here. The reference wave function is composed of a linear combination of Slater determinants, D_i, spanning a space, S:

$$\psi_{\mathrm{CIPSI}} = \sum_{D_i \in S} c_i D_i \,. \tag{1}$$

Typically, this reference wave function is initially the single determinant with the lowest energy, which is also the case here. At every iteration, the many-body Hamiltonian, $\hat{\mathcal{H}}$, is diagonalized in the reference space S to obtain the variationally minimized energy E^{ref} and the corresponding c_i coefficients. Then, all the determinants D_j outside of this reference space that are connected to S by $\hat{\mathcal{H}}$ are generated, and their individual contributions to the energy are estimated with the Epstein-Nesbet[31,32] perturbation theory,

$$\delta E_j^{(2)} = \frac{|\langle D_j | \hat{\mathcal{H}} | \psi_{\mathrm{CIPSI}} \rangle|^2}{E^{\mathrm{ref}} - \langle D_j | \hat{\mathcal{H}} | D_j \rangle} \,. \tag{2}$$

If $|\delta E_j^{(2)}|$ is greater than a given threshold, determinant D_j is selected for the next iteration. Summing all these contributions gives E^{PT2}, namely, the second-order perturbative energy correction to E^{ref}, and the CIPSI energy of the current iteration is given by

$$E^{\mathrm{CIPSI}} = E^{\mathrm{ref}} + E^{\mathrm{PT2}}. \tag{3}$$

Finally, all the selected determinants are added to the reference space S for the next iteration.

For the generation of different lengths of the CIPSI component of the QMC wave function, one could either follow (a) an "expansion" scheme where we repeat the above process as many times as necessary, thereby systematically increasing the size of the wave function, or (b) a "truncation" scheme after generating a very large CIPSI wave function, typically to a point where the CIPSI energy (Eq. 3) is reasonably converged. The large wave function created for the second scheme can contain millions of determinants and one needs to reorder them in decreasing order of the absolute value of their CI coefficients and then truncate the expansion, keeping their coefficients the same as in the large CIPSI wave function. This strategy is said to be a better starting point for VMC optimization as the determinant coefficients are from a near FCI calculation[25]. We employ the expansion scheme while keeping the wave function an eigenstate of $\mathbf{S}^2$ while no such constraint is used when following the truncation scheme. Hence, the comparison of the two schemes is not exactly one to one. We use complete active space SCF (CASSCF) orbitals for the majority of these calculations, however, (c) the use of natural orbitals obtained from a large CIPSI calculation is another aspect of our investigation.

After generating the CIPSI expansion, we introduce a positive Jastrow correlation factor $\mathcal{J}$ and construct the Jastrow-CIPSI wave function as

$$\psi = \mathcal{J}\psi_{\mathrm{CIPSI}} = \mathcal{J} \sum_{i=0}^{N_{\mathrm{det}}} c_i D_i \,, \tag{4}$$

where $\mathcal{J}$ explicitly describes electron-electron and electron-nucleus (two-body) and electron-electron-nucleus (three-body) correlations[33] while imposing the electron-electron cusp conditions. $N_{\rm det}$ represents the number of determinants in the CIPSI wave function spanning the space *S*.

III. COMPUTATIONAL DETAILS

The QMC calculations are carried out with the program package CHAMP[34]. We employ scalar-relativistic energy-consistent Hartree-Fock pseudopotentials and the cc-pVTZ Gaussian basis set specifically constructed for our pseudopotentials[35,36]. In particular, we perform all calculations with the cc-pVTZ basis set and test the convergence of the results with the cc-pVQZ basis set. To generate the Slater component of the QMC wave functions, CIPSI calculations are performed in Quantum Package[37] using canonical orbitals obtained from a CAS(10,10) MCSCF calculation for the ground state of butadiene carried out with the program GAMESS(US)[38,39].

