



**HAL**  
open science

# Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F<sub>2</sub> using selected configuration interaction trial wavefunctions

Emmanuel Giner, Anthony Scemama, Michel Caffarel

## ► To cite this version:

Emmanuel Giner, Anthony Scemama, Michel Caffarel. Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F<sub>2</sub> using selected configuration interaction trial wavefunctions. *Journal of Chemical Physics*, 2015, 142 (4), pp.044115. 10.1063/1.4905528 . hal-01136398

**HAL Id: hal-01136398**

**<https://hal.science/hal-01136398>**

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.

# Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule $F_2$ using selected configuration interaction trial wavefunctions

Cite as: J. Chem. Phys. **142**, 044115 (2015); <https://doi.org/10.1063/1.4905528>

Submitted: 12 August 2014 . Accepted: 23 December 2014 . Published Online: 29 January 2015

Emmanuel Giner, Anthony Scemama, and Michel Caffarel 



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions](#)

The Journal of Chemical Physics **58**, 5745 (1973); <https://doi.org/10.1063/1.1679199>

[A deterministic alternative to the full configuration interaction quantum Monte Carlo method](#)

The Journal of Chemical Physics **145**, 044112 (2016); <https://doi.org/10.1063/1.4955109>

[Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space](#)

The Journal of Chemical Physics **131**, 054106 (2009); <https://doi.org/10.1063/1.3193710>



Lock-in Amplifiers

Zurich Instruments

Watch the Video

# Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule $F_2$ using selected configuration interaction trial wavefunctions

Emmanuel Giner, Anthony Scemama, and Michel Caffarel

*Laboratoire de Chimie et Physique Quantiques, CNRS-Université de Toulouse, Toulouse, France*

(Received 12 August 2014; accepted 23 December 2014; published online 29 January 2015)

The potential energy curve of the  $F_2$  molecule is calculated with Fixed-Node Diffusion Monte Carlo (FN-DMC) using Configuration Interaction (CI)-type trial wavefunctions. To keep the number of determinants reasonable and thus make FN-DMC calculations feasible in practice, the CI expansion is restricted to those determinants that contribute the most to the total energy. The selection of the determinants is made using the CIPSI approach (Configuration Interaction using a Perturbative Selection made Iteratively). The trial wavefunction used in FN-DMC is directly issued from the deterministic CI program; no Jastrow factor is used and no preliminary multi-parameter stochastic optimization of the trial wavefunction is performed. The nodes of CIPSI wavefunctions are found to reduce significantly the fixed-node error and to be systematically improved upon increasing the number of selected determinants. To reduce the non-parallelism error of the potential energy curve, a scheme based on the use of a  $R$ -dependent number of determinants is introduced. Using Dunning's cc-pVDZ basis set, the FN-DMC energy curve of  $F_2$  is found to be of a quality similar to that obtained with full configuration interaction/cc-pVQZ. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4905528>]

## I. INTRODUCTION

Fixed-Node Diffusion Monte Carlo (FN-DMC) and its diverse variants<sup>1,2</sup> are known to be accurate approaches for evaluating ground-state energies of molecules. The only fundamental error in FN-DMC is the fixed-node error resulting from the use of a trial wavefunction,  $\psi_T$ , with approximate nodes (or zeros). Other sources of error are also present, but they can be easily kept under control (mainly, the statistical, finite time-step, and population control errors). A great variety of FN-DMC applications have shown that the fixed-node error on total energies can be made small with respect to the exact total correlation energy, typically less than 10%. The common strategy to decrease the fixed-node error is to use trial wavefunctions of the best possible quality and to resort to multi-parameter stochastic optimization techniques to optimize their parameters (minimization of the total energy and/or its variance in a preliminary variational Monte Carlo run). Taking advantage of the great flexibility in the trial wavefunctions that can be used in quantum Monte Carlo (QMC), a great variety of functional forms have been considered (e.g., Refs. 3–13), the most popular form being certainly the multideterminant Slater-Jastrow trial wavefunction<sup>3,14</sup> combining a short determinantal expansion and a Jastrow prefactor describing explicitly electron-electron and electron-electron-nucleus correlations. Regarding the optimization step, several optimization techniques designed to be efficient in a Monte Carlo context have been developed, e.g., Ref. 15.

Although the fixed-node error is small compared to total energies, unfortunately, in most cases, it is (much) larger than the (small) energy differences to compute (e.g., ionization potentials, electron affinities, atomization energies, electronic excitations, and reaction barriers). A key issue in QMC is,

therefore, to control the accuracy on such differences, an accuracy directly related to the quality of nodes and to the level of error compensation occurring when computing the difference of total energies. In this work, we address this problem in the important case of the FN-DMC calculation of potential energy surfaces (PES). Instead of using highly correlated trial wavefunctions, we propose instead to employ configuration-interaction-type (CI) expansions, as recently proposed.<sup>16</sup> Using such trial wavefunctions, it has been shown in a series of applications to atoms (oxygen atom,<sup>16</sup> atoms from B to Ar,<sup>17</sup> and 3d transition metal series<sup>18</sup>) that (very) accurate nodes can be obtained.

As with any approximate electronic structure method, the main issue when computing PES is to get smooth and uniformly accurate potential surfaces (small non-parallelism error). Here, we shall illustrate with the example of the  $F_2$  molecule that using CI expansions of moderate sizes may represent an appealing alternative to the use of explicitly correlated trial wavefunctions. In short, the strategy proposed is as follows.

- (i) Use as trial wavefunction a multi-determinant CI expansion. To avoid the prohibitive size of standard CI expansions and thus make DMC calculations feasible in practice,<sup>19</sup> we propose to use selected CI expansions where only the determinants contributing the most to the energy are included (Configuration Interaction using a Perturbative Selection done Iteratively (CIPSI)-like algorithm, Refs. 20 and 21).
- (ii) Use the multi-determinant expansion directly *as it comes* from the output of the CI program. In particular, we propose to avoid any stochastic optimization step as it is usually done, for example, by introducing a Jastrow factor and re-optimizing both the determinantal and Jastrow part.

The main advantages of using selected CI-type expansions may be summarized as follows.

- (i) The construction of the trial wavefunction can be made fully automatic.
- (ii) When the number of determinants is large enough, the selected CI expansion approaches the Full Configuration Interaction (FCI) limit and the near-FCI nodes are expected to be of a good quality.
- (iii) When approaching the near-FCI limit, the dependence of the nodes on the type of molecular orbitals used (Hartree-Fock (HF), Kohn-Sham, natural orbitals, etc.) is weak (no dependence at all in the FCI limit) and the critical parameters reduce to the coefficients of the determinants. Thanks to linearity, the set of optimal coefficients is unique and easily obtained from the (deterministic) diagonalization of the CI Hamiltonian matrix.
- (iv) The variation of the CI nodes across the PES is smooth by construction. At the variational level, the non-parallelism error is of a near-FCI quality and a similar quality (or better) may be expected at the FN-DMC level. Furthermore, by using CI-type expansions, simple schemes based on the use of a variable number of determinants across the PES to minimize the non-parallelism error can be defined, see, for example, our proposal in Sec. IV C.

At this point, a few remarks are in order. First, in contrast with most FN-DMC studies, no Jastrow factor is employed here. Of course, despite the fact that we propose not to re-optimize the CI nodes in presence of such a term, it could nevertheless be used to decrease the statistical error. However, to get a significant improvement, a stochastic optimization of the Jastrow parameters is required, a step that we prefer to avoid to keep the FN-DMC scheme fully deterministic and automatic. Second, most of the advantages cited above can be achieved using standard correlated trial wavefunctions (e.g., Ref. 15). However, the necessity of optimizing in a Monte Carlo context a large number of linear/nonlinear parameters makes the problem of getting smooth and uniformly accurate PES not so simple, particularly for large systems. Finally, a last important remark concerns the fact that in our first applications to atoms, the fixed-node error was found to systematically decrease both as a function of the number of determinants selected and of the size of the one-particle basis set used.<sup>22</sup> A similar behavior is observed here for the  $F_2$  molecule. Such a systematic improvement of CIPSI nodes is much attractive since it allows a simple linear control of the fixed-node error: the size of the linear expansion may be stopped when a given level of convergence for the fixed-node error is reached. At this point, increasing the size of the expansion is no longer necessary.

Of course, there is a price to pay for such attractive features. When using Jastrow-free CI expansions, the variance of the trial wavefunction is (much) increased with respect to those of fully optimized explicitly correlated wavefunctions and simulation times are also significantly increased. However, in all applications performed up to now, this aspect turns out not to be critical thanks to our efficient implementation of the

calculation of the multi-determinant part. Using computational platforms of intermediate size (a few thousands of compute cores), it appeared possible in all cases to run simulations leading to a statistical error systematically smaller than the fixed-node one. For example, in the case of the all-electron FN-DMC calculation for the copper atom<sup>18</sup> more than 40 000 determinants have been used and a factor of less than one hundred in computing time is lost with respect to a single-determinant simulation. Such a performance for large multi-determinant expansions is possible thanks to an efficient implementation of the calculation of determinants and their derivatives, both at the algorithmic and compute core levels, see Refs. 23 and 24. From a more general perspective, we would like to advocate the idea that in view of the embarrassingly parallel property of QMC and the rapid development of ever more powerful parallel machines, to have to our disposal a simple, automatic, and systematic way of decreasing and controlling the fixed-node bias is a key point, even at the cost of more CPU-intensive simulations.

