A new scheme for fixed node diffusion quantum Monte Carlo with pseudopotentials: Improving reproducibility and reducing the trial-wave-function bias FREE

Andrea Zen ; Jan Gerit Brandenburg ; Angelos Michaelides ; Dario Alfè

Check for updates

View Online

Export Citation

A new scheme for fixed node diffusion quantum Monte Carlo with pseudopotentials: Improving reproducibility and reducing the trial-wave-function bias

View Online   Export Citation   CrossMark

Andrea Zen,[1,2,3,4,a)] (ID) Jan Gerit Brandenburg,[1,5] (ID) Angelos Michaelides,[1,2,3,b)] (ID) and Dario Alfè[1,2,4,6,c)] (ID)

AFFILIATIONS

[1] Thomas Young Centre, University College London, London WC1H 0AH, United Kingdom
[2] London Centre for Nanotechnology, Gordon St., London WC1H 0AH, United Kingdom
[3] Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom
[4] Department of Earth Sciences, University College London, London WC1E 6BT, United Kingdom
[5] Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 205A, 69120 Heidelberg, Germany
[6] Dipartimento di Fisica Ettore Pancini, Università di Napoli Federico II, Monte S. Angelo, I-80126 Napoli, Italy

[a)] Electronic mail: a.zen@ucl.ac.uk
[b)] Electronic mail: angelos.michaelides@ucl.ac.uk
[c)] Electronic mail: d.alfe@ucl.ac.uk

ABSTRACT

Fixed node diffusion quantum Monte Carlo (FN-DMC) is an increasingly used computational approach for investigating the electronic structure of molecules, solids, and surfaces with controllable accuracy. It stands out among equally accurate electronic structure approaches for its favorable cubic scaling with system size, which often makes FN-DMC the only computationally affordable high-quality method in large condensed phase systems with more than 100 atoms. In such systems, FN-DMC deploys pseudopotentials (PPs) to substantially improve efficiency. In order to deal with nonlocal terms of PPs, the FN-DMC algorithm must use an additional approximation, leading to the so-called localization error. However, the two available approximations, the locality approximation (LA) and the T-move approximation (TM), have certain disadvantages and can make DMC calculations difficult to reproduce. Here, we introduce a third approach, called the determinant localization approximation (DLA). DLA eliminates reproducibility issues and systematically provides good quality results and stable simulations that are slightly more efficient than LA and TM. When calculating energy differences—such as interaction and ionization energies—DLA is also more accurate than the LA and TM approaches. We believe that DLA paves the way to the automation of FN-DMC and its much easier application in large systems.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5119729

I. INTRODUCTION

A wide range of scientific topics greatly benefit from computer simulations, such as crystal polymorph prediction, molecular adsorption on surfaces, the assessment of phase diagrams, phase transitions, nucleation, and more. The accuracy of the computational methods employed for such simulations is of fundamental importance. In a wide range of physical-chemical problems, many important static, dynamic, and thermodynamic properties are related to the potential energy surface. Thus, one of the grand challenges of computational modeling is the evaluation of accurate energetics for molecules, surfaces, and solids. This challenge is far

04 June 2024 14:39:52

from straightforward because various types and strengths of inter-atomic and intermolecular interactions are relevant, and a method must describe all of them correctly.

There are various computational methods acknowledged for having high accuracy. For condensed phase systems, one very promising methodology is quantum Monte Carlo (QMC),[1] often in the fixed-node (FN) diffusion Monte Carlo (DMC) flavor. FN-DMC has favorable scaling with system size (between the 3-rd and the 4-th power of the system size), and it can be efficiently deployed on high performance computer facilities. Nowadays, there is an increasing amount of benchmark data for solids and surfaces obtained via FN-DMC.[2–20] Data provided by DMC are of use in tackling interesting materials science problems and also to help the improvement of density functional theory (DFT)[21] and other cheaper computational approaches.

DMC implements a technique to project out the exact ground state wave function Φ from a trial wave function Ψ_T by performing a propagation according to the imaginary time-dependent Schrödinger equation.[1] However, an unconstrained projection leads to a bosonic wave function, so in fermionic systems, the fixed-node (FN) approximation is typically employed to keep the projected wave function antisymmetric. FN-DMC constrains the projected wave function Φ_{FN} to have the same nodes as a trial wave function Ψ_T.[22,99–105] Thus, Φ_{FN} is as close as possible to the exact (unknown) fermionic ground state Φ, given the nodal constraints, and the equality is reached if Ψ_T has exact nodes. In addition to the FN approximation, in most practical DMC simulations, pseudopotentials (PPs) are used because the core electrons in atoms significantly increase the computational cost of FN-DMC simulations.[23–25] There are also a few other technical aspects of FN-DMC that can affect its accuracy and efficiency, such as the actual implementation of the imaginary time-dependent Schrödinger propagation for a finite time step.[26–28] However, in general the most important and sizeable approximations in FN-DMC arise from the fixed-node constraint and the use of pseudopotentials.

The use of pseudopotentials in FN-DMC brings a twofold approximation. The first and trivial source of approximation is due to the fact that no pseudopotential is perfect. Pseudopotentials (PPs) can represent implicitly the influence of the core electrons only approximatively, and any method employing pseudopotentials will be affected by this issue. The second source of approximation is more subtle and tricky. PPs have nonlocal operators, i.e., terms $\hat{V}_{NL}$, such that their application on a generic function $\xi(\mathbf{R})$ of the electronic coordinates $\mathbf{R}$ gives $\hat{V}_{NL}\xi(\mathbf{R}) = \int \langle \mathbf{R}|\hat{V}_{NL}|\mathbf{R}'\rangle \xi(\mathbf{R}')d\mathbf{R}'$. The nonlocal PP operators are a big issue in FN-DMC simulations. Indeed, one of the quantities that should be evaluated in the FN-DMC projection is the value of the nonlocal terms applied to the projected wave function, $\hat{V}_{NL}\Phi_{FN}$, which cannot be calculated as we do not know the functional form of Φ_{FN}. There have been attempts to circumvent the difficulty, but the issue persists.[29] There has been no satisfactory way to deal with the nonlocal PP terms within FN-DMC exactly, and it is necessary to rely on some approximation. So far, there are two alternatives: the locality approximation[30] (LA) and the T-move approximation[29,31] (TM). In the former, the trial wave function Ψ_T is used to localize the nonlocal PP terms, so the unknown term $\Phi_{FN}^{-1}\hat{V}_{NL}\Phi_{FN}$ is approximated with $\Psi_T^{-1}\hat{V}_{NL}\Psi_T$. In the latter, only the terms in $\hat{V}_{NL}$ yielding a sign-problem are localized using Ψ_T. In both LA and TM, there is a *localization error*.[32] Both LA and

TM are used in production calculations, and it is unclear which is better.[33,106–109]

Both LA and TM have a big problem: reproducibility. As discussed above, the localization error arises from the projection of all or part of $\hat{V}_{NL}$ on Ψ_T. With either LA or TM, FN-DMC will produce different results with different Ψ_T, even if the nodes are unchanged. Unfortunately, Ψ_T has a level of arbitrariness in the way it can be defined because it is not straightforward to tell what is the optimal choice for Ψ_T. Ideally, we want a Ψ_T as close as possible to the exact ground state (which is unknown) and for which the ratio $\Psi_T(\mathbf{R})/\Psi_T(\mathbf{R}')$ is quickly evaluated computationally. In practice, Ψ_T can have many different functional forms, and different QMC packages often use different forms, exacerbating the reproducibility problem. Even within a specific implementation of Ψ_T, there are parameters to be set in some way, and there is no unique way to do this. Moreover, in large systems the number of parameters increases rapidly, leading to additional difficulties. For this reason, it is not uncommon to find differences in FN-DMC results from nominally similar studies. This contrasts with other electronic structure techniques where agreement between different studies is relatively straightforward to achieve. Resolving issues such as this are critical to reducing the labor intense nature of the DMC simulations and making DMC easier to use in general.

In this paper, we introduce a new approach to deal with the nonlocal potential terms in FN-DMC which resolves the reproducibility problem. We call this method the determinant localization approximation (DLA). In DLA, as explained below, the key point is to do the localization on just the determinant part of the trial wave function. The localization error in DLA is not eliminated, but it can be reproduced systematically across different implementations of Ψ_T and in different QMC packages. To this aim, the localization only uses the part of the trial wave function that can be obtained deterministically: the determinant part, which fixes the nodes. In this way, the reproducibility of DLA is guaranteed by construction.

It needs to be tested if DLA yields results competitive to LA and TM. It has to be noticed that with PPs, we are always interested in energy differences and not in absolute energies. So, the most accurate method is not necessarily the one with the smallest absolute localization error, but the method that makes consistently the same localization error across different configurations of the same system, such that there is the largest error cancellation in the energy difference. By construction, DLA makes very consistent localization errors. Indeed, we observe in all the representative cases considered in this paper that DLA always yields accurate results, which are systematically better than LA and TM whenever the Ψ_T is not optimal. Moreover, we notice that DLA produces very stable simulations, in contrast to LA. In terms of efficiency, DLA appears slightly more efficient than both LA and TM. All these features make DLA the best candidate to perform FN-DMC calculations in large systems, where the quality of Ψ_T could be hard to assess and a stable and efficient simulation is highly needed.

The outline of the paper is the following: we provide a short overview of the FN-DMC method in Sec. II; we describe our determinant localization approximation in Sec. III; we illustrate the results produced by DLA, compared with LA and TM, in Sec. IV, with a specific focus on interaction energy evaluations (Sec. IV A),

04 June 2024 14:39:52

ionization energies (Sec. IV B), stability (Sec. IV C), and efficiency (Sec. IV D). A reader already familiar with DMC can skip to Sec. III. We draw our final conclusions in Sec. V.

II. OVERVIEW ON FIXED NODE DIFFUSION MONTE CARLO

A. The trial wave function

The trial wave function has a critical role in determining the accuracy of FN-DMC. A QMC trial wave function is the product $\Psi_T(\mathbf{R}) = \mathscr{D}(\mathbf{R}) * \exp \mathscr{J}(\mathbf{R})$ of an antisymmetric function $\mathscr{D}(\mathbf{R})$ and a symmetric (bosonic) function $\exp \mathscr{J}(\mathbf{R})$, called the Jastrow factor, where $\mathbf{R}$ is the electronic configuration. The function $\mathscr{D}(\mathbf{R})$ is typically a single Slater determinant, especially when large systems are simulated. However, it is worth mentioning that if the system under consideration is not too large (say, generally not more than a few atoms) better functions can be used, such as multideterminant expansions of Slater determinants,[34–39] valence-bond wave functions,[40] the antisymmetrized geminal product,[41,42] the Pfaffian,[43] and others (see, for instance, the review by Austin *et al.*[44]). Moreover, the backflow transformation[45–47] can be employed to further improve any of the aforementioned ansätze, at the price of a significantly larger computational cost. The Jastrow factor describes the dynamical correlation between the electrons, by including explicit functions of the electron-electron distances. In DMC, a property of interest is the nodal surface of Ψ_T, which is the hypersurface corresponding to $\Psi_T(\mathbf{R}) = 0$, for real wave functions, or the complex phase of Ψ_T for complex wave functions. They are both determined by $\mathscr{D}(\mathbf{R})$ as the Jastrow factor can only alter the amplitude of Ψ_T. The Jastrow factor $\mathscr{J}$ is implemented differently in different QMC packages.

When large and complex systems are simulated, such as adsorption on surfaces or molecular crystals, the most common practice is to obtain $\mathscr{D}$ from a deterministic approach, usually DFT, and to decide a functional form for $\mathscr{J}$ and optimize, within the variational Monte Carlo (VMC) scheme,[1] the parameters minimizing either the energy or the variance. Since $\mathscr{D}$ comes from a deterministic method, there is no reproducibility problem here, and in taking energy differences we can usually expect a large cancellation of the FN error. On the other hand, $\mathscr{J}$ is optimized stochastically, so its parameters are affected by an optimization uncertainty. Dealing with this uncertainty becomes increasingly challenging as the system gets larger. Moreover, a new optimization of $\mathscr{J}$ is needed for every distinct orientation of the molecular systems, and optimizing $\mathscr{J}$ so frequently is tedious and time-consuming, and, due to the stochastic nature of the optimization procedure, can lead to Jastrow factors of different qualities, resulting in less than optimal cancellation of errors. A human supervision of the optimization is always highly recommended, if not necessary. The optimization is responsible for making QMC labor intense and nonautomatic.