All parameters (Jastrow, orbital, and CI coefficients) are variationally optimized in VMC using the stochastic reconfiguration method[40] in a conjugate gradient implementation[41]. Most calculations are performed with a two-body Jastrow factor and the impact of the electron-electron-nucleus terms on the energy is tested for a few cases. Exploiting the low-numerical-scaling computation of energy and wave function derivatives[2], the ground-state geometry of butadiene is simultaneously optimized with the wave function following the path of steepest descent and an appropriate rescaling of the inter-atomic forces. To avoid spikes in forces, an improved estimator of the inter-atomic forces is used in all calculations, which is obtained by sampling the square of a modified wave function close to nodes[42]. In the DMC calculations, we treat the pseudopotentials beyond the locality approximation[43] and use an imaginary time step of 0.015 a.u. As shown in the Supplementary Information (SI), this time step yields DMC energies converged to better than 0.1 mHartree for a simple Jastrow-CIPSI wave function with only two determinants and is therefore appropriate for all wave functions of higher quality considered in this work.

We employ Gaussian09[44] to perform CCSD(T) geometry optimization in combination with our pseudopotentials and in all-electron calculations in the frozen core (FC) approximation with the cc-pVXZ (X = D, T, Q, and 5) and aug-cc-pVXZ (X = D, T, and Q) basis sets. For CCSD(T) geometry optimization without the FC approximation, we use the the PSI4 code[45] with the cc-pCVXZ, cc-pwCVXZ, and corresponding augmented (X = D, T, and Q) basis sets. The results of these optimizations and their extrapolations to the CBS limit are detailed in the SI.

IV. RESULTS AND DISCUSSION

We investigate the merits of the choice of a CIPSI expansion to describe the Slater component of a QMC wave function when optimizing the ground-state geometry of butadiene over the conventional CAS description. We start all structural optimizations with the Jastrow-CIPSI wave functions from the same initial MP2/cc-pVQZ optimized geometry and, as described above, variationally optimize the Jastrow parameters, CI coefficients, and orbitals simultaneously with the geometry in VMC. Post convergence of the VMC energy and stabilization of the bond lengths, we perform 40 additional iterations and average these geometries. A final VMC and DMC energy calculation is done on this average geometry with the wave function obtained in the last iteration. As detailed in the SI, reoptimizing the wave function on the average geometry leads to equivalent energies within the statistical error.

We primarily focus our discussion on the use of the "truncation" scheme for the CIPSI-determinant selection since, unlike the expansion scheme, no added selection criterion has been used. An initial CIPSI wave function is constructed including as many as 1.17 million determinants and then truncated to generate a set of determinantal expansions of increasing size. The results are summarized in Table I and illustrated in Figs. 1 and 2, where the final VMC and DMC energies, and corresponding structural parameters are plotted against the size of the CIPSI expansions. We also compare our results with previous QMC calculations[2] of the structural optimization of butadiene, which employed various CAS expansions: a CAS(4,4), CAS(4,16), and CAS(4,20) correlating 4 π electrons in the orbitals constructed from the $2p_z$, $3p_z$, $3d_{xz}$, $3d_{yz}$, and $4p_z$ atomic orbitals; a CAS(10,10) consisting of 6 σ and 4 π electrons in 10 bonding and antibonding orbitals; a truncated RAS(10,22) that includes single and double excitations in additional 12 π and δ orbitals over the CAS(10,10) space. Since the cc-pVTZ and cc-pVQZ basis sets yield VMC and DMC energies compatible to better than 2 and 0.2 mHartee, respectively, and structural parameters differing by less 0.4 mÅ (see SI), we can directly compare our cc-pVTZ calculations with the results obtained with the optimization of Jastrow-CAS wave functions in the cc-pVQZ basis.

