

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

Monika Dash,[†] Saverio Moroni,^{*,‡} Anthony Scemama,^{*,¶} and Claudia
Filippi^{*,†}

[†]*MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede,
The Netherlands*

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

[¶]*Laboratoire de Chimie et Physique Quantiques, Université de Toulouse, CNRS, UPS, France*

E-mail: moroni@democritos.it; scemama@irsamc.ups-tlse.fr; c.filippi@utwente.nl

1 Full CI extrapolation of CIPSI energies

We extrapolate the variational CI energy of the CIPSI expansions to the full CI limit, i.e to $E^{\text{PT}2} = 0$ obtained with both canonical MOs and natural MOs. In the fitting procedure, the values of $E^{\text{PT}2}$ are quite large, so a quadratic fit is used. We consider points where $E^{\text{PT}2} < -0.4$ a.u., and impose that the two extrapolated curves (natural and canonical) cross at $E^{\text{PT}2} = 0$ a.u. The extrapolated FCI energy is -26.2753 a.u., in good agreement with the CCSD(T)/cc-pVTZ value of -27.2773 a.u. (see Table 5 below).

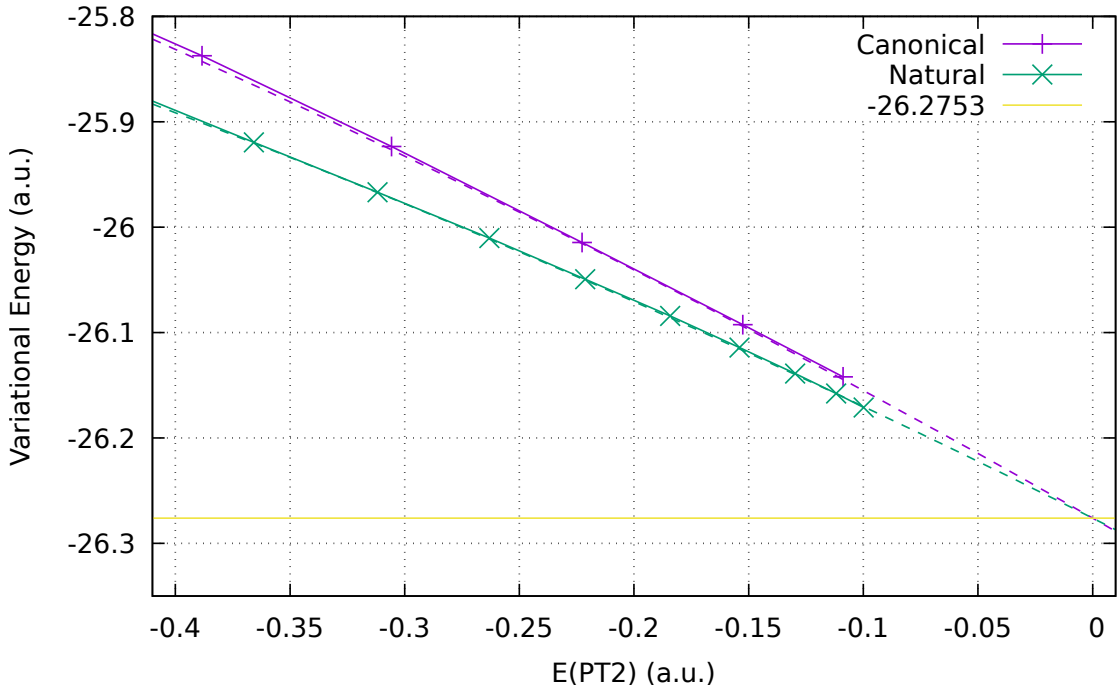


Figure 1: Full CI extrapolation of the variational CI (E^{ref}) energy with respect to perturbation correction, $E^{\text{PT}2}$, obtained with canonical and natural orbitals. We use the BFD pseudopotentials and the corresponding cc-pVTZ basis set.

2 Averaging of converged geometries and use of the wave function from the last iteration

We validate our approach of averaging the last 40 geometries to determine the optimal structure, and using the wave function obtained at the last iteration to compute the final VMC and DMC energies on the average geometry here. To this aim, we select the Jastrow-CIPSI wave function with 5114 determinants obtained with the “truncation” scheme and symmetrize our average geometry imposing C_{2h} symmetry. Starting with the wave function at the last iteration, we optimize all wave function parameters in the Jastrow-CIPSI wave function on the average geometry. The optimization yields a wave function characterized by a VMC energy of $-26.29957(24)$ Hartree, which is equivalent within statistical error to the energy of $-26.29980(24)$ Hartree obtained using the wave function at the last iteration on top of the average geometry (also reported in Table I in the main text).