B. Diffusion Monte Carlo

The DMC algorithm with importance sampling performs a time evolution of $f(\mathbf{R}, t) = \Psi_T(\mathbf{R})\psi(\mathbf{R}, t)$, where $\Psi_T(\mathbf{R})$ is a trial wave function (described in Sec. II A), $\mathbf{R}$ are the

$3N$-dimensional electronic coordinates, and $\psi(\mathbf{R}, t)$ is the solution at time t of the imaginary time Schödinger equation,

$$-\frac{\partial}{\partial t}\psi(\mathbf{R}, t) = (\hat{H} - E_T)\psi(\mathbf{R}, t), \qquad (1)$$

where $\hat{H}$ is the Hamiltonian and E_T is a trial energy, with initial condition $\psi(\mathbf{R}, 0) = \Psi_T(\mathbf{R})$ and converging exponentially to the exact ground state $\Phi(\mathbf{R})$ for $t \rightarrow \infty$. Thus, $\lim_{t \rightarrow \infty} f(\mathbf{R}, t) = \Psi_T(\mathbf{R})\Phi(\mathbf{R})$. Since Φ is an eigenstate for $\hat{H}$, the ground state energy E_0 can be calculated using the mixed estimator,

$$E_0 = \frac{\langle\Phi|\hat{H}|\Psi_T\rangle}{\langle\Phi|\Psi_T\rangle} = \frac{\int \Phi(\mathbf{R})\Psi_T(\mathbf{R})E_L(\mathbf{R})d\mathbf{R}}{\int \Phi(\mathbf{R})\Psi_T(\mathbf{R})d\mathbf{R}}, \qquad (2)$$

where $E_L(\mathbf{R}) = \frac{\langle\mathbf{R}|\hat{H}|\Psi_T\rangle}{\langle\mathbf{R}|\Psi_T\rangle}$ is the local energy in the electronic configuration $\mathbf{R}$ for the trial wave function Ψ_T.

The time evolution of $f(\mathbf{R}, t)$ follows from the imaginary time Schödinger equation (1), which in the integral form leads to

$$f(\mathbf{R}', t + \tau) = \int G(\mathbf{R}' \leftarrow \mathbf{R}, \tau)f(\mathbf{R}, t)d\mathbf{R}, \qquad (3)$$

where τ is the time step and $G(\mathbf{R}' \leftarrow \mathbf{R}, \tau)$ is the Green function for the importance sampling, which is defined (symbolically) as

$$G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \equiv \frac{\Psi_T(\mathbf{R}')}{\Psi_T(\mathbf{R})}\langle\mathbf{R}'|\exp(-\tau\hat{H})|\mathbf{R}\rangle. \qquad (4)$$

Thus, by starting from $f(\mathbf{R}, 0) = \Psi_T(\mathbf{R})^2$ and performing an evolution according to the Green function $G(\mathbf{R}' \leftarrow \mathbf{R}, t)$, we are able to assess expectation values of the exact ground state Φ,

$$\Phi(\mathbf{R}')\Psi_T(\mathbf{R}') = \lim_{t \rightarrow \infty} \int G(\mathbf{R}' \leftarrow \mathbf{R}, t)\Psi_T(\mathbf{R})^2 d\mathbf{R}. \qquad (5)$$

This is the process implemented in the DMC algorithm. In fermionic systems, the fixed-node (FN) approximation is typically introduced, so the FN Hamiltonian $\hat{H}_{FN} \equiv \hat{H} + \hat{V}_{FN}$, where $\hat{V}_{FN}$ is an infinite wall at the nodal surface of Ψ_T, is used. Further details are reported in Appendix A.

The Hamiltonian $\hat{H}$ is the sum of the kinetic and potential operators $\hat{K}$ and $\hat{V}$, respectively. In all-electron calculations, the potential operator $\hat{V}$ is local, $\hat{V} = \hat{V}_L$. However, in general, there is the need to deploy pseudopotentials to represent the core electrons of the atoms and reduce the computational cost of the calculation, see Appendix B. In this case, the potential term has both local and nonlocal operators: $\hat{V} = \hat{V}_L + \hat{V}_{NL}$. The presence of nonlocal operators in the potential complicates the formulation of the DMC algorithm and forces the introduction of a further approximation. In the following, we will first consider the simple case of a potential with only local operators, Sec. II C, and later we will consider the case of potential term with nonlocal operators, Sec. II D.

C. Green's function for $\hat{H} = \hat{K} + \hat{V}_L$

The simplest case is when the Hamiltonian has only a local potential term; thus, it can be written as $\hat{H} = \hat{K} + \hat{V}_L$, with

$\hat{K} = -\frac{1}{2}\nabla^2$. By substitution in the imaginary time Schrödinger equation [Eq. (1)], multiplication by $\Psi_T(\mathbf{R})$, and some algebraic operations, we obtain

$$\frac{\partial}{\partial t}f(\mathbf{R},t) = \frac{1}{2}\nabla^2 f(\mathbf{R},t) - \nabla \cdot (\mathbf{V}(\mathbf{R})f(\mathbf{R},t)) \\ - [E_L(\mathbf{R}) - E_T]f(\mathbf{R},t), \qquad (6)$$

where $\mathbf{V}(\mathbf{R}) = \nabla \log|\Psi_T(\mathbf{R})|$. Thus, the time evolution of $f(\mathbf{R},t)$ is given on the right-hand side (RHS) of Eq. (6). If the RHS only had the first two terms, we would have a pure drift-diffusion process, having a Green's function that for a small time step τ and for N electrons in the system can be approximated as $G_{DD}(\mathbf{R}' \leftarrow \mathbf{R},\tau) = \frac{1}{(2\pi\tau)^{3N/2}}\exp\left\{-\frac{[\mathbf{R}'-\mathbf{R}-\tau\mathbf{V}(\mathbf{R})]^2}{2\tau}\right\}$. The last term on the RHS of Eq. (6) is the branching term, and its associated Green's function is $G_B(\mathbf{R}' \leftarrow \mathbf{R},\tau) = \exp\left\{-\tau\frac{E_L(\mathbf{R}')+E_L(\mathbf{R})-2E_T}{2}\right\}$. Green's function of $f(\mathbf{R},t)$ for a small time interval τ can be approximated[48,49] as

$$G_{BDD}(\mathbf{R}' \leftarrow \mathbf{R},\tau) \approx G_B(\mathbf{R}' \leftarrow \mathbf{R},\tau)G_{DD}(\mathbf{R}' \leftarrow \mathbf{R},\tau), \qquad (7)$$

which is exact for $\tau \to 0$. $G_{BDD}(\mathbf{R}' \leftarrow \mathbf{R},\tau)$ can be used to approximate Green's function for an arbitrarily large time interval t.[50] G_{BDD} defines a branching-drift-diffusion process, as described, for instance, in Ref. 1. The algorithms implemented in QMC packages are usually a little more involved.[51,110,111] However, there is no need here to complicate further the picture. We will be concerned with the results of DMC in the continuous limit $\tau \to 0$. In this limit, the only bias in the DMC energy evaluation E_{FN} is given by the FN approximation. In particular, $E_{FN} \geq E_0$, with the equality reached if the nodes of Ψ_T are exact.

D. Green's function for $\hat{H} = \hat{K} + \hat{V}_L + \hat{V}_{NL}$

When pseudopotentials are used, the potential term has nonlocal operators $\hat{V}_{NL}$, and the Hamiltonian can be written as $\hat{H} = \hat{K} + \hat{V}_L + \hat{V}_{NL}$. If we consider the imaginary time Schrödinger equation (1) and substitute $\hat{H}$, we obtain the following time evolution of $f(\mathbf{R},t)$:

$$\frac{\partial}{\partial t}f(\mathbf{R},t) = \frac{1}{2}\nabla^2 f(\mathbf{R},t) - \nabla \cdot (\mathbf{V}(\mathbf{R})f(\mathbf{R},t)) \\ - \left[\frac{(\hat{K}+\hat{V}_L)\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} + \frac{\hat{V}_{NL}\psi(\mathbf{R},t)}{\psi(\mathbf{R},t)} - E_T\right]f(\mathbf{R},t). \quad (8)$$

The drift and diffusion terms on the RHS are identical to Eq. (6), but there is a complication in the branching term. Indeed, we cannot calculate $\frac{\hat{V}_{NL}\psi(\mathbf{R},t)}{\psi(\mathbf{R},t)}$, as we do not know the analytical form of $\psi(\mathbf{R},t)$.

There is an alternative approach, which is to write the Green's function $G(\mathbf{R}' \leftarrow \mathbf{R},\tau)$ for $\hat{H}$. Using the Zassenhaus formula, for small τ, we can approximate $e^{-\tau(\hat{K}+\hat{V}_L+\hat{V}_{NL})}$ with $e^{-\tau\hat{V}_{NL}}e^{-\tau(\hat{K}+\hat{V}_L)}$, and by substituting it into Eq. (4), we obtain

$$G(\mathbf{R}' \leftarrow \mathbf{R},\tau) \sim \int T_{NL}(\mathbf{R}' \leftarrow \tilde{\mathbf{R}},\tau) * G_L(\tilde{\mathbf{R}} \leftarrow \mathbf{R},\tau)d\tilde{\mathbf{R}}, \quad (9)$$

where $G_L(\mathbf{R}' \leftarrow \mathbf{R},\tau) \equiv \frac{\Psi_T(\mathbf{R}')}{\Psi_T(\mathbf{R})}\langle\mathbf{R}'|e^{-\tau(\hat{K}+\hat{V}_L)}|\mathbf{R}\rangle$ is Green's function for the local part of the Hamiltonian, which has been discussed in Sec. II C, and $T_{NL}(\mathbf{R}' \leftarrow \mathbf{R},\tau) \equiv \frac{\Psi_T(\mathbf{R}')}{\Psi_T(\mathbf{R})}\langle\mathbf{R}'|e^{-\tau\hat{V}_{NL}}|\mathbf{R}\rangle$ is Green's

function of the nonlocal part of the potential. For small τ, we have that $T_{NL}(\mathbf{R}' \leftarrow \mathbf{R},\tau) \sim \delta_{\mathbf{R}',\mathbf{R}} - \tau V_{\mathbf{R}',\mathbf{R}}$, where $\delta_{\mathbf{R}',\mathbf{R}}$ is the Dirac's delta and

$$V_{\mathbf{R}',\mathbf{R}} \equiv \frac{\Psi_T(\mathbf{R}')}{\Psi_T(\mathbf{R})}\langle\mathbf{R}'|\hat{V}_{NL}|\mathbf{R}\rangle. \qquad (10)$$

Notice that $V_{\mathbf{R}',\mathbf{R}}$ can be either positive or negative depending on Ψ_T, $\hat{V}_{NL}$, $\mathbf{R}$, and $\mathbf{R}'$. Whenever $V_{\mathbf{R}',\mathbf{R}} > 0$ for some $\mathbf{R}' \neq \mathbf{R}$, then $T_{NL}(\mathbf{R}' \leftarrow \mathbf{R},\tau) < 0$. The DMC algorithm needs to interpret the Green's function as a transition probability, but if $T_{NL}(\mathbf{R}' \leftarrow \mathbf{R},\tau) < 0$ for some $\mathbf{R}$ and $\mathbf{R}'$, it cannot be a transition probability from $\mathbf{R}$ to $\mathbf{R}'$ (sign problem). Thus, the presence of $\hat{V}_{NL}$ yields a sign problem in the DMC algorithm[29,31] because it gives a Green's function $G(\mathbf{R}' \leftarrow \mathbf{R},\tau)$ which can have negative terms.

There is no direct solution to this problem, and as a consequence, an approximation is introduced. As noted earlier, two approaches are available: either to use the locality approximation (LA)[30] or Casula's T-move approximation (TM).[29,31] They are summarized in Secs. II D 1 and II D 2.

1. Locality approximation in FN-DMC

The approach taken in LA is to approximate the unknown quantity $\frac{\hat{V}_{NL}\Phi(\mathbf{R},t)}{\Phi(\mathbf{R},t)}$ with $\frac{\hat{V}_{NL}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}$, which is the value of the nonlocal potential localized on the trial wave function $\Psi_T(\mathbf{R})$. By using this approximation in Eq. (8), we obtain that the 3rd term on the RHS is $-[E_L(\mathbf{R}) - E_T]f(\mathbf{R},t)$, and the equation becomes identical to Eq. (6). Thus, Green's function in LA is given by Eq. (7) and the DMC algorithm is a branching-drift-diffusion process.

