

# TurboRVB: A many-body toolkit for *ab initio* electronic simulations by quantum Monte Carlo

Cite as: J. Chem. Phys. **152**, 204121 (2020); <https://doi.org/10.1063/5.0005037>

Submitted: 19 February 2020 . Accepted: 20 March 2020 . Published Online: 29 May 2020

Kousuke Nakano , Claudio Attaccalite , Matteo Barborini , Luca Capriotti , Michele Casula , Emanuele Coccia , Mario Dagrada, Claudio Genovese , Ye Luo , Guglielmo Mazzola , Andrea Zen , and Sandro Sorella 

## COLLECTIONS

Paper published as part of the special topic on [Electronic Structure Software](#)

Note: This article is part of the JCP Special Topic on Electronic Structure Software.



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Siesta: Recent developments and applications](#)

The Journal of Chemical Physics **152**, 204108 (2020); <https://doi.org/10.1063/5.0005077>

[The CRYSTAL code, 1976–2020 and beyond, a long story](#)

The Journal of Chemical Physics **152**, 204111 (2020); <https://doi.org/10.1063/5.0004892>

[Modern quantum chemistry with \[Open\]Molcas](#)

The Journal of Chemical Physics **152**, 214117 (2020); <https://doi.org/10.1063/5.0004835>

Lock-in Amplifiers  
up to 600 MHz



Watch



# TURBORVB: A many-body toolkit for *ab initio* electronic simulations by quantum Monte Carlo

Cite as: J. Chem. Phys. 152, 204121 (2020); doi: 10.1063/5.0005037

Submitted: 19 February 2020 • Accepted: 20 March 2020 •

Published Online: 29 May 2020



View Online



Export Citation



CrossMark

Kousuke Nakano,<sup>1,2,a)</sup> Claudio Attaccalite,<sup>3</sup> Matteo Barborini,<sup>4</sup> Luca Capriotti,<sup>5,6</sup> Michele Casula,<sup>7,b)</sup> Emanuele Coccia,<sup>8</sup> Mario Dagrada,<sup>9</sup> Claudio Genovese,<sup>1</sup> Ye Luo,<sup>10,11</sup> Guglielmo Mazzola,<sup>12</sup> Andrea Zen,<sup>13,14</sup> and Sandro Sorella<sup>1,c)</sup>

## AFFILIATIONS

<sup>1</sup> International School for Advanced Studies (SISSA), Via Bonomea 265, 34136 Trieste, Italy

<sup>2</sup> Japan Advanced Institute of Science and Technology (JAIST), Asahidai 1-1, Nomi, Ishikawa 923-1292, Japan

<sup>3</sup> Aix-Marseille Université, CNRS, CINaM UMR 7325, Campus de Luminy, 13288 Marseille, France

<sup>4</sup> CNR-NANO, Via Campi 213/a, 41125 Modena, Italy

<sup>5</sup> New York University, Tandon School of Engineering, 6 MetroTech Center, Brooklyn, New York 11201, USA

<sup>6</sup> Department of Mathematics, University College London, Gower Street, London WC1E 6BT, United Kingdom

<sup>7</sup> Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (IMPIC), Sorbonne Université, CNRS UMR 7590, IRD UMR 206, MNHN, 4 Place Jussieu, 75252 Paris, France

<sup>8</sup> Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy

<sup>9</sup> Forescout Technologies, John F. Kennedylaan 2, 5612AB Eindhoven, The Netherlands

<sup>10</sup> Computational Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, Illinois 60439, USA

<sup>11</sup> Argonne Leadership Computing Facility, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, Illinois 60439, USA

<sup>12</sup> IBM Research Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

<sup>13</sup> Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, United Kingdom

<sup>14</sup> Thomas Young Centre and London Centre for Nanotechnology, 17-19 Gordon Street, London WC1H 0AH, United Kingdom

**Note:** This article is part of the JCP Special Topic on Electronic Structure Software.

<sup>a)</sup> Author to whom correspondence should be addressed: kousuke\_1123@icloud.com

<sup>b)</sup> Electronic mail: michele.casula@upmc.fr

<sup>c)</sup> Electronic mail: sorella@sissa.it

## ABSTRACT

TURBORVB is a computational package for *ab initio* Quantum Monte Carlo (QMC) simulations of both molecular and bulk electronic systems. The code implements two types of well established QMC algorithms: Variational Monte Carlo (VMC) and diffusion Monte Carlo in its robust and efficient lattice regularized variant. A key feature of the code is the possibility of using strongly correlated many-body wave functions (WFs), capable of describing several materials with very high accuracy, even when standard mean-field approaches [e.g., density functional theory (DFT)] fail. The electronic WF is obtained by applying a Jastrow factor, which takes into account dynamical correlations, to the most general mean-field ground state, written either as an antisymmetrized geminal power with spin-singlet pairing or as a Pfaffian, including both singlet and triplet correlations. This WF can be viewed as an efficient implementation of the so-called resonating valence bond (RVB) *Ansatz*, first proposed by Pauling and Anderson in quantum chemistry [L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, 1960)] and condensed matter physics [P.W. Anderson, *Mat. Res. Bull.* **8**, 153 (1973)], respectively. The RVB *Ansatz* implemented in TURBORVB has a large variational freedom, including the Jastrow correlated Slater determinant as its simplest, but nontrivial case. Moreover, it has the remarkable advantage of remaining with an affordable computational cost, proportional to the one spent for the evaluation of a single Slater determinant. Therefore, its application to large systems is computationally feasible. The WF is expanded in a localized basis set. Several basis set functions are implemented, such as Gaussian, Slater, and mixed types, with no restriction on the choice of their contraction. The code implements the adjoint algorithmic differentiation that enables a very efficient evaluation of energy derivatives, comprising the ionic forces.

Thus, one can perform structural optimizations and molecular dynamics in the canonical NVT ensemble at the VMC level. For the electronic part, a full WF optimization (Jastrow and antisymmetric parts together) is made possible, thanks to state-of-the-art stochastic algorithms for energy minimization. In the optimization procedure, the first guess can be obtained at the mean-field level by a built-in DFT driver. The code has been efficiently parallelized by using a hybrid MPI-OpenMP protocol, which is also an ideal environment for exploiting the computational power of modern Graphics Processing Unit accelerators.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0005037>

## NOMENCLATURE

$\mathcal{A}$	antisymmetrization operator
$a, b, I, J$	indices of atoms
$\mathbf{F}_a$	atomic force for an atom $a$ , $\mathbf{F}_a \equiv -\frac{dE}{d\mathbf{R}_a}$
$f_i$	general force for the $i$ th variational parameter, $f_i \equiv -\frac{\partial E}{\partial \alpha_i}$
$f(\mathbf{r}_i, \mathbf{r}_j)$	spatial part of the pairing function
$G$	$N \times N$ matrix with elements $G_{i,j} = g(\mathbf{i}, \mathbf{j})$
$g(\mathbf{i}, \mathbf{j})$	pairing function between electrons $\mathbf{i} \leftrightarrow (\mathbf{r}_i, \sigma_i)$ and $\mathbf{j} \leftrightarrow (\mathbf{r}_j, \sigma_j)$
$\mathbf{i}, \mathbf{j}$	compact notation for $(\mathbf{r}_i, \sigma_i)$ and $(\mathbf{r}_j, \sigma_j)$ , respectively
$i, j$	indices of electrons/orbitals/Monte Carlo sampling
$J$	Jastrow factor, composed of $J_1, J_2$ , and $J_{3/4}$
$l, m$	indices of orbitals
$M$	the number of configurations of electrons (the number of Monte Carlo samplings)
$N$	the number of electrons
$N_{\text{at}}$	the number of atoms
$N_d$	the number of electrons with spin down
$N_u$	the number of electrons with spin up
$O_k(\mathbf{x})$	the logarithmic derivative of a many-body WF $O_k(\mathbf{x}) \equiv \frac{\partial \ln \Psi(\mathbf{x})}{\partial \alpha_k}$
$\mathbf{R}_a$	$a$ th ion coordination
$\mathbf{r}_i$	$i$ th electron coordination
$\mathbf{S}$	overlap matrix with elements $S_{\mu,\nu} \equiv \langle \psi_\mu   \psi_\nu \rangle$
$\mathcal{S}$	variance-covariance matrix of the logarithmic derivative of a many-body WF
$\mathcal{S}^{\text{QMC}}(\mathbf{R})$	variance-covariance matrix of QMC forces
$\mathbf{x}$	$N$ set of electron coordinations including spins, the same as $\mathbf{x}_i$
$\mathbf{x}_i$	$N$ set of electron coordinations including spins $\equiv (\mathbf{r}_1\sigma_1, \mathbf{r}_2\sigma_2, \dots, \mathbf{r}_N\sigma_N)_i$ . The index $i$ refers to the Monte Carlo sampling index
$\Phi$	$N \times k$ matrix with elements $\Phi_{i,k} = \phi_k(\mathbf{i})$
$\Phi_{\text{AGP}}$	antisymmetrized geminal power (AGP)
$\Phi_{\text{AS}}$	antisymmetric (AS) part
$\Phi_{\text{Pf}}$	Pfaffian (Pf)
$\Phi_{\text{SD}}$	single Slater determinant (SD)
$\Psi$	many-body WF Ansatz, $\Psi = \Phi_{\text{AS}} \times \exp J$
$\Psi_G$	many-body guiding function
$\Psi_T$	many-body trial WF
$\alpha_1, \dots, \alpha_p$	variational parameters of the trial WF
$\alpha_k$	$k$ th variational parameter
$\mu, \nu$	indices of orbitals
$\sigma_i$	spin of $i$ th electron
$\chi_i(\mathbf{r})$	primitive atomic orbital for the Jastrow part
$\psi_i(\mathbf{r})$	primitive atomic orbital for the antisymmetric part

$\phi_k(\mathbf{r})$	$k$ th molecular orbital for the antisymmetric part
$\int d\mathbf{i}$	compact notation for $\sum_{\sigma_i} \int d\mathbf{r}_i$

## I. INTRODUCTION

The solution of the many-body Schrödinger equation, which describes the interaction between electrons and ions at the quantum mechanical level, represents a fundamental challenge in computational chemistry, condensed matter, and materials science. Since about a century ago, there has been a relentless theoretical and computational effort to find an accurate solution to this problem, which also features many recent interdisciplinary applications in machine learning<sup>1</sup> and materials informatics.<sup>2</sup> While computationally the scaling of the problem is exponential with the number of electrons, several numerical approximate methods have been put forward in the last decades. Among them, the Density Functional Theory (DFT) method, proposed by Kohn and Sham<sup>3</sup> in 1965, is one of the most successful approaches. In this framework, the original interacting  $3N$  many-body problem ( $N$  being the number of electrons in a system) is mapped to a non-interacting electron system, defined by an effective mean-field potential, to be determined self-consistently.<sup>4</sup> While DFT is an exact theory in principle, the exact form for the exchange-correlation functional, which is an essential part of the DFT mean-field potential, remains unknown. Unfortunately, the progress in generating increasingly successful approximations of this functional is rather slow,<sup>5</sup> partly because there is no established strategy for systematic improvement,<sup>6</sup> while maintaining an efficient scaling with the system size. The commonly adopted approximations for the exchange-correlation DFT functional have well-known limitations, especially in describing weak dispersive interactions, strongly correlated materials, and extreme environments (e.g., high pressure).<sup>7,8</sup>

Alternative strategies are represented by the so-called wave function (WF)-based approaches popular in quantum chemistry applications.<sup>9</sup> In these methods, electronic correlations are captured either variationally or perturbatively by post-Hartree-Fock theories, such as the Møller-Plesset perturbation theory (MP),<sup>10</sup> configuration interaction (CI) and Full-CI (FCI),<sup>11</sup> multi-configurational self-consistent field (MCSCF),<sup>12</sup> and Coupled-Cluster (CC) theory,<sup>13</sup> to name a few. Among them, coupled-cluster with single, double, and perturbative triple excitations, or CCSD(T), is considered to be the *gold standard* in quantum chemistry as it typically provides results in good agreement with experiments and a reasonable balance between accuracy and computational affordability (despite its cost growing as the seventh power of the number of electrons). While quantum chemistry methods have the advantage of treating electronic exchange and correlation effects

in a systematically improvable fashion, they are also much more computationally demanding compared to DFT, and their applicability to large or periodic systems is often computationally prohibitive.

Another way to tackle the problem of the electron correlation and the huge dimension of a many-body WF, exponentially large in the number of electrons, is by means of stochastic approaches, which are in this context referenced with the widely used expression of quantum Monte Carlo (QMC) methods.<sup>14,15</sup> Since the invention of Markov Chain Monte Carlo (MCMC) in the 1940s,<sup>16</sup> stochastic approaches to numerical algorithms had a pervasive influence in a wide range of fields, from physics and engineering to finance. In this respect, the QMC framework is qualitatively different from the deterministic one mentioned above, and it represents, therefore, an original and alternative approach for the solution of the Schrödinger equation, overcoming some of the drawbacks of DFT and the deterministic WF-based approaches of quantum chemistry. In particular, QMC does not rely on uncontrolled approximations; its accuracy can be systematically improved, the scaling with system size is good, and it is straightforwardly applicable to both isolated and periodic systems. While the scaling with the system size is comparable with standard DFT methods, the prefactor is typically much larger. However, methods based on stochastic sampling are well suited for massively parallel computing architectures as the algorithms can sustain an almost ideal scaling with the number of cores. As a result, the feasibility and popularity of QMC are expected to increase with the foreseeable substantial improvements in high-performance computing (HPC) facilities.

`TURBORVB` includes two of the most popular ground-state QMC algorithms: variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC).<sup>14</sup> In VMC, a systematically improvable approximation for the ground state is obtained by direct minimization of the energy, evaluated by a parametrized many-body WF. In an explicitly correlated WF, the electron coordinates are not separable, and the expectation value of the Hamiltonian has to be calculated with Monte Carlo integration, hence the name. In realistic calculations, a faithful and nevertheless compact parametrization for the trial state is essential for a successful energy optimization. The first property allows an unbiased treatment of the electronic correlations across different regimes, such as bond dissociation and electronic phase transitions. The second is needed for a stable optimization of the variational parameters, which may become a too difficult task when the chosen parametrization is redundant or unnecessarily detailed. Therefore, a central effort, in the `TURBORVB` project, has been devoted to the development of efficient and systematic parametrizations of correlated WFs.

As we will see in the following, an accurate trial WF is also fundamental in DMC, which is the second main algorithm present in `TURBORVB`. DMC is an imaginary-time projection technique,<sup>15</sup> which, when combined with the so-called *fixed-node* (FN) approximation, represents a powerful route to electronic ground-state calculations. In this approximation, the nodal-surface, which is kept fixed during the projection, can be determined by an accurate variational optimization.

So far, several groups have implemented the above algorithms and established excellent QMC codes such as QMC-PACK,<sup>17</sup> CASINO,<sup>18</sup> QWALK,<sup>19</sup> CHAMP,<sup>20</sup> and HANDE-QMC.<sup>21</sup> We also remark that other QMC algorithms have also been developed recently, such as the auxiliary field quantum Monte Carlo

(AF-QMC),<sup>22,23</sup> full-configuration interaction quantum Monte Carlo (FCI-QMC),<sup>24,25</sup> coupled cluster Monte Carlo (CCMC),<sup>26,27</sup> density matrix quantum Monte Carlo (DMQMC),<sup>28,29</sup> model space quantum Monte Carlo (MSQMC),<sup>30–32</sup> clock quantum Monte Carlo (CQMC),<sup>33</sup> or driven-dissipative quantum Monte Carlo (DDQMC).<sup>34</sup>

A typical workflow for a QMC calculation performed with `TURBORVB` is shown in Fig. 1. It shall be noticed that the choice of the most suited *Ansatz* for the WF, as well as the optimization of its parameters, is a prerequisite of both VMC and DMC evaluations. Ionic forces, if evaluated, can be used to perform QMC-based structural relaxations, Langevin molecular dynamics (MD), or path integral molecular dynamics (PIMD). Moreover, several post-processing tools are available to analyze the outcomes. The `TURBORVB` code is peculiar because it relies on the assumption that a complex electronic problem, though, until now, cannot be solved exactly with a black-box computer software, can be described very accurately by appropriate variational wave functions, guided by robust physical and chemical requirements and human ingenuity. In particular, `TURBORVB` is different from other codes in the following features: (i) It employs the Resonating Valence Bond (RVB) WF, first proposed by Pauling and Anderson in quantum chemistry<sup>35</sup> and condensed matter physics,<sup>36</sup> respectively. It includes static and dynamical correlation effects beyond the commonly used Slater determinant, while keeping the computational cost at the single-determinant level, thanks to its efficient implementation. (ii) The code implements a VMC algorithm based on localized orbitals (e.g., Gaussians) and state-of-the-art optimization methods, such as the stochastic reconfiguration. Therefore, at the VMC level, one can optimize not only the amplitude of the WF (i.e., the Jastrow factor) but also the nodal surfaces (e.g., the Slater determinant). This leads to a better variational energy in general and also improves the corresponding FN-DMC energy. (iii) The energy derivatives (e.g., atomic forces) are calculated very efficiently, thanks to an implementation based on the Adjoint Algorithmic Differentiation (AAD). As a consequence, one can perform structural optimizations and Langevin molecular dynamics. (iv) The code implements the newly developed Lattice Regularized Diffusion Monte Carlo (LRDMC), a stable DMC algorithm that, very recently, has shown to have a better scaling with the atomic number  $Z$ , compared with standard DMC.<sup>37</sup>

This review is organized as follows: in Sec. II, we briefly explain the fundamental QMC algorithms implemented in `TURBORVB`, namely, the VMC and LRDMC; in Sec. III, we describe the RVB WF; in Sec. IV, we extend the WF to treat periodic systems; in Sec. V, we introduce the DFT algorithm efficiently implemented in `TURBORVB`; in Sec. VI, we describe the way energy derivatives are computed (e.g., atomic forces) by means of AAD; in Sec. VII, we list the `TURBORVB` steps to optimize a many-body WF; in Sec. VIII, we introduce the first and the second-order Langevin molecular dynamics implemented in `TURBORVB`; in Sec. IX, we summarize the typical QMC workflow from the WF generation to the final LRDMC calculation; in Sec. X, we show weak and strong scaling results of `TURBORVB`, measured on the Marconi/CINECA supercomputer; in Sec. XI, we list the physical properties that can be calculated by `TURBORVB`; in Sec. XII, we review the major `TURBORVB` applications done so far; and in Sec. XIII, we introduce a Python-based workflow system, named `TURBO-GENIUS`.

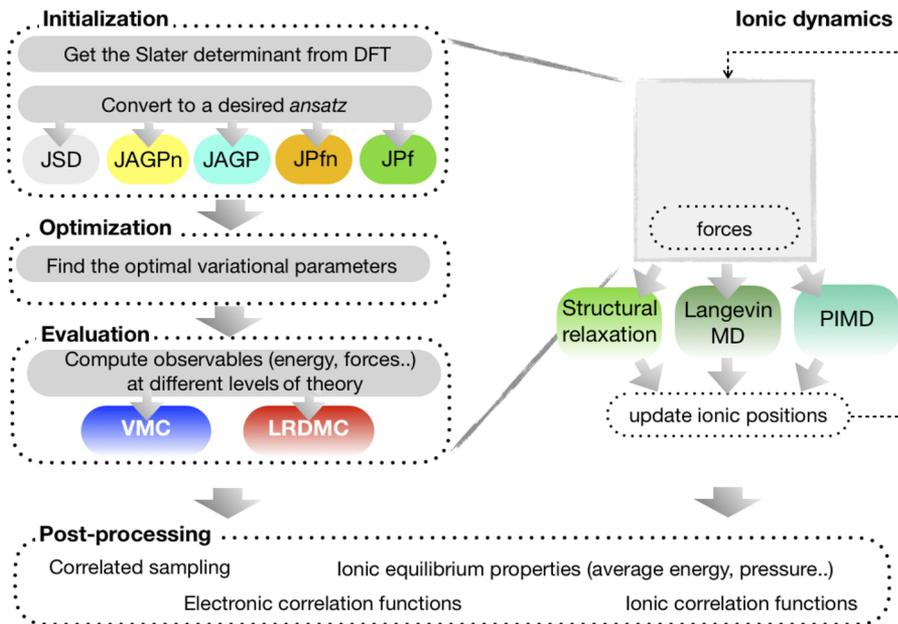


FIG. 1. Schematic picture of some of the possible tools available in TurboRVB. The arrows indicate typical workflow calculations.

## II. METHODS

TURBORVB implements two types of well established quantum Monte Carlo methods: Variational Monte Carlo (VMC) and Lattice Regularized Diffusion Monte Carlo (LRDMC). We summarize these methods in this section. The interested readers should also refer to the comprehensive review of QMC<sup>14</sup> for details. TURBORVB also implements an original DFT engine to generate trial WFs, which is more suitable for the VMC and LRDMC calculations, as shown in Sec. V.

### A. Variational Monte Carlo

Starting from the variational principle, the expectation value of the energy evaluated for a given WF  $\Psi$  can be written as

$$\langle E \rangle = \frac{\int d\mathbf{x} \Psi^2(\mathbf{x}) \cdot \hat{H}\Psi(\mathbf{x}) / \Psi(\mathbf{x})}{\int d\mathbf{x} \Psi^2(\mathbf{x})} = \int d\mathbf{x} e_L(\mathbf{x}) \pi(\mathbf{x}), \quad (1)$$

where  $\mathbf{x} = (\mathbf{r}_1\sigma_1, \mathbf{r}_2\sigma_2, \dots, \mathbf{r}_N\sigma_N)$  here and henceforth is a shorthand notation for the  $N$  electron coordinates and their spins, whereas

$$e_L(\mathbf{x}) \equiv \frac{\hat{H}\Psi(\mathbf{x})}{\Psi(\mathbf{x})} \quad \text{and} \quad \pi(\mathbf{x}) \equiv \frac{\Psi^2(\mathbf{x})}{\int d\mathbf{x}' \Psi^2(\mathbf{x}')}$$

are the so-called local energy and the probability of the configuration  $\mathbf{x}$ , respectively. This multidimensional integration can be evaluated stochastically by generating a set  $\{\mathbf{x}_i\}$  according to the distribution  $\pi(\mathbf{x})$  using the Markov chain Monte Carlo such as the (accelerated<sup>38,39</sup>) Metropolis method and by averaging the obtained local energies  $e_L(\mathbf{x}_i)$ ,

$$E_{\text{VMC}} = \langle e_L(\mathbf{x}) \rangle_{\pi(\mathbf{x})} \approx \frac{1}{M} \sum_{i=1}^M e_L(\mathbf{x}_i), \quad (2)$$

which has an associated statistical error of  $\sqrt{\text{Var}[e_L(\mathbf{x}_i)]/\bar{M}}$ , where  $\text{Var}[e_L(\mathbf{x}_i)]$  is the variance of the sampled local energies and  $\bar{M}$  is the sampling size  $M$  divided by the autocorrelation time. This indicates that the precision of the VMC evaluation is inversely proportional to the square root of the number of samplings (i.e., of the computational cost). It is worth to note that if  $\Psi(\mathbf{x})$  is an eigenfunction of  $\hat{H}$ , say with eigenvalue  $E_0$ , then  $e_L(\mathbf{x}) = E_0$  for each  $\mathbf{x}$ , implying that the variance of the local energy is zero and  $E_{\text{VMC}} = E_0$  with no stochastic uncertainty. This feature is known as the zero-variance property.

The probability distribution used for the importance sampling can also differ from  $\pi(\mathbf{x})$ . Indeed, one can use an arbitrary probability distribution function  $\pi'(\mathbf{x}) = \Psi_G^2(\mathbf{x}) / \int d\mathbf{x} \Psi_G^2(\mathbf{x})$  and estimate a generic local observable  $O(\mathbf{x})$  either by using  $\bar{O}_{\text{VMC}} = \langle O(\mathbf{x}) \rangle_{\pi(\mathbf{x})}$  or

$$\bar{O}_{\text{VMC}} = \frac{\langle O(\mathbf{x}') \mathcal{W}(\mathbf{x}') \rangle_{\pi'(\mathbf{x}')}}{\langle \mathcal{W}(\mathbf{x}') \rangle_{\pi'(\mathbf{x}')}} \approx \frac{\sum_{i=1}^{M'} O(\mathbf{x}'_i) \mathcal{W}(\mathbf{x}'_i)}{\sum_{i=1}^{M'} \mathcal{W}(\mathbf{x}'_i)}, \quad (3)$$

where  $\mathcal{W}(\mathbf{x}) = |\Psi(\mathbf{x})/\Psi_G(\mathbf{x})|^2$ , and the points  $\mathbf{x}'_i$  are distributed according to  $\pi'$ . This *reweighting* scheme is very important when evaluating atomic forces, as discussed in Sec. VI B. Since the evaluations of the standard deviations are nontrivial in this case, TURBORVB employs the bootstrap and jackknife methods in order to estimate the mean value and the statistical error,<sup>15</sup> which are also used when evaluating those of the local energy, forces, and so on. Indeed, the code outputs the history of the local energies, forces, or other properties in appropriate files (when the corresponding option is true), thus allowing the error bar estimates by simple post-processing. The user can also use the reblocking (binning) technique to remove the autocorrelation bias.<sup>40</sup>

One can optimize the WF based on the *variational theorem* by introducing a set of parameters  $(\alpha_1, \alpha_2, \dots, \alpha_p)$  to the WF  $\Psi(\mathbf{x}, \alpha)$ ,

$$E_{\text{VMC}}(\alpha) = \int d\mathbf{x} e_L(\mathbf{x}, \alpha) \pi(\mathbf{x}, \alpha) \geq E_{\text{exact}}. \quad (4)$$

However, the optimization of a many-body WF remains a difficult challenge not only because optimizing a cost function containing many parameters is a complex numerical task, due to the presence of several local minima in the energy landscape, but also because this difficult task is further complicated by the presence of statistical errors in the QMC evaluation of any quantity. Nevertheless, a great improvement in this field has been achieved when the QMC optimization techniques have made use of the explicit evaluation of energy derivatives with finite statistical errors. In particular, in TURBORVB, the adjoint algorithmic differentiation (AAD) has been implemented, by allowing the efficient calculation of generalized forces ( $f_i = -\frac{\partial E}{\partial \alpha_i}$ )<sup>41</sup> and very efficient optimization methods, the so-called “stochastic reconfiguration”<sup>42,43</sup> and “the linear method.”<sup>44–46</sup> These methods are discussed in Secs. VII A and VII B.