3 Basis-set convergence

To test the dependence of the structural parameters on the choice of basis set, we perform a calculation following the “truncation” scheme where we employ the cc-pVQZ basis set and optimize a Jastrow-CIPSI expansion with 5000 determinants obtained from a large expansion of 4.9 million determinants.

Table 1: Basis-set dependence of the optimal structural parameters (\AA) and VMC and DMC energies (a.u.) computed with the cc-pVTZ and cc-pVQZ basis sets and CIPSI expansions of comparable sizes obtained with the truncation scheme.

Basis	No. det	No. param	C-C	C=C	BLA	E_{VMC}	E_{DMC}
cc-pVTZ	5114	12147	1.45549(08)	1.33434(11)	0.12115(07)	-26.29980(24)	-26.32424(09)
cc-pVQZ	5000	21582	1.45576(15)	1.33475(06)	0.12102(14)	-26.30154(22)	-26.32408(21)

The use of the larger cc-pVQZ basis set does not yield any significant difference in the optimal bond lengths and energies with respect to the values obtained with the cc-pVTZ basis set. There is an expected lowering of the VMC energy but only by 2 mHartree, and the DMC energy does not change. The bond lengths and BLA computed with the two basis sets are compatible to better than 0.5 m \AA . This good agreement justifies our comparison of the CIPSI optimization results obtained with the cc-pVTZ basis and the CAS calculations in the cc-pVQZ basis.

4 Calculations following the “expansion” scheme for the CIPSI determinant selection

In Figure 2 and Table 2, we present the optimal structural parameters and corresponding VMC energies obtained with Jastrow-CIPSI wave functions where the CIPSI component is obtained following the “expansion” procedure. We perform these calculations with the additional constraint that the wave function is an eigenstate of \mathbf{S}^2 .

The VMC energies obtained with comparable number of determinants in the “expansion” and “truncation” schemes are rather compatible but the convergence of the structural parameters is