The major difference from Sec. II C is that we approximate the Hamiltonian, which is no longer given by the FN Hamiltonian $\hat{H}_{FN} \equiv \hat{H} + \hat{V}_{FN}$, but by

$$\hat{H}_{FN}^{LA} \equiv \hat{K} + \hat{V}_L + \frac{\hat{V}_{NL}\Psi_T}{\Psi_T} + \hat{V}_{FN}, \qquad (11)$$

where the notation $\frac{\hat{V}_{NL}\Psi_T}{\Psi_T}$ is used to indicate that the nonlocal potential $\hat{V}_{NL}$ has been localized using the function $\Psi_T(\mathbf{R})$. So, given a generic function $\xi(\mathbf{R})$, we have $\frac{\hat{V}_{NL}\Psi_T}{\Psi_T}\xi(\mathbf{R}) = \int d\mathbf{R}'\Psi_T(\mathbf{R}')\langle\mathbf{R}'|\hat{V}_{NL}|\mathbf{R}\rangle\frac{\xi(\mathbf{R})}{\Psi_T(\mathbf{R})}$. Notice that $\hat{H}_{FN}^{LA}$ has no nonlocal potential term, i.e., the action of $\hat{H}_{FN}^{LA}$ on the generic function ξ at point $\mathbf{R}$ only depends on the value of ξ at $\mathbf{R}$.

The ground state for $\hat{H}_{FN}^{LA}$ is the projected wave function Φ_{FN}^{LA}. The expectation value of the energy E_{FN}^{LA} can be evaluated using the mixed estimator because $\frac{\hat{V}_{NL}\Psi_T}{\Psi_T}\Psi_T(\mathbf{R}) = \hat{V}_{NL}\Psi_T(\mathbf{R})$, so $\hat{H}_{FN}^{LA}|\Psi_T\rangle = \hat{H}|\Psi_T\rangle$. However, in general Φ_{FN}^{LA} is different from the (unknown) ground state Φ_{FN} for $\hat{H}_{FN}$; thus, $E_{FN}^{LA} \neq E_{FN}$. In other words, with LA we have lost the variationality of the approach because the error introduced by this approximation can be either positive or negative, and E_{FN}^{LA} is not, in general, an upper bound for E_0. Only in the (ideal) case of $\Psi_T = \Phi_{FN}$, we do have $\hat{H}_{FN}^{LA}|\Phi_{FN}\rangle = E_{FN}|\Phi_{FN}\rangle$, so $E_{FN}^{LA} = E_{FN}$. As a corollary, with the exact trial wave function, $\Psi_T = \Phi$, then we have that $E_{FN}^{LA} = E_0$. However, the trial wave function having exact nodes is not a sufficient condition for having $E_{FN}^{LA} = E_0$ as the LA depends on the overall trial wave function Ψ_T and not just on its nodes.[52] In other words, E_{FN}^{LA} has both a FN error and a

localization error. Mitas *et al.*[30] showed that the error (including both FN and localization) on the energy evaluation is quadratic in the wave function error, i.e., $E_{FN}^{LA} - E_0 = \mathcal{O}[(\Psi_T - \Phi)^2]$. Notice that FN and localization errors add to the time step error and are present also in the limit of zero time step.

2. T-move approximation in FN-DMC

In the T-move approach, the nonlocal Green's function T_{NL} includes only the terms without sign-problems and the remaining part of the nonlocal potential is localized in a similar manner to LA. To this aim, the positive, $V_{\mathbf{R'},\mathbf{R}}^+$, and negative, $V_{\mathbf{R'},\mathbf{R}}^-$ parts, $V_{\mathbf{R'},\mathbf{R}}^{\pm} = \frac{V_{\mathbf{R'},\mathbf{R}} \pm |V_{\mathbf{R'},\mathbf{R}}|}{2}$, of the term $V_{\mathbf{R'},\mathbf{R}}$ defined in Eq. (10) are used. The sign problem arises from terms $V_{\mathbf{R'},\mathbf{R}}^+$, which have to be localized, while the terms $V_{\mathbf{R'},\mathbf{R}}^-$ yield a T_{NL} with non-negative sign. Details about how this algorithm can be implemented are discussed in Refs. 29 and 31.

The corresponding T-move Hamiltonian is

$$\hat{H}_{FN}^{TM} \equiv \hat{K} + \hat{V}_L + \hat{V}_{NL}^- + \frac{\hat{V}_{NL}^+ \Psi_T}{\Psi_T} + \hat{V}_{FN}, \tag{12}$$

where the operators $\hat{V}_{NL}^+$ and $\hat{V}_{NL}^-$ correspond to $V_{\mathbf{R'},\mathbf{R}}^+$ and $V_{\mathbf{R'},\mathbf{R}}^-$, respectively. The projected wave function Ψ_{FN}^{TM} is the ground state of $\hat{H}_{FN}^{TM}$, and since $\hat{H}_{FN}^{TM}|\Psi_T\rangle = \hat{H}|\Psi_T\rangle$), the expectation value of the energy E_{FN}^{TM} can be evaluated using the mixed estimator. Similar to the LA approach, the projected function Φ_{FN}^{TM} is, in general, different from the fixed node ground state Φ_{FN}, but if the trial wave function $\Psi_T = \Phi_{FN}$, then $E_{FN}^{TM} = E_{FN}$. If $\Psi_T = \Phi$, then $E_{FN}^{TM} = E_0$, but if Ψ_T has exact nodes but differs from Φ, then $E_{FN}^{TM} \neq E_0$. Similar to LA, TM depends on the overall trial wave function Ψ_T.[53] Note that the FN and localization errors add to the time step error and are present also in the limit of zero time step.

The TM approach is computationally slightly more expensive than LA and often has a larger time step error. However, it has two advantages over LA: E_{FN}^{TM} is an upper bound of the exact ground state E_0,[54] and it is a more stable algorithm than the LA.

III. NEW APPROACH: DETERMINANT LOCALIZATION APPROXIMATION IN FN-DMC

The major practical disadvantage of both LA and TM is that the results are highly dependent on the Jastrow factor $\mathcal{J}$. This might result in problems of reproducibility, especially between results from different QMC packages, as the Jastrow factor is often expressed in different and nonequivalent functional forms across the codes. Moreover, the parameters of the Jastrow are affected by stochastic uncertainty. In contrast, it is much easier to control the reproducibility of the determinant part of the wave function $\mathcal{D}$, which is generally obtained from a deterministic method, e.g., DFT.

Therefore, we propose to use only the determinant part $\mathcal{D}$ of the trial wave function to localize the nonlocal potential $\hat{V}_{NL}$.[55] If we bear in mind that pseudopotentials are tested by developers using deterministic methods—density functional theory[56,57] or coupled cluster with single, double, and perturbative triple excitations [CCSD(T)][58–61]—our suggestion seems also quite reasonable because they are not tested widely and systematically in the presence of a Jastrow and within a DMC scheme.

In DLA, the FN Hamiltonian is

$$\hat{H}_{FN}^{DLA} \equiv \hat{K} + \hat{V}_L + \frac{\hat{V}_{NL}\mathcal{D}}{\mathcal{D}} + \hat{V}_{FN}, \tag{13}$$

and the associated projected wave function is Φ_{FN}^{DLA}. In order to be able to use the mixed estimator, we need to define the Hamiltonian

$$\hat{H}^{DLA} \equiv \hat{K} + \hat{V}_L + \frac{\hat{V}_{NL}\mathcal{D}}{\mathcal{D}} \tag{14}$$

such that $\hat{H}_{FN}^{DLA}|\Psi_T\rangle = \hat{H}^{DLA}|\Psi_T\rangle$ and

$$E_{FN}^{DLA} = \frac{\int \Phi_{FN}^{DLA}(\mathbf{R})\Psi_T(\mathbf{R})E_L^{DLA}(\mathbf{R})d\mathbf{R}}{\int \Phi_{FN}^{DLA}(\mathbf{R})\Psi_T(\mathbf{R})d\mathbf{R}}, \tag{15}$$

where the local energy is $E_L^{DLA}(\mathbf{R}) = \frac{\langle \mathbf{R}|\hat{H}_{FN}^{DLA}|\Psi_T\rangle}{\langle \mathbf{R}|\Psi_T\rangle}$. DLA becomes exact in the limit of $\mathcal{D} \to \Phi$, as $H_{FN}^{DLA}|\Phi\rangle = \hat{H}|\Phi\rangle$ and $E_{FN}^{DLA} = E_0$. It can be shown, along the lines of the argument of Mitas *et al.*,[30] that the error $E_{FN}^{DLA} - E_0$ on the energy evaluation (including both FN and localization) is $\mathcal{O}[(\mathcal{D} - \Phi)^2]$. Comparing with the corresponding error within LA, it implies than the DLA error on the absolute energy is typically expected to be larger than the LA error as $\Psi_T = \mathcal{D}\exp \mathcal{J}$ is typically closer than $\mathcal{D}$ to Φ. On the other hand, we suggest an alternative point of view: the DLA approach should be seen as a modification of the PPs. PPs introduce an approximation in the Hamiltonian, whereas the remaining part of $\hat{H}$ comes from first principles (see Appendix B). There is no proof showing that PPs provide a better approximation of the core if the localization is performed using a wave function with or without the Jastrow factor. However, without the Jastrow, there are clear advantages in terms of reproducibility of results and better error cancellation in energy differences, as discussed below. In other words, we are not concerned that E_{FN}^{DLA} might be further from E_0 than E_{FN}^{LA} (or E_{FN}^{TM}) as long as $\hat{V}_L + \frac{\hat{V}_{NL}\mathcal{D}}{\mathcal{D}}$ yields a good representation of the ionic potential energy.

Within DLA, the quality of the fixed node energy E_{FN}^{DLA} depends exclusively on $\mathcal{D}$. The Jastrow factor $\mathcal{J}$ does not affect the accuracy; the only influence of the Jastrow is on the efficiency, as it will affect the time step errors and the variance of H^{DLA}. In the limit of zero time step, all calculations which use the same $\mathcal{D}$ will provide the same energy, no matter what (if any) Jastrow factor is used.[62]

The implementation of DLA is straightforward as it is a simplification of the LA algorithm. It implies the numerical integration of $\mathcal{D}$ (instead of Ψ_T in LA) over a sphere to determine the nonlocal potential energy $\frac{\hat{V}_{NL}\mathcal{D}}{\mathcal{D}}$; see Eq. (C1) in Appendix C. The numerical integration scheme employs quadrature rules; hence, a number of wave function ratios on the integration grids are evaluated at every energy measurement.[30] While in the calculations reported in this manuscript, we used this simple implementation, it should be noticed that DLA allows a more involved but much more efficient implementation. Whenever $\mathcal{D}$ is used instead of Ψ_T, the integrals in Eq. (C1) can be factorized into simpler integrals involving the molecular orbitals defining $\mathcal{D}$.[24] Thus, all the nonlocal integrals can be done analytically (e.g., analytical expressions have been obtained by Hammond *et al.*[24] under the assumption that the basis

functions are gaussian type orbitals) or numerically, becoming a local potential which, for instance, can be precomputed on a grid at the beginning of a QMC simulation. This approach prevents the evaluation of many wave function ratios and possibly yields an appreciable speedup.

IV. RESULTS

In Secs. II and III, we have outlined that both LA and TM yield total energies, E_{FN}^{LA} and E_{FN}^{TM}, affected by the quality of the trial wave function $\Psi_T = \mathscr{D} * \exp \mathscr{J}$. Within the DLA scheme introduced here, the total energy E_{FN}^{DLA} is affected only by the determinant part $\mathscr{D}$ of the trial wave function Ψ_T. Therefore, DLA eliminates the uncertainty due to the Jastrow factor on the DMC results performed with pseudopotentials. We are going to show here, in a few examples, the amount of uncertainty that the Jastrow can introduce in LA and TM, in contrast to DLA which is not affected by this uncertainty.

### A. DLA is good for interaction energy evaluations

The first system that we considered is water bound to benzene, as shown in the inset of Fig. 1. This is a simple example of the calculation of an interaction energy $E_{int} \equiv E_{bound} - E_{far}$, which is the difference between the energy (E_{bound}) of the system in the bound configuration and the energy (E_{far}) of the molecules far away. Many of these calculations are performed to evaluate a binding energy curve, which are needed for example in adsorption energy calculations of molecules on surfaces.[3–6,19,20] Whereas in this small system, it is not overly burdensome to optimize $\mathscr{J}$ at every different geometry, in a larger and more complex adsorption system, this would be tedious and time-consuming, notwithstanding the variability of the quality of the optimization, due to its stochastic nature.