## B. Lattice regularized Monte Carlo

Lattice regularized diffusion Monte Carlo (LRDMC), which was initially proposed by Casula *et al.*,<sup>47</sup> is a projection technique that allows us to improve a variational *Ansatz* systematically. This method is based on Green’s function Monte Carlo (GFMC),<sup>48–50</sup> filtering out the ground state WF  $|\Upsilon_0\rangle$  from a given trial WF  $|\Psi_T\rangle$ : since the eigenstates of the Hamiltonian have the completeness property, the trial WF can be expanded as

$$|\Psi_T\rangle = \sum_n a_n |\Upsilon_n\rangle, \quad (5)$$

where  $a_n$  is the coefficient for the  $n$ th eigenvectors ( $\Upsilon_n$ ). Therefore, by applying  $\mathcal{G}^M = (\Lambda - \hat{\mathcal{H}})^M$ , one can obtain

$$\begin{aligned} |\Upsilon_0\rangle &\propto \lim_{M \rightarrow \infty} (\Lambda - \hat{\mathcal{H}})^M |\Psi_T\rangle \\ &= \lim_{M \rightarrow \infty} (\lambda - E_0)^M \left[ a_0 |\Upsilon_0\rangle + \sum_{n \neq 0} \left( \frac{\lambda - E_n}{\lambda - E_0} \right)^M a_n |\Upsilon_n\rangle \right], \end{aligned} \quad (6)$$

where  $\Lambda$  is a diagonal matrix with  $\Lambda_{x',x} = \lambda \delta_{x',x}$  ( $\lambda$  should be sufficiently large to obtain the ground state) and  $E_n$  is the  $n$ th eigenvalue of  $\hat{\mathcal{H}}$ . Since  $\frac{\lambda - E_n}{\lambda - E_0} < 1$ , the projection filters out the ground state WF  $\Upsilon_0$  from a given trial WF  $|\Psi_T\rangle$ , as long as the trial WF is not orthogonal to the true ground state (i.e.,  $a_0 \equiv \langle \Upsilon_0 | \Psi_T \rangle \neq 0$ ). To apply the GFMC for *ab initio* electron calculations, the original continuous Hamiltonian is regularized by allowing electron hopping with step size  $a$ , in order to mimic the electronic kinetic energy. The corresponding Hamiltonian  $\hat{\mathcal{H}}^a$  is then defined such that  $\hat{\mathcal{H}}^a \rightarrow \hat{\mathcal{H}}$  for  $a \rightarrow 0$ . Namely, the kinetic part is approximated by a finite difference form

$$\begin{aligned} \Delta_i f(x_i, y_i, z_i) &\approx \Delta_i^a f(x_i, y_i, z_i) \\ &\equiv \frac{1}{a^2} \{ [f(x_i + a) - f(x_i)] + [f(x_i - a) - f(x_i)] \} \\ &\leftrightarrow y_i \leftrightarrow z_i, \end{aligned} \quad (7)$$

and the potential term is modified as

$$V^a(\mathbf{x}) = V(\mathbf{x}) + \frac{1}{2} \left[ \frac{\sum_i (\Delta_i^a - \Delta_i) \Psi_G(\mathbf{x})}{\Psi_G(\mathbf{x})} \right]. \quad (8)$$

The corresponding Green’s function matrix elements are

$$\mathcal{G}_{x',x} = \langle x' | (\Lambda - \hat{\mathcal{H}}^a) | x \rangle \equiv (\Lambda - \mathcal{H}^a)_{x',x} \quad (9)$$

and the single LRDMC iteration step is given by the following equation:

$$\Psi_{n+1}(x') = \sum_x \mathcal{G}_{x',x} \Psi_n(x). \quad (10)$$

The sketch of the LRDMC algorithm, a Markov chain that evolves the many-body WF according to Eq. (10), is as follows:<sup>47</sup> (STEP 1) Prepare a walker with configuration  $x$  and weight  $w$  ( $w_0 = 1$ ). (STEP 2) A new configuration  $x'$  is generated by the transition probability,

$$p_{x',x} = \mathcal{G}_{x',x} / b_x, \quad (11)$$

where

$$b_x = \sum_{x'} \mathcal{G}_{x',x} \quad (12)$$

is a normalization factor. By applying the discretized Hamiltonian to a given configuration  $(\hat{\mathcal{H}}^a | x)$ ,  $(6N + 1)$  configurations  $|x'\rangle$  are determined according to the probability  $p_{x',x}$  in Eq. (11), where  $N$  is the number of electrons in the system.<sup>51</sup> This allows the evaluation of the normalization factor  $b_x$  in Eq. (12) even in a continuous model. Note that  $6N$  comes from the diffusion of each electron in two directions ( $\pm a$ ) and the remaining 1 stands for the starting configuration  $x$  before the possible hoppings (all  $N$  electrons) (i.e.,  $x' = x$ ). (STEP 3) Finally, update the weight with  $w_{n+1} = w_n b_x$  and return to STEP I. After a sufficiently large number of iterations (the Markov process is equilibrated), one can calculate the ground state energy  $E_0$ ,

$$E_0 \approx \frac{\langle w_n e_L(x_n) \rangle}{\langle w_n \rangle}, \quad (13)$$

where  $\langle \dots \rangle$  denotes the statistical average over many independent samples generated by the Markov chain and  $e_L(x)$  is called the (bare) local energy that reads

$$e_L(x) = \sum_{x'} \mathcal{H}_{x',x} = \lambda - b_x. \quad (14)$$

Indeed, the ground state energy can be calculated after many independent  $n$ -step calculations. A more efficient computation can be realized by using the so-called “correcting factor” technique: after a single simulation that is much larger than the equilibration time, one can imagine starting a projection of length  $p$  from each  $(n - p)$ th iteration. The accumulated weight for each projection is

$$\mathcal{G}_n^p = \prod_{j=1}^p b_{n-j}. \quad (15)$$

Then, the ground state energy  $E_0$  can be estimated by

$$E_0 \approx \frac{\sum_n \mathcal{G}_n^p e_L(x_n)}{\sum_n \mathcal{G}_n^p}. \quad (16)$$

This straightforward implementation of the above simple method is not suitable for realistic simulations due to fluctuations of weights, large correlation times, sign problems, and so on. `TURBORVB` implements the following state-of-art techniques for real electronic structure calculations.

If the potential term [Eq. (8)] is unbounded (it is the case in *ab initio* calculations), the bare weight  $b_x$  [Eq. (12)] and the local energy  $e_L(x)$  [Eq. (14)] significantly fluctuate, making the numerical simulation very unstable and inefficient. To overcome this difficulty, the code employs the importance sampling scheme,<sup>15</sup> in which the original Green's function is modified using the so-called guiding function  $\Psi_G$  as

$$\tilde{\mathcal{G}}_{x',x} = \mathcal{G}_{x',x} \frac{\Psi_G(x')}{\Psi_G(x)}, \quad (17)$$

and the projection is modified as

$$\Psi_G(x') \Psi_{n+1}(x') = \sum_x \tilde{\mathcal{G}}_{x',x} \Psi_G(x) \Psi_n(x). \quad (18)$$

In practice, the guiding function is prepared by a VMC calculation. The modified Green's function for importance sampling  $\tilde{\mathcal{G}}_{x',x}$  has the same eigenvalues as the original one, and this transformation does not change the formalism of LRDMC. The weight is updated by

$$\tilde{b}_x = \sum_{x'} \tilde{\mathcal{G}}_{x',x}, \quad (19)$$

and the local energy with importance sampling is

$$\tilde{e}_L(x) = \frac{\langle \Psi_G | \hat{\mathcal{H}} | x \rangle}{\langle \Psi_G | x \rangle} = \sum_{x'} \mathcal{H}_{x',x} \frac{\Psi_G(x')}{\Psi_G(x)}. \quad (20)$$

Equation (20) implies that if the guiding function  $\Psi_G$  is an exact eigenstate of the Hamiltonian, there are no statistical fluctuations, implying the zero variance property, namely, the computational efficiency to obtain a given statistical error on the energy improves with the quality of the variational WF. In this respect, it is also important to emphasize that a meaningful reduction in the statistical fluctuations is obtained by satisfying the so-called cusp conditions. As long as they are satisfied, the resulting local energy does not diverge at the coalescence points where two particles are overlapped, despite the singularity of the Coulomb potential term [ $V(x)$  in Eq. (8)].<sup>51</sup> In addition, the importance sampling maintains the electrons in a region away from the nodal surface since the guiding function vanishes there [i.e., the RHS of Eq. (17)  $\rightarrow 0$ ]. This clearly enhances the efficiency of the sampling because the local energy diverges at the nodal surface.

The Green's function cannot be made strictly positive for fermions; therefore, the fixed-node (FN) approximation has to be introduced<sup>15</sup> in order to avoid the sign problem. Indeed, the Hamiltonian is modified using the spin-flip term  $\mathcal{V}_{sf}(x) = \sum_{x':s_{x',x}>0} \mathcal{H}_{x',x} \Psi_G(x') / \Psi_G(x)$ ,

$$\mathcal{H}_{x',x}^{\text{FN},\gamma} = \begin{cases} \mathcal{H}_{x,x} + (1+\gamma)\mathcal{V}_{sf}(x) & \text{for } x' = x \\ \mathcal{H}_{x',x} & \text{for } x' \neq x, s_{x',x} < 0 \\ -\gamma\mathcal{H}_{x',x} & \text{for } x' \neq x, s_{x',x} > 0, \end{cases} \quad (21)$$

where  $s_{x',x} = \Psi_G(x') \mathcal{H}_{x',x} / \Psi_G(x)$  and  $\gamma \geq 0$  is a real parameter. The use of the fixed-node Green's function

$$\tilde{\mathcal{G}}_{x',x}^{\text{FN}} = (\Lambda - \mathcal{H}^{\text{FN}})_{x',x} \frac{\Psi_G(x')}{\Psi_G(x)} \quad (22)$$

can prevent the crossing of regions where the configuration space yields a sign flip of the Green's function; therefore, the walkers are constrained in the same nodal pockets and avoid the sign problem.

`TURBORVB` also implements the many-walker technique and the branching (denoted as reconfiguration<sup>49</sup> in `TURBORVB`) scheme for a more efficient computation.<sup>15</sup> The code performs the branching as follows: (1) Set the new weights equal to the average of the old ones,

$$w'_{\alpha,n} = \tilde{w} \equiv \frac{1}{N_w} \sum_{\beta} w_{\beta,n}. \quad (23)$$

(2) Select the new walkers from the old ones with a probability that is proportional to the old walkers' weights,

$$p_{\alpha,n} = \frac{w_{\alpha,n}}{\sum_{\beta} w_{\beta,n}}, \quad (24)$$

which does not change the statistical average of weights but suppresses the fluctuations by dropping walkers having small weights. The code performs branching (reconfiguration) after a projection time  $\tau$ , which can be chosen as a user input parameter. In practice, within the many-walker and branching schemes, the average weights are stored and are set to one for all walkers after each branching. The user can retrieve the accumulated weights at the end of the simulation,

$$\mathcal{G}_n^p = \prod_{j=1}^p \tilde{w}_{n-j}, \quad (25)$$

and calculate the ground state energy,

$$E_0 \approx \frac{\sum_n \mathcal{G}_n^p e_L(x_n)}{\sum_n \mathcal{G}_n^p}, \quad (26)$$

where  $e_L(x_n)$  is the mean local energy averaged over the walkers, which reads

$$e_L(x_n) = \frac{\sum_{\alpha} w_{\alpha,n} e_L(x_{\alpha,n})}{\sum_{\alpha} w_{\alpha,n}} \quad (27)$$

and is evaluated just before each reconfiguration. Note that  $p$  is also an input parameter, which has to be carefully chosen by the user to allow energy convergence.

When  $\lambda$  is sufficiently large, the correlation time also becomes large because the diagonal terms of the Green's function become very close to one (i.e., a walker remains in the same configuration), which causes a very large correlation time. In `TURBORVB`, this difficulty is solved by considering in a different way the diagonal and non-diagonal moves. In a given interval of  $M$  iterations, the non-diagonal updates are efficiently calculated using a random number according to the probability that the configuration remains in the same one (diagonal updates). This technique can be generalized to the continuous-time limit, namely,  $M \rightarrow \infty$ , at  $\frac{M}{\Lambda} = \tau$  fixed. In the

$M \rightarrow \infty$  limit, the projection  $(\Lambda - \hat{\mathcal{H}})^M$  is equal to the imaginary time evolution  $\exp(-\tau\hat{\mathcal{H}})$ , apart from an irrelevant constant  $\Lambda^M$ . Thus, the user should specify only  $\tau$  as the input parameter. Indeed, the walker weight is updated by  $w \rightarrow w \exp(-\tau_\xi e_L(x))$  and the imaginary time is updated by  $\tau_{\text{left}} \rightarrow \tau_{\text{left}} - \tau_\xi$  at each non-diagonal update until  $\tau_{\text{left}}$  becomes 0, where  $\tau_\xi = -\log(1 - \xi)/b_x$  is a diagonal move time step determined by a uniform random number  $0 \leq \xi < 1$ . The branching (reconfiguration) is performed after each projection time of length  $\tau$  that a user puts in the input file within the many walker and the branching implementation.

In practice, there are three important features in LRDMC. First, there is not a time step error in LRDMC because the Suzuki–Trotter decomposition is not necessary, unlike the standard DMC algorithm.<sup>15</sup> Instead, there is a finite-size lattice error due to the discretization of the Hamiltonian ( $a$ ). Therefore, in order to obtain an unbiased FN energy, it is important to extrapolate the LRDMC energy to the  $a \rightarrow 0$  limit by using several results corresponding to different lattice spaces.<sup>47</sup> This is then consistent with the standard DMC energy estimate (Fig. 2) obtained in the limit of an infinitely small time step. Probably one of the most important advantages of the LRDMC method is that the extrapolation to the  $a \rightarrow 0$  limit is very smooth and reliable so that unbiased FN energies are easily obtained with low order polynomial fits. Second, LRDMC can straightforwardly handle different length scales of a WF by introducing different mesh sizes ( $a$  and  $a'$ ) so that electrons in the vicinity of the nuclei and those in the valence region can be appropriately diffused,<sup>37,47</sup> which defines the so-called double-grid LRDMC. This scheme saves a substantial computational cost in all-electron calculations, especially for a system including large atomic number atoms,<sup>37</sup> with a typical computational cost scaling with  $Z^{-6}$ , where  $Z$  is the maximum atomic number. TURBORVB makes use of an appropriate ratio of the mesh sizes (i.e.,  $ala'$ ), the smaller one  $a$  used when electrons are close to the nuclei and the larger one

$a' \gg a$  adopted in the valence region. By choosing a proper Thomas–Fermi characteristic length around the nuclei, where short hops of lengths  $a$  mostly occur, a significant improvement of the scaling (i.e.,  $Z^{-5.6} \rightarrow Z^{-5}$ ) has recently been reported.<sup>37</sup> Finally, the inclusion of non-local pseudopotentials in this framework is straightforward by means of an additional spherical grid defined in an appropriate mesh.<sup>51</sup> As described in Refs. 47 and 52, LRDMC provides an upper bound for the true ground-state energy and allows the estimation of  $E_{\text{FN}}$ , even in the presence of non-local pseudopotentials. Note that this variational property has also been extended to the standard DMC framework,<sup>53</sup> with the introduction of the so-called T-moves. Moreover, the recently introduced determinant locality approximation (DLA)<sup>54</sup> to deal with non-local pseudopotentials is also implemented in TURBORVB and can be optionally used in LRDMC.

### III. WAVE FUNCTIONS

Both the accuracy and the computational efficiency of QMC approaches crucially depend on the WF *Ansatz*. The optimal *Ansatz* is typically a trade-off between accuracy and efficiency. On the one side, a very accurate *Ansatz* can be involved and cumbersome, having many parameters and being expensive to evaluate. On the other hand, an efficient *Ansatz* is described only by the most relevant parameters and can be quickly and easily evaluated. In particular, in Sec. II, we have seen that QMC algorithms, both at the variational and fixed-node levels, imply several calculations of the local energy  $e_L(\mathbf{x})$  and the ratio  $\Psi(\mathbf{x})/\Psi(\mathbf{x}')$  for different electronic configurations  $\mathbf{x}$  and  $\mathbf{x}'$ . The computational cost of these operations determines the overall efficiency of QMC and its scaling with system size.

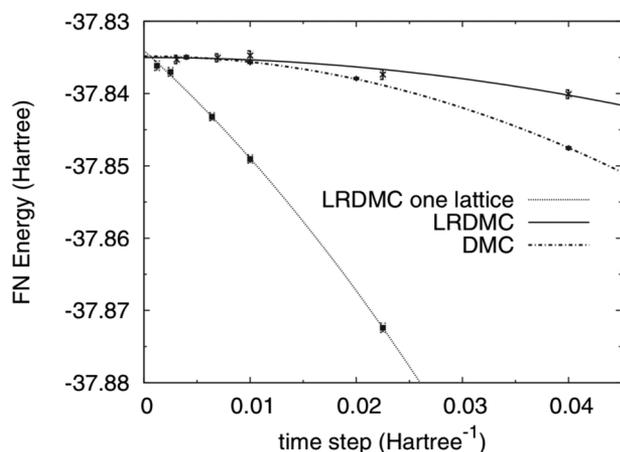
TURBORVB employs a many-body WF *Ansatz*  $\Psi$ , which can be written as the product of two terms,

$$\Psi = \Phi_{\text{AS}} \times \exp J, \quad (28)$$

where the term  $\exp J$ , conventionally dubbed Jastrow factor, is symmetric under electron exchange and the term  $\Phi_{\text{AS}}$ , also referred to as the determinant part of the WF, is antisymmetric. The resulting WF  $\Psi$  is antisymmetric, thus fermionic.

In the majority of QMC applications, the chosen  $\Phi_{\text{AS}}$  is a single Slater determinant (SD)  $\Phi_{\text{SD}}$ , i.e., an antisymmetrized product of single-electron WFs. Clearly, SD alone does not include any correlation other than the exchange. However, when a Jastrow factor, explicitly depending on the inter-electronic distances, is applied to  $\Phi_{\text{SD}}$ , the resulting *Ansatz*  $\Psi_{\text{JSD}} = \Phi_{\text{SD}} * \exp J$  often provides over 70% of the correlation energy<sup>55</sup> at the variational level. Thus, the Jastrow factor proves very effective in describing the correlation, employing only a relatively small number of parameters and therefore providing a very efficient way to improve the *Ansatz*.<sup>56</sup> A Jastrow correlated SD (JSD) function yields a computational cost for QMC simulations—both VMC and FN—about  $\propto N^3$ , namely, the same scaling of most DFT codes. Therefore, although QMC has a much larger prefactor, it represents an approach much cheaper than traditional quantum chemistry ones, at least for large enough systems.

While the JSD *Ansatz* is quite satisfactory in several applications, there are situations where very high accuracy is required, and a more advanced *Ansatz* is necessary. The route to improve JSD is



**FIG. 2.** Fixed-node energies for the all-electron carbon atom computed within DMC, single-grid LRDMC (one lattice), and double-grid LRDMC. The lattice discretization parameter  $a$  is mapped to the time step  $\tau$  as  $a = \sqrt{\tau}$ . Reprinted with permission from M. Casula, C. Filippi, and S. Sorella, Phys. Rev. Lett. **95**, 100201 (2005). Copyright 2005 APS.

not unique, and different approaches have been attempted within the QMC community. First, it should be mentioned that improving the Jastrow factors is not an effective approach to achieve higher accuracy at the FN level as the Jastrow is positive and cannot change the nodal surface. A popular approach is through the employment of backflow,<sup>57</sup> which is a remapping of the electronic configurations that enter into  $\Phi_{AS}$  (SD as a special case) where each electron position is appropriately changed depending on nearby electrons and nuclei. Backflow is an effective way to recover correlation energy, both at the variational and FN levels. However, it can be used at a price to increase significantly an already large computational cost.<sup>58</sup> Another possibility is to improve  $\Psi_{SD}$  similar to conventional quantum chemistry approaches, namely, by considering  $\Phi_{AS}$  as a linear expansion of several Slater determinants. While this second approach can provide very high accuracy, it may be extremely expensive as the number of determinants necessary to remain with a pre-defined accuracy grows combinatorially with the system size.

The vision embraced in TURBORVB is that the route toward an improved *Ansatz* should not compromise the efficiency and good scaling of QMC. For this reason, neither backflow nor explicit multideterminant expansions are implemented in the code. Within the TURBORVB project, the main goal is instead to consider an *Ansatz* that can be implicitly equivalent to a multideterminant expansion but remains in practice as efficient as a single determinant. There are five alternatives for the choice of  $\Phi_{AS}$ , which correspond to (i) the Pfaffian (Pf), (ii) the Pfaffian with a constrained number of molecular orbitals (Pfn), (iii) the Antisymmetrized Geminal Power (AGP), (iv) the Antisymmetrized Geminal Power with a constrained number of molecular orbitals (AGPn), and (v) the single Slater determinant. It is interesting to observe that the latter four WFs are all obtained by introducing specific constraints on the most general Pf *Ansatz*. The hierarchy of the five *Ansätze* is represented in the Venn diagram of Fig. 3. Clearly, a more general *Ansatz* is more accurate in the total energy but not necessarily in the energy differences.

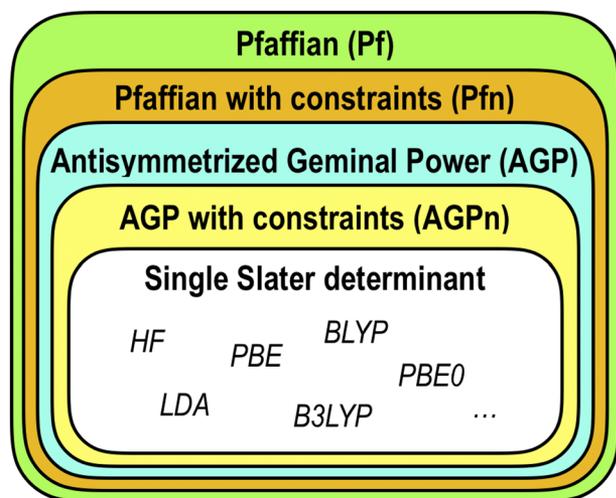


FIG. 3. *Ansatz* hierarchy. The output of Hartree–Fock (HF) or DFT simulations with different exchange–correlation functionals are special instances of the SD *Ansatz*.

Moreover, it is described by more variational parameters that could imply a more challenging optimization and a slightly higher cost. TURBORVB includes several tools to go from one *Ansatz* to another, as represented in Fig. 4. Typically, the starting point is SD, which can be obtained from a Hartree–Fock (HF) or a DFT calculation with different exchange–correlation functionals. Both methods are not expected to provide the optimal parameters when the Jastrow factor is included in the WF. Indeed, in QMC, the WFs are always meant to include the Jastrow factor, which proves fundamental to improve the properties of the overall WF. For instance, AGP carries more correlation than SD. However, it is not size-consistent unless it is multiplied by a Jastrow factor. Thus, a fundamental step to take advantage of the WF *Ansatz* is the possibility to perform reliable optimizations of the parameters. Optimization will be discussed in Sec. VII. In this section, we will describe the functional form of the Jastrow factor implemented in TURBORVB (Sec. III A), the Pfaffian (Sec. III B), the AGP (Sec. III C), the AGPn and Pfn (Sec. III D), the SD (Sec. III E), the multiconfigurational character of the AGP (Sec. III F), the basis set (Sec. III G), the pseudopotentials (Sec. III H), the contractions of the orbitals (Sec. III I), and the conversion tool (Sec. III J).

#### A. Jastrow factor (J)

The Jastrow factor ( $\exp J$ ) plays an important role in improving the correlation of the WF and in fulfilling Kato's cusp conditions.<sup>59</sup> TURBORVB implements the Jastrow term composed of one-body, two-body, and three/four-body factors ( $J = J_1 + J_2 + J_{3/4}$ ). The one-body and two-body factors are used to fulfill the electron–ion and electron–electron cusp conditions, respectively, and the three/four-body factors are employed to consider a systematic expansion, in principle converging to the most general electron pair contribution. The one-body Jastrow factor  $J_1$  is the sum of two parts, the

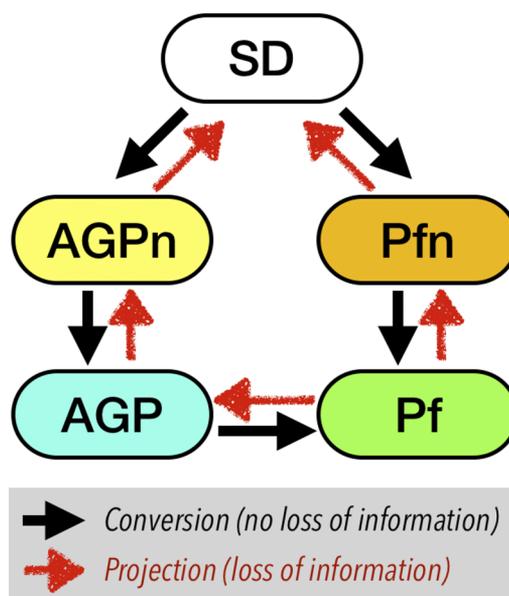


FIG. 4. *Ansatz* conversion.

homogeneous part (enforcing the electron–ion cusp condition),

$$J_1^h(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i=1}^N \sum_{a=1}^{N_{at}} \left( -(2Z_a)^{3/4} u_a \left( (2Z_a)^{1/4} |\mathbf{r}_i - \mathbf{R}_a| \right) \right), \quad (29)$$

and the corresponding inhomogeneous part,

$$J_1^{inh}(\mathbf{r}_1\sigma_1, \dots, \mathbf{r}_N\sigma_N) = \sum_{i=1}^N \sum_{a=1}^{N_{at}} \left( \sum_l M_{a,l}^{\sigma_i} \chi_{a,l}(\mathbf{r}_i) \right), \quad (30)$$

where  $\mathbf{r}_i$  are the electron positions,  $\mathbf{R}_a$  are the atomic positions with corresponding atomic number  $Z_a$ ,  $l$  runs over atomic orbitals  $\chi_{a,l}$  (e.g., GTO) centered on the atom  $a$ ,  $\sigma_i$  represents the electron spin ( $\uparrow$  or  $\downarrow$ ),  $\{M_{a,l}^{\sigma_i}\}$  are variational parameters, and  $u_a(\mathbf{r})$  is a simple bounded function. In TURBORVB, the most common choice for  $u_a$  is

$$u_a(r) = \frac{1}{2b_{ea}} \left( 1 - e^{-rb_{ea}} \right), \quad (31)$$

depending on a single variational parameter  $b_{ea}$ , which may be optimized independently for each atomic species.