The contents of this paper are as follows. In Sec. II, a few words about the Fixed-Node Diffusion Monte Carlo method employed here are given. In Sec. III, the CIPSI algorithm used for building the selected configuration interaction expansion is described. In Sec. IV, CIPSI and FN-DMC results for the  $F_2$  ground-state potential energy curve and the corresponding spectroscopic quantities are presented in detail. The important problem of keeping a coherent nodal error along the potential energy curve is addressed and a practical solution is proposed. Finally, the main results of this work are summarized in Sec. V.

## II. THE FIXED-NODE DIFFUSION MONTE CARLO

In this work, the FN-DMC method—the standard quantum Monte Carlo electronic-structure approach for molecules—is employed. For a detailed presentation of its theoretical and practical aspects, the reader is referred to the literature, e.g., Refs. 1, 25, and 26. Here, we just recall that the central quantity of such approaches is the trial wavefunction  $\Psi_T$  determining both the magnitude of the fixed-node error through its approximate nodes as discussed in the Introduction and the quality of the statistical convergence (good trial wavefunctions = small statistical fluctuations). In the present case, let us remark that the molecule is sufficiently small and the trial wavefunction sufficiently good to obtain statistical errors much smaller than fixed-node ones. Thus, in practice, we will only be concerned with the problem of reducing as much as possible the fixed-node bias. The computational cost of FN-DMC is almost entirely determined by the evaluation at each Monte Carlo step of the value of  $\Psi_T$  and its first (drift vector) and second derivatives (Laplacian needed for the local energy). In view of the very large number of MC steps usually required (typically at least millions and often much more) to be able of computing such quantities very rapidly is essential.

In the present work, as presented in detail in Sec. III, the trial wavefunction will be obtained from a truncated configuration interaction expansion, that is, a finite sum of determinants. Typically, the size of the expansion considered will range from a few thousands up to a few hundred thousands of determinants. As a consequence, some care is required

when computing such expansions to keep the computational cost reasonable. The calculation of the Laplacian of the wavefunction and the drift term involves the computation of the inverse of the Slater matrices corresponding to each determinant. At first glance, the CPU cost is expected to be proportional to the number of determinants  $N_{\text{dets}}$  involved in the expansion of the trial wavefunction. Actually, it is not true since in the spin-free formalism used in QMC (Ref. 27 and also Refs. 1, 25, and 26) each Slater determinant expressed in terms of spin-orbitals decomposes into a product of two determinants, each of them corresponding to a given occupation of a set of *purely spatial* molecular orbitals. In practice, only two inverse Slater matrices (one of each spin) are computed with an  $O(N^3)$  algorithm. All the other matrices are built using  $O(N^2)$  Sherman-Morrison updates.<sup>28</sup> Therefore, the computational cost scales as  $O(N_\alpha^2 \times \sqrt{N_{\text{dets}}})$  where  $N_\alpha$  is the number of  $\alpha$  electrons and  $N_{\text{dets}}$  is the number of determinants (products of  $\alpha$  and  $\beta$  determinants).

### III. PERTURBATIVELY SELECTED CONFIGURATION INTERACTION

In multi-determinant expansions, the ground-state wavefunction  $|\Psi_0\rangle$  is written as a linear combination of Slater determinants  $\{|D_i\rangle\}$ , each determinant corresponding to a given occupation by the  $N_\alpha$  and  $N_\beta$  electrons of  $N = N_\alpha + N_\beta$  electrons among a set of  $M$  spin-orbitals  $\{\phi_1, \dots, \phi_M\}$  (restricted case). When no symmetries are considered, the maximum number of such determinants is  $\binom{M}{N_\alpha} \binom{M}{N_\beta}$ , a number that grows factorially with  $M$  and  $N$ . The best representation of the exact wavefunction in the entire determinantal basis is the FCI wavefunction written as

$$|\Psi_0\rangle = \sum_i c_i |D_i\rangle, \quad (1)$$

where  $c_i$  are the ground-state coefficients obtained by diagonalizing the Hamiltonian matrix,  $H_{ij} = \langle D_i | H | D_j \rangle$ , within the orthonormalized set,  $\langle D_i | D_j \rangle = \delta_{ij}$ , of determinants  $|D_i\rangle$ .

As well known, the main problem with FCI is the exponential increase of the wavefunction, leading to unfeasible calculations except for small systems. However, an important feature of FCI expansion is that a vast majority of determinants have negligible coefficients due to their unphysical meaning and, in practice, only a tiny fraction of the FCI space is expected to be important.

To avoid handling the prohibitive size of the expansion, one may try to select determinants by order of excitations with respect to the HF reference determinant, e.g., by taking all single and double excitations (CISD), triple and quadruple (CISDTQ), but, here also, we are faced with the problem of handling a formidable number of determinants. For example, in the case of a CISD calculation, the expansion size is about  $(N_\alpha + N_\beta)^2 n_{\text{virt}}^2$  where  $n_{\text{virt}}$  is the number of virtual orbitals (unoccupied orbitals in the HF determinant), while for CISDTQ, this size is of order  $(N_\alpha + N_\beta)^4 n_{\text{virt}}^4$ . However, in both cases, we are still managing unnecessarily a great number of determinants having a negligible weight in the expansion.

A natural idea to make such CI wavefunction much more compact (in practice, a most important property for DMC calculations) is to select among the FCI expansion only those determinants that contribute in a non-negligible way to the total energy. Such an idea and similar ones have been developed by several groups during the last decades (see, among others Refs. 20, 21, and 29–35). Here, we shall employ an approach close to that introduced by Huron *et al.*<sup>20</sup> and Evangelisti *et al.*<sup>21</sup> Referred to as the CIPSI method, it is based on a selection process constructed by using a perturbative estimate of the energy contribution of each determinant to a reference wave function built iteratively. More details can be found in Refs. 20 and 21. Starting from this idea, we have implemented a CIPSI-like algorithm to build compact trial wavefunctions to be used in FN-DMC calculations.

In its simplest form, the multi-determinant wavefunction is iteratively built as follows:

- Step 0: Start from a given determinant (e.g., the Hartree-Fock determinant) or set of determinants, thus defining an initial reference subspace:  $S_0 = \{|D_0\rangle, \dots\}$ . Diagonalize  $H$  within  $S_0$  and get the ground-state energy  $E_0^{(0)}$  and eigenvector

$$|\Psi_0^{(0)}\rangle = \sum_{i \in S_0} c_i^{(0)} |D_i\rangle. \quad (2)$$

Here and in what follows, the superscript on various quantities is used to indicate the iteration number.

Then, do iteratively ( $n = 0, \dots$ ):

- Step 1: Collect all *different* determinants  $|D_{i_c}\rangle$  connected by  $H$  to  $|\Psi_0^{(n)}\rangle$ , namely,

$$\langle \Psi_0^{(n)} | H | D_{i_c} \rangle \neq 0. \quad (3)$$

- Step 2: Compute the second-order change to the total energy resulting from each connected determinant

$$\delta e(|D_{i_c}\rangle) = - \frac{\langle \Psi_0^{(n)} | H | D_{i_c} \rangle^2}{\langle D_{i_c} | H | D_{i_c} \rangle - E_0^{(n)}}. \quad (4)$$

- Step 3: Add the determinant  $|D_{i_c}^*\rangle$  associated with the largest  $|\delta e|$  to the reference subspace

$$S_n \rightarrow S_{n+1} = S_n \cup \{|D_{i_c}^*\rangle\}.$$

- Step 4: Diagonalize  $H$  within  $S_{n+1}$  to get

$$|\Psi_0^{(n+1)}\rangle = \sum_{i \in S_{n+1}} c_i^{(n+1)} |D_i\rangle \text{ with } E_0^{(n+1)}. \quad (5)$$

- Go to step 1 or stop if the target size  $N_{\text{dets}}$  for the reference subspace has been reached.

Let us denote  $|\Psi_0\rangle$  the wavefunction issued from the previous selection process and  $E_0$  its variational energy. In all what follows,  $|\Psi_0\rangle$  will be referred to as the CIPSI reference wavefunction and  $E_0$  the variational CIPSI energy. Having constructed a zeroth order wavefunction, an improved estimate of the FCI energy can be obtained by adding to the variational energy the second-order correction,  $E_{\text{PT2}}$

$$E_{\text{PT2}} = - \sum_{i \in M} \frac{\langle \Psi_0 | H | D_i \rangle^2}{\langle D_i | H | D_i \rangle - E_0}, \quad (6)$$



where  $M$  denotes the set of all determinants not belonging to the reference space and connected to the reference wavefunction  $|\Psi_0\rangle$  by the Hamiltonian  $H$  (single and double excitations). Finally, the total energy obtained is given by

$$E_0(\text{CIPSI}) = E_0 + E_{\text{PT}2}. \quad (7)$$

In this work, this latter energy will be referred to as the full CIPSI energy.