This specific water-benzene configuration has a reference E_{int} of -128 ± 1 meV,[3] as obtained from basis set converged CCSD(T)

calculations.[63] A standard setup for FN-DMC was used, with TN-DF pseudopotentials,[57] a Slater-Jastrow Ψ_T with determinant $\mathscr{D}$ obtained from a DFT calculation.[64] A Jastrow factor $\mathscr{J}$ having explicit electron-electron (e-e), electron-nucleus (e-n), and electron-electron-nucleus (e-e-n) terms was used here. Within this specific functional form of $\mathscr{J}$, we obtained two different Jastrow factors, which we call J.bound and J.far. The former, J.bound, is the Jastrow obtained when we optimize the parameters by minimizing the variational variance $\mathrm{Var}[E_L]_{VMC}$ of the local energy for the bound configuration. The latter, J.far, is instead obtained by optimizing the parameters on a configuration where the water and the benzene are far away (around 10 Å) and are effectively noninteracting. In Fig. 1, we compare the reference value with the FN-DMC evaluations obtained with LA, TM, and DLA, and the different Jastrow factors, whereas Fig. 2 reports the FN-DMC total energies.

The left panel of Fig. 1 shows results for J.far used for both the bound and the far-away configuration. With this setup, DLA is the only method that provides a reliable interaction energy, which we can estimate to be -131 ± 2 meV for the $\tau \to 0$ limit from a quadratic fit of the values obtained at finite values of τ.[65] The estimated $\tau \to 0$ limit for LA and TM is -187 ± 3 meV and -76 ± 1 meV.[66] So, with this nonoptimal Jastrow factor, LA severely overbinds and TM underbinds.[67]

A different choice, which is indeed the standard procedure adopted in DMC, is to optimize the Jastrow factor specifically for each configuration, i.e., we use J.bound for the bound configuration and J.far for the far configuration. We named this scheme J.mix, and the results obtained with LA, TM, and DLA are shown in the right panel of Fig. 1. In this case, all three methods are in decent agreement with the CCSD(T) reference, from a quadratic fit, we obtain the $\tau \to 0$ limit: -135 ± 3 meV for LA, -127 ± 1 meV for TM, and -129 ± 2 meV for DLA.[68] The figure also shows the time step error associated with the three different methods. The first consideration is that the better choice of the Jastrow has greatly improved the accuracy for any finite τ evaluation with respect to the case with J.far. The
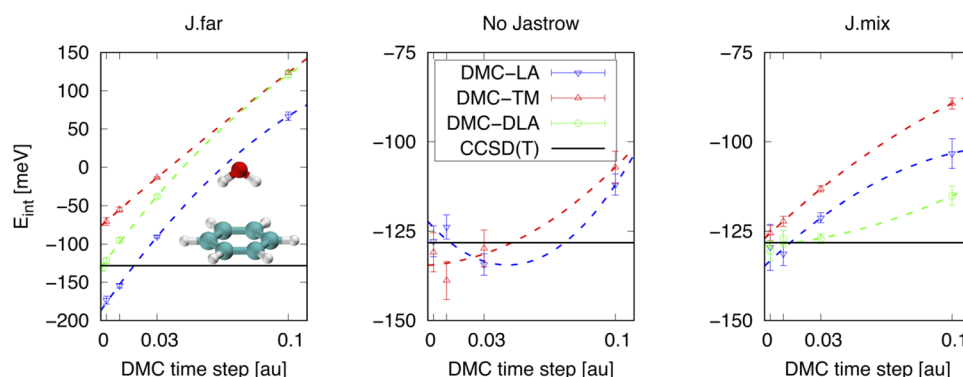
FIG. 1. Interaction energy E_{int} for a water-benzene complex in the so-called "2-leg" configuration, for the geometry shown in the inset. The plots show FN-DMC evaluations vs the time step τ, using LA, TM, and DLA. (Left plot) Results obtained using the Jastrow factor J.far in both the bound and the far configurations. J.far has been optimized for the far configuration, i.e., for noninteracting water-benzene molecules. (Middle plot) Results obtained using a trial wave function without any Jastrow factor. In this case, LA and DLA are equivalent. (Right plot) Results obtained with mixed Jastrow factor: for the noninteracting configuration, we used J.far, and for the bound configuration, we used J.bound. The reference CCSD(T) value is -128 ± 1 meV.[3]
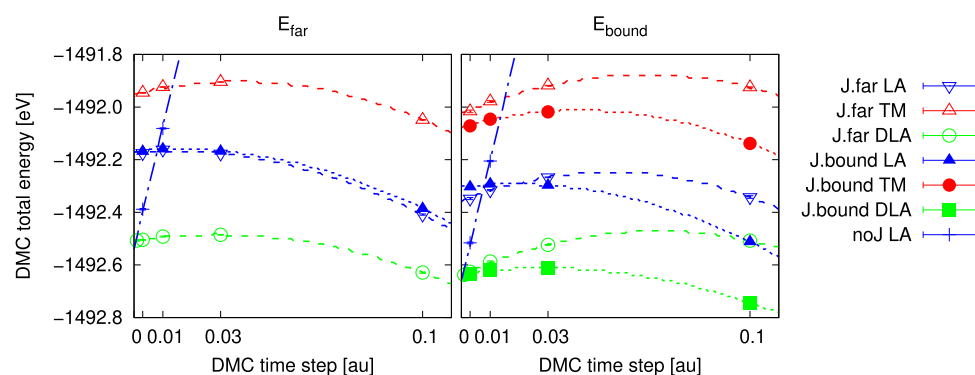
FIG. 2. Total energy for the far (E_{far}, left plot) and bound (E_{bound}, right plot) configurations. The plots show FN-DMC evaluations vs the time step τ, using LA, TM, and DLA, and different choices of the Jastrow factor: J.far, J.bound, and no Jastrow (noJ), as defined in the text. FN-DMC-TM noJ points fall outside the interval range [E_{far} = −1491.225(7) meV and E_{bound} = −1491.358(1) meV in the limit $\tau \to 0$].

best time step dependence is obtained for the DLA approach, where the interaction energy evaluation for τ = 0.03 a.u. is −127 ± 1 meV, which appears already converged (2 meV difference with respect to the $\tau \to 0$ limit).

At this point, one could wonder what is the outcome if the Jastrow is not used at all. We performed the FN-DMC simulation with LA and no Jastrow (so, it is equivalent to DLA), and the outcome is reported on the middle panel of Fig. 1. The $\tau \to 0$ limit of the interaction energy is −125 ± 4 meV, in excellent agreement with the other DMC-DLA evaluations with J.far and J.mix (as it has to be by construction) and also with the reference CCSD(T). Quite unexpectedly, we also notice that the time step error is quite small, much smaller than the case with J.far, and similar to the case with J.mix. This happens despite the huge time step error on the total energy evaluations (see Fig. 2) when a Jastrow is not used and indicates an unexpectedly good error cancellation of the finite step bias in the energy difference. We do not know if this behavior of the no Jastrow case is transferable to other systems. If it was, one would be tempted to do simulations without a Jastrow. However, this is not recommended because the variance of the local energy Var[E_L] is much larger without a Jastrow, around ten times the variance of the Slater-Jastrow Ψ_T. Since the computational cost is proportional to the variance, then a simulation with a given τ and a target precision will cost, computationally, around an order of magnitude more in the absence of a Jastrow. If this extra cost could be recovered by using time steps ten times as large is something that would have to be checked on each system.

B. Evaluation of ionization energies

Ionization energies (IEs) are typical quantities that are evaluated when new PPs are developed or tested.[56–61,69–74] The nth ionization energy, IE(n) is the energy necessary to remove one electron from an atomic (or molecular) species X and charge (n − 1), i.e., to have $X^{n-1} \to X^n + e^-$. Good PPs are expected to yield IEs estimations close to the corresponding all-electron (AE) evaluations. Typically, these checks are not performed at the QMC level, even for PPs specifically developed for QMC, but using Hartree-Fock, DFT,[75] or CCSD(T).[76] Here, we test the FN-DMC evaluations using LA, TM, and DLA.

We considered the first three ionization energies of the carbon atom, and we performed calculations with the energy consistent correlated electron pseudopotential (eCEPP),[60] which has a He-core for

the carbon atom. These pseudopotentials perform very well at the CCSD(T) level of theory. This can be seen in Table I, where the absolute difference in the IE between AE and eCEPP evaluations is <0.07 eV, and the relative difference is <0.5%.[77]

When PPs are deployed in FN-DMC, the difference from the AE results can be much larger than what is found for CCSD(T) because of the localization error. As usual, the localization error depends on the trial wave function. To have an idea of the magnitude of the localization error, we performed two sets of FN-DMC calculations, both with eCEPP and a Slater-Jastrow function with the Slater determinant obtained from a DFT calculation with local-density approximation (LDA) exchange–correlation functional.[78] The difference among the two sets is the Jastrow factor, which in one case (Ju) includes only e-e terms and in the other case (Juxf) includes e-e, e-n, and e-e-n terms. The parameters of the Jastrow factors have been optimized for each atomic ion considered. We have calculated the difference between the FN-DMC evaluations of IE(n) in two sets, for LA, TM, and DLA. Notice that the difference between the corresponding Juxf and the Ju wave functions is solely due to the localization error because the same $\mathscr{D}$ is used for Ju and Juxf, so the nodal surfaces are the same. Results are reported in Fig. 3. By construction, DLA is not affected by the Jastrow uncertainty, whereas the uncertainty on both TM and LA is larger than 0.2 eV. So, we see in this case that in LA and TM, the choice of the Jastrow yields a localization error that can be more than two times larger than the CCSD(T) error of the eCEPP pseudopotential. Given this large uncertainty in LA and TM, it is unclear what the most suitable Jastrow is. Juxf is variationally better than Ju (smaller VMC

TABLE I. First three ionization energies of the carbon atom, in eV. We give experimental results (Expt.) and compare CCSD(T) results based on all electron (AE) and PP (eCEPP[60]).

	Expt.	AE[a]	eCEPP[b]	$\Delta_{eCEPP−AE}$
IE(1)	11.26	11.26	11.22	−0.04
IE(2)	24.38	24.36	24.29	−0.07
IE(3)	47.89	47.86	47.90	0.04

[a]Performed with Orca,[48] using an aug-cc-pV(T,Q)Z basis set.
[b]Performed with Orca,[48] using the aug-cc-pV5Z-TN basis.[60] Differences with respect to aug-cc-pVQZ-TN are below 0.01 eV.
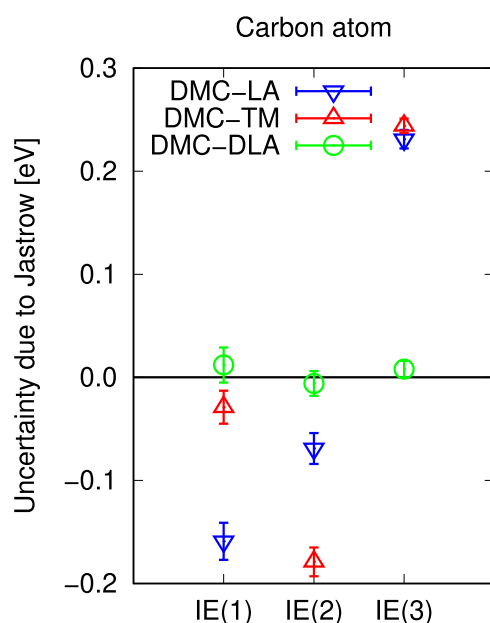
FIG. 3. Uncertainty of the FN-DMC estimation of the nth ionization energy IE(n) of the carbon atom due to the Jastrow, using an eCEPP pseudopotential[60] and either LA, TM, or DLA. Uncertainty is evaluated as the difference between the estimation using two different sets of Slater-Jastrow wave functions having the same $\mathscr{D}$ and different $\mathscr{J}$ (see text). Error bars correspond to a standard deviation due to the DMC sampling.

energy and variance). However, in Appendix D, we compare the difference between the PPs and the AE results, and it is unclear if Ju or Juxf is better.[79] DLA solves the issue because results from Ju and Juxf are equivalent.

In which way can we improve the DMC accuracy with DLA? It can be improved systematically by changing the determinant part, for instance using a multideterminant $\mathscr{D}$ term obtained using a method such as the complete active space self-consistent field (CASSCF),[35–37] the Configuration Interaction using a Perturbative Selection made Iteratively (CIPSI),[38,39] or the antisymmetrized geminal power.[41,42] DLA appears convenient in this case because we can anticipate improvements in the nodes and in the wave function, and we do not need to be concerned about the unpredictable effects of the Jastrow factor.