The two-body Jastrow factor is defined as

$$J_2(\mathbf{r}_1\sigma_1, \dots, \mathbf{r}_N\sigma_N) = \sum_{i < j} v_{\sigma_i, \sigma_j}(|\mathbf{r}_i - \mathbf{r}_j|), \quad (32)$$

where  $v_{\sigma_i, \sigma_j}$  is another simple bounded function. There are several possible choices for  $v_{\sigma_i, \sigma_j}$  implemented in TURBORVB (all listed in the file input.tex in the doc folder), and one of them is, for instance, the following spin-dependent form:

$$v_{\sigma_i, \sigma_j}(r_{ij}) = \begin{cases} \frac{r_{ij}}{4} \cdot (1 + b_{ee}^{\text{para}} \cdot r_{ij})^{-1} & (\sigma_i = \sigma_j) \\ \frac{r_{ij}}{2} \cdot (1 + b_{ee}^{\text{anti}} \cdot r_{ij})^{-1} & (\sigma_i \neq \sigma_j), \end{cases} \quad (33)$$

where  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ , and  $b_{ee}^{\text{para}}$  and  $b_{ee}^{\text{anti}}$  are variational parameters.

The three/four-body Jastrow factor reads

$$J_{3/4}(\mathbf{r}_1\sigma_1, \dots, \mathbf{r}_N\sigma_N) = \sum_{i < j} \left( \sum_{a,l} \sum_{b,m} M_{a,l,b,m}^{\sigma_i, \sigma_j} \chi_{a,l}(\mathbf{r}_i) \chi_{b,m}(\mathbf{r}_j) \right), \quad (34)$$

where the indices  $l$  and  $m$  again indicate different orbitals centered on corresponding atoms  $a$  and  $b$ , and  $\{M_{a,l,b,m}^{\sigma_i, \sigma_j}\}$  are variational parameters. Sometimes, it is convenient to set to zero part of the coefficients of the four-body Jastrow factor, namely, those corresponding to  $a \neq b$ , as they increase the overall variational space significantly and make the optimization more challenging, without being much more effective in improving the variational WF.

## B. Pfaffian wave function (Pf)

The SD is an antisymmetrized product of single-electron WFs. Thus, SD neglects almost entirely the correlation between electrons. A natural way to improve this description is to include explicitly in the Ansatz pairwise correlations among electrons. This is precisely what the Pfaffian WF does.<sup>60–62</sup> The building block of the Pfaffian WF is the pairing function  $g(\mathbf{i}, \mathbf{j})$  between any pair of electrons  $i$  and  $j$ . Henceforth, we denote with the generic bolded index  $\mathbf{i}$  both the

space  $\mathbf{r}_i$  coordinates and the spin values  $\sigma_i$ ,

$$\mathbf{i} \leftrightarrow (\mathbf{r}_i, \sigma_i), \quad (35)$$

corresponding to the  $i$ th electron.

For simplicity, let us first consider a system with an even number  $N$  of electrons. The WF, written in terms of pairing functions, is

$$\Phi_{AS}(\mathbf{1}, \dots, \mathbf{N}) = \mathcal{A} \{ g(\mathbf{1}, \mathbf{2}) g(\mathbf{3}, \mathbf{4}) \dots g(\mathbf{N}-\mathbf{1}, \mathbf{N}) \}, \quad (36)$$

where  $\mathcal{A}$  is the antisymmetrization operator,

$$\mathcal{A} \equiv \frac{1}{N!} \sum_{P \in S_N} \epsilon_P \hat{P}, \quad (37)$$

$S_N$  is the permutation group of  $N$  elements,  $\hat{P}$  is the operator corresponding to the generic permutation  $P$ , and  $\epsilon_P$  is its sign.

Let us define  $G$  the  $N \times N$  matrix with elements  $G_{i,j} = g(\mathbf{i}, \mathbf{j})$ . Note that

$$g(\mathbf{i}, \mathbf{j}) = -g(\mathbf{j}, \mathbf{i}) \quad (\text{and } G_{i,j} = -G_{j,i}), \quad (38)$$

as a consequence of the statistics of fermionic particles; thus,  $G$  is skew-symmetric (i.e.,  $G^T = -G$ ,  $T$  being the transpose operator), so the diagonal is zero and the number of independent entries is  $\sum_{n=1}^{N-1} n = N(N-1)/2$ .

The Pfaffian<sup>63</sup> of a  $N \times N$  skew-symmetric matrix  $G$  is defined as

$$\text{Pf}(G) \equiv \frac{1}{2^{N/2} (N/2)!} \sum_{P \in S_N} \epsilon_P G_{P(1), P(2)} \dots G_{P(N-1), P(N)} \quad (39)$$

if  $N$  is even, and it is zero if  $N$  is odd.<sup>64</sup> Therefore, the WF  $\Phi_{AS}$  defined in the right-hand side (RHS) of Eq. (36) equals  $\frac{1}{N!} \text{Pf}(G)$ , where the semifactorial  $N!! \equiv \frac{N!}{2^{N/2} (N/2)!}$  is irrelevant in QMC as it affects only the normalization of the WF. Thus, we can define our electronic WF as

$$\Phi_{\text{Pf}} = \text{Pf}(G). \quad (40)$$

Note that the  $\Phi_{\text{Pf}}$  here defined allows the description of any system with  $N_u$  electrons with spin-up and  $N_d$  electrons with spin-down, provided that  $N = N_u + N_d$  is even. Indeed, with no loss of generality, we can assume that electrons  $i = 1, \dots, N_u$  have  $\sigma_i = 1/2$  and electrons with  $i = N_u + 1, \dots, N$  have  $\sigma_i = -1/2$ . Thus, the  $N \times N$  skew-symmetric matrix  $G$  is written as

$$G = \begin{bmatrix} G_{uu} & G_{ud} \\ G_{du} & G_{dd} \end{bmatrix}, \quad (41)$$

where  $G_{uu}$  is a  $N_u \times N_u$  skew-symmetric matrix with elements  $g_{uu}(\mathbf{i}, \mathbf{j})$ ,  $G_{dd}$  is a  $N_d \times N_d$  skew-symmetric matrix with elements  $g_{dd}(\mathbf{i}, \mathbf{j})$ ,  $G_{ud}$  is a  $N_u \times N_d$  matrix with elements  $g_{ud}(\mathbf{i}, \mathbf{j})$ , and  $G_{du} = -G_{ud}^T$ , i.e.,  $g_{du}(\mathbf{i}, \mathbf{j}) = -g_{ud}(\mathbf{j}, \mathbf{i})$ .  $g_{uu}$  describes the pairing between a pair of electrons with spin-up,

$$g_{uu}(\mathbf{i}, \mathbf{j}) = f_{uu}(\mathbf{r}_i, \mathbf{r}_j) |\uparrow\uparrow\rangle, \quad (42)$$

where the function  $f_{uu}$  describes the spatial dependence on the coordinates  $\mathbf{r}_i, \mathbf{r}_j$  for  $i, j \leq N_u$ . Note that  $f_{uu}(\mathbf{r}_j, \mathbf{r}_i) = -f_{uu}(\mathbf{r}_i, \mathbf{r}_j)$  as a consequence of the properties of  $g$ . The spin part  $|\uparrow\uparrow\rangle$  describes a system with unit total spin and spin projection along the  $z$ -axis and will

be indicated by  $|1, +1\rangle$ . Similarly,  $g_{dd}$  describes the pairing between pairs of electrons with spin-down for  $i, j > N_u$ ,

$$g_{dd}(\mathbf{i}, \mathbf{j}) = f_{dd}(\mathbf{r}_i, \mathbf{r}_j)|\downarrow\downarrow\rangle \quad (43)$$

with  $f_{dd}(\mathbf{r}_j, \mathbf{r}_i) = -f_{dd}(\mathbf{r}_i, \mathbf{r}_j)$ , and the spin part  $|\downarrow\downarrow\rangle$  describes a system with total unit spin and negative spin projection along the  $z$ -axis, indicated with  $|1, -1\rangle$ .  $g_{ud}$  describes the pairing between pairs of electrons with unlike spins. Since two electrons with unlike spins can form a singlet  $|0, 0\rangle = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}}$  or a triplet  $|1, 0\rangle = \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}}$ , in the general case, the pairing function  $g_{ud}$  will be a linear combination of the two components,

$$g_{ud}(\mathbf{i}, \mathbf{j}) = f_S(\mathbf{r}_i, \mathbf{r}_j) \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}} + f_T(\mathbf{r}_i, \mathbf{r}_j) \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}}, \quad (44)$$

where  $f_S(\mathbf{r}_i, \mathbf{r}_j) = f_S(\mathbf{r}_j, \mathbf{r}_i)$  describes the spatial dependence of the singlet part of  $g_{ud}$  and  $f_T(\mathbf{r}_i, \mathbf{r}_j) = -f_T(\mathbf{r}_j, \mathbf{r}_i)$  describes the spatial dependence of the triplet part. Therefore, the generic pairing function  $g(\mathbf{i}, \mathbf{j})$  is the sum of all four components mentioned above, namely,

$$g(\mathbf{i}, \mathbf{j}) = f_S(\mathbf{r}_i, \mathbf{r}_j)|0, 0\rangle + f_T(\mathbf{r}_i, \mathbf{r}_j)|1, 0\rangle + f_{uu}(\mathbf{r}_i, \mathbf{r}_j)|1, +1\rangle + f_{dd}(\mathbf{r}_i, \mathbf{r}_j)|1, -1\rangle. \quad (45)$$

The pairing functions  $f_S$ ,  $f_T$ ,  $f_{uu}$ , and  $f_{dd}$  are expanded over atomic orbitals (see Sec. III G). Say, for a generic pairing function  $f$ , we have

$$f(\mathbf{r}_i, \mathbf{r}_j) = \sum_{l,m,a,b} A_{\{a,l\},\{b,m\}} \psi_{a,l}(\mathbf{r}_i) \psi_{b,m}(\mathbf{r}_j), \quad (46)$$

where  $\psi_{a,l}$  and  $\psi_{b,m}$  are primitive or contracted atomic orbitals, their indices  $l$  and  $m$  indicate different orbitals centered on atoms  $a$  and  $b$ , while  $i$  and  $j$  label the electron coordinates. Symmetries on the system or properties of the underlying pairing function  $f$  imply constraints on the coefficients. For instance, the coefficients of  $f_S$  are such that  $A_{\{a,l\},\{b,m\}} = A_{\{b,m\},\{a,l\}}$  because  $f_S(\mathbf{r}_i, \mathbf{r}_j) = f_S(\mathbf{r}_j, \mathbf{r}_i)$ , whereas  $A_{\{a,l\},\{b,m\}} = -A_{\{b,m\},\{a,l\}}$  for  $f_T$ ,  $f_{uu}$ , and  $f_{dd}$ .

Let us consider now the remaining case of a system with an odd number of electrons  $N$ . The simplest way to handle this case is to consider a system with an extra fictitious electron  $\mathbf{N} + 1 \leftrightarrow (\mathbf{r}_{N+1}, \sigma_{N+1})$  that we set at infinity  $\mathbf{r}_{N+1} \rightarrow \infty$ , thus non-interacting with all  $N$  physical electrons. The extra matrix elements are easily computed,

$$\begin{aligned} g(\mathbf{N}+1, \mathbf{N}+1) &= 0, \\ g(\mathbf{i}, \mathbf{N}+1) &= -g(\mathbf{N}+1, \mathbf{i}) = \phi(\mathbf{i}) \quad \text{for } i \leq N, \end{aligned} \quad (47)$$

where  $\phi(\mathbf{i}) \equiv \phi(\mathbf{r}_i, \sigma_i)$  can be considered an extra spin dependent unpaired orbital vanishing at infinity.

The antisymmetric part of the WF of the overall system is then

$$\Phi_{AS}(\mathbf{1}, \dots, \mathbf{N}) = \mathcal{A}\{g(\mathbf{1}, \mathbf{2})g(\mathbf{3}, \mathbf{4})\dots g(\mathbf{N}, \mathbf{N}+1)\}, \quad (48)$$

where the antisymmetrization operator acts on the (now even)  $N + 1$  electrons. This is a perfectly allowed  $N$  electron fermionic WF because by definition (i) it is antisymmetric for all permutations of the  $N + 1$  particles and, in particular, for the  $N$ -physical ones and (ii) it does not depend on the extra  $N + 1$ th coordinates

and spin as the fictitious particle is at infinity. On the other hand, it is easy to show that this is a nontrivial and nonvanishing WF because we can use the basic Pfaffian formula in Eq. (39) for the antisymmetrization of the  $N + 1$  particles and we readily obtain that

$$\Phi_{Pf} = \text{Pf}(\tilde{G}), \quad (49)$$

where

$$\tilde{G} = \begin{bmatrix} G & \Phi \\ -\Phi^T & 0 \end{bmatrix} \quad (50)$$

is a  $(N + 1) \times (N + 1)$  skew-symmetric matrix and  $\Phi$  is a  $N$ -dimensional vector whose elements are  $\phi(\mathbf{i})$ ,  $i = 1, \dots, N$ . Its Pfaffian is quite generally nonvanishing for at least some configuration, provided that the diagonalization of the Pfaffian (see later) has at least  $(N + 1)/2$  non-zero eigenvalues (likewise for the Slater determinant, it is enough that the  $N$ -molecular orbitals are linearly independent), and we obtain in this way that the Pfaffian is perfectly defined even in the odd number of electrons case.

Finally, we would like to emphasize that the above argument can be generalized to arbitrary number  $k$  of unpaired orbitals and arbitrary boundary conditions including calculations on periodic supercells. We can indeed assume at the beginning that when  $N$  is odd (even), we add an odd (even) number  $k$  of unpaired orbitals such that there are  $N$  electrons coupled by the matrix  $G$  plus  $k$  unpaired electrons. The  $k$  unpaired electrons are paired with  $k$  fictitious electrons [with a pairing function satisfying conditions analogous to those in Eq. (47)]. This yields a Pfaffian of even linear dimension  $N + k$  of the same form as the one in Eq. (50) but with  $\Phi$  being a block  $N \times k$  rectangular matrix, determined by  $k$  different spin-dependent orbitals. The same argument holds that the antisymmetrized WF is written as the Pfaffian of the corresponding  $(N + k) \times (N + k)$  skew-symmetric matrix and does not depend on the fictitious particle coordinates introduced and therefore is an allowed physical WF antisymmetric over the  $N$  physical particles.

A special case that is of interest is when all  $N$  electrons are set to be unpaired (i.e., when  $k = N$  and  $\Phi$  is a generic  $N \times N$  matrix) and it is assumed that there is no pairing among them [i.e.,  $G = 0$  in the  $N \times N$  block matrix in Eq. (50)]. Thus, using a well-known relation of a Pfaffian

$$\text{Pf} \left( \begin{array}{c|c} 0 & \Phi \\ \hline -\Phi^T & 0 \end{array} \right) = (-1)^{\frac{N \times (N-1)}{2}} \det(\Phi), \quad (51)$$

we recover the standard Slater determinant. In this way, we can clearly see that the Pfaffian represents a quite remarkable generalization of the single determinant picture and that a larger variational freedom is exploited by allowing pairing correlations with a non-zero matrix  $G$ .

Recently, we have noticed that, following the above-mentioned derivation, it is possible to generalize further the Pfaffian WF *Ansatz* by introducing fictitious particles and relaxing the constraints on the  $\tilde{G}$  matrix, even in the lower block diagonal part, which can be assumed to be an arbitrary non-zero skew-symmetric matrix. Work is in progress in order to understand whether this extended formulation can further improve the accuracy within a single Pfaffian WF.

### C. Antisymmetrized geminal power (AGP)

If we consider only the case of a pairing function  $g(\mathbf{i}, \mathbf{j})$  that is a spin singlet [namely,  $f_{uu}, f_{dd}, f_T$  in Eq. (45) are set to zero, yielding  $g(\mathbf{i}, \mathbf{j}) = f_S(\mathbf{r}_i, \mathbf{r}_j)|0, 0\rangle$ ], then we obtain the singlet Antisymmetrized Geminal Power (denoted as AGP).

Let us consider first an unpolarized system, having an even number  $N$  of electrons, and without loss of generality, we can assume that the electrons  $i = 1, \dots, N/2$  have spin up and electrons  $j = N/2 + 1, \dots, N$  have spin down. Then, the matrices  $G_{uu}$  and  $G_{dd}$  defined in Sec. III B are both zero matrices of size  $N/2 \times N/2$ , and the matrix  $G_{ud}$  has only the contribution coming from the singlet, which we dub  $G_S$ . The antisymmetrization operator implies the computation of

$$\text{Pf}\left(\begin{array}{c|c} 0 & G_S \\ \hline -G_S^T & 0 \end{array}\right) = (-1)^{\frac{N/2 \times (N/2 - 1)}{2}} \det(G_S), \quad (52)$$

where the equality follows from a property of the Pfaffian [see Eq. (51)]. The overall sign is arbitrary for a WF; thus, the antisymmetrized product of singlet pairs (geminals) is indeed equivalent to the computation of the determinant of the matrix  $G_S$ ,

$$\Phi_{\text{AGPs}} = \det(G_S). \quad (53)$$

It should be noted that it is not necessary that the matrix  $G_{ud}$  be symmetric to reduce the Pfaffian to a single determinant evaluation. As long as the matrices  $G_{uu}$  and  $G_{dd}$  are zero, the Pfaffian is indeed equivalent to  $\det(G_{ud})$ , and describes an antisymmetric WF. However, if  $G_{ud}$  is not symmetric, the function

$$\Phi_{\text{AGP}} = \det(G_{ud}) \quad (54)$$

is not an eigenstate of the spin. In other terms, there is a spin contamination, similar to the case of unrestricted HF calculations.

The AGP *Ansatz* can be generalized to describe polarized systems, i.e., systems where the number  $N_u$  of electrons with spin up is different from the number  $N_d$  of electrons with spin down. With no loss of generality, we can assume that  $N_u > N_d$ ; thus, the system is constituted by a number  $p = N_d$  of electron pairs and a number  $k = N_u - N_d$  of unpaired electrons (clearly,  $N = N_u + N_d = 2p + k$ ). We aim at evaluating

$$\mathcal{A}\{g(\mathbf{1}, \mathbf{2}) \cdots g(\mathbf{2p} - \mathbf{1}, \mathbf{2p}) \phi_1(\mathbf{2p} + \mathbf{1}) \cdots \phi_k(\mathbf{N})\}. \quad (55)$$

As mentioned above, we assume that the pairing function is zero for like-spin pairs. With no loss of generality, we can assume that electrons  $i = 1, \dots, N_u$  have spin up and electrons  $j = N_u + 1, \dots, N$  have spin down. Moreover, as was done at the end of Subsection III B, we can add  $k$  fictitious entries to  $g(\mathbf{i}, \mathbf{j})$  such that  $g(\mathbf{i}, \mathbf{N} + \mathbf{1}) = \phi_l(\mathbf{i})$  for  $l = 1, \dots, k$  and  $i = 1, \dots, N_u$ . Thus, the antisymmetrization implies the use of Eq. (51), i.e.,

$$\text{Pf}\left(\begin{array}{c|c} 0 & \tilde{G} \\ \hline -\tilde{G}^T & 0 \end{array}\right) = (-1)^{\frac{N_u \times (N_u - 1)}{2}} \det(\tilde{G}), \quad (56)$$

with the  $N_u \times N_u$  matrix  $\tilde{G} = [G_{ud} | \Phi]$ , the  $N_u \times N_d$  matrix  $G_{ud}$  describing the pairing between the  $N_u$  spin up electrons and the  $N_d$  spin down electrons, and the  $N_u \times k$  matrix  $\Phi$  describing the

$k$  unpaired orbitals. Thus, we need to evaluate

$$\Phi_{\text{AGP}} = \det(\tilde{G}). \quad (57)$$

One of the most important advantages of the AGP *Ansatz* is that it is equivalent to a linear combination of Slater determinants (i.e., multi configurations), but the computational cost remains at the level of a single-determinant one<sup>65–67</sup> (see Sec. III F). This is especially important for large systems because the naive multireference approach requires an exponentially large number of Slater determinants, which drastically increases the computational cost. The AGP *Ansatz* was applied to the *ab initio* calculation by Casula and Sorella<sup>65</sup> for the first time in 2003; then, it has also been implemented in other QMC codes.

### D. AGP and Pf with a constrained number of molecular orbitals (AGPn and Pfn)

A convenient way to impose constraints on the variational parameters defining the AGP or Pf WF is obtained by rewriting the expansion of the geminal in terms of molecular orbitals (MOs). As shown in Eq. (46), a geminal  $g(\mathbf{i}, \mathbf{j})$  is natively expressed in terms of the atomic orbitals  $\{\psi_{a,l}(\mathbf{r})\}$ , by summing over all the atoms  $a$ , the corresponding orbitals  $l$ , and the spin index  $\sigma$  (as the elements of the basis may in principle also depend on the spin, even though the most common choice is to take the same orbital for each of the two possible values of the spins). In order to simplify the notation here, let us merge the atomic orbital and spin indices in a unique one that is indicated with a Greek symbol [e.g.,  $\mu \leftrightarrow (a, l)$ ] running from 1 to the total dimension of the  $2L$  atomic orbitals used,  $L$  for each spin component. Therefore,

$$g(\mathbf{i}, \mathbf{j}) = \sum_{\mu, \nu}^{2L} A_{\mu, \nu} \psi_{\mu}(\mathbf{i}) \psi_{\nu}(\mathbf{j}), \quad (58)$$

and clearly, the symmetry of  $g$  implies that  $A_{\mu, \nu} = -A_{\nu, \mu}$ . The coefficients  $A_{\mu, \nu}$  define a  $2L \times 2L$  skew-symmetric matrix  $A$ . If we define the  $2L$  dimensional vector  $\Psi_i = (\psi_1(\mathbf{i}), \dots, \psi_{2L}(\mathbf{i}))^T$ , Eq. (58) rewrites as  $g(\mathbf{i}, \mathbf{j}) = \Psi_i^T A \Psi_j$ . Moreover, the overlaps  $S_{\mu, \nu} \equiv \langle \psi_{\mu} | \psi_{\nu} \rangle$  between atomic orbitals define the overlap matrix  $S$ , which in the case of a spin-dependent basis is block diagonal,

$$S = \left[ \begin{array}{c|c} S_{uu} & 0 \\ \hline 0 & S_{dd} \end{array} \right], \quad (59)$$

with  $S_{uu}$  and  $S_{dd}$  being positive definite  $L \times L$  square matrices ( $S_{uu} = S_{dd}$  when orbitals are the same for spin up and spin down). In TURBORVB, the overlap matrix  $S$  is computed on a suitable uniform mesh with an efficient and general parallel algorithm. Then, an orthonormal basis is defined,

$$\tilde{\psi}_{\mu}(\mathbf{i}) = \sum_{\nu} S_{\mu, \nu}^{-1/2} \psi_{\nu}(\mathbf{i}), \quad (60)$$

where  $S^{-1/2}$  is well defined since  $S$  is strictly positive definite.<sup>68</sup> The matrix elements of  $g$  can be recasted in the new orthonormal basis  $\{\tilde{\psi}_{\mu}(\mathbf{i})\}$ , yielding  $g(\mathbf{i}, \mathbf{j}) = \sum_{\mu, \nu}^{2L} \tilde{A}_{\mu, \nu} \tilde{\psi}_{\mu}(\mathbf{i}) \tilde{\psi}_{\nu}(\mathbf{j}) = \tilde{\Psi}_i^T \tilde{A} \tilde{\Psi}_j$ , with the matrix  $\tilde{A} \equiv S^{1/2} A S^{1/2}$  that is skew symmetric.

At this point, from the spectral theory of skew symmetric matrices, it is possible to perform the Youla decomposition of  $\tilde{A}$  (see

Ref. 69), which can be written in the form  $\bar{A} = Q\Sigma Q^T$ , where  $Q$  is unitary (also real if  $\bar{A}$  is real) and the matrix  $\Sigma$  is block diagonal with  $\Sigma_{2k-1,2k} = \lambda_k = -\Sigma_{2k,2k-1}$  for  $k = 1, \dots, L$  and zero everywhere else, with  $\lambda_k \geq 0$ .<sup>70</sup> So, the pairing function  $g(\mathbf{i}, \mathbf{j})$  can be written as  $\Phi_i^T \Sigma \Phi_j$ , where  $\Phi_i = Q^T \tilde{\Psi}_i$  for each  $i$ . This defines a basis of  $L$  MOs  $\{\phi_k(i)\}$  and corresponding twinned ones  $\{\tilde{\phi}_k(i)\}$ , forming together a basis of  $2L$  mutually orthonormal elements for which the original geminal function reads

$$g(\mathbf{i}, \mathbf{j}) = \sum_{k=1}^L \lambda_k [\phi_k(\mathbf{i}) \tilde{\phi}_k(\mathbf{j}) - \tilde{\phi}_k(\mathbf{i}) \phi_k(\mathbf{j})] \quad (61)$$

with  $\lambda_k \geq 0$ . After these transformations, these MOs can be finally written in the chosen (hybrid) atomic basis,

$$\begin{aligned} \phi_k(\mathbf{i}) &= \sum_{\nu=1}^{2L} P_{\mu,\nu} \Psi_{\nu}(\mathbf{i}), \\ \tilde{\phi}_k(\mathbf{i}) &= \sum_{\nu=1}^{2L} \bar{P}_{\mu,\nu} \Psi_{\nu}(\mathbf{i}) \end{aligned} \quad (62)$$

by appropriate  $p \times 2L$  rectangular matrices  $P$  and  $\bar{P}$ . Then, with no loss of generality, we can assume that the molecular orbitals  $\{\phi_k(\mathbf{i}), \tilde{\phi}_k(\mathbf{i})\}$  are ranked such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ . The above expression highlights that the most important MOs are those corresponding to the larger values of  $\lambda_k$ . Therefore, it is possible to constrain the variational freedom by neglecting all orbitals with  $k > n$ , yielding the pairing function

$$g^n(\mathbf{i}, \mathbf{j}) = \sum_{k=1}^n \lambda_k [\phi_k(\mathbf{i}) \tilde{\phi}_k(\mathbf{j}) - \tilde{\phi}_k(\mathbf{i}) \phi_k(\mathbf{j})], \quad (63)$$

where  $n$  is conveniently chosen and is  $\ll L$ . This yields the AGPn *Ansatz* and the  $\Phi_{\text{AGPn}}$  WF, which can be useful to improve the stability of the wave-function optimization. The corresponding algorithm (enabled by setting `molopt = -1` in the optimization section input of `TURBORV6`), based on projection operators in the space of the  $n$  molecular orbitals considered, has been described extensively in Ref. 15. Moreover, in the original paper<sup>71</sup> introducing the AGPn, a precise recipe was given to improve the evaluation of the binding energies. Indeed, despite a constraint on the variational parameters necessarily increasing the variational energy expectation value, energy differences may actually improve by an appropriate choice of  $n$ . In the mentioned work,<sup>71</sup> this promising approach was applied with an AGP containing only singlet correlations, but the binding energies were defined without using a rigorous size consistent criterion. This drawback can now be removed, by exploiting the full variational freedom of the Pf WF combined with a general spin-dependent Jastrow factor (see, e.g., Fig. 14). Work is in progress in this interesting research direction.