At this point, a number of remarks are in order:

(i) Although the selection scheme is presented here for computing the ground-state eigenvector only, no special difficulties arise when generalizing the scheme to a finite number of states (see, e.g., Ref. 21).

(ii) The decomposition of the Hamiltonian  $H$  underlying the perturbative second-order expression introduced in step 2

is given by

$$H = H_0 + \langle D_{i_c} | H | D_{i_c} \rangle | D_{i_c} \rangle \langle D_{i_c} |,$$

where  $H_0$  is the restriction of  $H$  to the reference subspace. This decomposition known as the Epstein-Nesbet partition<sup>36,37</sup> is not unique, other possible choices are the Møller-Plesset partition<sup>38</sup> or the barycentric one,<sup>20</sup> see discussion in Ref. 21.

(iii) Instead of calculating the energetic change perturbatively, expression (4), it can be preferable to employ the non-perturbative expression resulting from the diagonalization of  $H$  into the two-dimensional basis consisting of the vectors  $|\Psi_0^{(n)}\rangle$  and  $|D_{i_c}\rangle$ . Simple algebra shows that the energetic change is given by

$$\delta e(|D_{i_c}\rangle) = \frac{[\langle D_{i_c} | H | D_{i_c} \rangle - E_0(N_{\text{dets}})] \left[ 1 - \sqrt{1 + \frac{4\langle \Psi_0^{(n)} | H | D_{i_c} \rangle^2}{[\langle D_{i_c} | H | D_{i_c} \rangle - E_0(N_{\text{dets}})]^2}} \right]}{2}. \quad (8)$$

In the limit of small transition matrix elements,  $\langle \Psi_0^{(n)} | H | D_{i_c} \rangle$ , both expressions (4) and (8) coincide. In what follows the non-perturbative formula will be used.

(iv) In step 3, a unique determinant is added at each iteration. Adding a few of them simultaneously is also possible, a feature particularly desirable when quasi-degenerate low-lying determinants are showing up. In the applications to follow, this possibility has been systematically used by keeping at each iteration all determinants associated with an energetic change whose absolute value is greater than a given threshold.

(v) The implementation of this algorithm can be performed using limited amount of central memory. On the other hand, the CPU time required is essentially proportional to  $N_{\text{dets}} n_{\text{occ}}^2 n_{\text{virt}}^2$ , where  $n_{\text{occ}}$  and  $n_{\text{virt}}$  are the number of occupied and virtual molecular orbitals, respectively.

#### IV. APPLICATION TO THE F<sub>2</sub> MOLECULE

In this section, calculations of the potential energy curve of the F<sub>2</sub> molecule both at the deterministic CIPSI and stochastic Fixed-Node DMC levels are presented. In Subsection IV A, we first present and discuss the results obtained with CIPSI. The dependence of the variational and full CIPSI energy curves on the number of selected determinants and on the basis set (Dunning's cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ) is analyzed. Following the standard implementation of CIPSI, the number of determinants selected is kept constant along the potential energy curve; this approach is referred to as CIPSI at constant number of determinants. In Subsection IV B, FN-DMC results using CIPSI reference functions as trial wavefunctions are presented. To improve the global description of the potential energy curve (decrease of the non-parallelism error), we propose a strategy based on the

use of CIPSI reference functions with a variable number of determinants along the energy curve; this approach presented in Subsection IV C is referred to as CIPSI at constant  $E_{\text{PT}2}$ . FN-DMC results obtained with such trial wavefunctions are presented in Subsection IV D. Finally, a graphical summary of the potential energy curves obtained here at various levels of approximation is presented in Subsection IV E.

#### A. CIPSI at constant number of determinants

In Figure 1, variational CIPSI energy curves for increasing numbers of selected determinants are presented. The curves have been drawn from the energy values calculated at 27 interatomic distances. Interpolation between points is made using standard cubic splines. The number of determinants is kept constant along the potential energy curve. The atomic basis set used is the Dunning cc-pVDZ (VDZ) basis set.<sup>39</sup> For all basis sets considered in this work, CIPSI calculations are done with the molecular orbitals obtained from a minimal Complete-Active-Space Self-Consistent-Field (CASSCF) calculation (two electrons in two orbitals); no re-optimization of the molecular orbitals is performed. In addition, the core electrons are kept frozen. The curves of Fig. 1 are obtained by stopping the CIPSI iterative process for a number of determinants  $N_{\text{dets}} = 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 7.5 \times 10^4$ , and  $10^5$ . At the scale of the figure, the 500- and 1000-determinant energy curves are not yet converged to the full CI. In both cases, a fictitious dissociation barrier at intermediate internuclear distances is observed. This artefact associated with the lack of convergence of the multi-determinant expansion disappears for larger numbers of determinants. A convergence of the whole curve at the kcal/mol level is reached for a number of determinants between  $5 \times 10^4$  and  $7.5 \times 10^4$ . Using the cc-pVDZ basis

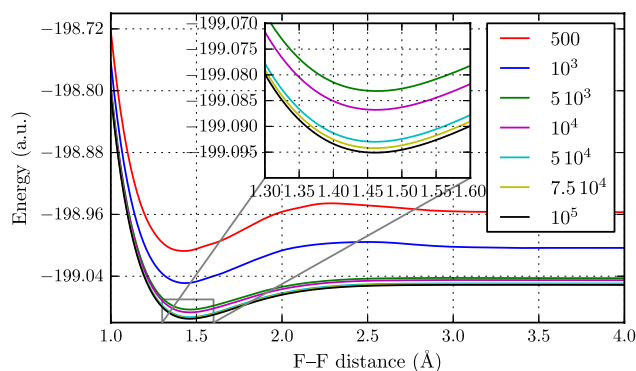


FIG. 1. cc-pVDZ basis set. Convergence of the variational CIPSI energy as a function of the number of selected determinants. Inset = blow up of the equilibrium region.

set (28 atomic orbitals), the size of the FCI space is about  $10^{12}$  determinants (1s orbitals frozen and no symmetry taken into account). The rapid convergence of variational CIPSI energy curves for such a large size illustrates the benefit of considering selected CI instead of more conventional CI schemes based on the use of full subspaces corresponding to multi-excitations of increasing order (all single-, all single- and double-, etc.) whose sizes become rapidly too large. Here, a CISD calculation (all single- and double- excitations) leads to a subspace size of about  $2 \times 10^4$ , while all SDTQ excitations (CISDTQ) generate about  $10^9$  determinants.

Figure 2 shows the full CIPSI energy curves obtained by adding to each variational energy the second-order perturbation energy correction, Eq. (6). The improvement of the convergence in the number of determinants is striking. At about  $5 \times 10^3$  determinants, the convergence of the whole energy curve is reached with chemical accuracy ( $\sim 1$  kcal/mol). At  $10^5$  determinants, the full CIPSI curve is expected to be an accurate estimate of the exact nonrelativistic full valence CI-VDZ potential energy curve. To quantify this latter aspect, we report in Table I our CIPSI energies together with the values of Bytautas *et al.*<sup>40</sup> calculated with the Correlation Energy Extrapolation by Intrinsic Scaling (CEEIS) approach. The CEEIS energies are believed to coincide with the exact FCI values with an accuracy of about 0.3 mEh. At the experimental equilibrium distance, we also report the total energies obtained by Cleland *et al.* using *i*-FCIQMC.<sup>41</sup> The 13 internuclear distances of Table I are those considered by Bytautas *et al.* At the VDZ level, the differences between CIPSI and CEEIS values are very small. Around the equilibrium energy, the maximum error between them is about 0.2 mEh and is slightly larger at larger distances with a maximum of about 0.4 mEh. At the equilibrium distance, both CIPSI and CEEIS values almost coincide with the *i*-FCIQMC results.