As shown in Fig. 1, a CIPSI wave function with roughly 100 determinants yields a VMC energy which is only about 1 and 4 mHartree higher than the energies of the CAS(10,10) and RAS(10,22) expansions consisting of 15,912 and 45,644 Slater determinants, respectively. Unlike the Jastrow-CAS case, where the energies are scattered around a relatively flat value, CIPSI expansions of increasing size yield a monotonic decrease in the VMC energy, with our largest considered expansion of 32,768 determinants amounting to an energy which is about 40 mHartree lower than the best RAS(10,22) value. In fact, the VMC energy obtained with our largest CIPSI expansion is within 5 mHartree of the converged DMC energy obtained with the Jastrow-CAS wave functions. Therefore, a smart selection of determinants from the approximate FCI space helps us attain much lower VMC energies in comparison to CAS expansions over conventionally used active-space definitions, which instead lead to the inclusion of many determinants with little contribution to the energy.

The behavior of DMC parallels the VMC results with the energy of the Jastrow-CAS wave functions being lowered by about 8 mHartree when the size of the wave function is increased from 1 to 20 determinants, and stagnating afterwards

TABLE I. Optimal ground-state structural parameters (Å) of butadiene and corresponding VMC and DMC energies (a.u.) with increasing number of CIPSI determinants obtained in the truncation scheme. The total number of optimized parameters in the wave function is listed. The statistical error is given in brackets.

No. det	No. param	C-C	C=C	BLA	VMC	DMC
1	749	1.45595(28)	1.32415(11)	0.13180(30)	-26.24310(32)	-26.30426(28)
2	782	1.44596(20)	1.33025(14)	0.11571(29)	-26.24912(32)	-26.30681(14)
8	822	1.45244(28)	1.33245(15)	0.11999(50)	-26.25644(31)	-26.31044(26)
128	1594	1.45778(12)	1.33564(20)	0.12214(20)	-26.26562(30)	-26.31223(24)
1024	5514	1.45632(22)	1.33493(08)	0.12139(25)	-26.28829(26)	-26.31908(20)
2048	7726	1.45626(15)	1.33456(09)	0.12170(17)	-26.29386(25)	-26.32147(16)
5114	12147	1.45549(08)	1.33434(11)	0.12115(07)	-26.29980(24)	-26.32424(09)
15469	24818	1.45491(06)	1.33406(08)	0.12085(10)	-26.30880(22)	-26.32873(22)
32768	44265	1.45487(25)	1.33414(21)	0.12072(37)	-26.31194(12)	-26.32928(20)

as shown in Figure 1. The Jastrow-CIPSI wave function yields a DMC energy which is comparable with the Jastrow-CAS case when the CIPSI expansion has only about 100 determinants, while the largest CIPSI expansion gives a DMC energy 12 mHartree lower than the truncated RAS(10,22) case. We also note that the estimate of the FCI limit in the current basis set on the initial geometry is about -26.275 Hartree (see SI) and, therefore, as much as 37 and 54 mHartree higher than our best VMC and DMC energies. Our best DMC energies are also superior to the CCSD(T) optimized values obtained with a quintuple-ζ basis set (see SI).

Importantly, the use of CIPSI expansions (in combination with a Jastrow factor and the optimization of all wave function parameters) is not only beneficial in terms of the quality of the final total energies: the smooth and monotonic convergence of the VMC and DMC energies with the number of determinants demonstrates the effectiveness of such an approach in identifying energetically relevant determinants in a systematic manner. These important excitations are not easily accessible through manual selection as demonstrated by the energy plateau one reaches in constructing very large expansions based on an *a priori* choice of an apparently reasonable set of active orbitals. Besides being completely automated, this feature of the CIPSI scheme is also crucial for obtaining a smooth convergence of the structural properties (to better than 1 mÅ) with the number of determinants as discussed next.