The variational optimization of an AGP with a fixed number  $n$  of molecular orbitals can be easily generalized to the Pf case, by exploiting that the constrained Pf WF, dubbed Pfn, can be written either in the canonical form with MOs, as in Eq. (63), or in the localized basis set expansion, as in Eq. (58), with a corresponding matrix  $A_{\mu,\nu}^n = \sum_{k=1}^n \lambda_k [P_{\mu,k} \bar{P}_{\nu,k} - P_{\nu,k} \bar{P}_{\mu,k}]$ . According to Eq. (63), an arbitrary small variation  $\delta g_n$  of the constrained pairing function  $g_n$  reads

$$\begin{aligned} \delta g_n(\mathbf{i}, \mathbf{j}) &= \sum_{k=1}^n \delta \lambda_k [\phi_k(\mathbf{i}) \tilde{\phi}_k(\mathbf{j}) - \tilde{\phi}_k(\mathbf{i}) \phi_k(\mathbf{j})] \\ &+ \sum_{k=1}^n \lambda_k [\delta \phi_k(\mathbf{i}) \tilde{\phi}_k(\mathbf{j}) - \delta \tilde{\phi}_k(\mathbf{i}) \phi_k(\mathbf{j})] \\ &+ \sum_{k=1}^n \lambda_k [\phi_k(\mathbf{i}) \delta \tilde{\phi}_k(\mathbf{j}) - \tilde{\phi}_k(\mathbf{i}) \delta \phi_k(\mathbf{j})] \end{aligned} \quad (64)$$

and therefore satisfies the following property, as will be shown later,

$$(\hat{I} - \mathbf{L}) \delta g_n (\hat{I} - \mathbf{R}) = 0, \quad (65)$$

where  $\hat{I}$  is the identity operator,  $\mathbf{L}$  and  $\mathbf{R}$  are projection operators over the occupied MOs, i.e.,  $\mathbf{L}^2(\mathbf{i}, \mathbf{j}) = \int d\mathbf{k} \mathbf{L}(\mathbf{i}, \mathbf{k}) \mathbf{L}(\mathbf{k}, \mathbf{j}) = \mathbf{L}(\mathbf{i}, \mathbf{j})$ , and similarly,  $\mathbf{R}^2 = \mathbf{R}$ , where here and henceforth the shorthand integration symbol  $\int d\mathbf{k} = \sum_{\sigma_k} \int d\mathbf{r}_k$  contains implicitly also the spin summation. These operators are then defined as follows:

$$\begin{aligned} \mathbf{L}(\mathbf{i}, \mathbf{j}) &= \sum_{k=1}^n [\phi_k(\mathbf{i}) \phi_k^*(\mathbf{j}) + \tilde{\phi}_k(\mathbf{i}) \tilde{\phi}_k^*(\mathbf{j})], \\ \mathbf{R}(\mathbf{i}, \mathbf{j}) &= \sum_{k=1}^n [\phi_k^*(\mathbf{i}) \phi_k(\mathbf{j}) + \tilde{\phi}_k^*(\mathbf{i}) \tilde{\phi}_k(\mathbf{j})]. \end{aligned} \quad (66)$$

With the above definitions, Eq. (64) is easily verified because each term of Eq. (64) is annihilated either by the left  $(\hat{I} - \mathbf{L})$  or the right  $(\hat{I} - \mathbf{R})$  projection over the unoccupied MOs. Note that  $\mathbf{L} = \mathbf{R}$  in the real case and  $\mathbf{L} = \mathbf{R}^*$  in the most general complex case. In this way, in order to implement a constrained variation  $\delta g_n$  of the Pfn WF, corresponding to an appropriate variation of its matrix  $\delta A_{\mu,\nu}^n$ , it is useful to work with a small free variation  $\delta g$  (with corresponding  $\delta A_{\mu,\nu}$ ). This is then projected onto the chosen restricted *Ansatz* by means of the following equation:

$$\delta g_n = \delta g - (\hat{I} - \mathbf{L}) \delta g (\hat{I} - \mathbf{R}). \quad (67)$$

Indeed, it is easy to show that the RHS of the above equation vanishes if we apply  $\hat{I} - \mathbf{L}$  and  $\hat{I} - \mathbf{R}$  to its left and its right, respectively, just because  $\hat{I} - \mathbf{R}$  and  $\hat{I} - \mathbf{L}$  are projection operators, being such  $\mathbf{R}$  and  $\mathbf{L}$ , yielding  $(\hat{I} - \mathbf{R})^2 = (\hat{I} - \mathbf{R})$  and  $(\hat{I} - \mathbf{L})^2 = (\hat{I} - \mathbf{L})$ , from which Eq. (67) fulfills Eq. (65). Equation (67) represents, therefore, a *linear relation* applied to the variational parameter matrix change  $\delta A_{\mu,\nu}$  corresponding to the unconstrained geminal  $g$  in Eq. (58), yielding the new constrained variation  $\delta A_{\mu,\nu}^n$ . Indeed, by using the definitions of the projector operators in Eq. (66) and the expansion of the MOs in the atomic (hybrid) basis [see Eq. (62)], Eq. (67) turns to a number of matrix-matrix operations acting on  $\delta A$ ,  $P$ ,  $\bar{P}$ , and the overlap matrix  $S$  that can be easily and efficiently implemented.<sup>15</sup>

This linear relation between  $A$  and  $A^n$  can be therefore easily implemented together with the corresponding derivatives necessary for the optimization of the energy<sup>72</sup> and allows the explicit calculation of the new matrix  $A_{\mu,\nu}^n$ , yielding the new constrained geminal  $g_n + \delta g_n$ . Then, the new geminal can be recasted in the form of Eq. (63) by the mentioned diagonalization of skew-symmetric matrices, in this way implicitly neglecting nonlinear contributions that are irrelevant close to convergence, when  $\delta g_n \rightarrow 0$ . After employing several iterations of this type, the lowest energy *Ansatz* of the JPfn type can be obtained in a relatively simple and very efficient way.

Finally, it is also important to emphasize that this constrained optimization algorithm allows the reduction of the number of parameters by efficiently exploiting locality, namely that variational parameters  $A_{\mu,\nu}$  corresponding to atoms at a distance larger than a reasonable cutoff (an input named `rmax` in `TURBORVB`) can be safely disregarded with negligible error.<sup>15</sup>

### E. Single Slater determinant (SD)

An important special case of the AGPn and Pfn *Ansätze* discussed in Sec. III D is when we constrain the pairing function  $g_n$  to use the minimum possible number of molecular orbitals providing a non-zero WF. The minimum number for an unpolarized system with  $N$  electrons is equal to the number of electron pairs, that is,  $n = N/2$ . The WF obtained in this way starting from the AGP is indeed the single Slater determinant (SD),<sup>15,71</sup> and we dub it  $\Phi_{\text{SD}}$ . In principle, also the Pfn WF with  $n = N/2$  corresponds to a single Slater determinant with spin dependent molecular orbitals, a case that has never been considered so far, but it represents an available option within the most recent versions of `TURBORVB`.

Similarly, in a polarized system having  $N_u$  spin up electrons and  $N_d$  spin down electrons (we assume  $N_u > N_d$ ), the SD *Ansatz* is obtained by using  $N_u - N_d$  unpaired electrons and using a constrained pairing function  $g_n$  with  $n = N_d/2$ .

### F. Implicit multiconfigurational character of the AGP

In Secs. III B and III C, it has been mentioned that the Pfaffian and the AGP *Ansätze* have a multiconfigurational character despite the fact that they can be evaluated at the cost of a single determinant. In this section, we expand the AGP *Ansatz* in terms of Slater determinants to show this aspect explicitly. In order to simplify the derivation, we will consider here a simplified case, while the most general case could be studied with a similar approach but involving more cumbersome expressions.

We consider the real AGPs *Ansatz* [Eq. (53)] for an unpolarized system of  $N = 2N_p$  electrons (i.e.,  $N_u = N_d = N/2$ ). The symmetry implies that the twin molecular orbitals  $\{\phi_k(\mathbf{i}), \tilde{\phi}_k(\mathbf{i})\}$  appearing in the RHS of Eq. (61) have the same spatial part, which we denote  $\tilde{\phi}_k(\mathbf{r})$ , modulus a sign (because they are orthonormal), and they have an opposite spin part. Without the loss of generality, we can assume that  $\phi_k$  is spin up and  $\tilde{\phi}_k$  is spin down. Given this convention, the scalar product  $l_k$  between the spatial parts of  $\phi_k$  and  $\tilde{\phi}_k$  will either be +1 or -1. We define  $\tilde{\lambda}_k = l_k \lambda_k$ , where  $\lambda_k$  is the same as the one in the RHS of Eq. (61). Note that  $|\tilde{\lambda}_k| = \lambda_k$ , so  $\{\tilde{\lambda}\}$  are ranked in decreasing order of their absolute value (i.e.,  $|\tilde{\lambda}_k| \geq |\tilde{\lambda}_{k+1}|$ ). The pairing function can then be written as

$$g(\mathbf{i}, \mathbf{j}) = \sum_{k=1}^L \tilde{\lambda}_k \tilde{\phi}_k(\mathbf{r}_i) \tilde{\phi}_k(\mathbf{r}_j) (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \quad (68)$$

This expression is useful for comparing with the standard CI expansion.

It is convenient now to use the second quantization notations in order to simplify the derivation. In particular, we indicate with  $\hat{a}_{k,\uparrow}^\dagger$  ( $\hat{a}_{k,\downarrow}^\dagger$ ) the operator that creates an electron of spin up (down) in the orbitals  $\tilde{\phi}_k$  and satisfies the canonical anticommutation relations. We can write the pairing functions  $g(\mathbf{i}, \mathbf{j}) \equiv \langle \mathbf{i}, \mathbf{j} | \hat{g} | 0 \rangle$ , where  $|0\rangle$  is the vacuum,  $|\mathbf{i}, \mathbf{j}\rangle$  is the WF of a system with one electron with

coordinates  $(\mathbf{r}_i, \sigma_i)$  and another with coordinates  $(\mathbf{r}_j, \sigma_j)$ , and  $\hat{g}$  is the operator

$$\hat{g} = \sum_{k=1}^L \tilde{\lambda}_k \hat{a}_{k,\uparrow}^\dagger \hat{a}_{k,\downarrow}^\dagger, \quad (69)$$

where  $\tilde{\lambda}_k$  is the one defined above. Using this notation, it is easy to show (see Appendix of Ref. 67) that the pairing function  $g$  is equivalent to the complete active space of two electrons on the  $L$  molecular orbitals  $\{\tilde{\phi}_k\}$ .

Within this notation, the AGP WF is

$$|\Phi_{\text{AGPs}}\rangle = \hat{g}^p |0\rangle, \quad (70)$$

where  $p \equiv N/2$  is the number of electron pairs. If we substitute in  $\hat{g}^p$  the expansion for  $\hat{g}$  in Eq. (69), after having conveniently defined the operator  $\hat{b}_k \equiv \hat{a}_{k,\uparrow}^\dagger \hat{a}_{k,\downarrow}^\dagger$ , which created an electron pair on the orbital  $\tilde{\phi}_k$ , and having noted that  $\hat{b}_k \hat{b}_l = \hat{b}_l \hat{b}_k$  and  $\hat{b}_k^2 = 0$  (as following from the anticommutation relations of the  $\hat{a}_k^\dagger$  operators), we obtain that

$$\hat{g}^p = p! \sum_{1 \leq i_1 < i_2 < \dots < i_p \leq L} \tilde{\lambda}_{i_1} \tilde{\lambda}_{i_2} \dots \tilde{\lambda}_{i_p} \hat{b}_{i_1} \hat{b}_{i_2} \dots \hat{b}_{i_p}. \quad (71)$$

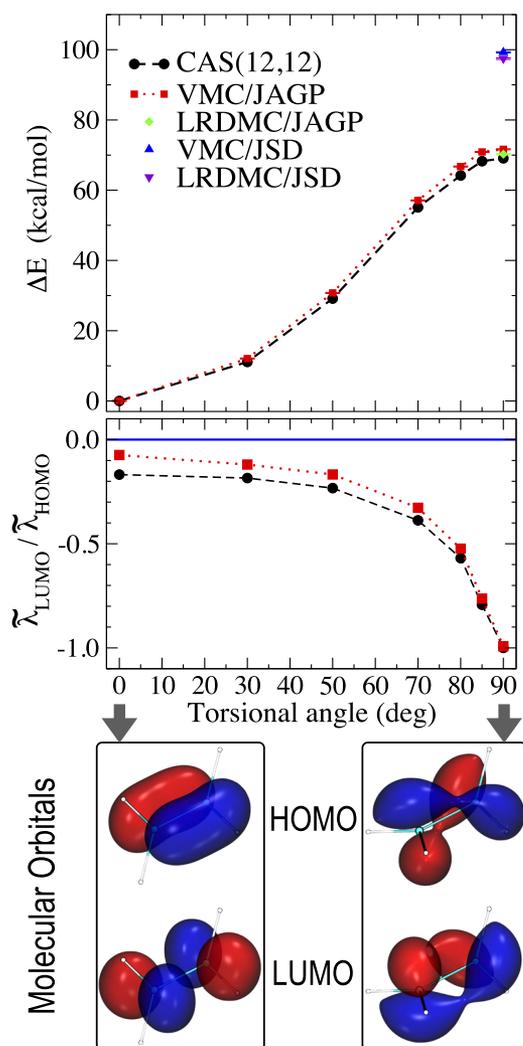
The chosen order of the  $\tilde{\lambda}_k$  coefficients implies that the leading term in  $\hat{g}^p$  is given by the term with  $(i_1, \dots, i_p) = (1, \dots, p)$ .<sup>73</sup> Thus, we can rewrite  $\hat{g}^p$  in terms of this leading term and excitations with respect to it,

$$\begin{aligned} \hat{g}^p \frac{|0\rangle}{p! \prod_{i=1}^p \tilde{\lambda}_i} &= \prod_{i=1}^p \hat{b}_i |0\rangle + \sum_{j=1}^p \sum_{q=p+1}^L \frac{\tilde{\lambda}_q}{\tilde{\lambda}_j} \left( \prod_{i=1}^p \hat{b}_i \right) \hat{b}_q |0\rangle \\ &+ \sum_{1 \leq j < k \leq p} \sum_{p < q < r \leq L} \frac{\tilde{\lambda}_q \tilde{\lambda}_r}{\tilde{\lambda}_j \tilde{\lambda}_k} \left( \prod_{\substack{i=1 \\ i \neq k \wedge i \neq j}}^p \hat{b}_i \right) \hat{b}_q \hat{b}_r |0\rangle + \dots. \end{aligned} \quad (72)$$

In the above expansion, we shall recognize that the first element is a single Slater determinant of the molecular orbitals  $\{\tilde{\phi}_i\}_{i=1, \dots, N/2}$ ; the second element is the summation of all possible double excitations going from an orbital  $j \leq N/2$  to an orbital  $q > N/2$ ; the third element is the summation of a subset of the possible quadruple excitations and so on. In other terms, AGPs can be written as the zero seniority<sup>74</sup> subset of the CI expansion, having some constraints on the coefficients of the expansion.

It is worth mentioning that AGPs are also reliable in cases where a single Slater determinant is not a good reference, as, for instance, in the case of a broken covalent bond, where the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are degenerate (see Fig. 5). In fact, we can have that  $|\tilde{\lambda}_p| = |\tilde{\lambda}_{p+1}|$ , so the two determinants  $\prod_{i=1}^p \hat{b}_i |0\rangle$  and  $\left( \prod_{i=1}^{p-1} \hat{b}_i \right) \hat{b}_{p+1} |0\rangle$  have the same weight, and they both are the leading terms of the constrained zero-seniority expansion in Eq. (72).

Note that in the AGPn *Ansatz*, we can perform an expansion similar to Eq. (72), but the excitations stop at the orbital  $n$  rather than  $L$ . Thus, it is straightforward to see that for  $n = N/2$ , there are no excitations and the only term is a single Slater determinant.



**FIG. 5.** The twisted ethylene molecule ( $C_2H_4$ ) is a prototypical system requiring a multireference method. The ground state geometry is planar (torsional angle equal to zero); C atoms have a double bond, and the HOMO and LUMO orbitals are separated. As the molecule is twisted, one bond is broken and the HOMO–LUMO orbitals become degenerate for a torsional angle of  $90^\circ$ . The upper panel shows evaluations of the torsional barrier as obtained using JSD and JAGP at the VMC and LRDMC level and using a complete active space (CAS) of 12 orbitals in 12 electrons. JAGP provides reliable results, whereas JSD (which neglects the LUMO orbital) overestimates the barrier both at the variational and diffusion levels. The middle panel shows the ratio  $\bar{\lambda}_{LUMO}/\bar{\lambda}_{HOMO}$  in the JAGP *Ansatz* and the corresponding value in the CAS calculations. The bottom panel shows the HOMO and LUMO orbitals for the planar and twisted geometries. Readapted with permission from Zen *et al.*, *J. Chem. Theory Comput.* **10**, 1048 (2014). Copyright 2014 ACS.

### C. Atomic basis set for the pairing function and the Jastrow factor

TURBORVB employs localized atomic orbitals such as the Gaussian type,

$$\psi_{l,\pm m,l}^{\text{Gaussian}}(\mathbf{r}; \zeta) = |\mathbf{r} - \mathbf{R}_I|^l e^{-\zeta|\mathbf{r}-\mathbf{R}_I|^2} \cdot \mathfrak{R}\left[(-i)^{\frac{l+m}{2}} Y_{l,m,l}(\theta, \varphi)\right], \quad (73)$$

or the Slater type,

$$\psi_{l,\pm m,l}^{\text{Slater}}(\mathbf{r}; \zeta) = |\mathbf{r} - \mathbf{R}_I|^l e^{-\zeta|\mathbf{r}-\mathbf{R}_I|} \cdot \mathfrak{R}\left[(-i)^{\frac{l+m}{2}} Y_{l,m,l}(\theta, \varphi)\right], \quad (74)$$

where the real and the imaginary part ( $m > 0$ ) of the spherical harmonic function  $Y_{l,m,l}(\theta, \varphi)$  centered at  $\mathbf{R}_I$  are taken and rewritten in Cartesian coordinates in order to work with real defined and easy to compute orbitals,  $l$  is the corresponding angular momentum, and  $m \geq 0$  is its projection number along the  $z$ -quantization axis. The localized atomic orbitals are also present in the one-body and three/four-body Jastrow parts [i.e., denoted as  $\chi(\mathbf{r})$  in Eqs. (30) and (34)]. One can use standard basis sets with exponents  $\zeta$  (also coefficients for a contracted basis set) taken from the available standard database such as the Basis Set Exchange<sup>75</sup> or from other more specific references when using pseudopotential.<sup>76–81</sup> A Python wrapper, named TURBO-GENIUS, makes the above procedure much easier, as shown later.

### H. Pseudopotential

TURBORVB supports pseudopotential calculations both in VMC and LRDMC calculations. Many ECPs have been generated and successfully used in quantum chemistry codes, but they are usually tuned to match Density Functional Theory (DFT) or Hartree–Fock (HF) all-electron (AE) calculations, which are not expected to be optimal for state-of-the-art many-body techniques. Recently, some progress has been made in this direction, and pseudopotentials determined by correlated many-body techniques are also available.<sup>76–83</sup> All the pseudopotentials used in QMC employ the standard semi-local form

$$\hat{V}_{pp}^I(\mathbf{r}_i) = V_{loc}^I(r_{i,I}) + \sum_{l=0}^{l_{max}} V_l^I(r_{i,I}) \sum_{m=-l}^l |Y_{l,m}\rangle \langle Y_{l,m}|, \quad (75)$$

where  $r_{i,I} = |\mathbf{r}_i - \mathbf{R}_I|$  is the distance between the  $i$ th electron and the  $I$ th ions,  $l_{max}$  is the maximum angular momentum of the ion  $I$ , and  $\sum_{l=0}^{l_{max}} \sum_{m=-l}^l |Y_{l,m}\rangle \langle Y_{l,m}|$  is a projection operator on the spherical harmonics centered at the ion  $I$ . In TURBORVB, the angular momentum projector is calculated by using standard polyhedral quadrature formulas for the angular integrations.<sup>84</sup> As it is now becoming a common practice not only in QMC, both the local  $V_{loc}^I(r_{i,I})$  and the non-local  $V_l^I(r_{i,I})$  functions are expanded over a simple Gaussian basis parametrized by coefficients (e.g., effective charge  $Z_{eff}$  and other simple constants), multiplying simple powers of  $r$  and a corresponding gaussian term

$$r^2 V_l(r) = \sum_k \alpha_{k,l} r^{\beta_{k,l}} \exp(-\gamma_{k,l} r^2), \quad (76)$$

where  $\alpha_{k,l}$ ,  $\beta_{k,l}$  (usually small positive integers), and  $\gamma_{k,l}$  are the parameters obtained by appropriate fitting. Several published pseudopotentials have already been tabulated in TURBORVB. Of course, one can also use any pseudopotential employing the semi-local form in the mentioned Gaussian basis with a straightforward little extra work for the input preparation.

### I. Contraction of the primitive atomic basis

The contraction of atomic orbitals has been widely used in quantum chemistry and DFT codes, and was originally introduced

to define a pseudo-Slater orbital by combining several primitive Gaussian orbitals. This is also important in the QMC context because it decreases the number of variational parameters by a large factor, as will be shown in this section. Although one can obtain a contracted basis set directly from a database, TURBORVB allows us to prepare a high-quality hybrid basis set (contraction) starting from a given primitive basis, by the so-called “geminal embedding scheme.”<sup>85</sup> To this purpose, let us decompose the geminal function in terms of atomic contribution, as far as the dependence over  $\mathbf{i}$  (the left argument) is concerned,

$$g(\mathbf{i}, \mathbf{j}) = \sum_I U_{\text{proj}}^I(\mathbf{i}, \mathbf{j}) = \sum_I \sum_{\mu, \nu} A_{\mu, \nu}^I \phi_{\mu}(\mathbf{i}) \phi_{\nu}(\mathbf{j}), \quad (77)$$

where  $I$  represents an atom in a system;  $U_{\text{proj}}^I(\mathbf{i}, \mathbf{j})$  is the pairing function projected on the atom  $I$ ;  $A_{\mu, \nu}^I$  is  $A_{\mu, \nu}$  if  $\mu \in I$  and  $A_{\mu, \nu}^I = 0$  otherwise, where  $A_{\mu, \nu}$  is assumed to be given for the system under consideration, e.g., obtained by a standard DFT calculation, where in this case, by Eq. (63),

$$A_{\mu, \nu} = \sum_{k=1}^n (c_{\mu}^k c_{\nu}^k - \tilde{c}_{\mu}^k \tilde{c}_{\nu}^k), \quad (78)$$

where  $c_{\mu}^k$  and  $\tilde{c}_{\mu}^k$  are the coefficients of the DFT molecular orbitals  $\phi_k(\mathbf{i}) = \sum_{\mu} c_{\mu}^k \phi_{\mu}(\mathbf{i})$  and  $\tilde{\phi}_k(\mathbf{i}) = \sum_{\nu} \tilde{c}_{\nu}^k \phi_{\nu}(\mathbf{i})$  in the atomic basis expansion, respectively.

Quite generally, the projected pairing function can be expanded in a truncated space spanned by  $q$  terms,

$$\tilde{U}_{\text{proj}}^I(\mathbf{i}, \mathbf{j}) = \sum_{k=1}^q \sigma_k \psi_k^{\text{GEO}}(\mathbf{i}) \psi_k^{\text{Env}}(\mathbf{j}). \quad (79)$$

In other words, the Schmidt decomposition is applied to the matrix  $\tilde{U}_{\text{proj}}^I(\mathbf{i}, \mathbf{j})$  describing the coupling between a given atom  $I$  and the environment, within the geminal *Ansatz*. This procedure defines the so-called geminal embedded orbitals (GEOs), which are determined in terms of an expansion over all atomic orbitals used for the atom  $I$ ,

$$\psi_k^{\text{GEO}}(\mathbf{i}) = \sum_{\mu \in I} \mu_{k, \mu}^{\text{GEO}} \psi_{\mu}(\mathbf{i}), \quad (80)$$

where  $\psi_k^{\text{GEO}}(\mathbf{i})$  are orthonormal. Following the Schmidt decomposition, it is possible to determine the best GEOs by minimizing the Euclidian distance between the original and the truncated geminal functions,

$$d^2 = |U_{\text{proj}}^I - \tilde{U}_{\text{proj}}^I|^2. \quad (81)$$

Considering all possible unconstrained functions  $\psi_k^{\text{Env}}(\mathbf{j})$  and employing the steady condition  $\frac{\delta d^2}{\delta \psi_k^{\text{Env}}(\mathbf{j})} = 0$ ,  $d^2$  reads

$$d^2 = |U_{\text{proj}}^I|^2 - \sum_{k=1}^p \int d\mathbf{i} d\mathbf{j} D_{\text{proj}}^I(\mathbf{i}, \mathbf{j}) \psi_k^{\text{GEO}}(\mathbf{i}) \psi_k^{\text{GEO}}(\mathbf{j}), \quad (82)$$

where  $|U_{\text{proj}}^I|^2 = \int d\mathbf{i} d\mathbf{j} |U_{\text{proj}}^I(\mathbf{i}, \mathbf{j})|^2$  and  $D_{\text{proj}}^I(\mathbf{i}, \mathbf{j})$  is the density matrix that reads

$$D_{\text{proj}}^I(\mathbf{i}, \mathbf{j}) = \int d\mathbf{k} U_{\text{proj}}^I(\mathbf{i}, \mathbf{k})^* U_{\text{proj}}^I(\mathbf{j}, \mathbf{k}). \quad (83)$$

Since the GEOs  $\psi_k^{\text{GEO}}(\mathbf{r})$  are orthonormal and the density matrix is Hermitian, Eq. (82) becomes minimum when the GEOs coincide with the  $p$  eigenvectors of the density matrix with the

maximum eigenvalues (denoted as  $w_i$ ). The original atomic basis  $\psi_{\mu}(\mathbf{i})$  is usually non-orthogonal, so the problem turns into the generalized eigenvalue equation,

$$\sum_{j \in J} [(A^{I*} S A^I) S]_{i,j} \mu_{k,j}^{\text{GEO}} = \omega_k \mu_{k,i}^{\text{GEO}}. \quad (84)$$

The truncation error is readily estimated by the summation of the eigenvalues  $d^2 = |U_{\text{proj}}^I|^2 - \sum_{i=1}^p w_i$ . Since the eigenvalues are sorted in ascending order, a suitably chosen value of  $p$  allows the user to neglect the most irrelevant vector components with small eigenvalues  $w_i$  and work with enough accuracy even with a few GEOs per atom, thus minimizing the number of variational parameters necessary to describe well the system, as will be shown below. With this construction, a new geminal is defined in the GEO basis, namely,

$$\tilde{g}(\mathbf{i}, \mathbf{j}) = \sum_{\mu, \nu} \tilde{A}_{\mu, \nu} \psi_{\mu}^{\text{GEO}}(\mathbf{i}) \psi_{\nu}^{\text{GEO}}(\mathbf{j}), \quad (85)$$

where the matrix coefficients  $\tilde{A}_{\mu, \nu}$  are given by maximizing the normalized overlap ( $Q$ ) between the original and the new geminals,

$$Q = \frac{\langle \tilde{g} | g \rangle^2}{\langle \tilde{g} | \tilde{g} \rangle \langle g | g \rangle}, \quad (86)$$

where  $\langle \tilde{g} | g \rangle = \int d\mathbf{i} d\mathbf{j} \tilde{g}(\mathbf{i}, \mathbf{j})^* g(\mathbf{i}, \mathbf{j})$ . It turns out that the overlap remains large even for small GEO basis set size  $p$ , implying that by using this scheme, one can decrease the number of variational parameters corresponding to the matrix  $A$ , i.e., from  $4L^2$  to  $4p^2 \ll 4L^2$ .