When increasing the size of the basis set, the variational and CIPSI curves have a very similar behavior as a function of the number of determinants (see supplementary information<sup>42</sup>). The only difference is that the convergence is slowed down due to the increase of the size of the Hilbert space. We have performed calculations with the cc-pVTZ (VTZ) (60 atomic orbitals) and aug-cc-pVTZ (AVTZ) (92 atomic orbitals) basis sets. The sizes of the Full CI space are much increased, about  $10^{20}$  and  $10^{23}$  for the cc-pVTZ and aug-cc-

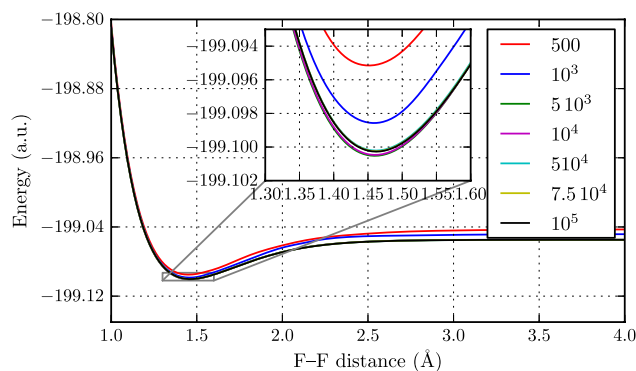


FIG. 2. cc-pVDZ basis set. Convergence of the full CIPSI energy as a function of the number of selected determinants. Inset = blow up of the equilibrium region.

pVTZ basis sets, respectively. At the variational level, the convergence at the chemical level with VTZ is attained for a number of determinants of about  $10^5$ . With AVTZ, this level of convergence is not reached even with  $10^5$  determinants. As for the VDZ basis set, the convergence is greatly enhanced when the second-order perturbative correction is added up. The full CIPSI-VTZ energy curve is converged with a maximum error of about 1 kcal/mol for about  $2 \times 10^4$  determinants. As seen from Table I, the errors with respect to the accurate values of Ref. 40 are small. Around the equilibrium distance, the error is of order 1 mEh and about two times larger in the long-distance regime. The convergence of the full CIPSI with the largest AVTZ basis set is still satisfactory and is obtained here at the kcal/mol level with a number of determinants greater than  $4 \times 10^4$  determinants. Once again, we emphasize that obtaining such good quality FCI curves with a number of determinants representing only a tiny fraction of the whole Hilbert space is particularly remarkable (fractions of about  $10^{-7}$ ,  $10^{-15}$ , and  $10^{-18}$  for the VDZ, VTZ, and AVTZ basis sets, respectively).

In Figure 3, a comparison of the CIPSI energy curves with those obtained by more standard approaches using the VDZ basis set is presented. The CASSCF(2,2) potential energy curve is plotted in the upper part of the figure. Due to the absence of dynamical correlation contributions, CASSCF values are much too high in energy. However, as it should be the energy curve displays a correct dissociation behavior, the large-distance energy converging to the sum of HF energies of the two fluorine atoms. Note that the Hartree-Fock curve is not given here since at this level of approximation, the  $F_2$  molecule is not even bound. The Coupled-Cluster curve (green line) using Single and Double excitations (CCSD) gives much more satisfactory results. In the equilibrium geometry region, CCSD energies are close to CIPSI results but still slightly higher. However, at large separations, the CCSD curve dissociates with a large error of about 0.5 a.u. with respect to the FCI-VDZ atomic energies. This latter atomic limit is drawn on the figure as a horizontal line. As seen the variational CIPSI energy curve almost dissociates toward the exact value (error of about 0.013 a.u. at  $R = 4$  Å a small but discernible quantity on the figure) while the full CIPSI energy curve is in full agreement with the FCI atomic limit (indistinguishable on the figure).

TABLE I. Total nonrelativistic ground-state energies calculated using CIPSI (core electrons frozen). Basis sets = cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ. For VDZ and VTZ results are compared with the values of Bytautas *et al.*<sup>40</sup> and those of Cleland *et al.*<sup>41</sup> at the experimental distance. Energy in hartree.

R in Å	CIPSI-VDZ	CEEIS-VDZ <sup>a</sup>	<i>i</i> -FCIQMC-VDZ <sup>b</sup>	CIPSI-VTZ	CEEIS-VTZ <sup>a</sup>	<i>i</i> -FCIQMC-VTZ <sup>b</sup>	CIPSI-AVTZ
1.14	-199.007 16	-199.007 18	...	-199.212 7	-199.213 4	...	-199.228 3
1.20	-199.048 02	-199.048 11	...	-199.252 2	-199.253 0	...	-199.267 9
1.30	-199.084 94	-199.085 10	...	-199.286 3	-199.287 0	...	-199.302 1
1.36	-199.095 18	-199.095 17	...	-199.294 2	-199.295 0	...	-199.310 2
1.411 93 <sup>c</sup>	-199.099 28	-199.099 20	-199.099 41(9)	-199.296 5	-199.297 2	-199.297 7(1)	-199.312 3
1.50	-199.099 77	-199.099 81	...	-199.293 5	-199.294 4	...	-199.309 5
1.60	-199.095 08	-199.095 10	...	-199.285 2	-199.286 1	...	-199.301 2
1.80	-199.080 90	-199.080 90	...	-199.266 2	-199.267 6	...	-199.281 8
2.00	-199.069 07	-199.068 82	...	-199.252 7	-199.254 3	...	-199.267 4
2.20	-199.061 84	-199.061 65	...	-199.245 3	-199.247 1	...	-199.259 4
2.40	-199.058 06	-199.058 23	...	-199.241 7	-199.243 6	...	-199.255 6
2.80	-199.055 52	-199.055 77	...	-199.239 3	-199.241 2	...	-199.252 3
8.00	-199.055 06	-199.055 45	...	-199.238 4	-199.240 8	...	-199.250 0
Atomic limit F + F							
	VDZ			VTZ			AVTZ
CIPSI, this work	-199.055 53			-199.241 1			-199.255 9
<i>i</i> -FCIQMC	-199.055 44(8) <sup>b</sup>			-199.241 0(2) <sup>b</sup>			...

<sup>a</sup>Reference 40.<sup>b</sup>Reference 41.<sup>c</sup>Expt. equilibrium distance.

To get a more quantitative view of the dependence of the CIPSI potential energy curves on the number of determinants selected, we report the results obtained for the three basic spectroscopic quantities: The equilibrium distance,  $R_{eq}$ , the dissociation energy  $D_0$ , and the second derivative  $k$  at  $R_{eq}$  (curvature), a quantity directly related to the harmonic frequency. Data for the VDZ, VTZ, and AVTZ basis sets are given in Table II. In this work, an accurate approximation of the exact non-relativistic, infinite nuclear masses, potential energy curve of the fluorine molecule has been built from data given in Refs. 43 and 44. From  $R = 1.14$  Å to  $R = 2.4$  Å, total energies are reconstructed from the non-relativistic contributions of Table IV in Ref. 43, while for the long-range regime, from  $R = 2.8$  Å to 8 Å, the data used are taken from Table V of Ref. 44. The (F + F) dissociation limit is calculated using the atomic value of Davidson *et al.* in Ref. 45. Note that the maximum error in the total energy curve resulting from these data is expected to be much smaller than the millihartree, a

precision sufficient for our needs. To get the spectroscopic quantities, an accurate fit of the energies calculated at 26 interatomic distances via a 10-parameter generalized Morse potential representation has been performed. In the case of the VDZ basis set, we also report an accurate estimate of the VDZ dissociation energy obtained from the *i*-FCIQMC energy calculation of Cleland *et al.*<sup>41</sup> The value of 45.00(11) a.u. reported in the table is a distance-corrected value that we have obtained by adding to the value of 43.87(11) a.u. from Ref. 41 calculated at the experimental equilibrium distance of  $R = 1.4119$  Å, the correction needed to shift to the minimum of the VDZ energy curve at  $R = 1.463$  Å. As expected, CASSCF values for the three spectroscopic quantities are of low quality. With the biggest AVTZ basis set, the CASSCF dissociation energy is less than one-half (28.4 mEh) of the exact value, the equilibrium distance is much too large, and the curvature too small. Using the VDZ basis set, variational CIPSI results give an equilibrium distance and curvature essentially converged between  $10^4$  and  $5 \times 10^4$  determinants. In contrast, the dissociation energy is still varying from  $10^4$  ( $D_0 = 41.9$  mEh) to  $10^5$  determinants ( $D_0 = 43.97$  mEh). At the full CIPSI level, the convergence of the dissociation energy is much better but still slightly decreasing. The value for  $10^5$  determinants is 45.17 mEh, to be compared with the corrected value of 45.00(11) mEh of Ref. 41. We also report the spectroscopic quantities obtained from the CEEIS data. The agreement with our own data is excellent, thus confirming that at the VDZ level (see Table I), the results obtained with CIPSI are of a quasi-FCI quality. Upon increasing the basis to the larger cc-pVTZ and aug-cc-pVTZ basis sets, the spectroscopic quantities are significantly improved. The size of the Hilbert spaces greatly increasing, the convergence is slowed down. At the VTZ level, the CIPSI dissociation energy is still varying (decreasing) up to  $5 \times 10^4$  determinants, the value obtained

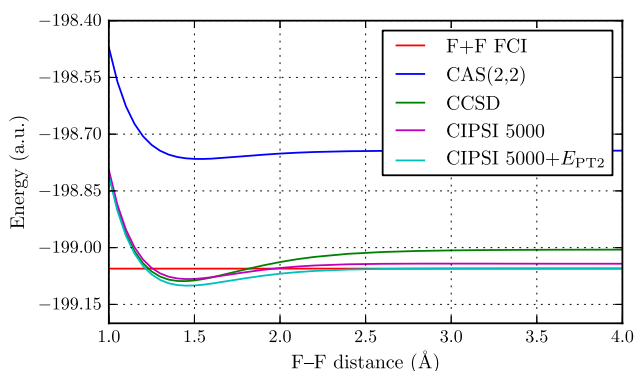


FIG. 3. cc-pVDZ basis set. Comparison of CIPSI results with CASSCF (upper part) and CCSD. The exact FCI (F + F) dissociation limit is given.