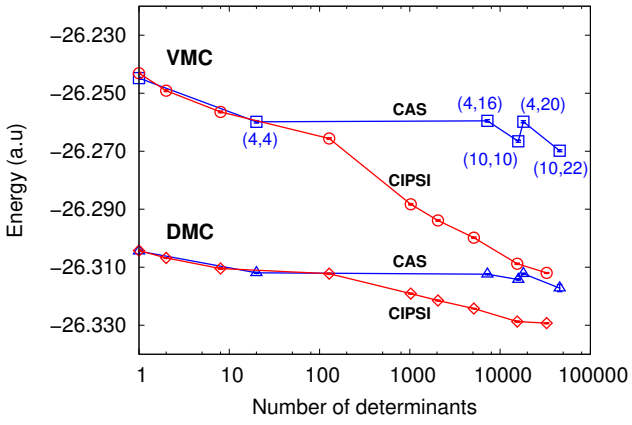

FIG. 1. Convergence of the VMC and DMC energies on the optimal geometry with the number of determinants in the CIPSI expansion (red). Previous results[2] obtained with Jastrow-CAS wave functions are also presented (blue).

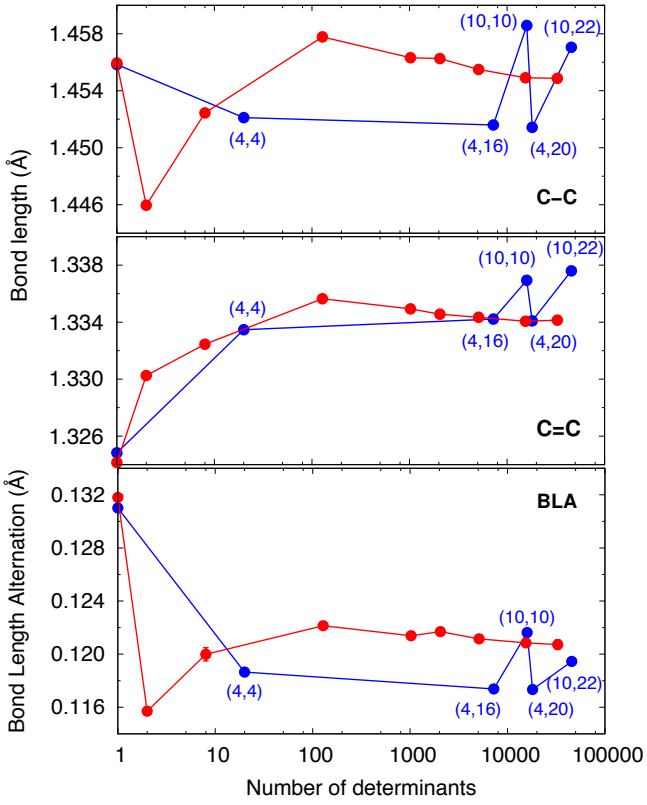

FIG. 2. Convergence of the single bond (C-C), double bond (C=C), and bond length alternation (BLA) (Å) with the number of determinants in the CIPSI wave function (red). Previous results[2] obtained with a Jastrow-CAS wave function are presented in blue.

In Figure 2, we illustrate the variation of the carbon-carbon single and double bonds and of the bond length alternation

(BLA) obtained with the Jastrow-CIPSI expansions. The convergence of the bond lengths and BLA is smooth with the increase in the size of the wave function. While the double bond is already converged to better than 0.5 mÅ with little over 2000 determinants, the single bond and, consequently, the BLA show a somewhat slower convergence and reach the same level of accuracy with about 5114 determinants. We stress however that we are here looking at extremely small differences while establishing the degree of convergence. For all practical purposes, the BLA is already converged to better than 1 mÅ within the limits of statistical error with just over 1000 determinants, again proving the ability of CIPSI expansions to obtain optimal geometry descriptions with a very small set of determinants. Besides overcoming the non-uniform convergence of the structural parameters when expanding the wave function on different active spaces, these calculations also confirm the need for a rather subtle multireference description for such conjugated systems[46,47]. The converged single bond length is in excellent agreement with the CCSD(T)/CBS value of 1.455 Å obtained with the same pseudopotentials used in the QMC calculations (see SI). Our double bond is instead less than 0.003 Å smaller than the corresponding CCSD(T)/CBS value, consequently resulting in a difference of about 0.002 Å on the BLA.