C. DLA yields stable DMC simulations

Stability is a very important practical aspect of DMC simulations. Instability in DMC is usually correlated with the quality of the trial wave function and of the pseudopotentials. In particular, a possible issue in the trial wave functions, which can generate instability, is the generation of the determinant part $\mathscr{D}$ of the trial wave function via plane-wave DFT packages using suboptimal Kleinman-Bylander projectors. In QMC (or in DFT calculations using localized basis), the nonlocal terms in the pseudopotential are evaluated using spherical harmonic projectors. In contrast, in plane-wave DFT calculations, the nonlocal terms in the pseudopotential are

evaluated using Kleinman-Bylander projectors[80] (i.e., projected into pseudo-atom wave functions). Thus, there is a possible inconsistency between the projection of the nonlocal terms at the DFT and DMC levels. The inconsistency is even worse if the Kleinman-Bylander projectors are not obtained with the same DFT functional used in the DFT preparation of $\mathscr{D}$. As a consequence, the method adopted to deal with the nonlocal part of the pseudopotential (LA, TM, or DLA) has a big impact on the stability of the DMC simulation. We have observed that DLA is not affected by the instability issues of LA and in all test simulations is as stable as TM and more stable than LA.

To illustrate this point, we consider carbon dioxide, CO_2, because in a previous study,[2] we noticed that the CO_2 molecule often leads to unstable DMC simulations. We consider here the case of the monomer and dimer of CO_2 (configurations taken from Ref. 2), and we use CEPP pseudopotentials[58] for both oxygen and carbon. We obtain $\mathscr{D}$ from a DFT/LDA calculation using the PWSCF code,[81] with a 600 Ry plane wave cutoff, and the molecular orbitals obtained were converted into splines.[82] The Kleinman-Bylander projectors used in DFT are from Ref. 58, and they were obtained from DFT with Perdew, Burke, and Ernzerhof (PBE) exchange–correlation functional so there is an inconsistency with the employed DFT functional.[83] In QMC, we used a Jastrow factor $\mathscr{J}$ with e-e, e-n, and e-e-n terms, and parameters were optimized by minimizing the local energy variance in VMC yielding $\text{Var}[E_L]_{\text{VMC}} = 0.956(2)$ a.u. in the molecule. The same identical trial wave function (i.e., no other optimization of $\mathscr{J}$) has a much smaller local energy variance in VMC if the DLA local energy E_L^{DLA} [as obtained from the Hamiltonian given in Eq. (14)] is used: $\text{Var}[E_L^{DLA}]_{\text{VMC}} = 0.685(5)$ a.u. Thus, the DLA Hamiltonian might already have advantages at the VMC level.

At the DMC level, LA simulations are not possible as population explosions happen so frequently that it was not possible to finish the DMC equilibration.[84] In contrast, both TM and DLA yield stable DMC simulations at all the attempted time steps (0.03, 0.01, and 0.003 a.u.). Moreover, the results of the DMC simulations seem reasonable despite the issue in the wave function. The most interesting quantity to consider is the interaction energy E_{int}. In this case, as the two carbon dioxide molecules are identical, the interaction can be evaluated as $E_{\text{int}} = E_{\text{dimer}} - 2 * E_{\text{monomer}}$, the difference between the energy of the dimer and twice the energy of the monomer. The FN-DMC results with TM and DLA are reported in Fig. 4. Both methods are in good agreement with the reference CCSD(T) evaluation[2] when we consider the infinitesimal time step limit. DMC-DLA is less affected by finite time step bias than DMC-TM. This is probably a consequence of the fact that the local energy variance in DLA is smaller than in TM.

D. Good efficiency for DLA

Efficiency is a fundamental property of a computational method. DLA appears to be more efficient than LA and TM. The efficiency of DMC can be estimated, as detailed in Appendix E. The choice among LA, TM, and DLA influences two quantities affecting the DMC efficiency: the computational cost T_{step} for a single DMC time step and the variance σ^2_{sys} of the local energy. The most efficient methods will have smaller T_{step} and σ^2_{sys}. DLA satisfies both the conditions. A detailed comparison of the efficiency of LA, TM, and DLA is provided in Appendix E. The outcome is that DLA is more
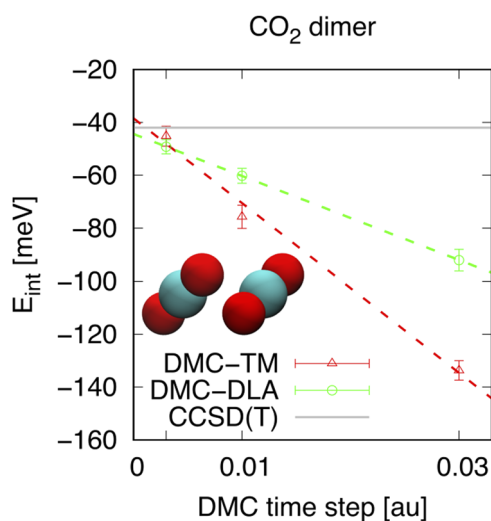
FIG. 4. Carbon dioxide dimer configuration (inset) and interaction energy E_{int} obtained from FN-DMC calculations and plotted as a function of the DMC time step τ. Dashed lines are obtained from a linear fit, and the extrapolated $\tau \to 0$ values are $-38(5)$ meV and $-44(1)$ meV for TM and DLA, respectively. The reference CCSD(T) value is -42 meV.[2] DMC-LA is not reported because the simulations are unstable.

efficient in most of the simulations involving organic molecules by roughly 30%.

V. CONCLUSION

In this paper, we have illustrated some drawbacks of FN-DMC in the presence of pseudopotentials. Specifically: (i) they generate unpredictable differences on results as a consequence of some arbitrary choices on the Jastrow functional form and the stochastic optimization procedure; (ii) they might affect the reproducibility of results if different QMC packages are used; (iii) the accuracy deteriorates whenever the Jastrow factor is not good, which is not easy to establish. These issues are particularly problematic for precisely where FN-DMC is most needed and offers most promise. We have shown that these issues arise essentially because the pseudopotentials have nonlocal terms. Within both LA and TM, the projection scheme is affected by a subtle interaction between the Jastrow factor and the pseudopotentials. In this paper, we have introduced a new alternative approximation, called the DLA. When FN-DMC deploys DLA, the projected wave function and the associated energy are not affected by the Jastrow factor. This solves the mentioned drawbacks. The advantages of DLA have been illustrated on a few examples, including the evaluation of an interaction energy and ionization energies. Moreover, the proposed algorithm appears as stable as TM and is much more stable than LA. In terms of efficiency, DLA performs better than both LA and TM.

An interesting perspective for DLA is that it allows the development of general purpose Jastrow factors, which do not need a system dependent optimization and yield a validated accuracy.[85,112–114] In this way, DLA opens the way to the automation of the FN-DMC, which can make DMC easier and less labor intensive to use.

The DLA method is already implemented in the CASINO,[86] TurboRVB,[87] and QMCPACK[88] packages.

APPENDIX A: DMC IMPLEMENTATION AND FIXED-NODE APPROXIMATION

In DMC, the non-negative and normalized function $f(\mathbf{R}, t)$ is interpreted as a probability density distribution, which is represented at each time t via a large number of electronic configurations $\mathbf{R}_i(t)$, also called walkers, and their associated weights $w_i(t)$. Walkers $\mathbf{R}_i(t)$ and weights $w_i(t)$ evolve in time according to a process that is ultimately determined by Green's function.[1] A key issue is that Green's function $G(\mathbf{R}' \leftarrow \mathbf{R}, t)$ in Eq. (4) does not impose any antisymmetry constraint to the system, so the ground state Φ obtained according to the imaginary time projection in Eq. (5) would be bosonic as it has a lower energy than the corresponding fermionic system. The traditional and most effective way to impose antisymmetry in Φ is to adopt the fixed node (FN) approximation: walkers are not allowed to cross the nodal surface of the trial wave function $\Psi_T(\mathbf{R})$. In other terms, the Hamiltonian $\hat{H}$ is replaced by the FN Hamiltonian $\hat{H}_{FN} \equiv \hat{H} + \hat{V}_{FN}$, where $\hat{V}_{FN}$ is an infinite wall at the nodal surface of Ψ_T. The projected function $\Phi_{FN}(\mathbf{R})$ obtained in this way has the same nodes as $\Psi_T(\mathbf{R})$, and it is the exact solution for the Hamiltonian $\hat{H}_{FN}$, namely, $\hat{H}_{FN}|\Phi_{FN}\rangle = E_{FN}|\Phi_{FN}\rangle$. This implies that $E_{FN} = \frac{\langle \Phi_{FN}|\hat{H}_{FN}|\Psi_T\rangle}{\langle \Phi_{FN}|\Psi_T\rangle}$, and noticing that $\hat{H}_{FN}|\Psi_T\rangle = \hat{H}|\Psi_T\rangle$ because $\hat{V}_{FN} \neq 0$ only in the nodal surface of Ψ_T, so we can evaluate E_{FN} using the mixed estimator,

$$E_{FN} = \frac{\int \Phi_{FN}(\mathbf{R})\Psi_T(\mathbf{R})E_L(\mathbf{R})d\mathbf{R}}{\int \Phi_{FN}(\mathbf{R})\Psi_T(\mathbf{R})d\mathbf{R}}. \quad (A1)$$

The FN energy E_{FN} is an upper bound to the exact energy E_0, namely, $E_{FN} \geq E_0$, because $\hat{H}_{FN} \geq \hat{H}$. The approach is exact if the nodes of $\Psi_T(\mathbf{R})$ are exact, because in this case, $\hat{H}_{FN}|\Phi_{FN}\rangle = \hat{H}|\Phi_{FN}\rangle$. Thus,

the quality of the FN approximation is determined by the quality of the nodes of the trial function $\Psi_T(\mathbf{R})$.

APPENDIX B: HAMILTONIAN FOR ALL-ELECTRONS VERSUS HAMILTONIAN FOR VALENCE ELECTRONS AND PSEUDOPOTENTIALS

The electronic Hamiltonian in Born-Oppenheimer approximation is

$$\hat{H} = -\frac{1}{2}\sum_i^N \nabla_i^2 + \sum_{i<j}^N \frac{1}{r_{ij}} - \sum_i^N \sum_\alpha^M \frac{Z_\alpha}{r_{i\alpha}}, \tag{B1}$$

where Roman letters are used to label the N electrons and Greek letters to label the M ions; $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between electrons i and j; $r_{i\alpha} = |\mathbf{r}_i - \mathbf{r}_\alpha|$ is the distance between electron i and ion α. The first term in the right-hand size is the kinetic energy $\hat{K}$, and the other two terms define a local potential $\hat{V}_L$.

The electrons can be separated into N_c core and N_v valence electrons such that $N = N_v + N_c$. Thus, Eq. (B1) can be recasted to

$$\hat{H} = \hat{H}_v + \hat{H}_c + \hat{H}_{cv}, \tag{B2}$$

$$\hat{H}_v = -\frac{1}{2}\sum_i^{N_v} \nabla_i^2 + \sum_{i<j}^{N_v} \frac{1}{r_{ij}} - \sum_i^{N_v}\sum_\alpha^M \frac{Z_\alpha}{r_{i\alpha}}, \tag{B3}$$

$$\hat{H}_c = -\frac{1}{2}\sum_i^{N_c} \nabla_i^2 + \sum_{i<j}^{N_c} \frac{1}{r_{ij}} - \sum_i^{N_c}\sum_\alpha^M \frac{Z_\alpha}{r_{i\alpha}}, \tag{B4}$$

$$\hat{H}_{cv} = +\sum_i^{N_c}\sum_j^{N_v} \frac{1}{r_{ij}}, \tag{B5}$$

where $\hat{H}_v$ and $\hat{H}_c$ are the components of the Hamiltonian involving only the valence electrons and the core electrons, respectively, and $\hat{H}_{cv}$ is the explicit pairing interaction between the core and the valence electrons.