TABLE II. Spectroscopic constants from variational and full CIPSI as a function of  $N_{\text{dets}}$  and of the basis set employed.  $R_{eq}$  in Å,  $D_0$  in millihartrees, and  $k$  in hartree/Å<sup>2</sup>. Spectroscopic constants ( $X = R_{eq}$ ,  $D_0$ , and  $k$ ) presented as  $X_1/X_2$  where  $X_1$  and  $X_2$  are the values obtained from the variational and full CIPSI energy curves, respectively. For comparison, the CASSCF, exact non-relativistic, and the corrected near-FCI value of Cleland *et al.* (see text) are also given.

VDZ basis set								
$N_{\text{dets}}$	CASSCF 2	CIPSI: Variational <sup>a</sup> /full <sup>b</sup>					CEEIS <sup>c</sup>	<i>i</i> -FCIQMC <sup>d</sup>
		$5 \times 10^3$	$10^4$	$5 \times 10^4$	$7.5 \times 10^4$	$10^5$		
$R_{eq}$	1.531	1.465/1.460	1.464/1.460	1.463/1.462	1.462/1.463	1.463/1.463	1.460	...
$D_0$	22.1	40.9/45.7	41.9/45.6	43.8/45.3	44.39/45.22	43.97/45.17	45.14	45.00(11)
$k$	0.43	0.73/0.78	0.75/0.78	0.77/0.77	0.78/0.76	0.76/0.76	0.80	...
VTZ basis set								
$N_{\text{dets}}$	CASSCF 2	CIPSI: Variational/full					CEEIS	
		$5 \times 10^3$	$10^4$	$5 \times 10^4$	$7.5 \times 10^4$	$10^5$		
$R_{eq}$	1.469	1.410/1.412	1.409/1.415	1.418/1.417	1.418/1.419	1.418/1.417	1.416	
$D_0$	26.5	56.8/61.0	49.8/58.6	52.9/57.6	49.8/58.6	54.14/57.6	56.7	
$k$	0.63	1.067/1.132	1.062/1.092	1.037/1.079	1.062/1.092	1.030/1.079	1.075	
AVTZ basis set								
$N_{\text{dets}}$	CASSCF 2	CIPSI: Variational/full					CEEIS	
		$5 \times 10^3$	$10^4$	$5 \times 10^4$	$7.5 \times 10^4$	$10^5$		
$R_{eq}$	1.463	1.405/1.413	1.392/1.414	1.415/1.418	1.418/1.419	1.418/1.419	...	
$D_0$	28.4	59.6/66.4	57.5/63.8	52.1/60.2	54./60.0	54./60.0	...	
$k$	0.66	1.158/1.151	1.108/1.117	1.013/1.078	1.03/1.073	1.03/1.073	...	
Infinite basis set (exact)								
$R_{eq}$					1.412			
$D_0$					62.35 <sup>e</sup>			
$k$					1.121			

<sup>a</sup>Spectroscopic data obtained from the variational CIPSI energy curve.

<sup>b</sup>Spectroscopic data obtained from the full CIPSI energy curve, Eq. (7).

<sup>c</sup>Reference 40.

<sup>d</sup>Reference 41.

<sup>e</sup>Estimation of the nonrelativistic full valence CI dissociation energy (no core correl.) from Table IV of Bytautas *et al.*<sup>43</sup>

with the largest calculation being  $D_0 = 57.6$  mEh. Compared to the VDZ value of 44.1 mEh, this value is much improved (exact value of 62.35 mEh). As it should be, the equilibrium distance is reduced and the curvature increased as a result of the deepening of the well. The comparison with CEEIS data is also very good, except for the dissociation energy which we found about 1 mEh larger with CIPSI (57.6 and 56.7). This difference clearly results from the increase as a function the distance of the CIPSI error with respect to the quasi-FCI (CEEIS) results: At  $R = 1.41193$ , the error is 0.7 mH and 2.4 mH at  $R = 8$ . At the AVTZ level, the CIPSI dissociation energy for  $10^5$  determinants is found to be 60.0 mEh. However, because of the increase of the CIPSI error with distance just discussed in the AVTZ case, this value must be taken with lot of caution and is very likely overestimated by one or two millihartrees. This is confirmed by the fact that using the CEEIS data of Bytautas *et al.*,<sup>43</sup> with the larger VQZ basis set the dissociation energy (expected to be larger than with AVTZ) is 59.8 mEh.

At this point, it would be desirable to increase further the basis set by considering higher cardinal numbers (QZ, 5Z, etc.). However, our purpose being to avoid to handle

exponentially increasing Hilbert spaces, we shall now turn our attention to FN-DMC methods.

## B. FN-DMC with CIPSI at constant number of determinants

We present all-electron FN-DMC calculations of the potential energy curve using CIPSI reference wavefunctions as trial wavefunctions. All FN-DMC simulations presented in this section have been performed by running first a CIPSI calculation at constant number of determinants and, then, using as trial wavefunction the CIPSI reference wave function *as it is*. No preliminary multi-parameter stochastic optimization of the trial wavefunction has been performed, the CASSCF molecular orbitals are kept unchanged for all distances and calculations, the CI coefficients are those issued from CIPSI, and no Jastrow prefactor has been employed (we just impose the electron-nucleus cusp conditions at very short distances). In this way, all calculations may be performed in a fully automatic way: (i) Choose a target number of determinants and of Monte Carlo steps, (ii) run CIPSI, and then (iii) run FN-DMC. From a computational point of view,

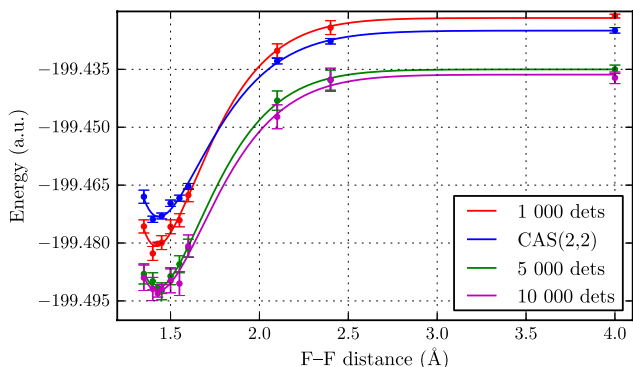


FIG. 4. FN-DMC with CIPSI-VDZ nodes. Convergence of the FN-DMC energy curve as a function of the number of determinants selected in the trial wavefunction.

an efficient implementation of the calculation of the multi-determinant expansion and its first and second derivatives has been used.<sup>23,24</sup> Only  $\alpha$ - and  $\beta$ -determinants that are different are computed. Their number is roughly given by the square root of the total number of determinants; for a wavefunction including  $10^4$  determinants (the maximum size considered in what follows), about 100  $\alpha$  and 100  $\beta$  determinants are to be computed. Moreover, an  $O(N^3)$  algorithm for computing determinants is only employed for the first  $\alpha$ - and  $\beta$ -determinants, then Sherman-Morrison updates<sup>28</sup> are used to compute all the other determinants as well as their  $3N$  gradient components and Laplacians. Finally, for  $10^4$  determinants, the calculation is about 39 times longer than for a single determinant.

The convergence of the FN-DMC energy curve as a function of the number of determinants for the cc-pVDZ basis set is presented in Figure 4. For the sake of comparison, the DMC curve obtained from CAS(2,2) nodes is also reported. Total energies have been calculated for a set of interatomic distances ranging from 1.35 to 4.00 Å. The statistical error on FN-DMC/CIPSI energies using 1000 and 5000 determinants is typically 0.002 a.u. For 10 000 determinants, a slight larger value of about 0.003 a.u. is obtained. To better estimate the dissociation energy (see below), the energy values at the distances of 1.428 and 4 have been computed using more statistics (Monte Carlo runs ten times longer). Quite remarkably, FN-DMC/CIPSI energies are found to systematically decrease for all interatomic distances as the number of selected determinants is increased. A similar property has been observed for the oxygen atom<sup>16</sup> and,

also, for bigger atoms.<sup>17,18</sup> To understand the origin of this remarkable mathematical property is not simple. However, a heuristic argument can be given as follows. When the total energy is lowered (the criterion used during the CIPSI selection process), the wavefunction is dominantly improved in the neighborhood of its maxima (regions contributing the most to the energy). In particular, the localization of such maxima is expected to be improved. Now, the positions of the maxima and zeros of a wavefunction being intimately correlated (as any solution of a wave-like equation), we can also expect an improved localization of the nodes. As seen on the figure, the convergence of the FN-DMC energy curve is approximately reached for  $5 \times 10^3$  determinants, a result coherent with the convergence of the variational CIPSI energy curve at roughly the same number of determinants, Fig. 1. Note that handling a few thousands of determinants in FN-DMC (the trial wavefunction and its derivatives are to be computed at each Monte Carlo step) is still feasible in practice.