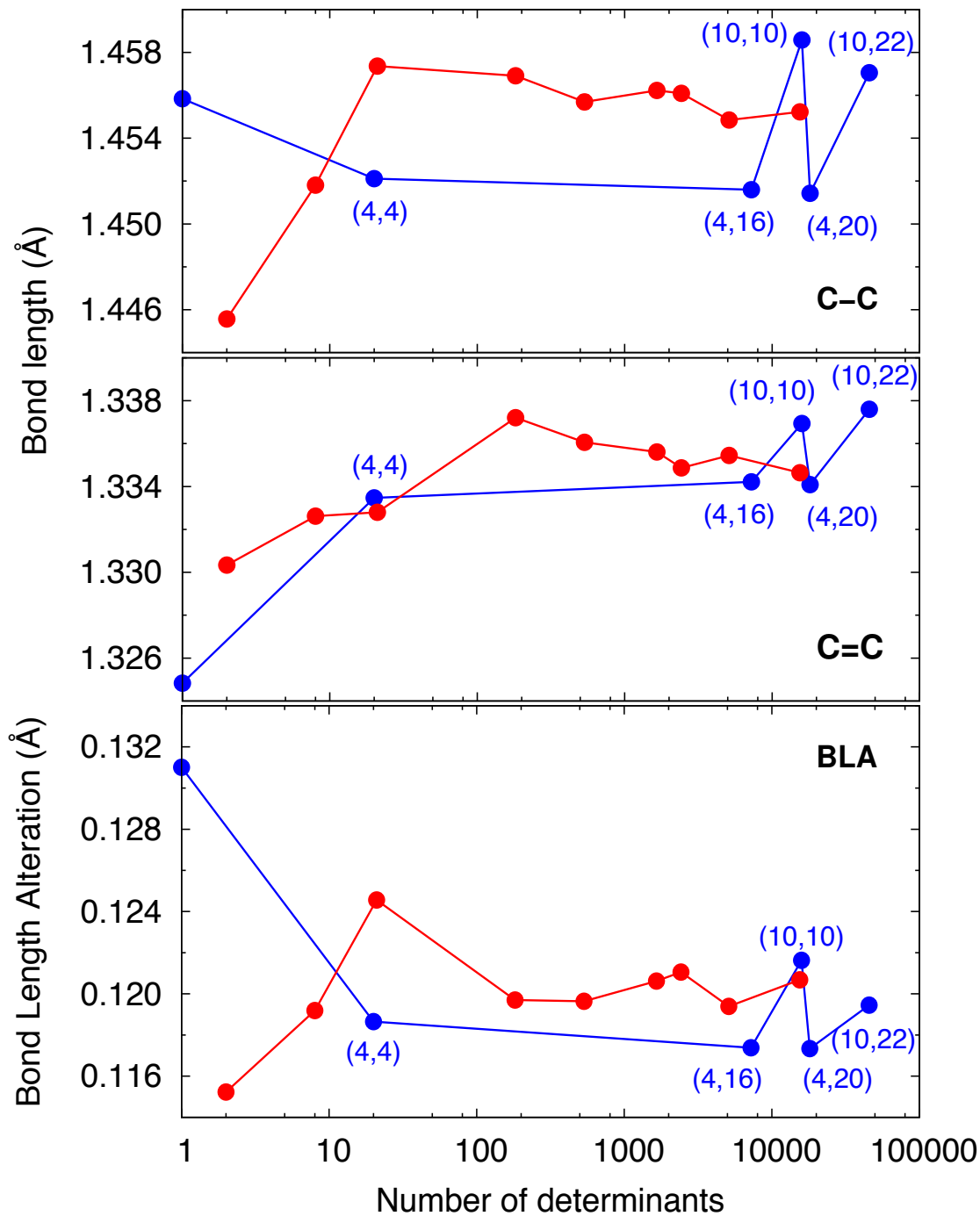


Figure 2: Convergence of the structural parameters (Å) of butadiene with the number of determinants obtained with Jastrow-CIPSI wave function and the “expansion” scheme (red). Previous results¹ obtained with Jastrow-CAS wave functions are presented in blue.

Table 2: Optimal ground-state structural parameters (Å) of butadiene and corresponding VMC energies (a.u.) with increasing number of determinants in the Jastrow-CIPSI wave function obtained with the “expansion” scheme. The cc-pVTZ basis set is used. The number of parameters optimized in the wave function is listed.

No. det	No. param	C-C	C=C	BLA	E_{VMC}
2	782	1.44557(18)	1.33034(08)	0.11523(20)	-26.24940(37)
8	822	1.45181(17)	1.33262(06)	0.11919(18)	-26.25682(31)
21	1067	1.45736(09)	1.33280(07)	0.12456(15)	-26.26068(30)
182	1629	1.45691(20)	1.33721(07)	0.11970(25)	-26.26586(30)
534	3761	1.45569(32)	1.33606(06)	0.11964(30)	-26.28079(27)
1654	6601	1.45623(24)	1.33561(05)	0.12062(27)	-26.29014(26)
2424	8003	1.45609(28)	1.33487(05)	0.12106(15)	-26.29264(25)
5114	12403	1.45485(12)	1.33545(06)	0.11939(11)	-26.30026(24)
15469	25652	1.45523(26)	1.33464(06)	0.12068(27)	-26.30881(22)

somewhat less smooth in the “expansion” scheme. In particular, the calculation with 5114 determinants following the “expansion” strategy breaks the monotonicity of the geometry convergence with the single bond being smaller than that obtained with 2424 and 15,469 determinants, and the double bond longer than the values obtained in the other two cases. The agreement with the “truncation” scheme improves upon removal of the constraint of having an eigenstate of \mathbf{S}^2 (see main text).

5 Use of natural orbitals

A better choice of starting orbitals for the selection scheme is expected to be a set of natural orbitals generated starting from a large CIPSI expansion. Therefore, starting from the CAS(10,10) canonical orbitals, we generate a very large CIPSI wave function comprising 2 million determinants and use it to construct the natural orbitals. Thereafter, we employ the truncation scheme on top of the natural orbitals to generate the expansions listed in Table 3. Since the package to generate natural orbitals from CIPSI wave functions lacks the symmetry information of the resulting orbitals, the number of variational parameters in the subsequent QMC optimization is much higher than in the canonical case.