## J. Conversion of the WF

As described in Sec. III, TURBORVB implements different types of *Ansätze*: (i) the Pfaffian (Pf), (ii) the Pfaffian with a constrained number of molecular orbitals (Pfn), (iii) the Antisymmetrized Geminal Power (AGP), (iv) the Antisymmetrized Geminal Power with a constrained number of molecular orbitals (AGPn), and (v) the single Slater determinant (SD). One can choose a proper *Ansatz* depending on a target system, considering the computational cost of a chosen *Ansatz* and the relevant physical and chemical properties of a target material. During the simulation, a user can go back and forth between the *Ansatz* using modules implemented in TURBORVB, with/without losing the information of an optimized *Ansatz* (see Fig. 4): The first case is to add molecular orbitals to an *Ansatz*, i.e., JAGP  $\Rightarrow$  JSD, JAGP  $\Rightarrow$  JAGPn, or JPf  $\Rightarrow$  JPfn. In TURBORVB, this is obtained by rewriting the expansion of the geminal in terms of molecular orbitals (see Secs. III D and III E). The corresponding tool is convertfort10mol.x. The second important case is to convert an *Ansatz* among the available ones, i.e., JSD, JAGP, or JAGPn  $\Rightarrow$  JAGP. In TURBORVB, this is the purpose of the convertfort10.x tool and is achieved by maximizing the overlap between the two WFs, one of them being the input (fort.10\_in) and the other being the type (fort.10\_out) to be filled by new geminal matrix coefficients (result written in fort.10\_new). In more details, in TURBORVB, the following overlap between two geminals is maximized:

$$\max Q = \frac{\langle g^{\text{new}} | g^{\text{ori}} \rangle^2}{\langle g^{\text{new}} | g^{\text{new}} \rangle \langle g^{\text{ori}} | g^{\text{ori}} \rangle}, \quad (87)$$

in order to obtain new geminal matrix coefficients  $A_{\mu,\nu}^{\text{new}}$ , defining the new pairing function as

$$g^{\text{new}}(\mathbf{i}, \mathbf{j}) = \sum_{\mu,\nu} A_{\mu,\nu}^{\text{new}} \psi_{\mu}^{\text{new}}(\mathbf{i}) \psi_{\nu}^{\text{new}}(\mathbf{j}), \quad (88)$$

while the original geminal was given in terms of the parameter matrix  $A_{\mu,\nu}^{\text{ori}}$ ,

$$g^{\text{ori}}(\mathbf{i}, \mathbf{j}) = \sum_{\mu,\nu} A_{\mu,\nu}^{\text{ori}} \psi_{\mu}^{\text{ori}}(\mathbf{i}) \psi_{\nu}^{\text{ori}}(\mathbf{j}). \quad (89)$$

Note that  $0 \leq Q \leq 1$ ; therefore, the larger the  $Q$ , the better the conversion, and  $Q$  approaches the unit value if the conversion is perfect. For this type of conversion, one can also apply the geminal embedding scheme to construct a hybrid basis set, as described in Sec. III I.

The final case is to convert a JAGP *Ansatz* to JPf. Since the JAGP *Ansatz* is a special case of the JPf one, where only  $G_{ud}$  and  $G_{du}$  terms are defined, as described in Sec. III C, the conversion can be realized just by direct substitution. Therefore, the main challenge is to find a reasonable initialization for the two spin-triplet sectors  $G_{uu}$  and  $G_{dd}$  that are not described in the JAGP and that otherwise have to be set to 0. There are two possible approaches:<sup>62,69</sup> (i) for polarized systems, we can build the  $G_{uu}$  block of the matrix by using an even number of unpaired orbitals  $\{\phi_i\}$  and build an antisymmetric  $g_{uu}$  by means of Eq. (63), where the eigenvalues  $\lambda_k$  are chosen to be large enough to occupy certainly these unpaired states, as in the standard Slater determinant used for our initialization. Again, we emphasize that this works only for polarized systems. (ii) The second approach that also works in a spin-unpolarized case is to determine a standard broken symmetry single determinant *Ansatz* (e.g., by the TURBORVB built-in DFT within the LSDA) and modify it with a global spin rotation. Indeed, in the presence of finite local magnetic moments, it is often convenient to rotate the spin moments of the WF in a direction perpendicular to the spin quantization axis chosen for our spin-dependent Jastrow factor, i.e., the  $z$  quantization axis. In this way, one can obtain reasonable initializations for  $G_{uu}$  and  $G_{dd}$ . TURBORVB allows for every possible rotation, including an arbitrary small one close to the identity. A particularly important case is when a rotation of  $\pi/2$  is applied around the  $y$  direction. This operation maps

$$|\uparrow\rangle \rightarrow \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) \text{ and } |\downarrow\rangle \rightarrow \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle). \quad (90)$$

One can convert from an AGP the pairing function that is obtained from a VMC optimization

$$g_{ud}(\mathbf{i}, \mathbf{j}) = f_S(\mathbf{r}_i, \mathbf{r}_j) \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}} + f_T(\mathbf{r}_i, \mathbf{r}_j) \frac{|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle}{\sqrt{2}} \quad (91)$$

to a Pf one

$$g_{ud}(\mathbf{i}, \mathbf{j}) \rightarrow g(\mathbf{i}, \mathbf{j}) = f_S(\mathbf{r}_i, \mathbf{r}_j) \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}} + f_T(\mathbf{r}_i, \mathbf{r}_j) (|\uparrow\uparrow\rangle - |\downarrow\downarrow\rangle). \quad (92)$$

This transformation provides a meaningful initialization to the Pfaffian WF that can then be optimized for reaching the best possible description of the ground state within this *Ansatz*.

## IV. BULK SYSTEMS

The application of TURBORVB is not limited to open systems such as atoms and molecules. TURBORVB can also simulate bulk systems in large supercells with arbitrary twisted boundary conditions. These are used to minimize finite-size effects and represent quite an important approach<sup>86,87</sup> in order to reach a meaningful and accurate thermodynamic limit.

### A. CRYSTAL basis set

For periodic system calculations, the many-body WF should satisfy the many-body Bloch condition,<sup>88,89</sup>

$$\Psi_{\mathbf{k}_s}(\mathbf{r}_1, \dots, \mathbf{r}_i + \mathbf{T}_s, \dots, \mathbf{r}_N) = e^{i\mathbf{k}_s \cdot \mathbf{T}_s} \Psi_{\mathbf{k}_s}(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N), \quad (93)$$

which follows from the property that the many-body Hamiltonian is invariant under the translation of *any electron coordinate* by a simulation-cell vector  $\mathbf{T}_s$ , where  $\mathbf{T}_s = l\mathbf{a} + m\mathbf{b} + n\mathbf{c}$  is determined by arbitrary integers  $l, n, m$  and the three vectors  $\mathbf{a}, \mathbf{b}$ , and  $\mathbf{c}$  define the supercell.

In TURBORVB, a single-particle basis set satisfies the following condition:

$$\psi_{l,m,I}^{\text{PBC}}(\mathbf{r} + \mathbf{T}_s; \zeta) = e^{i\mathbf{k}_s \cdot \mathbf{T}_s} \psi_{l,m,I}^{\text{PBC}}(\mathbf{r}; \zeta), \quad (94)$$

where  $\mathbf{k}_s$  is a twist vector [ $\mathbf{k}_s = (k_s^x, k_s^y, k_s^z)$ ] and  $\mathbf{T}_s$  represents an arbitrary simulation cell vector. Note that the use of a non-vanishing twist vector generally makes a many-body WF *complex*. TURBORVB implements the CRYSTAL periodic basis,<sup>15,90,91</sup>

$$\psi_{l,m,I}^{\text{PBC}}(\mathbf{r}; \zeta) = \sum_{\mathbf{T}_s} \psi_{l,m,I}(\mathbf{r} + \mathbf{T}_s; \zeta) e^{-i\mathbf{k}_s \cdot \mathbf{T}_s}, \quad (95)$$

where  $\psi_{l,m,I}$  is a non-periodic real atomic orbital such as Gaussian [Eq. (73)] and Slater [Eq. (74)]. The use of Gaussian or Slater orbitals that rapidly decay far from nuclei guarantees that the above summation converges fast with a finite small number of  $\mathbf{T}_s$ . Note that, in TURBORVB,  $\mathbf{T}_s$  are not limited to orthorhombic ones, but has recently been extended to include all possible crystal translation groups (e.g., rhombohedral, hexagonal triclinic).

The same procedure is applied to the basis set for the Jastrow part, although using simple periodic boundary conditions,<sup>15</sup> because the twists do not affect the Jastrow part of the WF, namely,

$$\chi_{l,m,I}^{\text{PBC}}(\mathbf{r}; \zeta) = \sum_{\mathbf{T}_s} \chi_{l,m,I}(\mathbf{r} + \mathbf{T}_s; \zeta), \quad (96)$$

which satisfies  $\chi_{l,m,I}^{\text{PBC}}(\mathbf{r} + \mathbf{T}_s; \zeta) = \chi_{l,m,I}^{\text{PBC}}(\mathbf{r}; \zeta)$ .

Moreover, the homogeneous one-body and two-body Jastrow factors have to be appropriately periodicized because they are not defined in terms of the above periodic basis. Namely, the homogeneous one-body Jastrow part [Eq. (30)] should satisfy

$$\tilde{J}_1(\mathbf{r} + \mathbf{T}_s) = \tilde{J}_1(\mathbf{r}), \quad (97)$$

and the two-body Jastrow part [Eq. (32)] should fulfill

$$J_2(\mathbf{r}_1\sigma_1, \dots, (\mathbf{r}_i + \mathbf{T}_s)\sigma_i, \dots, \mathbf{r}_N\sigma_N) = J_2(\mathbf{r}_1\sigma_1, \dots, \mathbf{r}_i\sigma_i, \dots, \mathbf{r}_N\sigma_N). \quad (98)$$

In order to satisfy the above constraints, we consider the relative electron–nuclei or electron–electron coordinate differences  $\mathbf{r}_d$ ,

necessary to evaluate  $J_1$  and  $J_2$ , respectively, and expand them as

$$\mathbf{r}_d = r_a \mathbf{a} + r_b \mathbf{b} + r_c \mathbf{c}, \quad (99)$$

where  $r_a$ ,  $r_b$ , and  $r_c$  are appropriate transformed coordinates, which are conveniently defined within a cube of unit length because of the assumed periodicity of the supercell, namely,  $|r_a|, |r_b|, |r_c| < 1/2$ . As a consequence, this mapping makes the physical electron–electron and electron–ion distance periodic by definition (i.e., they refer to the minimum distance image of the supercell). However, there may be divergences or singularity at the boundaries of this unit cube. Therefore, before computing the distance corresponding to  $\mathbf{r}_d$ , these coordinates are transformed  $(r_a, r_b, r_c) \rightarrow (\tilde{r}_a, \tilde{r}_b, \tilde{r}_c) = (p(r_a), p(r_b), p(r_c))$  by use of an appropriate function  $p(x)$ , with at least a continuous first derivative for  $|x| < 1/2$ . This function is chosen to preserve the physical meaning at short distances, i.e.,  $p(x) = x$  in these cases and nonlinear elsewhere, in order to satisfy not only the periodicity but also the requirement of continuous first derivatives of the many-body WF  $\Psi_{k_s}$ . We have, therefore, defined  $p(x)$  as follows:

$$p(x) = \begin{cases} x & (-\frac{1}{4} < x < \frac{1}{4}) \\ -\frac{1}{8(1+2x)} & (-\frac{1}{2} \leq x \leq -\frac{1}{4}) \\ \frac{1}{8(1-2x)} & (\frac{1}{4} \leq x \leq \frac{1}{2}). \end{cases} \quad (100)$$

Indeed, although the modified relative distance diverges (i.e.,  $|\mathbf{r}_d| \rightarrow \infty$ ) at the edges of the Wigner–Seitz cell (e.g.,  $\mathbf{r} = \pm \frac{1}{2} \mathbf{a}, \pm \frac{1}{2} \mathbf{b}, \pm \frac{1}{2} \mathbf{c}$ ), the exponential [Eq. (31)] and the Padé [Eq. (33)] functions remain finite,  $u(r) \rightarrow 1/2b_{ea}$  and  $v_{\sigma,\sigma'}(r_{ij}) \rightarrow 1/4b_{ee}^{\text{para}}$  or  $1/2b_{ee}^{\text{anti}}$ , respectively, thus preserving with continuity the periodicity of the one-/two-body homogeneous parts of the Jastrow factor. However, for the Padé form, one has to change the expression for  $p(x)$  in order to satisfy also the continuity in the WF derivatives, even when the modified relative distances diverge, i.e.,

$$p(x) = \begin{cases} x & (-\frac{1}{6} < x < \frac{1}{6}) \\ -\frac{1}{54(1/2+x)^2} & (-\frac{1}{2} \leq x \leq -\frac{1}{6}) \\ \frac{1}{54(1/2-x)^2} & (\frac{1}{6} \leq x \leq \frac{1}{2}). \end{cases} \quad (101)$$

In this case, the region where  $p(x) = x$  further shrinks. Thus, it is often more convenient to use the exponential form to obtain a more accurate variational WF because the long-range part is implicitly corrected by the inhomogeneous terms in Eq. (34).

Finally, we remark that the many-body WF also obeys the second Bloch condition,<sup>88,89</sup> namely,

$$\Psi_{k_p}(\{\mathbf{r}_i + \mathbf{T}_p\}) = e^{ik_p \cdot \mathbf{T}_p} \Psi_{k_p}(\{\mathbf{r}_i\}), \quad (102)$$

where  $\mathbf{T}_p$  represents a *unit-cell* (not supercell) vector and  $\mathbf{k}_p$  is the crystal momentum. This comes from the property that the many-body Hamiltonian is invariant under the *simultaneous translation* of all-electron coordinates by a unit-cell vector  $\mathbf{T}_p$ . Within TURBORVB, this condition can be employed by imposing the intra-unit cell translational symmetries on the Jastrow and the pairing function as simple linear constraints in the variational parameters. However, this option is restricted to the case  $\mathbf{k}_p = 0$ . On the other hand, it is possible to use different twists on each spin component, which has

proven very effective for implementing the mentioned translation symmetries within pairing WFs.<sup>92</sup>

## B. Finite-size effects

The systematic error induced by a finite simulation cell is a long-standing issue in the *ab initio* QMC calculation. There are two types of finite-size errors in QMC calculations: one is the so-called one-body effect that arises from the kinetic energy term of the Hamiltonian and the other is the so-called two-body effect that arises from the periodic Ewald contribution resulting from the electron–electron interaction. Note that, in the independent-particle calculations (i.e., DFT), only the former is present, which can be readily evaluated by  $k$  summation in the first Brillouin zone, but the two-body finite-size effects are not present because the exchange–correlation energy used in DFT usually derives from DMC results extrapolated to the infinite simulation cell size. To correct the one-body finite-size error, one can use the twisted averaged boundary conditions,<sup>86</sup> special  $k$ -points methods,<sup>88,89</sup> or the exact special twist (EST) method,<sup>87</sup> all of which have been implemented in TURBORVB. Within the TURBORVB implementation, the Jastrow part is independent of twists (i.e., TURBORVB uses a common Jastrow for all twists),

$$\Psi_{k_s}(\mathbf{r}_1, \dots, \mathbf{r}_N) = J(\mathbf{r}_1, \dots, \mathbf{r}_N) \Phi_{AS}^{k_s}(\mathbf{r}_1, \dots, \mathbf{r}_N). \quad (103)$$

As emphasized in Ref. 86, at variance with DFT, the QMC computational effort is independent of the number of  $k$ -points used. For the two-body finite-size effects, which cannot be alleviated by the above remedy, one can employ the model periodic Coulomb (MPC) interaction.<sup>93–95</sup> This method has not been implemented in TURBORVB yet. Nevertheless, simpler alternatives exist. For instance, one can alleviate the two-body finite size effects by directly increasing the supercell size, or one can estimate these effects by employing the KZK exchange–correlation function<sup>96</sup> at the DFT level. Moreover, it is also possible to employ systematic finite-size corrections based on the knowledge of the density structure factor in momentum space,<sup>97</sup> which can be readily computed within TURBORVB, with a short postprocessing computation.

## V. BUILT-IN DENSITY FUNCTIONAL THEORY (DFT) CODE

Although most QMC codes load their trial WFs from available DFT/quantum chemistry codes such as Gaussian,<sup>98</sup> CRYSTAL,<sup>91</sup> and Quantum Espresso,<sup>99</sup> TURBORVB is a self-consistent many-body package and does not require input from other codes. Indeed, in TURBORVB, an original DFT code is implemented that has the important feature to work in exactly the same basis and pseudopotentials (if any) used for the QMC many-body WF and, for instance, can work with mixed Slater–Gaussian basis or with arbitrary contraction, i.e., mixing angular momenta. On the other hand, the drawback is that only very few functionals are available so far because, within the TURBORVB approach, the DFT has to be used only to have a reasonable initialization of the antisymmetric part of the many-body WF. This is then optimized at best by direct minimization of the energy in the presence of the Jastrow factor and should very weakly depend on the initialization or, for instance, on the DFT functional used. At present, TURBORVB supports Perdew–Zunger (PZ81)

Local-Spin Density Approximation (LSDA)<sup>100</sup> and those combined with KZK exchange-correlation energy<sup>96</sup> but, for instance, does not allow LDA+*U* calculations. This will be the goal of a possible future work.

### A. Stabilization algorithm for large localized basis sets

An important issue in any DFT algorithm based on localized basis sets, such as Gaussian, is how to reach the so-called complete basis set limit. In the plane-wave method, this is systematically achieved by increasing the number of plane waves (e.g., kinetic cutoff energy). However, with a localized basis set, by increasing its dimension, numerical instabilities may occur as a consequence of the redundancy and non-orthonormality of the basis. On the other hand, the use of a small basis set implies too much biased results. In `TURBORVB`, an original and systematic algorithm to remove this redundancy<sup>101</sup> has been implemented. When the basis set is non-orthogonal, the key self consistent step for solving the DFT Kohn–Sham equations turns to a generalized eigenvalue problem,

$$\mathcal{H}\mathbf{c} = E\mathbf{S}\mathbf{c}, \quad (104)$$

where  $\mathcal{H}$  is the single-particle Hamiltonian matrix,  $\mathbf{c}$  is the vector of the coefficients of the orbitals,  $E$  is the single-particle energy, and the  $\mathbf{S}$  is the  $N_b \times N_b$  overlap matrix, where  $N_b$  is the dimension of the basis set,

$$S_{ij} = \langle \phi_i | \phi_j \rangle = \int d\mathbf{r} \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}). \quad (105)$$

The overlap matrix is strictly positive definite, namely, all its eigenvalues are positive. The problem mentioned above shows up when the overlap matrix  $\mathbf{S}$  becomes ill-conditioned. If the condition number, namely, the ratio between the largest (denoted as  $s_{N_b}$ ) and the smallest eigenvalues (denoted as  $s_1$ ), becomes very large, the CBS limit cannot be achieved by the standard procedure due to numerical instabilities. Therefore, in `TURBORVB`, the small eigenvalues and the corresponding eigenvectors are disregarded. Indeed, the original basis is modified as follows:

$$e_j^i = \frac{1}{\sqrt{s_i}} v_j^i \text{ for } s_i/s_{N_b} \geq \varepsilon_{\text{mach}}, \quad (106)$$

where  $s_i$  and  $v_j^i$  are the  $i$ th eigenvalues and eigenvectors of the overlap matrix,  $\varepsilon_{\text{mach}}$  is usually set to the machine precision,  $j$  runs from 1 to  $N_b$ ,  $i$  runs from 1 to  $M_b$ , and  $M_b$  represents the number of eigenvalues satisfying the above inequality. If infinite numerical accuracy were available, the matrix  $\mathbf{S}$  in the new basis would be just the identity and the generalized eigenvalue problem in Eq. (104) would turn to a standard one in this basis. Within finite precision arithmetic, it is better to iterate further the stabilization procedure and define a new basis,

$$\tilde{e}_j^i = \frac{1}{\sqrt{\tilde{s}_i}} \tilde{v}_j^i, \quad (107)$$

where  $\tilde{s}_i$  and  $\tilde{v}_j^i$  are the  $i$ th eigenvalues and eigenvectors of the recomputed overlap matrix  $\tilde{s}_{i,j} = \langle e_i | \mathbf{S} | e_j \rangle$  (now well conditioned since very close to the identity), respectively. For using the latter basis set, the  $N_b \times M_b$  global transformation matrix is stored as it takes into

account the projection from the original basis to the final modified one,

$$U_{i,j} = \sum_k e_j^{k-i} e_k^i. \quad (108)$$

On this basis set, the generalized eigenvalue problem becomes a well-conditioned one with a corresponding  $M_b \times M_b$  Hamiltonian matrix  $\tilde{\mathcal{H}} = \mathbf{U}\mathcal{H}\tilde{\mathbf{U}}$ . In this way, one can avoid the numerical instabilities induced by a too redundant large basis set. This stabilization introduces an error at most  $\sqrt{\varepsilon_{\text{mach}} \cdot N_b}$ , which is typically negligible compared with the target chemical accuracy.

### B. Electron-ion cusp condition

Another feature of the DFT code implemented in `TURBORVB` is that electron–nuclei cusp conditions are exactly fulfilled for any basis (i.e., Gaussian orbital) even within the DFT framework. This is achieved by an appropriate modification of the standard basis sets commonly used (e.g., ccpVTZ)<sup>102</sup> for WF based calculations: the new basis is obtained by multiplying each element of the original basis by a suitably chosen one-body Jastrow factor introducing the correct cusps, namely,

$$\tilde{\phi}_j^b(\mathbf{r} - \mathbf{R}_b) = \phi_j^b(\mathbf{r} - \mathbf{R}_b) \tilde{J}_1(\mathbf{r}), \quad (109)$$

where  $\tilde{J}_1(\mathbf{r})$  is the same as in Eq. (29) and the parameter  $b$  in Eq. (31) is optimized by direct minimization of the chosen DFT energy functional. In this way, each element of the reshaped basis set satisfies the so-called electron–ion cusp conditions, namely that, when  $\mathbf{r}$  is close to any atomic position  $\mathbf{R}_b$ , the following relation holds:

$$\lim_{\mathbf{r} \rightarrow \mathbf{R}_b} \frac{\nabla \tilde{\phi}_j^a}{\tilde{\phi}_j^a} = -Z_b \frac{\mathbf{r} - \mathbf{R}_b}{|\mathbf{r} - \mathbf{R}_b|} \quad (110)$$

for all  $a, b$ . This formulation allows us to reach the CBS limit extremely fast in DFT calculations, especially within all electrons. Indeed, we have experienced that a given target accuracy can be obtained with a smaller basis, e.g., our modified ccpVDZ basis is typically equivalent in accuracy to the much larger ccpVTZ (and the ccpVTZ equivalent to the ccpVQZ, and so on and so forth).

For QMC application, this tool is particularly useful because the Slater determinant obtained with this new basis does not have divergences in the local energy, which are instead present in the original basis, when electron positions are close to the nuclear ones. This feature emphasizes further the clear advantage of `TURBORVB` to allow the use of a special purpose DFT code, just devoted to the optimal initialization of a QMC many-body electronic WF.

### C. Double-grid remedy for all-electron calculations

As described in Sec. V B, `TURBORVB` employs standard atomic orbitals, modified by means of an appropriately chosen one-body factor. Thus, the electron–electron integrations cannot be evaluated analytically even when the Gaussian atomic basis set is employed. Therefore, `TURBORVB` calculates the electron–electron Hartree potential by solving Poisson’s equation with the fast Fourier transform (FFT) on a Cartesian grid both in open and periodic systems. The numerical solution becomes problematic in all-electron calculations when the atomic number  $Z$  becomes large. Indeed, in this case, the Kohn–Sham Hamiltonian matrix elements have to be computed in

the presence of a rapidly varying electron density in the vicinity of nuclei and therefore require an exceedingly large FFT mesh, with an almost prohibitive computational cost for its evaluation. `TURBORVB` alleviates this drawback by a double-grid scheme,<sup>103</sup> in which the Hartree potential is calculated with standard FFT convolution on a coarse mesh. Then, these values are interpolated on a much finer mesh in the vicinity of the nuclei. Although the DFT energy obtained with the above approximation is not exactly consistent with the one corresponding to a very dense uniform mesh, the corresponding VMC energies and the variances are almost indistinguishable to each other, at a much cheaper computational cost.

## VI. DERIVATIVES OF ENERGY

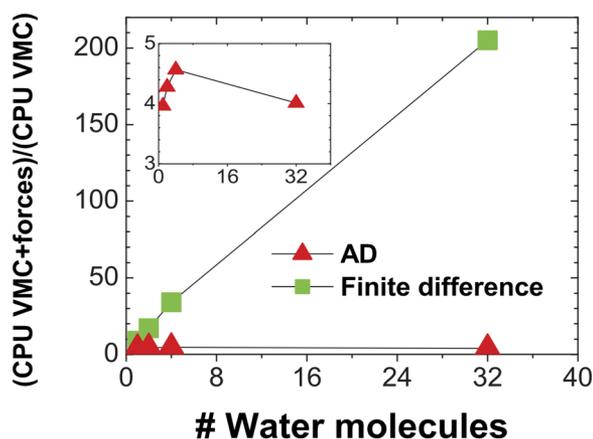
Derivatives of total energies with respect to variational parameters represent an essential ingredient for optimizing a many-body WF. Forces (derivative with respect to atomic positions) are also essential for performing structural optimization or molecular dynamics. However, in a complex code, and especially in QMC, the evaluation of the functional derivatives, necessary for the WF optimization, is very complicated, mainly for the complexity of the algorithm that, in turn, may lead to a very inefficient implementation, although recent progress has been done.<sup>104</sup> For instance, a simple approach is to compute them with finite difference expressions, leading to a too large computational time, as obviously proportional to the number of targeted derivatives.