To investigate the impact of the inclusion of three-body terms in the Jastrow factor, we add them to the CIPSI wave functions with 2048 and 5114 determinants and reoptimize all wave function parameters on the fixed average geometries we have determined with the two-body Jastrow factor. These optimizations result in the VMC and DMC energies presented in Table II. While there is an expected gain in the VMC energies, the DMC energies are equivalent within the statistical error of 0.2 mHartree to the energies obtained with a two-body Jastrow factor. Consequently, given the quality of our determinantal component, a two-body Jastrow recovers most of the missing dynamical correlation contribution and is sufficient for our purposes.

TABLE II. Effect of the inclusion of three-body Jastrow terms on the total VMC and DMC energies (a.u.). The structures optimized with the two-body Jastrow factor are used. ΔE denotes the gain in energy with respect to the values obtained with a two-body Jastrow factor.

No. det	E_{VMC}	E_{DMC}	ΔE_{VMC}	ΔE_{DMC}
2048	-26.29908(15)	-26.32162(20)	-0.00522(29)	-0.00015(26)
5114	-26.30333(23)	-26.32412(16)	-0.00353(33)	0.00012(18)

All results presented so far have been obtained by selecting the CIPSI determinants out of a much larger CIPSI wave function according to the "truncation" scheme. The VMC energies obtained with the "expansion" scheme where one constructs CIPSI expansions of increasing size are compared with the "truncation" scheme in Figure 3. Both schemes lead to very comparable convergence in the energy as a function of the number of determinants in the Jastrow-CIPSI wave function and to compatible converged structural parameters. However, as shown in the SI, the variation of the bond lengths and BLA is not as smooth as in the truncation case. This difference probably arises from the added requirement of having an eigenstate of $\mathbf{S}^2$ in the expansion scheme. In fact, if we follow the expansion scheme without this requirement and construct a wave function of 5114 determinants, we obtain an optimal VMC single and double bonds of 1.45564(26) and 1.33496(11) Å respectively, and, a BLA of 0.12068(26) Å. The VMC energy converges to -26.29960(24) a.u. These results are in good agreement with the corresponding values obtained with the truncation scheme.

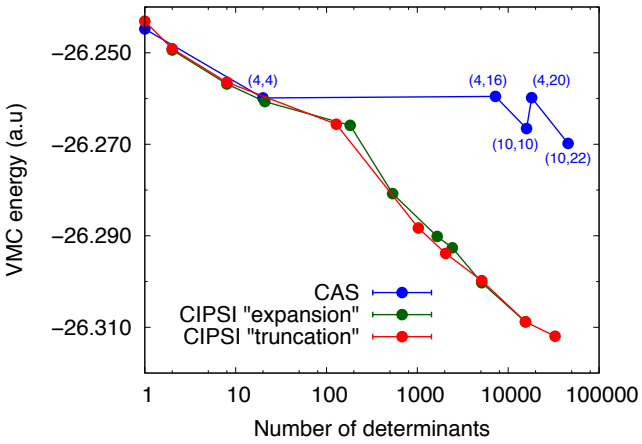

FIG. 3. Optimal ground-state VMC energies obtained with Jastrow-CAS and Jastrow-CIPSI wave functions constructed with the "expansion" and "truncation" schemes.

Finally, we explore a potentially better orbital description in terms of natural orbitals obtained from a CIPSI calculation instead of the canonical orbitals resulting from a CAS(10,10) self-consistent calculation as done above. To this aim, we determine the natural orbitals for a very large CIPSI wave function of 2 million determinants and construct three expansions of roughly 2000, 5000, and 15,000 determinants with the truncation scheme. Surprisingly, we consistently obtain higher VMC energies compared to the expansions generated with canonical orbitals. The reason behind this observation is unclear. The bond lengths and BLA obtained with 5000 determinants are in excellent agreement with the corresponding values for the truncated canonical set of comparable size but the other test cases do not provide as compatible values. The results of these calculations are given in the SI.