Comparing to the canonical case, we obtain a shorter carbon-carbon single bond length but a

Table 3: Optimal ground-state structural parameters of butadiene (\AA) and corresponding VMC energies (a.u.) with increasing number of determinants for CIPSI expansions over natural orbitals. The cc-pVTZ basis set is used.

No. det	No. param	C-C	C=C	BLA	E_{VMC}
2048	11851	1.45453(37)	1.33444(11)	0.12009(31)	-26.28733(26)
5000	18320	1.45536(14)	1.33444(09)	0.12091(20)	-26.29437(25)
15000	39325	1.45562(10)	1.33417(11)	0.12145(12)	-26.30379(23)

compatible double bond when using 2048 determinants. As a result, the predicted BLA is about 1.7 m\AA shorter than in the canonical case. Increasing the size of the expansion to 5000 brings all structural parameters in close correspondence with the canonical values. However, further increasing the expansion size to 15,000 determinants does not decrease the single bond as in the canonical case, thereby, resulting in a somewhat larger BLA. Rather surprisingly, the VMC energies obtained with natural orbitals are always about 5 mHartrees higher than those with canonical orbitals.

6 Time-step extrapolation of the DMC energies

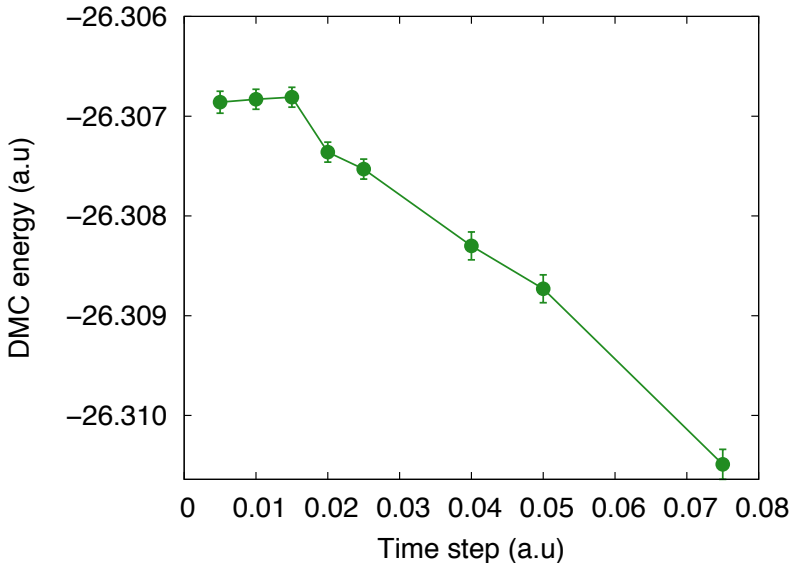


Figure 3: Convergence of the DMC energies with the time step. An optimal Jastrow-Slater wave function consisting of 2 determinants is used in combination with the cc-pVTZ basis set.

We perform a time-step extrapolation study of the DMC energies using a 2-determinant Jastrow-

Table 4: Dependence of the DMC energies (a.u.) on the time step (a.u.). A Jastrow-Slater wave function with 2 determinants is used.

Time step	E_{DMC}
0.005	-26.30686(11)
0.010	-26.30683(10)
0.015	-26.30681(10)
0.020	-26.30736(10)
0.025	-26.30753(10)
0.040	-26.30830(14)
0.050	-26.30873(14)
0.075	-26.31049(15)

Slater wave function with a two-body Jastrow factor. As illustrated in Figure 3 and Table 4, the DMC energy is converged to within 0.1 mHartree at an imaginary time step $\tau = 0.015$ a.u.. Therefore, we perform all our DMC calculations with $\tau = 0.015$ a.u.

7 CCSD(T) calculations and CBS extrapolation

As shown in Table 5, we perform CCSD(T) geometry optimization using our pseudopotentials and the corresponding cc-pVXZ ($X = \text{D, T, Q, and 5}$) and aug-cc-pVXZ ($X = \text{D, T, and Q}$) basis sets. We also optimize the geometry in all-electron calculations in the frozen core (FC) approximation with the all-electron cc-pVXZ and aug-cc-pVXZ basis sets. Finally, full all-electron calculations (no FC) are also carried out with the basis sets, cc-pCVXZ and aug-cc-pCVXZ, and cc-pwCVXZ and aug-cc-pwCVXZ.