Algorithmic differentiation (AD) is a method capable of solving all the above problems, essentially by a smart application of the differentiation chain rule. There are two types of algorithms: the forward algorithmic differentiation (FAD) that implements the chain rule straightforwardly from the beginning to the end of the algorithm and the adjoint algorithmic differentiation (AAD) that uses the chain rule starting from the end of the algorithm (also known as backward propagation). When the number of input parameters is much larger than the corresponding output ones, AAD is much more efficient than FAD and indeed allows the calculation of all possible derivatives in a computational time proportional (with a small prefactor; see, e.g., Fig. 6) to the one for computing the target function (i.e., the energy or the WF value). The former is therefore the ideal method for quantum Monte Carlo when the variational WF contains many variational parameters. The AAD was applied for the calculation of atomic forces in Ref. 41. To our knowledge, this was the first time that AAD was used within QMC. Now, `TURBORVB` implements all derivatives such as those of the local energy [ $\frac{\partial}{\partial \alpha_k} e_L(\mathbf{x}_i)$ ] or those corresponding to the WF logarithm [ $\frac{\partial}{\partial \alpha_k} \ln \Psi(\mathbf{x}_i)$ ] using the AAD, which drastically improves the efficiency (Fig. 6)<sup>41</sup> and reliability of the calculation.

### A. Derivatives with respect to variational parameters

The derivatives of the energy with respect to a given real variational parameter  $\alpha_k$  (one complex parameter can be thought to be composed by two real ones, its real and imaginary part, respectively) is represented as a generalized force,

$$f_k = -\frac{\partial E(\alpha)}{\partial \alpha_k} = -\frac{\partial}{\partial \alpha_k} \frac{\langle \Psi_\alpha | \hat{H} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle}. \quad (111)$$



**FIG. 6.** Ratio of CPU time required to compute energies and all force components referenced to the one required for the simple energy calculation within VMC. The calculations refer to 1, 2, 4, and 32 water molecules. The inset is an expansion of the lower part of the plot. Reprinted with permission from S. Sorella and L. Capriotti, *J. Chem. Phys.* **133**, 234111 (2010). Copyright 2010 AIP Publishing LLC.

In variational Monte Carlo, the derivative can be evaluated using  $M$  configurations of electron coordinates,<sup>15</sup>

$$f_k = -2\Re \left[ \frac{\sum_{\mathbf{x}} e_L^*(\mathbf{x})(O_k(\mathbf{x}) - \bar{O}_k) |\Psi_\alpha(\mathbf{x})|^2}{\sum_{\mathbf{x}} |\Psi_\alpha(\mathbf{x})|^2} \right] \approx -2\Re \left[ \frac{1}{M} \sum_{i=1}^M e_L^*(\mathbf{x}_i)(O_k(\mathbf{x}_i) - \bar{O}_k) \right], \quad (112)$$

where  $e_L(\mathbf{x})$  is the local energy,  $O_k(\mathbf{x})$  is the logarithmic derivative of the WF [i.e.,  $O_k(\mathbf{x}) = \frac{\partial \ln \Psi_\alpha(\mathbf{x})}{\partial \alpha_k}$ ], and  $\bar{O}_k$  is its average over  $M$  samples [i.e.,  $\bar{O}_k = \frac{1}{M} \sum_{i=1}^M O_k(\mathbf{x}_i)$ ]. In `TURBORVB`, the logarithmic derivative [ $O_k(\mathbf{x})$ ] is computed very efficiently by using the AAD algorithm. Note that the derivatives of the local energy are not needed here because the Hamiltonian does not depend on any variational parameter. Instead, these terms are necessary in order to calculate ionic forces (i.e., derivatives of the total energies with respect to atomic positions). If the WF is an exact eigenstate of the Hamiltonian, the generalized forces  $f_k$  exactly vanish without statistical errors because the local energy is no longer dependent on  $\mathbf{x}$ . In other words, the derivatives have the zero-variance property and therefore represent the fundamental ingredients for an efficient WF optimization, as described in Sec. VII.

In practice, during an optimization, the code monitors the variational energy [ $E(\alpha)$ ] and the maximum value of the signal to noise ratio among all the force components, which is denoted as `devmax` in the code,

$$\text{devmax} \equiv \max_k \left( \left| \frac{f_k}{\sigma_{f_k}} \right| \right), \quad (113)$$

where  $\sigma_{f_k}$  represents the estimated error bar of the force  $f_k$ . This value is used in `TURBORVB` as one of the convergence criteria of the

optimization. To our experience, after a reasonable number of iterations, devmax stabilizes to a value  $\leq 4.0$ , when a correct (possibly local) minimum is being approached.

## B. Atomic forces

In TURBORVB, a very useful space warp coordinate transformation (SWCT)<sup>41</sup> is employed for the calculation of atomic forces. This scheme was introduced several decades ago in Ref. 105, in order to decrease the statistical errors of forces. The finite-difference expression for ionic force calculations is

$$\mathbf{F}_a = -\frac{\Delta}{\Delta \mathbf{R}_a} E = -\frac{E(\mathbf{R}_a + \Delta \mathbf{R}_a) - E(\mathbf{R}_a)}{\Delta \mathbf{R}_a} + O(\Delta \mathbf{R}_a). \quad (114)$$

The SWCT is used to mimic the displacement of charges around the nucleus,

$$\bar{\mathbf{r}}_i = \mathbf{r}_i + \Delta \mathbf{R}_a \omega_a(\mathbf{r}_i), \quad (115)$$

$$\omega_a(\mathbf{r}) = \frac{\kappa(|\mathbf{r} - \mathbf{R}_a|)}{\sum_{b=1}^M \kappa(|\mathbf{r} - \mathbf{R}_b|)}, \quad (116)$$

where  $\kappa(r)$  is a function that decays sufficiently fast for large  $r$  because the charges far from the nuclei should not be affected by the SWCT, and  $\omega \rightarrow 0$ , as a consequence of this requirement. It turns out that the original<sup>105</sup> simple choice  $\kappa(r) = 1/r^4$  works very well and is indeed adopted in TURBORVB.

Starting from the finite-difference expression, one can straightforwardly derive the corresponding differential expression,<sup>41</sup>

$$\mathbf{F}_a = -\left\langle \frac{d}{d\mathbf{R}_a} e_L \right\rangle + 2 \left\langle \langle e_L \rangle \left\langle \frac{d}{d\mathbf{R}_a} \ln(\mathcal{J}^\dagger \Psi) \right\rangle - \left\langle e_L \frac{d}{d\mathbf{R}_a} \ln(\mathcal{J}^\dagger \Psi) \right\rangle \right\rangle, \quad (117)$$

where  $\mathcal{J}$  is the Jacobian of the above transformation and the brackets indicate a Monte Carlo average over the trial WF. All the terms above can be written by the partial derivatives of the local energy and those of the logarithm of the WF,<sup>41</sup>

$$\frac{d}{d\mathbf{R}_a} e_L = \frac{\partial}{\partial \mathbf{R}_a} e_L + \sum_i \omega_a(\mathbf{r}_i) \frac{\partial}{\partial \mathbf{r}_i} e_L, \quad (118)$$

$$\frac{d}{d\mathbf{R}_a} \ln(\mathcal{J}^\dagger \Psi) = \frac{\partial}{\partial \mathbf{R}_a} \ln(\Psi) + \sum_i \left[ \omega_a(\mathbf{r}_i) \frac{\partial}{\partial \mathbf{r}_i} \ln(\Psi) + \frac{1}{2} \frac{\partial}{\partial \mathbf{r}_i} \omega_a(\mathbf{r}_i) \right]. \quad (119)$$

In order to evaluate these differential expressions,  $6N + 6N_{\text{at}}$  components have to be evaluated, namely,  $\left\langle \frac{\partial}{\partial \mathbf{r}_i} e_L \right\rangle$ ,  $\left\langle \frac{\partial}{\partial \mathbf{r}_i} \ln \Psi \right\rangle$ ,  $\left\langle \frac{\partial}{\partial \mathbf{R}_a} \ln e_L \right\rangle$ , and  $\left\langle \frac{\partial}{\partial \mathbf{R}_a} \ln \Psi \right\rangle$ .<sup>41</sup> These values are very efficiently computed in TURBORVB, by using the aforementioned AAD, which works even in the presence of pseudopotentials.

The SWCT significantly decreases the statistical errors, but the forces still have infinite variance properties because  $\partial e_L$  and  $\partial \ln(\Psi)$  diverge in the vicinity of the nodal surfaces. Attaccalite and Sorella<sup>106</sup> developed a reweighting method to address this issue in an unbiased way. Within this scheme, the QMC sampling is not driven by the chosen trial WF  $\Pi_T(\mathbf{x}) = \Psi_T^2(\mathbf{x})$  but by a slightly different guiding

function  $\Psi_G(\mathbf{x})$  defined by

$$\Psi_G(\mathbf{x}) = \frac{R^\epsilon(\mathbf{x})}{R(\mathbf{x})} \Psi_T(\mathbf{x}), \quad (120)$$

where  $R(\mathbf{x})$  vanishes more weakly than the trial function does, when the configuration  $\mathbf{x}$  approaches the nodal surface, whereas  $R^\epsilon$  is an appropriately chosen regularization of  $R$ , depending on  $\epsilon$ , that never vanishes. In this way, the guiding function is larger when approaching the nodal surface, and this singular region can be more accurately sampled in order to avoid infinite variance problems in the calculation of forces, as will be shown later. For this purpose,  $R(\mathbf{x})$  is defined in a way to vanish as the antisymmetric part of the WF to some power  $2\theta_R \leq 1$ . However, in order to avoid too large fluctuations, a simple relation is used that  $\text{Pf}(G) \simeq \frac{1}{G_{ij}^{-1}}$ , where  $G_{i,j}$  is the pairing function [Eq. (45)], here referred to as the general AGP case (the other cases straightforwardly follows upon some restriction of the matrix  $G$ ). Indeed, if the Pfaffian goes to zero, most of the matrix elements of the inverse of  $G$  diverge inversely proportional to it. Thus, the quantity

$$R(\mathbf{x}) = \left( S \sum_{ij} |G_{ij}^{-1}|^2 \right)^{-\theta_R} \quad (121)$$

satisfies the requirement without depending explicitly on the full Pfaffian that has fluctuations exponentially large in the number of particles and would lead to an inefficient regularization.<sup>106</sup> Here, we have also introduced the scaling factor  $S$  that takes into account that the WF may vanish also when a single particle is going to infinity, and not because it is approaching the nodal surface. In this case, the scaling factor has to vanish in a proper way so as to allow a non-vanishing  $R$ . In this way, the guiding function can decay sufficiently fast in this limit and can be normalized in an open system. This represents a necessary condition to have a stable simulation; otherwise, all electrons will be kicked out at infinity, after a long Markov chain. Therefore, the factor  $S$  is chosen to vanish as long as one row of the pairing function vanishes because it is corresponding to an electron going to infinity (all matrix elements containing such an electron coordinate have to vanish),

$$S = \min_i \sum_j |G_{ij}|^2. \quad (122)$$

This definition is also useful because the regularization proposed is scale-invariant, namely,  $R(\mathbf{x})$  remains unchanged if  $G$  is scaled by an arbitrary constant and therefore is adopted as is also for bulk systems, and it is particularly important when, in these cases, there exist large regions of almost negligible electronic density.<sup>107</sup>

After that,  $R^\epsilon(\mathbf{x})$  is defined as

$$R^\epsilon(\mathbf{x}) = \max[R(\mathbf{x}), \epsilon], \quad (123)$$

which is much simpler than the original proposal reported in Ref. 106. By using the new probability, forces (the HF and the Pulay) can be evaluated *with finite variance* as

$$\langle \mathcal{W}(\mathbf{x}) \mathbf{F}(\mathbf{x}) \rangle_{\Pi_G(\mathbf{x})} / \langle \mathcal{W}(\mathbf{x}) \rangle_{\Pi_G(\mathbf{x})}, \quad (124)$$

where  $\mathcal{W}(\mathbf{x})$  is the new weight,

$$\mathcal{W}(\mathbf{x}) = (\Psi_T(\mathbf{x}) / \Psi_G(\mathbf{x}))^2 \equiv \left( \frac{R(\mathbf{x})}{\max[R(\mathbf{x}), \epsilon]} \right)^2 \quad (125)$$

and  $\mathbf{F}(\mathbf{x})$ , according to Eq. (117), indicates the average of an appropriate random variable containing either the derivative of the local energy (the HF contribution) or the product of the local energy and the log derivative of the WF (the Pulay one), both contributions diverging as  $\simeq \frac{1}{\delta^2}$ , where  $\delta$  is the distance from the nodal surface. The point is that since in the vicinity of the nodal surface,  $\mathcal{W}(\mathbf{x})$  is proportional to  $\delta^{4\theta_R}$ , whereas the probability vanishes much slower as  $\Pi_G(\mathbf{x}) \simeq \delta^{2-4\theta_R}$ , the reweighting scheme solves the infinite variance problem for  $1/4 < \theta_R \leq 1/2$  because the random variable  $\mathcal{W}(\mathbf{x})\mathbf{F}(\mathbf{x})$  remains with *finite variance*. This follows after simple inspection of its integrability in the small  $\delta$  region, i.e.,  $\int d\delta \delta^{4\theta_R-2} < \infty$ . Thus, both the HF and the Pulay terms can be evaluated with finite variances in the TURBORVB evaluation of atomic forces. In the most recent versions of TURBORVB, the default value of  $\theta_R$  is taken to be at the middle of the interval of stability, i.e.,  $\theta_R = 3/8$ , because the value  $\theta = 1/2$ , previously chosen,<sup>106</sup> leads to some instability; namely, during the QMC sampling, the nodal surface is approached too much closely, and the determinants or Pfaffians become very ill-conditioned with obvious problems of numerical accuracy. Moreover, the value of  $\varepsilon$  is automatically selected during the first steps of the simulation in a way that the average reweighting factor  $\langle \mathcal{W}(\mathbf{x}) \rangle \simeq 0.8$ , which represents empirically an almost optimal setting.

Finally, we remark that this method to compute the energy derivatives with finite variance is applied by default in TURBORVB even for the optimization of the WF variational parameters and that a similar approach can be used when computing atomic forces within the LRDMC algorithm.

## VII. OPTIMIZATION OF WFs

### A. Stochastic reconfiguration

Once the energy derivatives can be computed, the most straightforward strategy to optimize a WF is to employ the steepest descent method, where the WF parameters are iteratively updated as follows:

$$\alpha_k \rightarrow \alpha'_k = \alpha_k + \delta\alpha_k, \quad (126)$$

$$\delta\alpha_k = \Delta \mathbf{f}_k, \quad (127)$$

where  $\Delta$  is a small constant and  $f_k \equiv -\frac{\partial E}{\partial \alpha_k}$  is the generalized force already defined in Eq. (111). However, it does not work well when optimizing highly non-linear WF parameters because a small change in a given variational parameter may produce a very different WF, whereas another parameter change may weakly affect the WF. Of course, one can use more sophisticated methods such as the Newton–Raphson method, the conjugate gradient, and the quasi-Newton method, but the straightforward implementation of these optimizations does not work efficiently within a stochastic approach such as QMC. In order to overcome this difficulty, a more efficient change in the variational parameters has been defined by means of a positive-definite *preconditioning* matrix  $\mathcal{S}$  and the generalized force vector  $\mathbf{f}$ ,

$$\alpha_k \rightarrow \alpha_k + \Delta \cdot (\mathcal{S}^{-1}\mathbf{f})_k, \quad (128)$$

where the matrix  $\mathcal{S}$  is stochastically evaluated by means of  $M$  configuration samples  $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ ,

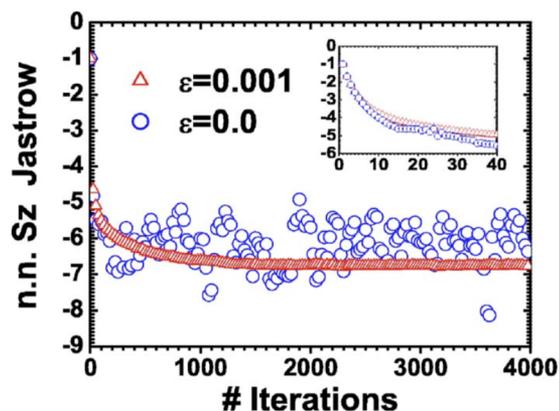
$$\mathcal{S}_{k,k'} = \left[ \frac{1}{M} \sum_{i=1}^M (O_k(\mathbf{x}_i) - \bar{O}_k)^* (O_{k'}(\mathbf{x}_i) - \bar{O}_{k'}) \right], \quad (129)$$

where  $O_k(\mathbf{x}_i) = \frac{\partial \ln \Psi(\mathbf{x}_i)}{\partial \alpha_k}$  and  $\bar{O}_k = \frac{1}{M} \sum_{i=1}^M O_k(\mathbf{x}_i)$ . The resulting approach is the so-called stochastic reconfiguration (SR) method.<sup>42</sup> Mazzola *et al.*<sup>108</sup> pointed out that the matrix  $\mathcal{S}$  is essentially a *metric* for the parameter space, measuring the distance of the underlying normalized WF. Therefore, Eq. (128) is simply the steepest descent in this curved manifold. This observation connects the SR method with the so-called natural gradient method, widely used in the context of deep learning.<sup>109</sup> In this context, for each parameter  $\alpha$ ,  $(\hat{O}_k - \bar{O}_k)$  can be interpreted as the score function (i.e., the gradient of the log-likelihood function) and the Fisher information matrix (FIM), respectively, while the WF square  $|\Psi(\mathbf{x})|^2$  plays the role of the likelihood function. In this sense, the stochastic reconfiguration method is essentially identical to the natural gradient optimization with the FIM that has been intensively used in the machine-learning community.

The straightforward implementation of the SR method is not stable mainly because the statistical noise sometimes makes the matrix  $\mathcal{S}$  ill-conditioned, which deteriorates the efficiency of the optimization method.<sup>43</sup> Therefore, in practice, the diagonal elements of the preconditioning matrix  $\mathcal{S}$  are shifted by a small positive parameter ( $\varepsilon$ ) as

$$s'_{i,i} = s_{i,i}(1 + \varepsilon). \quad (130)$$

This modification improves the efficiency of the optimization by several orders of magnitude, as shown in Fig. 7. Finally, the variational



**FIG. 7.** Optimization of the variational WF in the simple one-dimensional Heisenberg model  $\mathcal{H} = J \sum_i \tilde{S}_i \cdot \tilde{S}_{i+1}$  with the standard SR ( $\varepsilon = 0$ , open circles) and with the present regularization ( $\varepsilon = 0.001$ , open triangles). The evolution of the nearest neighbor spin–spin ( $S_2$ ) Jastrow parameter is plotted. The figure clearly shows that the SR method with regularization is several orders of magnitude more efficient than the standard SR for determining the variational parameter with a given statistical accuracy. The inset shows the first few iterations. Reprinted with permission from S. Sorella, M. Casula, and D. Rocca, *J. Chem. Phys.* **127**, 014105 (2007). Copyright 2007 AIP Publishing LLC.

parameters are updated as

$$\alpha_k \rightarrow \alpha_k + \Delta \cdot (\mathbf{S}^{-1} \mathbf{f})_k. \quad (131)$$

In practice, the user should provide two input parameters, namely,  $\Delta$  (denoted as `tpar` in the code input variables) and  $\varepsilon$  (denoted as `parr`).

## B. Linear method

In `TURBORVB`, the state-of-the-art QMC optimization method is also implemented, namely, the so-called linear method.<sup>44–46</sup> In this scheme, a many-body WF is expanded up the linear order (i.e., considering only the first derivatives for the variational parameters),

$$|\Psi_{\alpha+\delta\alpha}\rangle = |\Psi_\alpha\rangle + \sum_{k=1}^p \delta\alpha_k \frac{\partial}{\partial\alpha_k} |\Psi_\alpha\rangle \equiv |\Psi_0\rangle + \sum_{k=1}^p \delta\alpha_k |\Psi_k\rangle, \quad (132)$$

where  $\alpha_k$  is the  $k$ th variational parameter and  $\Psi_k(\mathbf{x})$  is the first derivative with respect to the  $k$ th variational parameter. Within this expansion, the expectation value of the energy reads

$$E_{\alpha+z} = \frac{\sum_{k,k'=0}^p z_k^* z_{k'} \langle \Psi_k | \hat{\mathcal{H}} | \Psi_{k'} \rangle}{\sum_{k,k'=0}^p z_k^* z_{k'} \langle \Psi_k | \Psi_{k'} \rangle}. \quad (133)$$

In order to obtain the vector  $\mathbf{z}$  that minimizes this expectation value, one should solve the generalized eigenvalue problem,

$$\mathcal{H}\mathbf{z} = E\mathbf{S}\mathbf{z}, \quad (134)$$

where  $\mathcal{H}_{k,k'} = \langle \Psi_k | \hat{\mathcal{H}} | \Psi_{k'} \rangle$  and  $\mathcal{S}_{k,k'} = \langle \Psi_k | \Psi_{k'} \rangle$ .

This algorithm is more conveniently implemented in the so-called semi-orthogonal basis,<sup>15</sup>

$$|\Psi_{\alpha+\delta\alpha}\rangle = z_0 |\tilde{\Psi}_\alpha\rangle + \sum_{k=1}^p z_k (\hat{O}_k - \bar{O}_k) |\tilde{\Psi}_\alpha\rangle, \quad (135)$$

where  $|\tilde{\Psi}_\alpha\rangle = |\Psi_\alpha\rangle / \|\Psi_\alpha\|$ ,  $\langle x | \hat{O}_k | x' \rangle = \delta_{x,x'} O_k(\mathbf{x})$ ,  $O_k(\mathbf{x}) = \frac{\partial}{\partial\alpha_k} \ln |\Psi_\alpha(\mathbf{x})|$ , and  $\bar{O}_k = \langle \Psi_\alpha | \hat{O}_k | \Psi_\alpha \rangle / \langle \Psi_\alpha | \Psi_\alpha \rangle$ . `TURBORVB` calculates  $\mathbf{z}$  by solving the generalized equation [Eq. (134)] with the matrices,

$$\mathcal{H}_{k,k'} \approx \frac{1}{M} \sum_{i=1}^M (O_k(x_i) - \bar{O}_k)^* \frac{\langle x_i | \hat{\mathcal{H}} (\hat{O}_{k'} - \bar{O}_{k'}) | \Psi_\alpha \rangle}{\langle x_i | \Psi_\alpha \rangle}, \quad (136)$$

$$\mathcal{S}_{k,k'} \approx \frac{1}{M} \sum_{i=1}^M (O_k(x_i) - \bar{O}_k)^* (O_{k'}(x_i) - \bar{O}_{k'}), \quad (137)$$

which can be readily evaluated by a Monte Carlo sampling, using not only the random variables  $O_k(x_i)$  necessary for the simpler SR technique but also the parameter derivatives of the local energy  $\partial_{\alpha_k} e_L(x_i)$ , which can be directly computed by AAD, and thus allow the calculation of the above Hamiltonian matrix elements, by straightforward algebra,

$$\partial_{\alpha_k} e_L(x_i) = \partial_{\alpha_k} \frac{\langle x_i | \hat{\mathcal{H}} | \Psi_\alpha \rangle}{\langle x_i | \Psi_\alpha \rangle} = \frac{\langle x_i | \hat{\mathcal{H}} (\hat{O}_k - \bar{O}_k) | \Psi_\alpha \rangle}{\langle x_i | \Psi_\alpha \rangle}. \quad (138)$$

Note that the code does not always take the eigenvector corresponding to the lowest eigenvalue (that in turn may be also complex, as in the linear method the matrix  $\mathcal{H}_{k,k'}$  is not Hermitian) but takes the one that maximizes  $|z_0|^2$  (i.e., the coefficient of the zeroth-order term) in order to have the most stable parameter change, namely,

the updated WF being as close as possible to the old one. Finally, the code updates the variational parameters by using the obtained  $\mathbf{z}$  and an input parameter  $\Delta$  (denoted as `tpar`) as

$$\alpha_k \rightarrow \alpha_k + \Delta \cdot z_k / z_0, \quad (139)$$

where  $\Delta \simeq 1/3$  is the default `TURBORVB` choice that is much more stable than the original algorithm ( $\Delta = 1$ ), but it is approximately 1/3 slower. The linear method is usually rather unstable for large systems and many parameters, and also, in such cases, each iteration requires a diagonalization of a huge matrix (a task that can be often prohibitive). In `TURBORVB`, a practical compromise has been devised by using the linear method for a restricted variational space containing up to a maximum number (`npbra`) of variational parameters with the largest signal to noise ratio [optionally larger than `parcutpar`, see Eq. (113)] and/or a number (`ncg`) of global line parameter directions spanned by the natural gradient ones (e.g.,  $\Delta$  is one variational parameter of this form in Eq. (131)) calculated at the given optimization iteration or the previous closest ones,<sup>43</sup> optionally neglecting the ones with a smaller signal-noise ratio (`<parcutmin`).

## C. Practical rule

In most optimizations, the number of samplings in VMC is much larger than the number of variational parameters and the optimization is stable and efficient. In the opposite case, the preconditioning matrix  $\mathbf{S}$  becomes rank-deficient singular; therefore, the optimization does not work properly (cf. the Cramér–Rao inequality). A practical rule is to set the number of samples  $M$  such that<sup>15</sup>

$$M \geq 5 \sim 10 \times p, \quad (140)$$

where  $p$  is the number of variational parameters. However, one can also employ a very small number of samplings with sufficiently large  $\varepsilon$  in the stochastic reconfiguration method because the regularization of the matrix  $\mathbf{S}$  in Eq. (130) can remove all instabilities as long as  $\varepsilon$  is sufficiently large, as in the infinite  $\varepsilon$  limit one recovers the very powerful stochastic gradient technique well-known within machine learning. This method seems promising for optimizing even a huge amount of variational parameters and could certainly represent an important development in QMC.

At present, the inversion in Eq. (131) can be done in `TURBORVB` without storing explicitly the matrix<sup>110</sup> as the associated linear problem is solved iteratively using conjugate gradients and implicit small matrix-vector operations, optimally distributed within the MPI protocol. Thus, the cost of this inversion becomes negligible in the large  $\varepsilon$  limit and linear with the number of variational parameters.