All electron calculations in QMC are computationally very expensive, having a scaling roughly proportional to Z_α^6,[23–25] because close to the nucleus, the local energy has large fluctuations, yielding a large variance and thus requiring very small time steps. As many properties of interest are determined by the behavior of the valence electrons, it is often convenient to use pseudopotentials to represent the core electrons, especially in heavy nuclei. The Hamiltonian for calculations with pseudopotentials is then

$$\hat{H}_{\mathrm{PP}} = -\frac{1}{2}\sum_i^{N_v} \nabla_i^2 + \sum_{i<j}^{N_v} \frac{1}{r_{ij}} + \sum_i^{N_v}\sum_\alpha^M \hat{V}_{\mathrm{PP}}^\alpha(\mathbf{r}_i), \tag{B6}$$

where the first two terms in the RHS are the kinetic and electron-electron interactions, respectively, as expressed also in $\hat{H}_v$. The interaction between electron i and ion α is described as follows:

$$\hat{V}_{\mathrm{PP}}^\alpha(\mathbf{r}_i) = V_{\mathrm{loc}}^\alpha(r_{i\alpha}) + \sum_{l=0}^{l_{\max}} V_l^\alpha(r_{i\alpha})\hat{P}_l^\alpha, \tag{B7}$$

where $V_{\mathrm{loc}}^\alpha(r_{i\alpha})$ is the local part of the pseudopotential of ion α and $V_l^\alpha(r_{i\alpha})$ are the nonlocal components, which are applied via the projector,

$$\hat{P}_l^\alpha = \sum_{m=-l}^{+l} \left|Y_{l,m}^\alpha\right\rangle\left\langle Y_{l,m}^\alpha\right|,$$

where $\left|Y_{l,m}^\alpha\right\rangle$ are spherical harmonics centered on nucleus α. The idea behind $\hat{V}_{\mathrm{PP}}^\alpha(\mathbf{r}_i)$ is that it represents an effective potential that reproduces the effects of both the nucleus and the core electrons on the valence electrons. However, there is not an exact mapping, or a thermodynamic integration, providing the pseudopotentials. Indeed, some criteria need to be chosen to produce them, and they need to be tested at some level of theory. Moreover, it has to be noticed that in independent-electron theories, such as Hartree Fock (HF) and DFT, the separation of the electrons among core and valence can be in principle exact, while it cannot be exact in QMC or in other many-body approaches, because of the electronic correlation. Thus, although the employment of pseudopotentials in QMC is most of the times necessary for efficiency reasons, it can yield errors which cannot be easily quantified. Accurate pseudopotentials for QMC are so far available only for a fraction of the periodic table. The few cases where QMC proves less accurate than DFT are typically related with poor quality PPs used in QMC.[89] A crucial property of a PP is its transferability, which is affected by the range of the nonlocal potential terms (the smaller the better) and the inclusion of higher angular momentum channels.[90]

APPENDIX C: JASTROW INTERACTION WITH PSEUDOPOTENTIAL NONLOCAL TERM

In Secs. II D 1 and II D 2, we have shown that the nonlocal operators needs to be localized using the trial wave function Ψ_T such that the localized nonlocal operator $\frac{\hat{V}_{NL}\Psi_T}{\Psi_T}$ acts on a generic function $\xi(\mathbf{R})$ as follows:

$$\frac{\hat{V}_{NL}\Psi_T}{\Psi_T}\xi(\mathbf{R}) = \mathcal{V}_{NL}^{loc}(\mathbf{R})\xi(\mathbf{R}).$$

By using the addition theorem for spherical harmonics and Eqs. (B6) and (B7), it can be shown that $\mathcal{V}_{NL}^{loc}(\mathbf{R})$ can be evaluated as follows:

$$\mathcal{V}_{NL}^{loc}(\mathbf{R}) = \sum_i^{N_v}\sum_\alpha^M\sum_{l=0}^{l_{\max}} \frac{2l+1}{4\pi}V_l^\alpha(r_{i\alpha})\int P_l(\cos\theta_{i\alpha}')$$
$$\times \frac{\Psi_T(\mathbf{r}_1,\ldots,\mathbf{r}_{i-1},\mathbf{r}_i',\mathbf{r}_{i+1}\ldots,\mathbf{r}_{N_v})}{\Psi_T(\mathbf{r}_1,\ldots,\mathbf{r}_{N_v})}d\Omega_{i\alpha}', \tag{C1}$$

where the angular integration is over the sphere passing through the ith electron and centered on the αth atom, P_l is the Legendre polynomial of degree l, and $\cos\theta_{i\alpha}' = \frac{\mathbf{r}_{i\alpha}'\cdot\mathbf{r}_{i\alpha}}{r_{i\alpha}'r_{i\alpha}}$.

Therefore, in the evaluation of $\mathcal{V}_{NL}^{loc}(\mathbf{R})$, we have to integrate the ratio between the trial wave functions for electrons displaced along spheres centered on every pseudo-atom. By considering that $\Psi_T = \mathcal{D}*\exp\mathcal{J}$, it is clear that this ratio has to be taken for both the determinant part $\mathcal{D}$ and the Jastrow factor $\exp\mathcal{J}$. This expression explains why both LA and TM yield to total energies that depend on the Jastrow factor: $\mathcal{J}$ effectively changes the potential term in the effective Hamiltonian, due to the localization of the nonlocal term.

We could wonder what parts of the Jastrow factor give rise to this issue, and if it is possible to use a Jastrow not affecting the ratio in the RHS of Eq. (C1). Although there are differences in the Jastrow factor implemented in the different codes, they typically can be expressed in the following way:

$$\mathscr{J}(\mathbf{R}) = \sum_{i<j} \delta_{\sigma_i,\sigma_j} u_{\uparrow\uparrow}(r_{ij}) + \sum_{i<j} (1 - \delta_{\sigma_i,\sigma_j}) u_{\uparrow\downarrow}(r_{ij})$$
$$+ \sum_{\alpha} \sum_{i} \chi_{\alpha}(r_{i\alpha}) + \sum_{\alpha} \sum_{i<j} f_{\alpha}(r_{i\alpha}, r_{j\alpha}, r_{ij}), \quad (C2)$$

where $u_{\uparrow\uparrow}$ and $u_{\uparrow\downarrow}$ are homogeneous two-electron correlation terms, or e-e, describing the interaction between like-spin and unlike-spin pairs, χ_{α} is a e-n term depending on the distance of any electron from the αth atom, and f_{α} is a three-body term, or e-e-n, describing the interaction between electron pairs in proximity of the αth atom. Different packages can provide different parameterization of these terms, and sometimes additional terms.[91,115,116] The important aspect is that in the displacement $\mathbf{r}_i \rightarrow \mathbf{r}'_i$ performed in the angular integration in Eq. (C1), we have that $r_{i\alpha} \rightarrow r'_{i\alpha} = r_{i\alpha}$, but $r_{ij} \rightarrow r'_{ij} \neq r_{ij}$. Thus, the e-n term χ_{α} is not responsible for the Jastrow dependence of $\mathscr{V}^{loc}_{NL}(\mathbf{R})$, while the e-e and e-e-n terms are at the source of the issue. One could in principle decide not to use the three-body term f_{α}, but the same cannot be done for the terms $u_{\uparrow\uparrow}$ and $u_{\uparrow\downarrow}$ as they are the most important in the Jastrow factor, responsible for most of the correlation captured by Ψ_T and the decrease in the variance of the local energy.[92]

APPENDIX D: HOW LARGE ARE ERRORS DUE TO PPs IN IE EVALUATIONS?

Ideally, a perfect PP delivers exactly the same energy differences as AE calculations. In FN-DMC, the situation is complicated by the fact that localization errors appear when PPs are used. In Sec. IV B, it is shown that these localization errors can be large and that with LA and TM, there is a big dependence on the Jastrow factor in the trial wave function. For instance, it is not clear if both the wave functions Ju and Juxf discussed in Sec. IV B show a sizable localization error, or if instead one wave function is responsible for most of the error.

In order to investigate this, we have considered the difference $\Delta_{eCEPP-AE}$ between the eCEPP and the AE results, the IE(n) evaluated with LA, TM, and DLA. In the AE calculation, we have used the same level of theory to generate the wave function, so a Slater-Jastrow function with the determinant from DFT/LDA.[93] Results are reported in Fig. 5. The plot highlights some unexpected behavior. We would have expected that the Juxf yields better results than Ju because Juxf has more parameters and include Ju as a special case, and indeed, it yields a lower variational energy and variance than Ju. However, Juxf does not show a smaller $\Delta_{eCEPP-AE}$ compared to Ju in any of the IEs. In fact, the worse wave function, Ju, generates the smallest $\Delta_{eCEPP-AE}$ in LA and TM. On the other hand, DLA yields $\Delta_{eCEPP-AE}$ quite similar in absolute value to the best LA or TM case, but in DLA, the sign is inverted. We notice a clear correlation between DLA and HF errors. This correlation suggests that the $\Delta_{eCEPP-AE}$ in DLA might just reflect the limitations of the determinant part of the wave function.

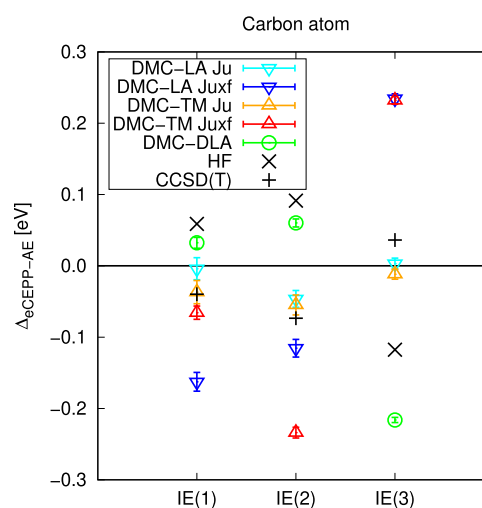**FIG. 5**. Difference $\Delta_{eCEPP-AE}$ between the IE(n) evaluation with eCEPP PPs and with all-electrons (AE). FN-DMC with PPs yields different results with different Ψ_T, and we consider two sets of Slater-Jastrow functions, dubbed Ju and Juxf (see text). In all reported DMC calculations, $\mathscr{D}$ is a single Slater determinant obtained from DFT/LDA. Results using LA, TM, and DLA are shown. As a comparison, outcomes from Hartree Fock (HF) and CCSD(T) are also reported.

APPENDIX E: ESTIMATION OF COMPUTATIONAL COST OF A FN-DMC SIMULATION WITH LA, TM, AND DLA

The efficiency of a DMC calculation depends on the computational resources required to achieve a target stochastic error σ_{target} on the quantity of interest. In most cases, we are interested in the energy, and in this case, the computational time spent T_{cost} is

$$T_{cost} = T_{step} * \frac{t_{ac}}{\tau} \left(\frac{t_{ac}}{t_{eq}} N_w + \frac{\sigma^2_{sys}}{\sigma^2_{target}} \right), \quad (E1)$$

where T_{step} is the computational time of a single DMC step (which clearly depends on the specific architecture where the calculation is performed), t_{ac} and t_{eq} are the autocorrelation and equilibration times (which are roughly the same and typically of the order of 1 a.u.), N_w is the number of walkers, and σ^2_{sys} is the variance of the local energy in the corresponding DMC scheme.[94] The first term in the parentheses is due to the equilibration time in DMC, which has to be removed from the sampling, and which typically is negligible compared to the second term in the parentheses, which instead comes from the statistical sampling. The quantities on the RHS of Eq. (E1) that are affected by the choice of LA, TM, or DLA are only T_{step} and σ^2_{sys}. In particular, σ^2_{sys} is $\text{Var}[E_L]_{DMC-LA}$, $\text{Var}[E_L]_{DMC-TM}$, and $\text{Var}[E_L^{DLA}]_{DMC-DLA}$ for LA, TM, and DLA, respectively.

The dependence on T_{step} is straightforward: if we name T_{step}^{LA}, T_{step}^{TM}, and T_{step}^{DLA} the cost per step in the LA, TM, and DLA approaches, respectively, then

$$T_{step}^{DLA} < T_{step}^{LA} < T_{step}^{TM}.$$
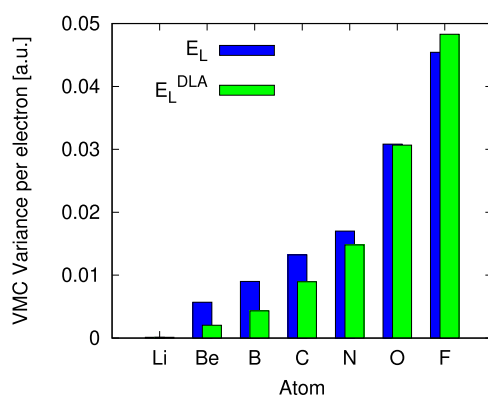
Indeed, in TM, the operations are just the same performed in LA, plus those to perform the moves connected to Green's function $T_{NL}(\mathbf{R}' \leftarrow \mathbf{R})$ (T-moves). On the other hand, in DLA, there are fewer operations than in LA because the Jastrow factor does not need to be considered when evaluating the nonlocal part of the pseudopotential (see details in Appendix C). However, the observed difference in the costs of LA, TM, and DLA is relatively small (just a few percent).