For the pseudopotential and FC cases where we have a reasonable number of data points, we extrapolate the bond lengths to the complete basis set (CBS) limit and use a five parameter fit as a combination of two three-parameter exponential fits² for the cc- and aug-cc- basis sets according to the following equation

$$f(X) = (a + b e^{-cX})\delta_{\text{cc-}} + (a + d e^{-eX})\delta_{\text{aug-cc-}}, \quad (1)$$

where X is the cardinality of the basis set. We employ Mathematica³ to perform the fit.

In the all-electron calculations, the use of augmentation appears to slow down the convergence of the bond lengths and lead to a seemingly smaller BLA. The same observation also holds for the pseudopotential case and the carbon-carbon bond length. The use of pseudopotentials leads to a somewhat larger BLA, bringing the CCSD(T) result in even better agreement with our QMC calculations.

We also report previous calculations,² where the CBS limit with a three-parameter exponential fit is computed on the CCSD(T) bond lengths obtained in the FC approximation with the aug-cc-pVXZ basis sets with X = D, T, and Q. The core-valence correlation error is then extrapolated (again in a three-parameter exponential fit) on the differences with the values computed without the FC approximation and with the cc-pwCVXZ basis set (with X = D, T, and Q). The addition of a CCSDT(Q)(FC)/cc-pVDZ correction leads to a lengthening of the double bond and a consequent shortening of the BLA.

Table 5: Optimal CCSD(T) ground-state structural parameters of butadiene and corresponding energies.

Basis	C-C	C=C	BLA	Energy
Burkatzki-Filippi-Dolg pseudopotentials				
cc-pVDZ	1.4743	1.3605	0.1138	-26.11741
cc-pVTZ	1.4575	1.3399	0.1176	-26.27737
cc-pVQZ	1.4561	1.3375	0.1186	-26.30670
cc-pV5Z	1.4546	1.3367	0.1179	-26.32791
aug-cc-pVDZ	1.4679	1.3579	0.1100	-26.14781
aug-cc-pVTZ	1.4570	1.3400	0.1170	-26.28746
aug-cc-pVQZ	1.4569	1.3381	0.1188	-26.31379
CBS limit	1.4556	1.3371	0.1185	–
All-electron calculations with FC				
cc-pVDZ	1.4725	1.3585	0.1140	-155.53877
cc-pVTZ	1.4612	1.3439	0.1173	-155.69593
cc-pVQZ	1.4587	1.3412	0.1175	-155.74073
cc-pV5Z	1.4580	1.3407	0.1173	-155.75411
aug-cc-pVDZ	1.4733	1.3609	0.1124	-155.56506
aug-cc-pVTZ	1.4608	1.3447	0.1161	-155.70597
aug-cc-pVQZ	1.4591	1.3417	0.1174	-155.74455
CBS limit	1.4581	1.3407	0.1174	–
All-electron calculations with no FC				
cc-pwCVDZ	1.4688	1.3541	0.1147	-155.67785
cc-pwCVTZ	1.4570	1.3396	0.1174	-155.89914
cc-pwCVQZ	1.4552	1.3379	0.1174	-155.96010
aug-cc-pwCVDZ	1.4704	1.3577	0.1127	-155.69915
aug-cc-pwCVTZ	1.4571	1.3407	0.1164	-155.90786
cc-pCVDZ	1.4689	1.3546	0.1143	-155.70116
cc-pCVTZ	1.4581	1.3407	0.1174	-155.89970
cc-pCVQZ	1.4554	1.3381	0.1173	-155.95485
aug-cc-pCVDZ	1.4708	1.3581	0.1126	-155.72248
aug-cc-pCVTZ	1.4580	1.3418	0.1162	-155.90907
Other calculations				
CCSD(T)/CBS ^a	1.4548	1.3377	0.1171	
CCSD(T)/CBS-corr ^b	1.4549	1.3389	0.1160	

^a Ref. 2; ^b Ref. 2, including a CCSDT(Q)(FC)/cc-pVDZ correction.

References

- (1) Assaraf, R.; Moroni, S.; Filippi, C. Optimizing the Energy with Quantum Monte Carlo: A Lower Numerical Scaling for Jastrow–Slater Expansions. *J. Chem. Theory Comput.* **2017**, *13*, 5273–5281.
- (2) 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.
- (3) Mathematica, Version 11.1.1, Wolfram Research, Inc., Champaign, Illinois. 2017.