In Table III, the FN-DMC spectroscopic quantities ( $R_{eq}$ ,  $D_0$ , and  $k$ ) obtained with CAS(2,2) and CIPSI/VDZ nodes for increasing numbers of determinants are presented. The values and error bars have been obtained by fitting a set of 20 energy curves using a 10-parameter generalized Morse potential representation. Each of these curves is obtained from different independent realizations of the statistical noise. Note that due to the absence of correlations for the noise, the error bars obtained in this way are certainly overestimated. For  $D_0$ , we also give the value obtained by directly computing the difference  $\Delta E$  between total FN-DMC energies at the equilibrium geometry and at the large-distance value of  $R = 4$  Å. Spectroscopic quantities are essentially converged within error bars for  $5 \times 10^3$  determinants. However, due to the magnitude of statistical fluctuations, to extract accurate values of the equilibrium distances and curvatures is impossible. Using CIPSI nodes, typical values are  $R_{eq} = 1.43(3)$  and  $k = 1.0(4)$ . Error bars are clearly too large to allow any detailed analysis. In contrast, the situation is more satisfactory for the dissociation energy. Coherent values with small enough statistical errors are obtained either from the statistical fit of the energy curves or the direct calculation of the energy gap. Using  $10^4$  determinants, the FN-DMC dissociation energy obtained is 56.1(12) mEh, a clear improvement with respect to the value of 41.9 mEh corresponding to the trial wavefunction (see Table II). Finally, remark that results obtained with FN-DMC/CIPSI-VDZ are of a comparable quality to those obtained at the variational CIPSI/VTZ level (see Table II).

TABLE III. FN-DMC spectroscopic quantities computed with CAS(2,2) and CIPSI-VDZ nodes (number of determinants =  $10^3$ ,  $5 \times 10^3$ , and  $10^4$ ). Values of ( $R_{eq}$ ,  $D_0$ , and  $k$ ) and error bars obtained from a statistical distribution of 20 different energy curves fitted with a generalized Morse potential. In the case of  $D_0$ , the dissociation energy directly obtained from the energy gap is also given. Equilibrium distances in Å, dissociation energies in millihartrees, and energy curvatures in hartree/Å<sup>2</sup>.

	Exact	CAS-nodes	$10^3$ -det nodes	$5 \times 10^3$ -det nodes	$10^4$ -det nodes
$R_{eq}$	1.412	1.434(8)	1.421(15)	1.427(26)	1.429(20)
$D_0$	62.0 <sup>a</sup>	48.3(7)/ $\Delta E = 48.3$ (10)	59.1(13)/ $\Delta E = 59.0$ (6)	56.6(16)/ $\Delta E = 56.5$ (13)	56.1(12)/ $\Delta E = 55.9$ (18)
$k$	1.121	1.0(2)	1.4(6)	1.0(5)	1.0(4)

<sup>a</sup> Value taken from Table IV of Bytautas *et al.*<sup>43</sup> and corresponding to the estimated non-relativistic full CI dissociation energy including the contribution of the core correlation (present in FN-DMC).

To improve the quality of the results beyond FN-DMC/CIPSI-VDZ, a natural solution is to increase the size of the basis set. However, such a strategy is doomed to failure since the number of determinants necessary to build converged nodes is expected to increase too rapidly. For example, in the case of the VTZ basis set for which the variational energy was found to converge around  $5 \times 10^4$  determinants, a similar number of determinants should be expected to get accurate enough nodes. In what follows, we want to avoid to follow this path and, instead, propose an alternative strategy based on the improvement of the global shape of the energy curve instead of on the search for increasing precision for total energies.

### C. CIPSI at constant $E_{PT2}$

Stopping the iterative CIPSI process at a given number of determinants identical for all geometries does not insure a coherent description of the energy curve. To construct a wavefunction of comparable quality in each region (short interatomic distances, equilibrium region, intermediate regime, and near-dissociation limit), it is natural to consider expansions involving a variable number of determinants as a function of the interatomic distance. This point is clearly illustrated in Fig. 1 where the intermediate region between  $R = 2 \text{ \AA}$  and  $2.5 \text{ \AA}$  is poorly described with a CIPSI-VDZ wavefunction having a small number of determinants. For  $5 \times 10^2$  and  $10^3$  determinants, a spurious local maximum is observed, this artefact disappearing for larger numbers of determinants. In contrast, in the equilibrium regime where the wavefunction is known to have a much less marked multi-configurational character, no qualitative change for the energy curve is observed when passing from a (very) small to a large number of determinants. Fig. 2 shows that after adding the second-order correction to the variational CIPSI energy, all curves become much better behaved and in particular no spurious maxima are observed. As already pointed out, this result shows that  $E_{PT2}$  represents most of the remaining differences between the variational CIPSI and FCI energies. It thus motivates us to consider the magnitude of  $E_{PT2}$  as an indicator for evaluating the difference between the CIPSI multi-determinant wavefunction and the FCI solution.

The new strategy is then as follows: At a given geometry, the CIPSI iterative process is stopped when a target value for the second-order contribution is reached and not when a fixed number of determinants is obtained. The target value is chosen identical along the PES and the FCI limit is recovered when this value is decreased down to zero. When  $E_{PT2}$  represents a good estimate of the difference between the variational CIPSI and FCI energies, the variational CIPSI curve obtained with the variable number of determinants is expected to be almost parallel to the unknown FCI one. Of course, when higher-order corrections  $E_{PTN}$  ( $N > 2$ ) are present, some deviation from the FCI curve is expected. Remark that some work with a similar idea has been done previously by Angeli and Persico.<sup>46</sup> The difference with the present work is that the norm of some approximate first-order perturbed wavefunction was considered instead. We believe that using directly an energetic criterion is more natural in this context.

To illustrate more quantitatively these ideas, in the case of the CIPSI-VDZ reference wavefunction with  $10^3$  determi-

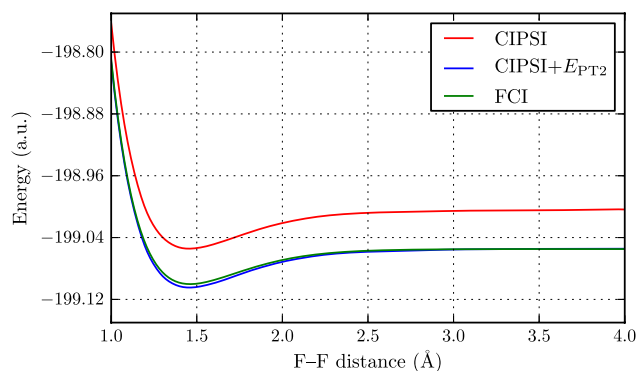


FIG. 5. CIPSI potential energy curve obtained by imposing a constant value  $E_{PT2}$  of  $-0.05$  hartree. The full CIPSI +  $E_{PT2}$  curve is also presented and compared to the FCI curve.

nants, the second-order energy correction is found to vary around a value of about  $-0.05$  a.u. in the equilibrium region and  $-0.045$  in the large-distance limit and to reach a maximal value of about  $-0.058$  in the transition region (around  $2.2 \text{ \AA}$ ). If now we consider a constant value for  $E_{PT2}$  of  $-0.05$  hartree, the number of determinants needed to get this correction varies approximately from 1000 determinants to a maximum of 1200 in the intermediate regime.

The potential energy curves obtained using this scheme are presented in Fig. 5 for a value of  $E_{PT2}$  of  $-0.05$  hartree. The variational CIPSI energy curve obtained with a variable number of determinants (upper curve), the full CIPSI curve obtained by adding the constant perturbative contribution of  $-0.05$ , and the FCI one accurately approximated by the  $10^5$ -determinant curve presented in Fig. 2 are plotted. Remarkably, the FCI and full CIPSI curves are almost identical. It means that higher-order perturbative contributions beyond the second-order are small. Their largest contributions lie in the equilibrium region but do not exceed  $0.005$  a.u. Imposing a constant  $E_{PT2}$ , thus, leads to a full CIPSI curve close to the FCI one and then to a variational CIPSI energy curve almost parallel to the FCI one. Note also that the spurious maximum observed when using a constant small number of determinants is no longer present.

Figure 6 presents the full CIPSI curves obtained with  $E_{PT2} = -0.2, -0.1, -0.05, -0.02,$  and  $-0.01$  (atomic units). As seen on this figure, the various curves, except for the largest

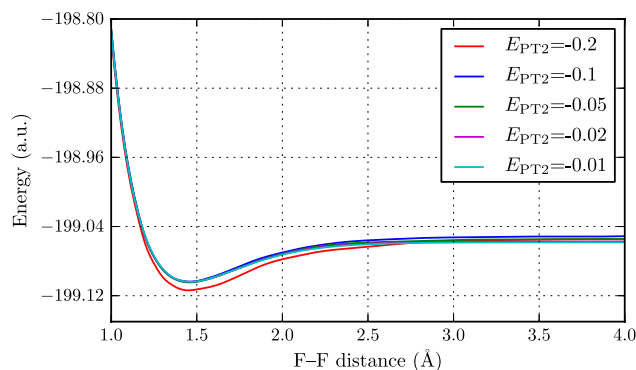


FIG. 6. Convergence of the full CIPSI energy curve for various values of constant  $E_{PT2}$ .