More important for the efficiency is the dependence on the variance σ_{sys}^2. First of all, we have to notice that here the variance under consideration is the one relative to the corresponding DMC sampling, which is $\text{Var}[E_L]_{\text{DMC-LA}}$, $\text{Var}[E_L]_{\text{DMC-TM}}$, and $\text{Var}[E_L^{DLA}]_{\text{DMC-DLA}}$ for LA, TM, and DLA, respectively. The LA and TM approaches are strictly related to the variational variance $\text{Var}[E_L]_{\text{VMC}}$; the local energy is calculated in the same way, but the underlying probability distribution is different, and in particular, different values of the time step change the sampling and the corresponding variance. In DMC-LA with large τ, the variance becomes the same as the variational variance (this is likely a consequence of the Metropolis step to enforce detailed balance after the drift-diffusion step), while at small τ, we notice that the DMC-LA variance is typically slightly smaller than the VMC variance. In DMC-TM, the variance converges to the same variance of DMC-LA for small τ, but for large τ, it is larger than the VMC and the DMC-LA variance.[95] In DMC-DLA, the variance has roughly the same relation with the variational DLA variance that the DMC-LA variance has with the VMC variance. Thus, the difference in the variance between DMC-LA and DMC-DLA is mostly captured by the difference in the corresponding variational variances, $\text{Var}[E_L]_{\text{VMC}}$ and $\text{Var}[E_L^{DLA}]_{\text{VMC}}$. Notice that the VMC sampling is precisely the same if Ψ_T is the same, and the difference only comes from the Hamiltonian, thus the local energy.

In order to investigate this difference, in Fig. 6, we report the variational variances ($\text{Var}[E_L]_{\text{VMC}}$ and $\text{Var}[E_L^{DLA}]_{\text{VMC}}$) per electron on the first-row atoms, using eCEPP pseudopotentials,[60] which are He-core for all the reported atoms. The $\mathscr{D}$ in Ψ_T was obtained

TABLE II. Variational variance, in a.u., $\text{Var}[E_L]_{\text{VMC}}$ and $\text{Var}[E_L^{DLA}]_{\text{VMC}}$ for the boron atom, simulated using He-core pseudopotentials, a single Slater determinant obtained from a DFT/LDA calculation and different parameterizations of the Jastrow factor $\mathscr{J}$, each with parameters optimized to minimize the variance. For the eCEPP pseudopotentials,[60] we used the aug-cc-pVQZ-eCEPP basis provided by Trail and Needs,[60] while for the ccECP pseudopotentials,[61] we used the QZ basis provided by Bennett *et al.*[61]

Pseudo	Jastrow	$\text{Var}[E_L]_{\text{VMC}}$	$\text{Var}[E_L^{DLA}]_{\text{VMC}}$
eCEPP	no	0.181(3)	0.181(3)
eCEPP	e-e	0.0401(1)	0.0317(1)
eCEPP	e-e, e-n	0.0285(3)	0.0135(1)
eCEPP	e-e, e-n, e-e-n	0.0278(2)	0.0133(1)
ccECP	e-e, e-n, e-e-n	0.0355(4)	0.0166(1)

from a DFT/LDA calculation performed with the ORCA package[96] and with the localized basis set aug-cc-pVQZ-eCEPP provided in Refs. 60 and 97. The Jastrow factor used has e-e, e-n, and e-e-n terms, with parameters optimized in any atom to minimize the variational variance.[98] Figure 6 shows that $\text{Var}[E_L^{DLA}]_{\text{VMC}}$ is smaller than $\text{Var}[E_L]_{\text{VMC}}$ for most elements with the exception of the fluorine atom where the DLA variance is +6% larger. In particular, in the carbon atom, the variance is −32% smaller, in the nitrogen atom, it is −13%, and in the oxygen atom, it is −0.5%.

A natural question at this point is which part of $\mathscr{J}$ is mostly involved in producing a difference between E_L and E_L^{DLA}? In Appendix C, we show that the e-n term drops out when we evaluate the nonlocal potential term; thus, it produces no difference between E_L and E_L^{DLA}. The terms to consider are thus the e-e and the e-e-n. In Table II, we report the variational variances on the boron atom with different parameterizations of the $\mathscr{J}$. It shows than the e-e term produces most of the difference. Moreover, one could wonder how much this difference is affected by the choice of the pseudopotentials. In Table II, we provide the variance also for the ccECP pseudopotential.[61] We notice that, although in absolute terms, eCEPP and ccECP yield different variances, in relative terms, the difference between E_L and E_L^{DLA} is the same.

The lower variance automatically translates into a smaller stochastic error in the DMC evaluations with same sampling. Thus, DLA is more efficient in most of the simulations involving organic molecules because they are characterized by the presence of many carbon atoms, where DLA is roughly 30% faster than LA and even more compared to TM (where it crucially depends on the time step value).

REFERENCES

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

[2] A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides, Proc. Natl. Acad. Sci. U. S. A. **115**, 1724 (2018).

[3] J. G. Brandenburg, A. Zen, M. Fitzner, B. Ramberger, G. Kresse, T. Tsatsoulis, A. Grüneis, A. Michaelides, and D. Alfè, J. Phys. Chem. Lett. **10**, 358 (2019).

[4] A. Zen, L. M. Roch, S. J. Cox, X. L. Hu, S. Sorella, D. Alfè, and A. Michaelides, J. Phys. Chem. C **120**, 26402 (2016).

FIG. 6. Local energy variances in VMC, $\text{Var}[E_L]_{\text{VMC}}$ in blue and $\text{Var}[E_L^{DLA}]_{\text{VMC}}$ in green, per valence electron on the first-row atoms, using eCEPP pseudopotentials[60] (He-core), a Jastrow factor with e-e, e-n, and e-e-n terms, and a single Slater determinant obtained from a DFT/LDA calculation with the aug-cc-pVQZ-eCEPP basis.[60]

04 June 2024 14:39:52

[5] Y. S. Al-Hamdani, M. Rossi, D. Alfè, T. Tsatsoulis, B. Ramberger, J. G. Brandenburg, A. Zen, G. Kresse, A. Grüneis, A. Tkatchenko, and A. Michaelides, J. Chem. Phys. **147**, 044710 (2017).

[6] Y. S. Al-Hamdani, D. Alfè, and A. Michaelides, J. Chem. Phys. **146**, 094701 (2017).

[7] Y. S. Al-Hamdani, M. Ma, D. Alfè, O. A. von Lilienfeld, and A. Michaelides, J. Chem. Phys. **142**, 181101 (2015).

[8] J. Trail, B. Monserrat, P. López Ríos, R. Maezono, and R. J. Needs, Phys. Rev. B **95**, 121108 (2017).

[9] Y. Luo, A. Benali, L. Shulenburger, J. T. Krogel, O. Heinonen, and P. R. C. Kent, New J. Phys. **18**, 113049 (2016).

[10] M. Dubecký, L. Mitas, and P. Jurecka, Chem. Rev. **116**, 5188 (2016).

[11] N. Devaux, M. Casula, F. Decremps, and S. Sorella, Phys. Rev. B **91**, 081101 (2015).

[12] Y. Wu, L. K. Wagner, and N. R. Aluru, J. Chem. Phys. **142**, 234702 (2015).

[13] H. Zheng and L. K. Wagner, Phys. Rev. Lett. **114**, 176401 (2015).

[14] L. K. Wagner and P. Abbamonte, Phys. Rev. B **90**, 125129 (2014).

[15] S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, Phys. Rev. Lett. **112**, 165501 (2014).

[16] A. Benali, L. Shulenburger, J. T. Krogel, X. Zhong, P. R. C. Kent, and O. Heinonen, Phys. Chem. Chem. Phys. **18**, 18323 (2016).

[17] H. Shin, Y. Luo, P. Ganesh, J. Balachandran, J. T. Krogel, P. R. C. Kent, A. Benali, and O. Heinonen, Phys. Rev. Mater. **1**, 073603 (2017).

[18] J. Ahn, I. Hong, Y. Kwon, R. C. Clay, L. Shulenburger, H. Shin, and A. Benali, Phys. Rev. B **98**, 085429 (2018).

[19] T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alf, A. Michaelides, and A. Grüneis, J. Chem. Phys. **146**, 204108 (2017).

[20] K. Doblhoff-Dier, J. Meyer, P. E. Hoggan, and G.-J. Kroes, J. Chem. Theory Comput. **13**, 3208 (2017).

[21] R. J. Maurer, C. Freysoldt, A. M. Reilly, J. G. Brandenburg, O. T. Hofmann, T. Björkman, S. Lebègue, and A. Tkatchenko, Annu. Rev. Mater. Res. **49**, 1 (2019).

[22] Alternative strategies to the FN approximation have been developed,[99–105] which are typically more accurate but much more demanding in terms of computational cost. Thus, typically only FN-DMC is affordable in large systems.

[23] D. M. Ceperley, J. Stat. Phys. **43**, 815 (1986).

[24] B. L. Hammond, P. J. Reynolds, and W. A. Lester, J. Chem. Phys. **87**, 1130 (1987).

[25] A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Phys. Rev. E **71**, 066704 (2005).

[26] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè, Phys. Rev. B **93**, 241118 (2016).

[27] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).

[28] M. F. DePasquale, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **89**, 3629 (1988).

[29] M. Casula, Phys. Rev. B **74**, 161102 (2006).

[30] L. Mitas, E. L. Shirley, and D. M. Ceperley, J. Chem. Phys. **95**, 3467 (1991).

[31] M. Casula, S. Moroni, S. Sorella, and C. Filippi, J. Chem. Phys. **132**, 154113 (2010).

[32] The localization error goes to zero (ideally) as we can find a Ψ_T closer to the (unknown) fixed-node solution Φ_{FN}, both in LA and TM.

[33] TM yields more stable simulations than LA.[29,106] However, if the trial wave function Ψ_T is good, LA has typically no stability issues. In these cases, TM is often affected by larger finite time step and stochastic errors than LA.[107] Moreover, LA satisfies a detailed balance condition, while TM does not, and the size-consistency condition can be harder to satisfy in TM (indeed, the first version of the TM algorithm[29] was size-inconsistent, an issue solved four years later with two alternative revised TM algorithms[31]). Recent investigations[108,109] of the localization errors in LA and TM do not show that one method is clearly superior to the other in terms of accuracy.

[34] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996).

[35] O. Valsson and C. Filippi, J. Chem. Theory Comput. **6**, 1275 (2010).

[36] P. M. Zimmerman, J. Toulouse, Z. Zhang, C. B. Musgrave, and C. J. Umrigar, J. Chem. Phys. **131**, 124103 (2009).

[37] J. Toulouse and C. J. Umrigar, J. Chem. Phys. **128**, 174101 (2008).

[38] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).

[39] A. Scemama, Y. Garniron, M. Caffarel, and P.-F. Loos, J. Chem. Theory Comput. **14**, 1395 (2018).

[40] B. Braida, J. Toulouse, M. Caffarel, and C. J. Umrigar, J. Chem. Phys. **134**, 084108 (2011).

[41] M. Casula and S. Sorella, J. Chem. Phys. **119**, 6500 (2003).

[42] A. Zen, E. Coccia, Y. Luo, S. Sorella, and L. Guidoni, J. Chem. Theory Comput. **10**, 1048 (2014).

[43] M. Bajdich, L. Mitas, G. Drobny, L. Wagner, and K. Schmidt, Phys. Rev. Lett. **96**, 130201 (2006).

[44] B. M. Austin, D. Y. Zubarev, and W. A. J. Lester, Chem. Rev. **112**, 263 (2012).

[45] R. Feynman and M. Cohen, Phys. Rev. **102**, 1189 (1956).

[46] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Phys. Rev. E **74**, 066701 (2006).

[47] M. Holzmann and S. Moroni, Phys. Rev. B **99**, 085121 (2019).

[48] B. L. Hammond, W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, 1994), https://www.worldscientific.com/doi/pdf/10.1142/1170.

[49] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, 1st ed. (Wiley, 2009).

[50] Given the values of t and τ, we can approximate $G(\mathbf{R}' \leftarrow \mathbf{R}, t) \approx \int \prod_{i=1}^{n} G(\mathbf{R}_i \leftarrow \mathbf{R}_{i-1}, \tau) d\mathbf{R}_{n-1} \ldots d\mathbf{R}_1$, where $n = t/\tau$, $\mathbf{R}' = \mathbf{R}_n$, and $\mathbf{R} = \mathbf{R}_0$.