V. CONCLUSION

We demonstrated the excellent performance of compact perturbatively-selected CIPSI determinantal expansions in obtaining significantly lower VMC and DMC energies as compared to conventional active space definitions for the challenging case of butadiene. With the use of these wave functions, we were able to obtain converged ground-state structural parameters with the use of only a few thousand determinants. We tested two different schemes for the selection of the CIPSI expansions either by constructing a large CIPSI wave function

and then truncating it ("truncation" scheme), or by considering successive sets of determinants in the CIPSI construction ("expansion" scheme). We found that the two representations are rather equivalent in terms of energy but that the truncation scheme possibly leads to a somewhat smoother convergence of the structural parameters with the size of the CIPSI expansion. CCSD(T) calculations with the same pseudopotentials yield a CBS estimate of the carbon-carbon single bond in very good agreement with our converged value but a double bond and corresponding BLA smaller by about 0.002 Å. We do not expect any significant change in the bond lengths upon inclusion of additional determinants, in view of their weak variations observed over a wide range of the number of determinants already considered.

Our study therefore shows that the automated selection of determinants from a CIPSI wave function is an extremely suitable and less cumbersome alternative for the fast optimization of ground-state geometries in QMC than a choice based on correlating electrons in active spaces constructed from energetically low-lying orbitals. The latter results in large expansions with many determinants which contribute little to the energy and to the convergence of the structural parameters. The use of CIPSI-based wave functions in combination with the low-scaling algorithms for simultaneous wave function and geometry optimization opens the way to the accurate and efficient QMC optimization of large molecular systems.

SUPPLEMENTARY INFORMATION

See Supplementary Information for CIPSI energies and FCI extrapolation; VMC and DMC energies of bare CIPSI wave functions; impact of reoptimizing the wave function on the average geometry; basis-set convergence of geometry and corresponding VMC and DMC energies; bond-length and BLA convergence obtained with the "expansion" scheme; energy and geometry convergence with the use of natural orbitals; DMC time-step extrapolation and CCSD(T) geometry optimizations with pseudopotential and all-electron basis sets.

ACKNOWLEDGEMENTS

This work is part of the Industrial Partnership Programme (IPP) "Computational sciences for energy research" of the Netherlands Organisation for Scientic Research (NWO-I, formerly FOM). This research programme is co-financed by Shell Global Solutions International B.V. This work was carried out on the Dutch national supercomputer Cartesius with the support of SURF Cooperative, and using HPC resources from CALMIP (Toulouse) under allocation 2017-0510. The authors declare no competing financial interest.