TABLE IV. Basis set = cc-pVDZ. Convergence of the spectroscopic quantities with CIPSI at constant  $E_{PT2}$  as function of  $E_{PT2}$ . For each  $E_{PT2}$ , results obtained from the variational and full CIPSI energy curves are given. Equilibrium distance in Å, dissociation energy in millihartree, and curvature in hartree/Å<sup>2</sup>.

$E_{PT2}$ (a.u.)	CIPSI: Variational/full		
	$R_{eq}$	$D_0$	$k$
-0.2 ( $\sim 1.6 \times 10^2$ dets)	1.458/1.458	57.5/57.3	0.82/0.83
-0.1 ( $\sim 5 \times 10^2$ dets)	1.450/1.451	50.0/51.0	0.83/0.82
-0.05 ( $\sim 1.1 \times 10^3$ dets)	1.455/1.454	49.2/48.6	0.81/0.82
-0.02 ( $\sim 3.5 \times 10^3$ dets)	1.459/1.459	44.6/45.5	0.78/0.78
-0.01 ( $\sim 2 \times 10^4$ dets)	1.460/1.460	44.0/43.9	0.78/0.78
-0.008 ( $\sim 3.5 \times 10^4$ dets)	1.461/1.461	43.8/43.7	0.77/0.77
CIPSI $10^5$ dets <sup>a</sup>	1.463/1.463	43.97/45.17	0.76/0.76
<i>i</i> -FCIQMC		45.00(11)	...
Exact NR	1.412	62.35	1.121

<sup>a</sup>See Table II.

value of  $E_{PT2}$  of  $-0.2$ , almost coincide with the FCI curve at the scale of the figure.

In Table IV, the convergence of the dissociation energy, equilibrium distance, second-derivative, and non-parallelism error for the variational and full CIPSI energy curves as a function of  $E_{PT2}$  is reported. A first remark is that the spectroscopic values obtained from the variational and the full CIPSI curves at constant  $E_{PT2}$  are almost identical (in contrast with CIPSI at constant number of determinants, see Table II). Of course, it is expected since the difference between the variational and full CIPSI energies is imposed to be constant. Note that the (very) small differences observed are because imposing a strict constant value of  $E_{PT2}$  is not possible due to the integer character of the number of determinants. A second remark is that the convergence of the spectroscopic quantities as a function of the number of determinants (for each  $E_{PT2}$ , an estimate of the average number of determinants along the energy curve is given in parentheses) is more rapid than when using CIPSI with a constant number of determinants, thus illustrating the efficiency of the constant  $E_{PT2}$  CIPSI approach.

#### D. FN-DMC with CIPSI at constant $E_{PT2}$

FN-DMC energy curves obtained from CIPSI-VDZ reference wavefunctions at constant  $E_{PT2}$  are presented in Fig. 7.

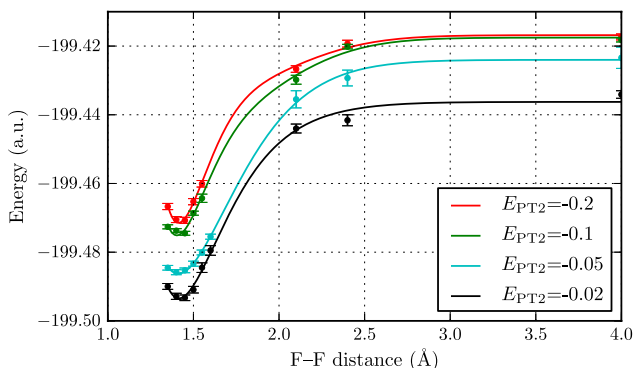


FIG. 7. FN-DMC energy curves for several values of  $E_{PT2}$  (VDZ basis set).

TABLE V. Basis set = cc-pVDZ. FN-DMC spectroscopic values using CIPSI-VDZ trial wavefunctions at constant  $E_{PT2}$ . Equilibrium distance in Å, dissociation energy in millihartree, and curvature in hartree/Å<sup>2</sup>.

$E_{PT2}$ (a.u.)	$R_{eq}$	$D_0$	$k$
-0.2 ( $\sim 1.6 \times 10^2$ dets)	1.414	53.1(4)	1.839
-0.1 ( $\sim 5 \times 10^2$ dets)	1.408	56.2(5)	1.271
-0.05 ( $\sim 1.1 \times 10^3$ dets)	1.414	60.3(6)	1.131
-0.02 ( $\sim 3.5 \times 10^3$ dets)	1.425	54.2(13)	1.125
$10^4$ nodes	1.428	55.3	1.117
Exact	1.412	62.0	1.121

Corresponding spectroscopic values are reported in Table V. To compare with, FN-DMC values using a constant number of  $10^4$  determinants are also given (taken from Table III). FN-DMC spectroscopic quantities using trial wavefunctions at constant  $E_{PT2}$  systematically improve when decreasing  $E_{PT2}$  from  $-0.2$  a.u. to  $-0.05$  a.u. For this latter value, results are much improved with respect to FN-DMC values at constant number of determinants: The error in the dissociation energy is greatly reduced from 6.7 to 1.7 mEh, the equilibrium distance and curvature are found of comparable quality. However, for the smaller value  $E_{PT2} = -0.02$ , results deteriorate and become close to those obtained with FN-DMC with a fixed number of determinants. This result is of course expected since in the limit of a vanishing value for the second-order energy correction, the CIPSI algorithm at constant  $E_{PT2}$  reduces to the standard one.

#### E. Graphical summary: $F_2$ curves at various levels of approximation

In Figure 8, the main results of this work are summarized by showing on the same graph the various energy curves obtained. For the sake of clarity, each energy curve has been shifted down by the constant leading to a dissociation toward the exact (infinite basis) nonrelativistic atomic limit.

All curves are located between the poor-quality CASSCF curve (upper curve) and the estimated exact nonrelativistic fixed nucleus curve (lowest solid line). For the sake of clarity, energy

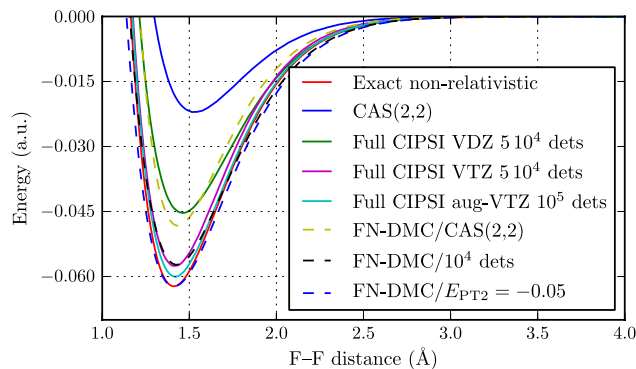


FIG. 8.  $F_2$  potential energy curves using the various approaches presented in this work. Comparison with the estimated exact nonrelativistic curve. Each energy curve has been shifted down to impose at large distance the exact atomic limit.



curves obtained with deterministic calculations are represented by solid lines, while dashed lines are used for FN-DMC energy curves. CIPSI energy curves using the VDZ, VTZ, and AVTZ basis sets are represented by the three curves (solid lines) between the two extreme CASSCF and exact curves. As discussed above, these curves are almost converged in number of determinants and may be considered as a good approximation of the full CI limit in each case. Increasing the basis set has an important impact on the quality of the results. With the largest aug-VTZ basis set, the energy curve obtained is the best one, except for the FN-DMC curve obtained with a CIPSI-constant  $E_{PT2}$  trial wavefunction which is the most accurate energy curve calculated in this work. With the larger QZ basis set, results continue to improve (not done here, our objective being not to increase indefinitely the basis set). The corresponding QZ energy curve is obtained in Ref. 40 at a quasi-FCI level and is found to be of a similar quality to the best FN-DMC-DZ curve presented here. The three FN-DMC curves are represented using a dashed line. By increasing order of quality, they correspond to using (i) the CASSCF nodes, (ii) the CIPSI/VDZ nodes at constant number of determinants, and (iii) the CIPSI/VDZ nodes at constant  $E_{PT2}$ . Using CASSCF nodes, the FN-DMC energy curve is much improved with respect to the variational CASSCF curve. Roughly speaking, it is of quality of the quasi-FCI curve obtained with the VDZ basis set. Now, taking CIPSI-VDZ nodes, DMC results are improved to a quality typical of a FCI-VTZ calculation. Only when considering a nodal construction based on a variable number of determinants in a CIPSI-VDZ framework, DMC results surpass the quality of a FCI-AVTZ calculations. Now, having a closer look at the overall shape of the two best curves, namely, the near FCI-AVTZ and the DMC/VDZ at constant  $E_{PT2}$  curves, the quality is not uniform as a function of the distance. The DMC curve is much better near the equilibrium geometry while at the intermediate region, it is no longer true. In this regime, the FCI calculation performs better thanks to its large number of determinants.