## VIII. MOLECULAR DYNAMICS

In `TURBORVB`, several types of *ab initio* molecular dynamics (MD) have been implemented for both classical and quantum nuclei. The MD is driven by ionic forces  $\mathbf{F} = -\nabla_{\mathbf{R}} V(\mathbf{R})$ , where  $\mathbf{F} \equiv \{\mathbf{F}_1, \dots, \mathbf{F}_N\}$ ,  $\mathbf{R} \equiv \{\mathbf{R}_1, \dots, \mathbf{R}_N\}$ , and the potential energy landscape  $V(\mathbf{R})$  is evaluated by VMC, namely,

$$V(\mathbf{R}) = \frac{\langle \Psi_{\mathbf{R}} | \hat{\mathcal{H}}(\mathbf{R}) | \Psi_{\mathbf{R}} \rangle}{\langle \Psi_{\mathbf{R}} | \Psi_{\mathbf{R}} \rangle}. \quad (141)$$

In Eq. (141),  $|\Psi_{\mathbf{R}}\rangle$  is the QMC WF that, according to the Born–Oppenheimer (BO) approximation, minimizes the expectation value of  $\mathcal{H}(\mathbf{R})$  at each ionic position  $\mathbf{R}$ . The expression for the ionic forces is quite complex. Computing these forces with finite variance and in a fast way, as done in TURBORVB (Sec. VI B), is of paramount importance to make a QMC-based MD possible. Moreover, in the QMC framework, the evaluation of  $V(\mathbf{R})$  and its corresponding force estimators are intrinsically noisy. The statistical noise must be kept under control if one wants to have an unbiased sampling of the phase space during the propagation of the trajectory. This issue has been solved by resorting to a Langevin type of molecular dynamics in both classical and quantum formulations, where the QMC noise becomes a controlled source of thermalization at the target temperature  $T$ .

Two types of Langevin dynamics (LD) have been implemented in TURBORVB: the second-order LD, in its classical (Sec. VIII A) and quantum (Sec. VIII B) variants, and the first-order LD accelerated by the covariance matrix of QMC forces (Sec. VIII C).

### A. Second-order Langevin dynamics

The equations of motion of the second-order LD read<sup>106</sup>

$$\dot{\mathbf{v}} = -\gamma(\mathbf{R}) \cdot \mathbf{v} + \mathbf{F}(\mathbf{R}) + \boldsymbol{\eta}(t), \quad (142)$$

$$\dot{\mathbf{R}} = \mathbf{v}, \quad (143)$$

$$\langle \eta_i(t) \eta_j(t') \rangle = S_{ij}(\mathbf{R}) \delta(t - t'), \quad (144)$$

where  $\mathbf{R}$ ,  $\mathbf{v}$ ,  $\mathbf{f}$ , and  $\boldsymbol{\eta}$  are the  $3N_{\text{at}}$ -dimensional vectors representing atomic positions, velocities, and deterministic and stochastic forces of  $N_{\text{at}}$  atoms, written in mass-scaled coordinates,

$$\begin{aligned} \mathbf{R}_i &= \mathbf{R}_i^0 \sqrt{m_i}, \\ \mathbf{F}_i &= \mathbf{F}_i^0 / \sqrt{m_i}, \\ \boldsymbol{\eta}_i &= \boldsymbol{\eta}_i^0 \sqrt{m_i} \end{aligned} \quad (145)$$

for  $i = 1, \dots, N_{\text{at}}$ . The stochastic forces are related to the friction matrix  $\gamma$  through the fluctuation–dissipation theorem, namely,

$$S(\mathbf{R}) = 2T\gamma(\mathbf{R}), \quad (146)$$

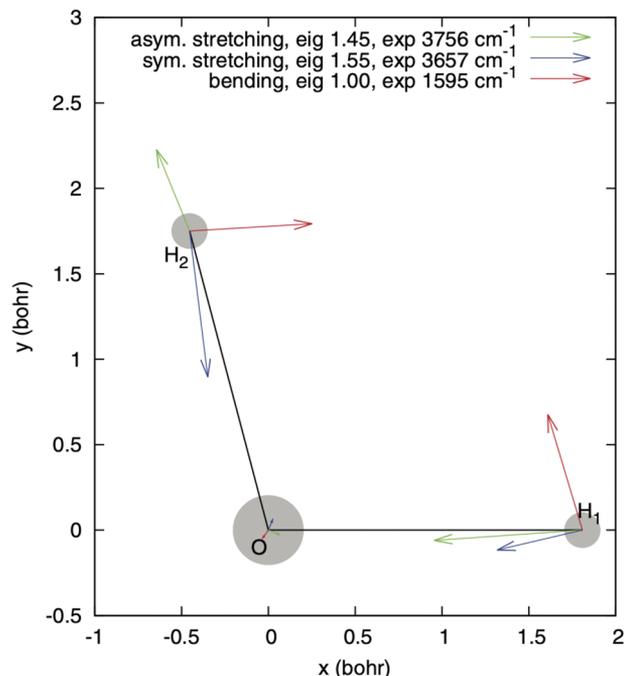
with the temperature  $T$  expressed in atomic units. TURBORVB exploits the freedom in Eq. (146), by assuming

$$S(\mathbf{R}) = s_0 \mathcal{I} + \Delta_0 S^{\text{QMC}}(\mathbf{R}), \quad (147)$$

where  $s_0$  and  $\Delta_0$  are constants,  $\mathcal{I}$  is the identity, and  $S^{\text{QMC}}(\mathbf{R})$  is the covariance matrix of QMC forces,

$$S_{ij}^{\text{QMC}}(\mathbf{R}) = \langle (\mathbf{F}_i(\mathbf{R}) - \langle \mathbf{F}_i(\mathbf{R}) \rangle) \langle (\mathbf{F}_j(\mathbf{R}) - \langle \mathbf{F}_j(\mathbf{R}) \rangle) \rangle. \quad (148)$$

In the above equation,  $\langle \dots \rangle$  refers to the average over the QMC sampling at the given MD step. The friction, a quantity that controls the nuclear sampling efficiency, is, therefore, position-dependent now. Luo *et al.* have shown that the force covariance matrix  $S_{ij}^{\text{QMC}}(\mathbf{R})$  is, within a good approximation, proportional to the dynamical matrix (Fig. 8). Therefore, the choice in Eq. (148) realizes the nearly optimal damping of the high-frequency modes. After discretizing Eqs. (142) and (143) in the time interval  $t_n - \tau/2 < t < t_n + \tau/2$ , where the index



**FIG. 8.** Eigenvectors of the  $3 \times 3$  correlation matrix of QMC forces  $S^{\text{QMC}}$  in the water monomer. The obtained eigenvectors correspond to the well-known three vibrational modes: bending (red arrows), symmetrical (blue arrows), and asymmetrical (green arrows) stretching. The smaller eigenvalue corresponds to the lowest-frequency vibrational mode. The eigenvalues in the plot are all rescaled by the lowest one. Reprinted with permission from Y. Luo, A. Zen, and S. Sorella, *J. Chem. Phys.* **141**, 194112 (2014). Copyright 2014 AIP Publishing LLC.

$n$  denotes the time slice  $t_n = n\tau$ , with time step  $\tau$ , and by integrating the above equations, one obtains

$$\mathbf{v}_{n+1}(t) = e^{-\gamma_n \tau} \cdot \mathbf{v}_n + \Gamma_n(\mathbf{F}_n + \tilde{\boldsymbol{\eta}}), \quad (149)$$

$$\mathbf{R}_{n+1} = \mathbf{R}_n + \tau \mathbf{v}_{n+1}, \quad (150)$$

with  $\mathbf{F}_n \equiv \mathbf{F}(\mathbf{R}_n)$ ,  $\gamma_n \equiv \gamma(\mathbf{R}_n)$ ,

$$\Gamma_n = \gamma_n^{-1} (I - e^{-\gamma_n \tau}), \quad (151)$$

and

$$\tilde{\boldsymbol{\eta}} = \frac{\gamma_n}{2 \sinh(\gamma_n \tau/2)} \int_{t_n - \tau/2}^{t_n + \tau/2} e^{-\gamma_n(t-t_n)} \cdot \boldsymbol{\eta}(t) dt. \quad (152)$$

By using Eq. (152), it is easy to show that the correlator defining the discrete noise is given by the following  $3N_{\text{at}} \times 3N_{\text{at}}$  matrix:

$$\langle \tilde{\eta}_i \tilde{\eta}_j \rangle \equiv S^* = T\gamma_n^2 \coth(\gamma_n \tau/2). \quad (153)$$

In this way,  $\tilde{\boldsymbol{\eta}}$  fulfill the fluctuation–dissipation theorem required by the Langevin thermostat. However, part of this noise is already present in the QMC force evaluation ( $\tilde{\boldsymbol{\eta}}^{\text{QMC}}$ ), as quantified by Eq. (148). Hence, TURBORVB adds the external random force  $\tilde{\boldsymbol{\eta}}^{\text{ext}}$  such that  $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}^{\text{ext}} + \tilde{\boldsymbol{\eta}}^{\text{QMC}}$ . This is generated according to the following

correlator:

$$\langle \tilde{\eta}_i^{\text{ext}} \tilde{\eta}_j^{\text{ext}} \rangle = S^* - S_{\text{QMC}}, \quad (154)$$

in order to fulfill the original fluctuation–dissipation theorem. Since the correlation matrix in Eq. (154) is positive definite as long as  $\tau < \Delta_0$ , the correlated external force can be readily generated by standard algorithms (e.g., the Box–Muller method<sup>112</sup> combined with the matrix diagonalization). This noise correction plays an important role in targeting a stable temperature even in the presence of statistical errors in the QMC forces. The original numerical integration detailed here has been further improved by Mazzola and Sorella (see the supplementary notes of Ref. 113).

## B. Path integral Ornstein–Uhlenbeck dynamics

Mouhat *et al.* have recently introduced the nuclear quantum effects into the second-order LD based on a path integral molecular dynamics (PIMD) approach,<sup>114</sup> driven by QMC nuclear forces. As a starting point of the Feynman path integral (PI) theory, the quantum partition function  $Z = \text{Tr}[e^{-\beta\mathcal{H}}]$  (with  $\beta = 1/k_B T$ ) can be written as<sup>115</sup>

$$Z \simeq \frac{1}{(2\pi\hbar)^{3N_{\text{at}}L}} \int d^{N_{\text{at}}L} \mathbf{p} d^{N_{\text{at}}L} \mathbf{q} e^{-\tau\beta\mathcal{H}_L(\mathbf{p}, \mathbf{q})}, \quad (155)$$

with  $\tau\beta = \beta/L$  denoting the imaginary time step. We have replaced here the true quantum particles by fictitious classical ring polymers, consisting of  $L$  replicas (beads) of the system. The beads in these necklaces are connected to each other by harmonic springs, evolving in the imaginary time. Without the loss of generality, we can extend the definition of classical vectors to the quantum case, by including all  $L$  replicas, such that the resulting vector  $\mathbf{q} \equiv (\mathbf{R}^{(1)}, \dots, \mathbf{R}^{(i)}, \dots, \mathbf{R}^{(L)})^T$  is  $3N_{\text{at}}L$ -dimensional. If interpreted classically, the partition function  $Z$  in Eq. (155) describes a system at the effective temperature  $LT$ . The Hamiltonian  $\mathcal{H}_L$  corresponding to this quantum-to-classical isomorphism reads as

$$\begin{aligned} \mathcal{H}_L(\mathbf{p}, \mathbf{q}) = & \sum_{i=1}^{N_{\text{at}}} \sum_{j=1}^L \left( \frac{1}{2} [\mathbf{p}_i^{(j)}]^2 + \frac{1}{2} \tilde{\omega}_L^2 (\mathbf{R}_i^{(j)} - \mathbf{R}_i^{(j-1)})^2 \right) \\ & + \sum_{j=1}^L V(\mathbf{R}_1^{(j)}, \dots, \mathbf{R}_{N_{\text{at}}}^{(j)}), \end{aligned} \quad (156)$$

written in mass-scaled coordinates [Eq. (145)], with  $\tilde{\omega}_L = L/\beta\hbar$ , and subjected to the ring boundary condition  $\mathbf{R}_i^{(0)} \equiv \mathbf{R}_i^{(L)}$ . The system is thermalized with a Langevin thermostat such that the related Liouville operator  $\mathcal{L}$  can be written as

$$\mathcal{L} = \sum_{i=1}^{3N_{\text{at}}L} \left( \mathcal{F}_i \partial_{p_i} + p_i \partial_{q_i} - \sum_{j=1}^{3N_{\text{at}}L} \gamma_{ij} (\partial_{p_i} p_j + k_B T M \partial_{p_i} \partial_{p_j}) \right) \quad (157)$$

in mass-scaled coordinates, with  $\partial_{q_i} \equiv \frac{\partial}{\partial q_i}$ .  $\mathcal{L}$  is built upon the Hamiltonian propagation, driven by the first two terms, and the Langevin thermostat, represented by the last two, deriving from the Fokker–Planck equation. Hereafter in this section,  $i$  and  $j$  run over all particles and beads indexed together. In Eq. (157),  $\mathcal{F}_i \equiv F_i^{\text{BO}} + F_i^{\text{harm}}$  is the total force acting on each replica, comprising the BO (intra-replica) and harmonic contributions (inter-replicas), where  $F_i^{\text{BO}} \equiv -\partial_{q_i} \tilde{V}$  and  $\tilde{V} = \sum_{j=1}^L V(\mathbf{R}_1^{(j)}, \dots, \mathbf{R}_{N_{\text{at}}}^{(j)})$ .

The quantum-to-classical isomorphism Hamiltonian in Eq. (156) includes very different energy scales. To cope with this, we split the Liouvillian in Eq. (157) into two operators, one containing only the physical (BO) modes and the other depending exclusively on the harmonic modes of the rings. To do so, we first separate the friction matrix into two contributions,

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\text{BO}} + \boldsymbol{\gamma}^{\text{harm}}. \quad (158)$$

We can then rewrite the total Liouvillian as the sum of two terms,  $\mathcal{L} = \mathcal{L}^{\text{BO}} + \mathcal{L}^{\text{harm}}$ , where

$$\mathcal{L}^{\text{harm}} = \sum_i \left( F_i^{\text{harm}} \partial_{p_i} + p_i \partial_{q_i} - \sum_j \gamma_{ij}^{\text{harm}} (\partial_{p_i} p_j + k_B T M \partial_{p_i} \partial_{p_j}) \right), \quad (159)$$

$$\mathcal{L}^{\text{BO}} = \sum_i \left( F_i^{\text{BO}} \partial_{p_i} - \sum_j \gamma_{ij}^{\text{BO}} (\partial_{p_i} p_j + k_B T M \partial_{p_i} \partial_{p_j}) \right), \quad (160)$$

in such a way that we can break up the evolution operator via a Trotter factorization<sup>115</sup> to get the following symmetric propagator:<sup>116,117</sup>

$$e^{-\mathcal{L}\delta t} \simeq e^{-\mathcal{L}^{\text{BO}}\delta t/2} e^{-\mathcal{L}^{\text{harm}}\delta t} e^{-\mathcal{L}^{\text{BO}}\delta t/2}. \quad (161)$$

The equations of motion corresponding to the propagator in Eq. (161) have been implemented in `TURBORVB`, where now  $F_i^{\text{BO}}$  can be computed in a QMC framework, and therefore, it can be affected by an intrinsic noise, as was in the classical scheme (Sec. VIII A).

We note that the equations of motion generated by  $\mathcal{L}^{\text{harm}}$  are *linear* in both  $\mathbf{p}$  and  $\mathbf{q}$ , thus exactly solvable in an analytic closed form. This is because the non-linear BO components of the total force have been put in the  $\mathcal{L}^{\text{BO}}$  factor. The dynamics generated by  $\mathcal{L}^{\text{harm}}$  is a quantum Ornstein–Uhlenbeck process, which describes the motion of Brownian quantum particles under the influence of friction.

The exact integration of  $\mathcal{L}^{\text{harm}}$  without further splitting the Langevin thermostat from the harmonic modes of the rings is a major achievement, which gives the algorithm an enhanced stability and better scaling with respect to the time step size  $\tau$  and the number of replica  $L$ . Moreover, the time step error is significantly smaller than the one of other PILD methods.<sup>118</sup> Thus, this is a key feature of the PI algorithm implemented in `TURBORVB`. This is also why the implemented method is called path integral Ornstein–Uhlenbeck dynamics (PIOUD).

If we now look at the  $\mathcal{L}^{\text{BO}}$  factor in Eq. (161), the corresponding equations of motion are the ones in Eq. (12) of Ref. 111. Indeed, only the momenta are evolved in  $\mathcal{L}^{\text{BO}}$ , and the resulting equations are equal to those of the simple classical Langevin algorithm introduced in Ref. 106, restricted to  $\mathbf{p}$ , which in the present case is a  $3N_{\text{at}}L$ -dimensional vector. As in the classical second-order QMC-driven LD, the QMC noise affecting the BO forces is controlled, thanks to the friction  $\boldsymbol{\gamma}^{\text{BO}}$  and the fluctuation–dissipation theorem. A noise reduction scheme similar to the one described in Eqs. (153) and (154) is also applied in this case, yielding a thermalization toward a steady temperature ensemble even for quantum particles.

We refer the interested reader to Ref. 114, where we report a detailed description of the PIOUD algorithm and the integrated equations of motion.

### C. First-order Langevin dynamics

TURBORVB also features an accelerated first-order dynamics to sample the (classical) ionic canonical distribution. This molecular dynamics scheme has been introduced recently by Mazzola and Sorella.<sup>119</sup> Here, the ions formally undergo the following dynamics:

$$\dot{\mathbf{R}} = \mathcal{S}^{-1}(\mathbf{R}) \cdot \mathbf{F}(\mathbf{R}) + \boldsymbol{\eta}, \quad (162)$$

$$\langle \eta_i \eta_j \rangle = 2T\delta(t-t')\mathcal{S}_{ij}^{-1}(\mathbf{R}), \quad (163)$$

where  $T$  is the temperature,  $\mathbf{F}(\mathbf{R})$  and  $\boldsymbol{\eta}$  are the deterministic and stochastic forces, and  $\mathcal{S}$  is the covariance matrix of the QMC forces, as introduced in Eq. (148). This dynamics generalizes the Newton method, which is valid for structural relaxation, to finite-temperature.

Indeed, the use of the matrix  $\mathcal{S}$ , which is empirically proportional to the Hessian matrix, can drastically decrease the auto-correlation time of the Markov chain. In other words, this *metric* plays an important role in treating the different time scales of a given system. In this way, one can still use a quite large integration time step  $\tau$ , which is beneficial in reducing the autocorrelation time of quantities that depend on slowly varying degrees of freedom (e.g., molecular diffusion), without introducing any bias in the sampling.

Discretization of Eq. (162) for a finite time step  $\tau$  at finite temperature is rather involved and has been described in Ref. 119 and reads

$$\begin{aligned} \mathbf{R}(t+\tau) = & \mathbf{R}(t) + \sqrt{2T\tau} \cdot \mathbf{z}(t) + \tau\mathcal{S}^{-1}(\mathbf{R})\mathbf{F}(\mathbf{R}) \\ & - \mathcal{S}^{-1}(\mathbf{R}) \left( \frac{\mathcal{S}(\mathbf{R}(t-\tau)) - \mathcal{S}(\mathbf{R})}{2} \right) (\mathbf{R}(t-\tau) - \mathbf{R}(t)), \end{aligned} \quad (164)$$

with

$$\langle z_i(t)z_j(t) \rangle = \mathcal{S}_{ij}^{-1}(\mathbf{R}(t)). \quad (165)$$

As the effective temperature depends on the finite integration time, the convergence to the target temperature  $T_{\text{target}}$  for  $\tau \rightarrow 0$  can be improved by a renormalization of the temperature  $T$  used in the simulation, namely, by using the above equations with an appropriate choice of  $T$ ,

$$T = T_{\text{target}}(1 - a\tau), \quad (166)$$

where  $a = 1/2$  in the case of the harmonic potential leads to the exact result for any discretization time  $\tau$ , as long as  $\mathcal{S}$  is the exact elastic matrix. In the general case, the parameter  $a$  should be tuned to speed up convergence to the  $\tau \rightarrow 0$  limit. Conversely, the target temperature of the simulation can be measured *a posteriori*, as is often done in the case of second-order Langevin dynamics, by measuring the average squared velocities of the particles.

The accelerated first-order Langevin dynamics has been successfully employed in the most recent studies of dense liquid hydrogen, providing equilibrated simulations even close to phase transitions.<sup>119,120</sup> Moreover, nuclear quantum effects can also be computed within this first-order LD, in exactly the same way as was implemented for the second-order case discussed in Subsection VIII B. However, so far, it is not clear which method is more efficient, and further studies are also necessary to clarify what is the optimal choice of the acceleration matrix  $\mathcal{S}(\mathbf{R})$  within the QMC framework.

### IX. IMPLEMENTATION AND OPERATIONS

Table I shows the main modules implemented in TURBORVB and a brief description of their functionalities. All programs reported in Table I are coded in FORTRAN90. The flow-chart of a typical QMC calculation (see Fig. 1) is detailed as follows: starting from the geometry and the chemical composition, the user first chooses the basis set for both Jastrow and antisymmetric factors and generates a JAGP-type Ansatz using makefort10.x (fort.10 is the wave function filename in TURBORVB).

TABLE I. Main modules in TURBORVB.

Module	Description
makefort10.x	Generates a JAGP WF from a given basis set and structure
assembling_pseudo.x	Prepares pseudopotentials
convertfort10mol.x	Adds molecular orbitals. If the number of molecular orbitals is equal to (larger than) half the number of electrons in a system, the resultant WF is the JSD (JAGPn)
convertfort10.x	Converts a wave function type (fort.10_in), i.e., JSD/JAGP/JAGPn/JPf/JPfn, to another one with a different type (fort.10_out). The output (fort.10_new) is as close as possible to the input and may also include an optimal contracted (hybrid) basis set.
convertfortpfaff.x	Converts a JAGP WF to a JPf one
prep.x	Performs a DFT calculation
turborvb.x	Performs VMC optimization, VMC evaluation, LRDMC, structural optimization, and molecular dynamics
readforward.x	Performs correlated samplings and calculates various physical properties

The antisymmetric part of the WF is initially determined at the DFT mean-field level. Since the DFT calculation is based on a single Slater determinant, the generated JAGP WF should be converted to a JSD one. This is performed by the `convertfort10mol.x` module, which adds  $M$  molecular orbitals to the WF, where  $M = N^\dagger$ .<sup>121</sup> The DFT molecular orbitals are then used to construct the initial trial WF for subsequent QMC calculations. In QMC, the user can choose to work with five *Ansätze*, namely, JSD, JAGP, JAGPn, JPfn, and JPf. The initial antisymmetric part of the JSD and JAGPn wave functions can be directly obtained from the DFT orbitals. In the JSD case, only the occupied orbitals are imported, while in the JAGPn WF also the unoccupied orbitals will be used in the geminal expansion, up to the  $n$ th orbital. Another possibility is to employ a JAGP *Ansatz*. In that case, one has to convert the initial JSD trial WF obtained by DFT to the JAGP one using `convertfort10.x`.

JAGPn can also be obtained by applying `convertfort10mol.x` to a previously determined JAGP *Ansatz*. Analogously, a JPf WF can be converted from a JAGP WF using `convertpfaff.x`. JPfn can be obtained by applying `convertfort10mol.x` to the JPf *Ansatz*. These conversions can be done either before or after a QMC optimization. The possibilities for *Ansatz* conversion are schematically drawn in Fig. 4.

After the initial determination of the antisymmetric part of the WF, one performs the WF optimization at the VMC level. This is an important step used to determine the best variational parameters of the Jastrow factor at fixed determinant or to fully relax the WF by optimizing both Jastrow and determinant parameters.

Finally, the user can proceed to VMC and LRDMC calculations using the optimized WF. In LRDMC, in order to remove the lattice discretization error, it is better to repeat the calculations with several lattice spaces  $a$  and to extrapolate the results with a quartic function  $E(a) = E_0 + k_1 a^2 + k_2 a^4$ , where  $E_0$  is the extrapolated ( $a \rightarrow 0$ ) energy.

The Python tool `TURBO-GENIUS` makes all the above steps more straightforward, as discussed later.

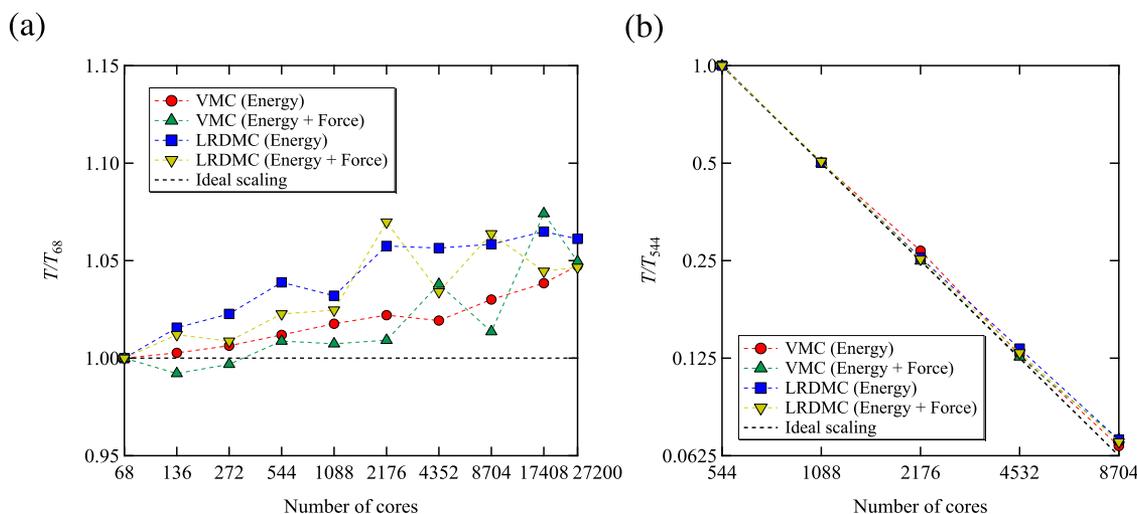
## X. PARALLELIZATION AND BENCHMARKS

`TURBORVB` has been parallelized using the MPI and OpenMP libraries, also supporting the hybrid MPI/OpenMP protocol. As well known, the QMC algorithm is readily parallelized by employing many independent random walks (called walkers), initialized with different seeds for random number generators. Figure 9 shows the strong and the weak scaling performances of `TURBORVB`, which are compared with the ideal scaling. The benchmarks have been measured using the conventional  $2 \times 2 \times 2$  diamond supercell containing 64 carbon atoms for a simulation with 256 electrons on Marconi KNL nodes (Cineca supercomputer/SCAI, Intel Xeon Phi 7250 CPU, 1.40 GHz, 68 cores per node without hyperthreading). For the weak scaling case, we set the number of walkers  $M$  to the minimum possible one (one walker per core) and increased the number of cores from 68 (one node) to 27 200 (400 nodes). On the other hand, for the strong scaling case, we fixed the number of walkers  $M = 2176$  and increased the number of cores from 544 (8 nodes) to 8704 (128 nodes) with the hybrid MPI/OpenMP protocol (the number of openMP threads was set to four). For both cases, the performances are excellent because all techniques are off from the ideal scaling by  $\sim 5\%$  at worst in all cases performed here. At present, the code does not support the use of Graphics Processing Unit (GPU) accelerators. This feature is currently under development.