[51] For instance, it has been observed that the time step error is largely decreased if a Metropolis step is included in order to enforce detailed balance[110,111] and that it is convenient to reformulate the algorithm in order to implement electron-by-electron updates instead of configuration-by-configuration updates as it improves the efficiency in large systems.[110,111] Moreover, in the proximity of the nodes of $\Psi_T(\mathbf{R})$, the local energy $E_L(\mathbf{R})$ and the drift vector $\mathbf{V}(\mathbf{R})$ diverge, yielding instabilities in the branching and drifting terms. This issue is substantially reduced by considering modified versions of the branching and drifting terms close to the nodes.[26,27] The Metropolis step, the modification to the branching and drifting terms, and the electron-by-electron DMC algorithm do affect the simulations for finite value of τ but do not change the limit for $\tau \to 0$. They are very important technical aspects as they enhance the stability and efficiency of the algorithm. For instance, the development of a size consistent version of the modification to the branching term, as discussed in Ref. 26, yields a speedup of up to two orders of magnitude in the evaluations of binding and cohesive energies.[2,26]

[52] If Ψ_{T1} and Ψ_{T2} are two different trial wave functions with the same nodes [e.g., having the same determinant $\mathcal{D}(\mathbf{R})$ but different Jastrow factors], the corresponding FN-DMC-LA energies, $E_{FN}^{LA}[\Psi_{T1}]$ and $E_{FN}^{LA}[\Psi_{T2}]$, will be in general different, $E_{FN}^{LA}[\Psi_{T1}] \neq E_{FN}^{LA}[\Psi_{T2}]$, because of the different localization errors.

[53] Two different trial wave functions Ψ_{T1} and Ψ_{T2} with the same nodes have in general different FN-DMC-TM energies $E_{FN}^{TM}[\Psi_{T1}] \neq E_{FN}^{TM}[\Psi_{T2}]$ because their localization errors are different.

[54] D. F. B. ten Haaf, H. J. M. van Bemmel, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, Phys. Rev. B **51**, 13039 (1995).

[55] Notice that this approach was used three decades ago by Hammond *et al.*[24] in one of the first works using PPs in FN-DMC, albeit the motivation was to make runs cheaper. Shortly after the LA was introduced[30] – with a prescription to solve the integrals involved in the localization [see Appendix C and Eq. (C1) therein]— and it quickly became the default approach.

[56] M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **126**, 234105 (2007).

[57] J. R. Trail and R. J. Needs, J. Chem. Phys. **122**, 174109 (2005).

[58] J. R. Trail and R. J. Needs, J. Chem. Phys. **139**, 014101 (2013).

[59] J. R. Trail and R. J. Needs, J. Chem. Phys. **142**, 064110 (2015).

[60] J. R. Trail and R. J. Needs, J. Chem. Phys. **146**, 204107 (2017).

[61] M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, J. Chem. Phys. **147**, 224106 (2017).

[62]If one is concerned with the variationality of the approach, it is also possible to use TM in conjunction with DLA. The corresponding Hamiltonian is then: $\hat{H}_{FN}^{DLTM} \equiv \hat{K} + \hat{V}_L + \hat{V}_{NL}^- + \frac{\hat{V}_{NL}^+ \mathscr{D}}{\mathscr{D}} + \hat{V}_{FN}$. Within this second approach, we have that $E_{FN}^{DLTM} \geq E_0$, with the equality obtained for $\mathscr{D} = \Phi$.

[63]The reference value of -128 ± 1 meV was obtained in Ref. 3 using all-electrons. We performed here a new CCSD(T) calculation using the TN-DF pseudopotentials[57]—the same used for the DMC evaluations—and we obtained an interaction of -130 ± 4 meV. The agreement between the two evaluations indicates that differences between CCSD(T) and DMC should not come from the quality of the employed PPs.

[64]DFT calculations were performed using the PWSCF package[81] with a plane wave cutoff of 600 Ry and the LDA functional. Molecular orbitals were later converted into splines[82] to enhance the efficiency in QMC calculations. QMC calculations were performed using the CASINO package.

[65]The error bar is the error of the fit, which is probably underestimated as we are fitting a quadratic function (so, having three parameters) with four points. As a comparison, the actual DLA evaluation with smallest τ, which is 0.001 a.u., has a stochastic error due to DMC sampling of ± 5 meV.

[66]The reported error comes from the fit.

[67]Notice that in both LA and TM, the bias is dominated by the error in the total energy of the bound system since the wave function is by construction much better for the unbound systems. Thus, the sign of the TM error can be rationalized: TM underbinds because of the variationality of the approach, which leads to an underestimate of the total energy of the bound system when a suboptimal Jastrow is used. In LA, the optimization bias could in principle have any sign because the approach is not variational. In this specific case, the worse Jastrow leads to a lower total energy (see Fig. 2).

[68]The reported errors are from the fit and are probably underestimated because we are fitting three parameters with four points, especially for the LA approach, where the simulation with the smallest time step, $\tau = 0.003$ a.u., was very unstable and we could not make the stochastic error smaller than 5 meV.

[69]J. R. Trail and R. J. Needs, J. Chem. Phys. **122**, 014112 (2005).

[70]A. Annaberdiyev, G. Wang, C. A. Melton, M. Chandler Bennett, L. Shulenburger, and L. Mitas, J. Chem. Phys. **149**, 134108 (2018).

[71]J. T. Krogel, J. A. Santana, and F. A. Reboredo, Phys. Rev. B **93**, 075143 (2016).

[72]P. Seth, P. L. Rios, and R. J. Needs, J. Chem. Phys. **134**, 084105 (2011).

[73]S. Fahy, X. W. Wang, and S. G. Louie, Phys. Rev. Lett. **61**, 1631 (1988).

[74]E. Ertekin, L. K. Wagner, and J. C. Grossman, Phys. Rev. B **87**, 155210 (2013).

[75]For instance, HF or DFT is used in the Burkatzki, Filippi, and Dolg's energy consistent pseudopotentials (BFD),[56] in the Trail and Needs' norm conserving Hartree-Fock pseudopotentials (TN-NC)[69] and smooth relativistic Hartree-Fock pseudopotentials (TN-DF).[57]

[76]For instance, CCSD(T) is used in Trail and Needs' correlated electron pseudopotentials (CEPPs)[58,59] and the shape and energy consistent pseudopotentials,[60] and in the Mitas and collaborators' correlation consistent effective core potentials (ccECPs).[61,70]

[77]PPs an order of magnitude more accurate for IEs have recently been reported.[61]

[78]DFT/LDA calculation performed using the Orca package,[96] which uses localized basis sets. Reported FN-DMC results are converged in the DMC time step and the size of the basis set.

[79]Note that here better means a better error cancellation.

[80]L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[81]P. Giannozzi et al., J. Phys.: Condens. Matter **21**, 395502 (2009).

[82]D. Alfè and M. J. Gillan, Phys. Rev. B **70**, 161101 (2004).

[83]In this case, the inconsistency among the Kleinman-Bylander projectors and the DFT functional used is likely the source of the instability of DMC simulations. Indeed, we can make the system much more stable by generating Kleinman-Bylander projectors using DFT/LDA atomic orbitals, for instance by using the Ld1 code included in the Quantum Espresso package.[81] However, in other systems, it can be harder to improve the wave function and enhance stability. Here, we are intentionally considering a case that amplifies instability issues.

[84]It should be mentioned that some modifications to Green's function can enhance slightly the stability. For instance, within the ZSGMA scheme[26] to cure the local energy and the drift velocity divergences, it is possible to perform some DMC-LA simulations with large time steps (e.g., $\tau > 0.03$ a.u.), but unfortunately the time step bias is too large.

[85]It should be mentioned that there have been previous attempts to use wave functions that do not need optimization in QMC[112–114] although they have focused on relatively simpler systems and do not deal with atomic PPs and localization errors in DMC.

[86]R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Rios, J. Phys.: Condens. Matter **22**, 023201 (2010).

[87]The TurboRVB Quantum Monte Carlo package includes a complete suite for variational, diffusion, and Green's function quantum Monte Carlo calculations on molecules and solids, for wave function and geometry optimization, and for quantum Monte Carlo based molecular dynamics simulations. It is developed by Sorella and co-workers. See http://people.sissa.it/~sorella/web/index.html.

[88]J. Kim et al., J. Phys.: Condens. Matter **30**, 195901 (2018).

[89]K. Saritas, T. Mueller, L. Wagner, and J. C. Grossman, J. Chem. Theory Comput. **13**, 1943 (2017).

[90]W. W. Tipton, N. D. Drummond, and R. G. Hennig, Phys. Rev. B **90**, 125110 (2014).

[91]For instance, in CASINO,[86] Jastrow terms involving arbitrary numbers of particles can be introduced,[115] and in TurboRVB,[87] there is also the possibility to use four-body e-e-n-n terms, and the e-e-n term is an explicit function of $(\mathbf{r}_{i\alpha}, \mathbf{r}_{j\alpha})$; see Ref. 116.

[92]A. Grüneis, S. Hirata, Y.-y. Ohnishi, and S. Ten-no, J. Chem. Phys. **146**, 080901 (2017).

[93]The employed Jastrow for the AE calculations has e-e, e-n, and e-e-n terms. Notice that in AE calculations, there is no localization error, i.e., the Jastrow factor does not affect the accuracy and does not lead to any uncertainty.

[94]For a derivation of Eq. (E1), see the supplementary material of Ref. 2.

[95]This is likely because the T-moves are performed after the Metropolis step, so the TM algorithm does not satisfy detailed balance.

[96]F. Neese, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73 (2012).

[97]Using a localized basis, we get rid of possible issues with the Kleinman-Bylander projectors in plane-wave DFT calculations. Moreover, we crosschecked in the carbon atom case that the $\mathscr{D}$ obtained from plane-wave DFT calculations with 600 Ry cutoff and from open-system DFT calculations with an aug-cc-pVQZ-eCEPP basis have similar quality. Indeed, the variance obtained when using the same Jastrow factor is roughly the same.

[98]Here, there is a possible ambiguity, as one could wonder if we have to minimize the variance $\mathrm{Var}[E_L]_{VMC}$ or $\mathrm{Var}[E_L^{DLA}]_{VMC}$, or provide different optimization if we intend to use the original Hamiltonian $\hat{H}$ or the DLA version $\hat{H}^{DLA}$. However, in all cases considered, we noticed that the parameters of $\mathscr{J}$ that minimize $\mathrm{Var}[E_L]_{VMC}$ are, within the stochastic uncertainty, the same as those that minimize $\mathrm{Var}[E_L^{DLA}]_{VMC}$, and vice-versa.

[99]D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**, 5833 (1984).

[100]R. Bianchi, D. Bressanini, P. Cremaschi, and G. Morosi, J. Chem. Phys. **98**, 7204 (1993).

[101]D. M. Arnow, M. H. Kalos, M. A. Lee, and K. E. Schmidt, J. Chem. Phys. **77**, 5562 (1982).

[102]J. B. Anderson, C. A. Traynor, and B. M. Boghosian, J. Chem. Phys. **95**, 7418 (1991).

[103]S. Zhang and M. H. Kalos, Phys. Rev. Lett. **67**, 3074 (1991).

[104]M. H. Kalos and F. Pederiva, Phys. Rev. Lett. **85**, 3547 (2000).

[105]F. A. Reboredo, R. Q. Hood, and P. R. C. Kent, Phys. Rev. B **79**, 195117 (2009).

[106]N. D. Drummond, J. R. Trail, and R. J. Needs, Phys. Rev. B **94**, 165170 (2016).

[107]M. Pozzo and D. Alfè, Phys. Rev. B **77**, 104103 (2008).

[108]J. T. Krogel and P. R. C. Kent, J. Chem. Phys. **146**, 244101 (2017).

[109]A. L. Dzubak, J. T. Krogel, and F. A. Reboredo, J. Chem. Phys. **147**, 024102 (2017).

[110]D. Ceperley, M. H. Kalos, and J. L. Lebowitz, Macromolecules **14**, 1472 (1981).

04 June 2024 14:39:52

[111] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).

[112] D. Ceperley, Phys. Rev. B **18**, 3126 (1978).

[113] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, Phys. Rev. E **68**, 046707 (2003).

[114] B. Wood and W. M. C. Foulkes, J. Phys.: Condens. Matter **18**, 2305 (2006).

[115] P. López Ríos, P. Seth, N. D. Drummond, and R. J. Needs, Phys. Rev. E **86**, 036703 (2012).

[116] A. Zen, Y. Luo, S. Sorella, and L. Guidoni, J. Chem. Theory Comput. **9**, 4332 (2013).

04 June 2024 14:39:52