## V. SUMMARY

In this work, we have investigated the use of selected CI (CIPSI) trial wavefunctions for the FN-DMC calculation of the  $F_2$  potential energy curve. A major advantage of CIPSI is to keep the size of the CI expansion limited since only those multiple-particle excitations that contribute the most in each sector of excitations (single, double, triple, etc.) are considered. As in the case of previous calculations for atomic systems,<sup>16–18</sup> it is found that using CIPSI wavefunctions directly issued from the output of the CI program, allows accurate FN-DMC calculations. In particular, no preliminary stochastic optimization step of the trial wavefunction parameters is performed. It is found that the CIPSI nodes obtained using the small cc-pVDZ basis set are much improved with respect to CASSCF(2,2) ones and are of a quality almost similar to that reached with much more complex and fully optimized Jastrow-Slater type wavefunctions.<sup>14</sup> Remarkably, and similarly to what has been observed for atoms, the fixed-node error is found to systematically decrease when increasing the number of selected determinants (see Fig. 4), a property allow-

ing a simple control of the error. In the case of the  $F_2$  molecule, we have shown how CIPSI expansions with a variable number of determinants as a function of the interatomic distance can be used to get a smooth and uniformly accurate potential energy curve (minimization of the non-parallelism error), a fundamental issue when computing potential energy surfaces whatever the electronic structure method used.

Finally, we would like to emphasize that the possibility of using CI expansions of moderate size generated in a simple and systematic way by a deterministic CI approach and avoiding any many-parameter stochastic optimization step is a particularly attractive feature since it should greatly facilitate the use of QMC in a “black box” way.

## ACKNOWLEDGMENTS

We would like to acknowledge interesting discussions with Nathalie Guihery and Jean-Paul Malrieu (Toulouse). A.S. and M.C. thank the Agence Nationale pour la Recherche (ANR) for support through Grant No. ANR 2011 BS08 004 01. This work has been made through generous computational support from CALMIP (Toulouse) under Allocation No. 2013-0510, and GENCI under Allocation No. x2014081738.

- <sup>1</sup>B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*, Volume 1 of Lecture and Course Notes in Chemistry, Volume 1: World Scientific Lecture and Course Notes in Chemistry (World Scientific, Singapore, 1994).
- <sup>2</sup>W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- <sup>3</sup>K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1990).
- <sup>4</sup>H. J. Flad, M. Caffarel, and A. Savin, “Quantum Monte Carlo calculations with multi-reference trial wavefunctions,” in *Recent Advances in Quantum Monte Carlo Methods* (World Scientific Publishing, 1997).
- <sup>5</sup>C. Filippi and C. J. Umrigar, *J. Chem. Phys.* **105**, 213 (1996).
- <sup>6</sup>B. Braïda, J. Toulouse, M. Caffarel, and C. J. Umrigar, *J. Chem. Phys.* **134**, 084108 (2011).
- <sup>7</sup>A. G. Anderson and W. A. Goddard III, *J. Chem. Phys.* **132**, 164110 (2010).
- <sup>8</sup>F. Fracchia, C. Filippi, and C. Amovilli, *J. Chem. Theory Comput.* **8**, 1943 (2012).
- <sup>9</sup>T. Bouabça, B. Braïda, and M. Caffarel, *J. Chem. Phys.* **133**, 044111 (2010).
- <sup>10</sup>A. Monari, A. Scemama, and M. Caffarel, “Large-scale quantum Monte Carlo electronic structure calculations on the EGEE grid,” in *Remote Instrumentation for eScience and Related Aspects* (Springer, 2012), pp. 195–207.
- <sup>11</sup>M. Bajdich, L. Mitáš, G. Drobný, and L. K. Wagner, *Phys. Rev. Lett.* **96**, 240402 (2006).
- <sup>12</sup>M. Casula, C. Attaccalite, and S. Sorella, *J. Chem. Phys.* **121**, 7110 (2004).
- <sup>13</sup>P. L. Rios, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- <sup>14</sup>J. Toulouse and C. J. Umrigar, *J. Chem. Phys.* **128**, 174101 (2008).
- <sup>15</sup>C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, *Phys. Rev. Lett.* **98**, 110201 (2007).
- <sup>16</sup>E. Giner, A. Scemama, and M. Caffarel, *Can. J. Chem.* **91**, 879 (2013).
- <sup>17</sup>E. Giner, “Méthodes d’interaction de configurations et Monte Carlo quantique: Marier le meilleur des deux mondes (Configuration interaction and QMC: The best of both worlds),” Ph.D. thesis, University of Toulouse, 2014, <https://hal.archives-ouvertes.fr/tel-01077016>.
- <sup>18</sup>A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, *J. Chem. Phys.* **141**, 244110 (2014).
- <sup>19</sup>For very small systems, the Full CI solution can still be enough compact to be used as FN-DMC trial wavefunction. It has been done in the case of the  $Li_2$  molecule (cc-pVQZ basis set leading to 16 138 determinants), see Ref. 10.
- <sup>20</sup>B. Huron, P. Rancurel, and J. P. Malrieu, *J. Chem. Phys.* **58**, 5745 (1973).
- <sup>21</sup>S. Evangelisti, J. P. Daudey, and J. P. Malrieu, *Chem. Phys.* **75**, 91 (1983).
- <sup>22</sup>From a mathematical point of view, this result is far from being trivial since lowering the variational energy does not necessarily lower the fixed-node energy. For example, it is known that using CI expansions with arbitrary



- truncation may lead to a non-monotonic evolution of the FN energy as a function of the number of the determinants when the molecular orbitals and determinant coefficients are not re-optimized, see, e.g., Ref. 4.
- <sup>23</sup>A. Scemama, M. Caffarel, E. Oseret, and W. Jalby, *Lect. Notes Comput. Sci.* **7851**, 118–127 (2013).
- <sup>24</sup>A. Scemama, M. Caffarel, E. Oseret, and W. Jalby, *J. Comput. Chem.* **34**, 938 (2013).
- <sup>25</sup>M. D. Towler, *Quantum Monte Carlo, or, How to Solve the Many-Particle Schrödinger Equation Accurately Whilst Retaining Favourable Scaling with System Size* (Wiley, 2011).
- <sup>26</sup>M. Caffarel, “Quantum Monte Carlo methods in chemistry,” in *Encyclopedia of Applied and Computational Mathematics* (Springer, 2012).
- <sup>27</sup>F. A. Matsen, *Adv. Quantum Chem.* **1**, 59 (1964).
- <sup>28</sup>W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, 1992), p. 73.
- <sup>29</sup>C. F. Bender and E. R. Davidson, *Phys. Rev.* **183**, 23 (1969).
- <sup>30</sup>R. J. Buenker and S. D. Peyerimhof, *Theor. Chim. Acta* **35**, 33 (1974).
- <sup>31</sup>R. J. Buenker and S. D. Peyerimhof, *Theor. Chim. Acta* **39**, 217 (1975).
- <sup>32</sup>R. J. Buenker, S. D. Peyerimhof, and W. Butscher, *Mol. Phys.* **35**, 771 (1978).
- <sup>33</sup>P. J. Bruna, D. S. Peyerimhof, and R. J. Buenker, *Chem. Phys. Lett.* **72**, 278 (1980).
- <sup>34</sup>R. J. Buenker, S. D. Peyerimhof, and P. J. Bruna, *Computational Theoretical Organic Chemistry* (Reidel, Dordrecht, 1981).
- <sup>35</sup>R. J. Harrison, *J. Chem. Phys.* **94**, 5021 (1991).
- <sup>36</sup>P. S. Epstein, *Phys. Rev.* **28**, 695 (1926).
- <sup>37</sup>R. K. Nesbet, *Proc. R. Soc. A* **230**, 312 (1955).
- <sup>38</sup>C. Møller and M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).
- <sup>39</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>40</sup>L. Bytautas, T. Nagata, M. S. Gordon, and K. Ruedenberg, *J. Chem. Phys.* **127**, 164317 (2007).
- <sup>41</sup>D. Cleland, G. H. Booth, C. Overy, and A. Alavi, *J. Chem. Theory Comput.* **8**, 4138 (2012).
- <sup>42</sup>See supplementary material at <http://dx.doi.org/10.1063/1.4905528> where variational and CIPSI curves using the cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ basis sets are given.
- <sup>43</sup>L. Bytautas, N. Matsunaga, T. Nagata, M. S. Gordon, and K. Ruedenberg, *J. Chem. Phys.* **127**, 204301 (2007).
- <sup>44</sup>L. Bytautas and K. Ruedenberg, *J. Chem. Phys.* **130**, 204101 (2009).
- <sup>45</sup>E. R. Davidson, S. A. Hagstrom, S. J. Chakravorty, V. M. Umar, and C. F. Fischer, *Phys. Rev. A* **47**, 3649 (1993).
- <sup>46</sup>C. Angeli and M. Persico, *Theor. Chem. Acc.* **98**, 117 (1997); *Remote Instrumentation for eScience and Related Aspects*, edited by F. Davoli, M. Lawenda, N. Meyer, R. Pugliese, J. Woglarz, and S. Zappatore (World Scientific Publishing, 2012).