[1] C. Filippi, R. Assaraf, and S. Moroni, J. Chem. Phys. **144**, 194105 (2016).
[2] R. Assaraf, S. Moroni, and C. Filippi, J. Chem. Theory Comput. **13**, 5273 (2017).
[3] B. Huron, J. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).
[4] J. Gouyet and M. Prat, J. Chem. Phys. **64**, 946 (1976).
[5] A. Povill, J. Rubio, and F. Illas, Theor. Chim. Acta **82**, 229 (1992).
[6] G. Trinquier and J. P. Malrieu, J. Am. Chem. Soc. **103**, 6313 (1981).
[7] M. Castex, M. Morlais, F. Spiegelmann, and J. Malrieu, J. Chem. Phys. **75**, 5006 (1981).
[8] M. Pelissier, J. Chem. Phys. **75**, 775 (1981).
[9] I. Nebot-Gil and J.-P. Malrieu, Chem. Phys. Lett. **84**, 571 (1981).
[10] R. Cimiraglia and M. Persico, J. Comput. Chem. **8**, 39 (1987).
[11] S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, Chem. Phys. **75**, 91 (1983).
[12] R. Cimiraglia, J. Chem. Phys. **83**, 1746 (1985).
[13] R. Cimiraglia, M. Persico, and J. Tomasi, J. Am. Chem. Soc. **107**, 1617 (1985).
[14] F. Illas, J. Rubio, J. Ricart, and P. Bagus, J. Chem. Phys. **95**, 1877 (1991).
[15] R. J. Harrison, J. Chem. Phys. **94**, 5021 (1991).
[16] F. Illas, J. Rubio, and J. Ricart, J. Chem. Phys. **89**, 6376 (1988).
[17] Y. Ohtsuka and J. Hasegawa, J. Chem. Phys. **147**, 034102 (2017).
[18] J. B. Schriber and F. A. Evangelista, J. Chem. Theory Comput. **13**, 5354 (2017).
[19] J. B. Schriber and F. A. Evangelista, J. Chem. Phys. **144**, 161106 (2016).
[20] A. A. Holmes, N. M. Tubman, and C. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).
[21] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).
[22] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[23] A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, J. Chem. Phys. **141**, 244110 (2014).
[24] E. Giner, A. Scemama, and M. Caffarel, J. Chem. Phys. **142**, 044115 (2015).
[25] E. Giner, R. Assaraf, and J. Toulouse, Mol. Phys. **114**, 910 (2016).
[26] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, J. Chem. Phys. **144**, 151103 (2016).
[27] A. Scemama, Y. Garniron, M. Caffarel, and P.-F. Loos, J. Chem. Theory Comput. **14**, 1395 (2018).
[28] M. Caffarel, E. Giner, A. Scemama, and A. Ramírez-Solís, J. Chem. Theory Comput. **10**, 5286 (2014).
[29] C. Angeli and M. Persico, Theor. Chem. Acc. **98**, 117 (1997).
[30] M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Using cipsi nodes in diffusion monte carlo," in *Recent Progress in Quantum Monte Carlo* (ACS Publications, 2016) pp. 15–46.
[31] P. S. Epstein, Phys. Rev. **28**, 695 (1926).
[32] R. Nesbet, in *Proc. R. Soc. Lond. A*, Vol. 230 (The Royal Society, 1955) pp. 312–321.
[33] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996), as Jastrow correlation factor, we use the exponential of the sum of three fifth-order polynomials of the electron-nuclear (e-n), the electron-electron (e-e), and of pure 3-body mixed e-e and e-n distances, respectively. The e-n and e-e-n terms have different coefficients for different atom types. The Jastrow factor is further adapted to deal with pseudo-atoms, and the scaling factor κ is set to 0.60 a.u. The 3-body e-e-n terms are not included in the 2-body Jastrow factor.
[34] C. Umrigar, C. Filippi, and S. Moroni, "Champ: Cornell-holland ab-initio materials package," (2018).
[35] M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **126**, 234105 (2007).
[36] For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. Dolg, M.; Filippi, C., private communication.
[37] A. Scemama, T. Applencourt, Y. Garniron, G. Giner, G. David, and M. Caffarel, "Quantum package v1.0." `https://github.com/LCPQ/quantum_package` (2016).
[38] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, and others, J. Comput. Chem. **14**, 1347 (1993).
[39] M. S. Gordon and M. W. Schmidt, in *Theory and applications of computational chemistry* (Elsevier, 2005) pp. 1167–1189.
[40] E. Neuscamman, C. Umrigar, and G. K.-L. Chan, Phys. Rev. B **85**, 045103 (2012).
[41] S. Sorella, M. Casula, and D. Rocca, J. Chem. Phys. **127**, 014105 (2007).
[42] C. Attaccalite and S. Sorella, Phys. Rev. Lett. **100**, 114501 (2008).
[43] M. Casula, Phys. Rev. B **74**, 161102 (2006).
[44] "Gaussian09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F.

Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT," (2016).

[45]R. M. Parrish, L. A. Burns, D. G. Smith, A. C. Simmonett, A. E. DePrince III, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, and others, J. Chem. Theory Comput. **13**, 3185 (2017).

[46]L. Tenti, E. Giner, J.-P. Malrieu, and C. Angeli, Comput. Theor. Chem. **1116**, 102 (2017).

[47]J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).