## XI. POSTPROCESSING

### A. Correlated samplings

A very efficient and easy-to-use correlated sampling technique has been implemented in `TURBORVB`. This is very useful in evaluating

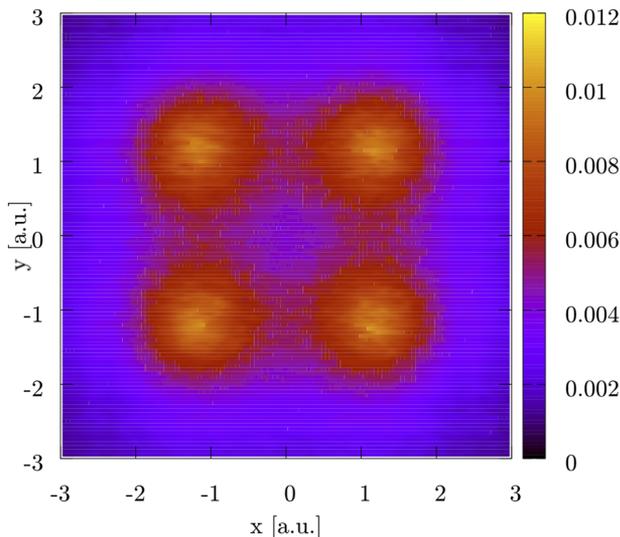


**FIG. 9.** (a) The weak and (b) strong scalings of `TURBORVB`. The benchmarks were measured using the conventional  $2 \times 2 \times 2$  diamond with the ccECP pseudopotential (256 electrons in the simulation cell) on Marconi KNL nodes (Cineca supercomputer/SCAI).

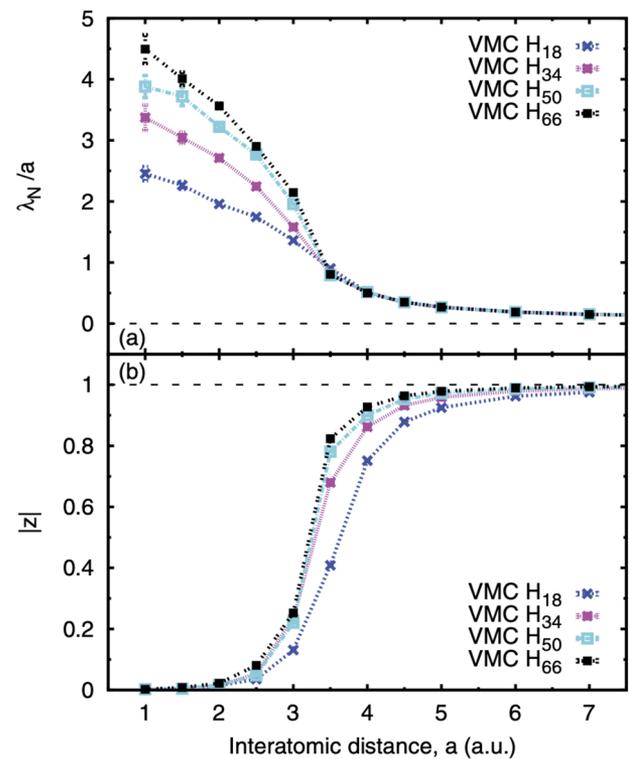
total energy differences between two very similar WFs. One of its practical sides is to investigate whether a conversion of a WF (e.g., JSD  $\rightarrow$  JAGP) has been successful. A user should run a short VMC calculation using one of the two mentioned WFs, during which the walkers' history is stored in an appropriate scratch area (turborvb.scratch). After that, a second run should be executed (by using the readforward.x module) to determine at the end the correlated energy difference and the overlap between the two WFs. Correlated samplings for LRDMC or for the difference in other quantities, e.g., forces, have not been implemented yet. The latter could be very important for the frozen phonon calculation based on the ionic force difference method, which represents a possible subject of research.

## B. Physical properties

TURBORVB enables us to calculate various properties using the optimized many-body WF at both the VMC and LRDMC levels by averaging local observables and applying the forward walking technique.<sup>15</sup> The observables presently available in the readforward.x module can be categorized as follows: (i) Charge density [ $\rho(\mathbf{r}) = \rho^\uparrow(\mathbf{r}) + \rho^\downarrow(\mathbf{r})$ ] and spin density [ $\rho^\sigma(\mathbf{r}) = (\rho^\uparrow(\mathbf{r}) - \rho^\downarrow(\mathbf{r}))/2$ ], fundamental properties obtained by a stochastic evaluation of multi-dimensional integrals, involving the many-body WF and improved estimators, such as those proposed by Assaraf, Caffarel, and Scemama.<sup>122</sup> Figure 10 shows the charge density of the square four hydrogen ( $H_4$ ) depicted by Genovese *et al.* (ii) Electronic correlation functions, such as the charge-charge and spin-spin structure factors, useful to study the critical behavior close to a phase transition or the physical properties of a given phase. (iii) The expectation values of the spin component along the quantization



**FIG. 10.** The  $xy$ -plane charge density of the square four hydrogen ( $H_4$ ), calculated using a JAGP WF. Reprinted with permission from C. Genovese, A. Meninno, and S. Sorella, *J. Chem. Phys.* **150**, 084102 (2019).<sup>124</sup> Copyright 2019 AIP Publishing LLC.

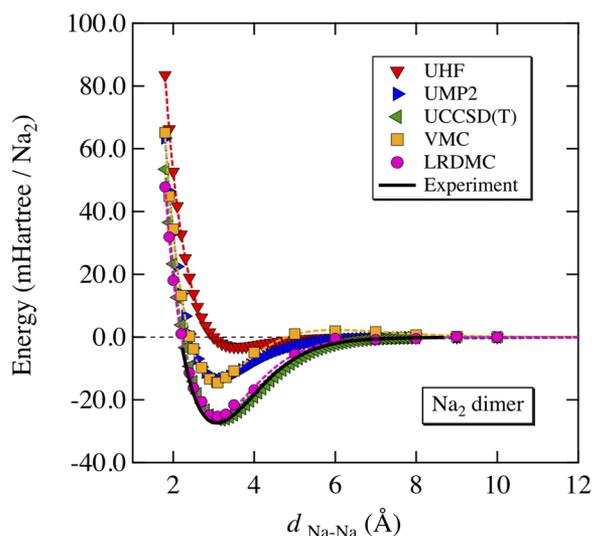


**FIG. 11.** (a) Electronic localization length  $\lambda_N$  divided by the interatomic distance  $a$  as a function of  $a$ , for the one-dimensional hydrogen chain, where  $\lambda_N$  is defined as  $\lambda_N = (L/2\pi)\sqrt{-\ln|z_N|^2/N}$ ,  $N$  is the number of electrons (hydrogen atoms), and  $L$  is the length of the simulation cell along the one dimensional chain. (b) Modulus of the complex polarization  $z_N$  as a function of the interatomic distance of the hydrogen atoms. According to the previous works,<sup>125,126</sup> in the thermodynamic limit ( $N \rightarrow \infty$ ), a metal can be characterized by a vanishing modulus of the complex polarization ( $|z_N| \rightarrow 0$ ,  $\lambda_N \rightarrow +\infty$ ), while it becomes unity in the insulating case ( $|z_N| \rightarrow 1$ ,  $\lambda_N \rightarrow 0$ ). Thus, one can discuss the metal-insulator transition. Reprinted with permission from Stella *et al.*, *Phys. Rev. B* **84**, 245117 (2011). Copyright 2011 APS.

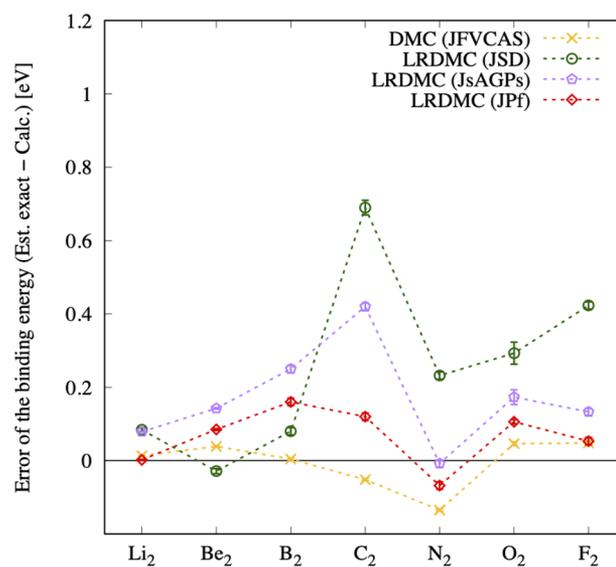
axis ( $S_z$ ) and the spin square ( $S^2$ ) operators inside a sphere centered on each atom,<sup>62,69</sup> whose radius can be specified by the user. (iv) The Berry phase:  $z_\alpha = \langle \Psi | e^{i(2\pi/L^\alpha) \sum_{i=1}^N r_i^\alpha} | \Psi \rangle$ , where the complex polarization is computed for the  $\alpha = \{x, y, z\}$  component  $r_i^\alpha$  of the position vector  $\mathbf{r}_i$  and  $L^\alpha$  is the supercell length in the same direction  $\alpha$ .<sup>123</sup> This property can be used to characterize the metal-insulator transition based on the many-body WF, as has been done in the one-dimensional hydrogen chain by Stella *et al.* (Fig. 11).

## XII. APPLICATIONS

Since the start of the TURBORVB project in 2003 by Casula and Sorella,<sup>65</sup> the code has been applied to study atomic species, molecular systems, and various materials, including challenging systems such as large complexes, surfaces, and liquids. Here, we provide a brief review of the applications done so far.



**FIG. 12.** PES of the sodium dimer. The dashed lines show the obtained Murrell-Sorbie function. The error bars are within the markers. The obtained LRDMC PES agrees with the CCSD(T) and experimental ones. Reprinted with permission from K. Nakano, R. Maezono, and S. Sorella, *J. Chem. Theory Comput.* **15**, 4044 (2019). Copyright 2019 ACS.



**FIG. 13.** Binding energies of the first-row dimers at their experimental bond length (see Table II). The vertical axis shows the difference between the LRDMC and the estimated exact binding energies. The results obtained from the DMC using a Jastrow correlated full-valence-CAS as trial WF [DMC(JFVCAS)] are taken from Ref. 138.

## A. Molecular systems

TURBORVB has been employed to study the properties of several molecular systems, including small diatomic systems (such as  $\text{Li}_2$ ,<sup>127</sup>  $\text{Be}_2$ ,<sup>71,128</sup>  $\text{B}_2$ ,<sup>71</sup>  $\text{C}_2$ ,<sup>62,71</sup>  $\text{N}_2$ ,<sup>71</sup>  $\text{O}_2$ ,<sup>69,71,127,129</sup>  $\text{F}_2$ ,<sup>71</sup>  $\text{Na}_2$ ,<sup>103</sup>  $\text{LiF}$ ,<sup>71</sup>  $\text{CN}$ ,<sup>71</sup> and  $\text{Fe}_2$ <sup>130</sup>), reactive oxygen species (singlet  $\text{O}_2$ ,  $\text{O}_2^-$ ,  $\text{OH}^*$ ,  $\text{OH}^-$ ,  $\text{NO}^*$ ,  $\text{NO}^-$ ,  $\text{HOO}^*$ ,  $\text{HOO}^-$ , and cis and trans  $\text{HOOO}^*$ ),<sup>129</sup> aromatic molecules (benzene<sup>127,131</sup> and oligoacene series,<sup>132</sup> see Sec. XII D), the water molecule,<sup>107,127</sup> dimer<sup>133,134</sup> and hexamer clusters,<sup>134</sup> the Zundel ion ( $\text{H}_5\text{O}_2^+$ ),<sup>114,135</sup>  $\text{H}_2\text{S}$ ,  $\text{SO}_2$ ,  $\text{NH}_3$ ,  $\text{PH}_3$ ,<sup>111</sup> and others. Most of the studies report both variational and FN diffusion Monte Carlo results, and several of the WF *Ansätze* available in TURBORVB are tested. Often, it is shown that quite good results are already obtained

at the variational level, and the FN DMC further improves the accuracy of the outcomes. For instance, see the case of the sodium dimer in Fig. 12.

Table II and Fig. 13 report the binding energies for the molecular dimers of the first-row atoms, evaluated using LRDMC with the JSD, JAGP, or JPF WF *Ansätze* as guiding functions. Some of them are close or reach the chemical accuracy (i.e., 1 kcal/mol or  $\sim 0.04$  eV). A systematic improvement going from the JSD  $\rightarrow$  JAGP  $\rightarrow$  JPF can be appreciated. It follows from the employment of a more flexible parametrization of the WF, which improves both the variational energy and the quality of the nodes. However, the improvement given by the JPF *Ansatz* over JAGP and JSD is even more fundamental than what Table II and Fig. 13 indicate, where the reported

**TABLE II.** Binding energy of the first-row dimers calculated by TURBORVB with LRDMC and different types of *Ansätze* as a guiding function (GF), at the experimental bond length (see Ref. 138). Js denotes the symmetric (i.e., spin-independent) Jastrow factor. LRDMC calculations are performed with several lattice spaces  $a$  and extrapolated for  $a \rightarrow 0$ . The spin-orbit coupling and the zero point energies<sup>71</sup> are not included in these evaluations.

Binding energy (eV)	$\text{Li}_2$	$\text{Be}_2$	$\text{B}_2$	$\text{C}_2$	$\text{N}_2$	$\text{O}_2$	$\text{F}_2$
LRDMC (GF = JSD)	0.976(6)	0.144(7)	2.83(1)	5.74(2)	9.67(1)	4.94(3)	1.27(1)
LRDMC (GF = JsAGPs)	0.9812(12)	-0.0270(12)	2.66(1)	6.01(1) <sup>a</sup>	9.91(1) <sup>a</sup>	5.06(2) <sup>a</sup>	1.56(1)
LRDMC (GF = JPF)	1.0580(12)	0.0304(22)	2.75(1)	6.31(1) <sup>a</sup>	9.97(1) <sup>a</sup>	5.127(6) <sup>a</sup>	1.64(1)
Est. exact	1.06(4) <sup>b</sup>	0.1153(3) <sup>c</sup>	2.91(6) <sup>d</sup>	6.43(2) <sup>e</sup>	9.902(3) <sup>e</sup>	5.233(3) <sup>e</sup>	1.693(5) <sup>e</sup>

<sup>a</sup>See Ref. 69.

<sup>b</sup>See Ref. 139.

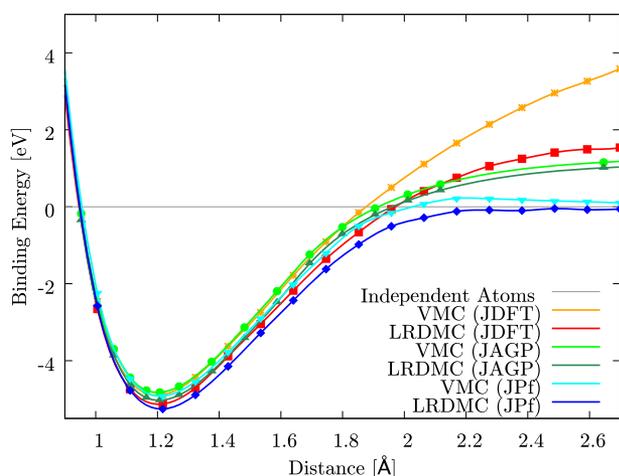
<sup>c</sup>See Ref. 140.

<sup>d</sup>See Ref. 141.

<sup>e</sup>See Ref. 142.

binding energies are evaluated as the difference between the energy of the molecule and twice the energy of the atom. This evaluation does not highlight that JPf is a size-consistent *Ansatz*, while both the JSD and the JAGP have a size-consistency issue with most of the molecules considered.<sup>136</sup> Let us consider the oxygen molecule as an example: the  $O_2$  ground state is a spin triplet ( $^3\Sigma_g^-$ ) which, in both JSD and JAGP, can be described using two unpaired electrons with the same spin. However, as we take the oxygen atoms far apart (say, the O–O system), they should both be in the ground state for the oxygen atom, which is also a spin triplet ( $^3P$ ). Neither JSD nor JAGP has the flexibility to describe the O–O system by the same WF parametrization used for  $O_2$ ;<sup>137</sup> thus,  $E_{O-O} > 2E_O$ . In contrast, JPf has the flexibility to describe the dissociation limit correctly, so  $E_{O-O}^{JPf} = 2E_O^{JPf}$ . Figure 14 shows that JPf is indeed able to provide a reliable dissociation curve of the oxygen molecule, whereas JSD and JAGP are reliable only in proximity of the minimum. Note that the FN projection scheme alleviates in part the limitations of the underlying *Ansatz* as both JSD and JAGP have a smaller size-inconsistency at the LRDMC than at the VMC level, but the correction is incomplete and the inconsistency is sizable ( $>1$  eV) for all but the JPf *Ansatz*. Therefore, in general, JPf will yield a better description of the overall potential energy surface (PES), including binding energies, vibrational properties, zero-point motion, and transition states.

The applications of TURBORVB are not limited to single point evaluations. As discussed in Sec. VI, a distinctive feature of TURBORVB among most QMC codes is the efficient evaluation of atomic forces. This capability has been extensively exploited to investigate several properties of the PES, such as vibrational frequencies, and to perform geometry optimizations and molecular dynamics simulations at finite temperature. It is important to remember that QMC atomic forces are, as the QMC energy, affected by a stochastic error. In a first attempt<sup>143</sup> to evaluate vibrational frequencies of a triatomic molecule (namely, water), a grid of points close to the guessed minimum geometry was used to evaluate energy and forces. The results



**FIG. 14.**  $O_2$  dissociation curve as obtained with TURBORVB at the variational (VMC) and fixed-node projected (LRDMC) levels of theory and using JSD with DFT orbitals (JDFT), JAGP, and JPf WF *Ansätze*.

were used to fit the PES around the minimum with a truncated Taylor expansion, yielding evaluations of the structural minimum and of both the harmonic and anharmonic vibrational frequencies (using the second-order perturbation theory). It was observed that the inclusion of the estimates of forces reduces the stochastic error on the vibrational frequencies by around an order of magnitude with respect to the fit using only the energy evaluations (on the same grid and using the same sampling size on each single point evaluation). This corresponds to about a 25 times larger speedup.<sup>144</sup> Moreover, it was observed that the optimal grid is a trade-off between maximizing the precision on the vibrational properties and minimizing the errors on the fit. A systematic study of the quality of the variational WF *Ansätze* in relation to the vibrational (and other) properties of the water molecule is reported in the work of Zen *et al.*<sup>167</sup> Afterward, Luo *et al.*<sup>111,145</sup> automated the process of finding the optimal grid by performing fits on samples generated by a finite temperature molecular dynamics and applied the approach to  $H_2O$ ,  $H_2S$ ,  $SO_2$ ,  $NH_3$ , and  $PH_3$ . The obtained vibrational frequencies are in good agreement with the experimental values.

Zen *et al.*<sup>134</sup> applied the second-order LD to study the structural properties of liquid water. They obtained radial distribution functions in good agreement with recent neutron scattering and X-ray experiments, especially for the position of the oxygen–oxygen peak, which is very sensitive to the quality of the water description and, in particular, to the inclusion of dispersive vdW forces. This is a stringent test for *ab initio* approaches as the shape of the peak can be directly compared with neutron scattering data.

Mouhat *et al.* developed a path integral LD formalism particularly suited for QMC and applied it to the Zundel ion ( $H_5O_2^+$ ), with the aim of studying nuclear quantum effects.<sup>114</sup> The outcome of their study shows how essential is the inclusion of nuclear quantum effects to properly deal with proton transfer in water and aqueous systems, even at room temperature.

## B. Weakly bonded systems and non-covalent interactions

An important application for QMC is the description of systems interacting through weak van der Waals (vdW) dispersive interactions, which are fundamental in supramolecular chemistry, layered and porous materials, molecular crystals, adsorption of molecules on surfaces, and so on. vdW interactions play a major role also in non-covalent intra-molecular bonds, which result from the interplay of forces of different nature, including electrostatic contributions. An accurate evaluation of the overall strength of these non-covalent bonds requires each component of the interaction to be modeled correctly. This is a big challenge for all *ab initio* approaches, and only a few methods prove reliable. QMC and CCSD(T) are considered to yield very accurate estimates, and there is a general agreement among the predictions from these two theories. Thus, they are typically used to produce benchmark values. Traditionally, mean-field schemes have difficulties in describing dispersive interactions because they arise from non-local electron correlations, completely missed by local density functionals. However, in the context of DFT, there has been an intensive effort in developing new exchange correlation functionals<sup>146</sup> to include vdW interactions. This is usually done in a semi-empirical way, often building on the benchmark provided by CCSD(T) or QMC.

The non-covalent interactions are typically a tiny fraction of the total energy of a compound. Similar to the cases discussed in Sec. XII A, also in the evaluation of weak interactions, it is important to use a size-consistent approach. However, very often, the molecules form a closed shell, and most of the times, the intermolecular interactions are among fragments having zero spin. In this case, a single Slater determinant is a size-consistent *Ansatz*, and so is the JSD. The AGP (without Jastrow) is not size-consistent due to the presence of unphysical charge fluctuations,<sup>43,147</sup> but a very flexible (i.e., large) Jastrow factor allows damping the charge fluctuations and recovering the size-consistency in the JAGP. This is of course true also in a subsequent FN-DMC projection, which is indeed equivalent to applying an infinitely flexible Jastrow factor to the trial WF. At the variational level, the JSD *Ansatz* occasionally leads to better results than the JAGP. Even though JAGP leads to a lower total energy, it is indeed difficult to use a Jastrow factor large enough to damp any charge fluctuation. So, there is a worse error cancellation when energy differences are considered in JAGP. This was observed, for instance, in the water dimer,<sup>134</sup> although at the FN level JSD and JAGP lead to the same binding energy. It should also be considered that in the widely used DMC scheme,<sup>148</sup> a size-consistency issue was present in any simulation with finite time step  $\tau$ , until the modification of the “standard” algorithm has been introduced by Zen *et al.*,<sup>149</sup> whose solution removes almost completely the size-consistency error. At variance with DMC, the LRDMC algorithm does not suffer from this bias, and it is always size consistent.

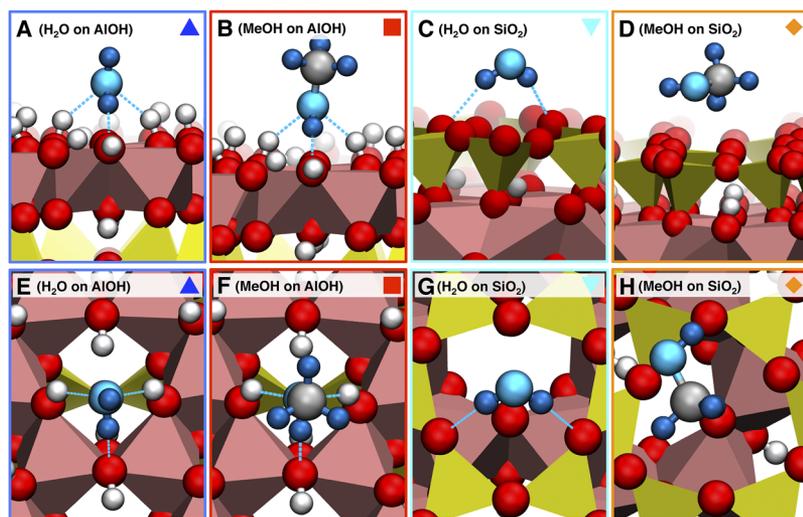
The DMC scheme is commonly used to project a single Slater determinant WF where only the Jastrow factor has been optimized, while the antisymmetric part is filled with DFT orbitals<sup>150</sup> (JDFT). The resulting FN energy is usually biased by the nodal surface, kept frozen from previous DFT or lower level mean-field calculations. Optimizing the Jastrow factor and determinant together, which is routinely done in `TURBORVB`, improves significantly the total energy. However, in the evaluation of non-covalent interactions, there is not yet a clearcut indication of the advantage of the full

optimization over JDFT. This is due to the fact that the error cancellation plays a major role. Indeed, it is expected that, for non-covalent interactions, JDFT has almost the same bias in the interacting and non-interacting systems. When pseudopotentials are used, DLA<sup>54</sup> can be used to remove any bias given by the Jastrow optimization.

`TURBORVB` with JDFT, JSD, and JAGP guiding functions has been applied to several interesting systems, such as molecular hydrogen adsorbed on benzene,<sup>151</sup> benzene dimers,<sup>43</sup> graphite layers,<sup>152</sup> water or methanol molecules adsorbed on a clay surface (see Fig. 15),<sup>153</sup> and metallic clusters,<sup>154</sup> and to evaluate the cohesive energies of  $B_2O_3$  polymorphs.<sup>155</sup>

### C. Strongly correlated and superconducting materials

Another important QMC application is the study of strongly correlated materials, which represents a challenge for any theoretical method, because in several cases a fully consistent description remains, so far, elusive. In this context, QMC methods have been widely used for solving, as accurately as possible, strongly correlated lattice Hamiltonians, such as, the Hubbard model. Real-space *ab initio* QMC methods could fill the gap between realistic Hamiltonians and lattice models, by providing an accuracy similar to the one reached in lattice Hamiltonians while keeping the complexity of real materials. `TURBORVB` has been originally conceived in this perspective, with the idea that one of the most flexible WF *Ansätze* tested on correlated lattice models,<sup>156–160</sup> namely, the generalized version of the resonating valence bond (RVB) WF, could be translated into a successful variational WF for *ab initio* systems. One of the most appealing features of the RVB WF is that it allows several broken-symmetry phases, such as antiferromagnetic insulators and superconductors, which are present—and sometimes coexisting—in the extremely rich phase diagrams of strongly correlated materials. As introduced in Sec. III, the RVB WF employed in `TURBORVB` takes the static and dynamic correlation effects into consideration beyond the commonly used Jastrow–Slater determinant (JSD) WF. Therefore, the

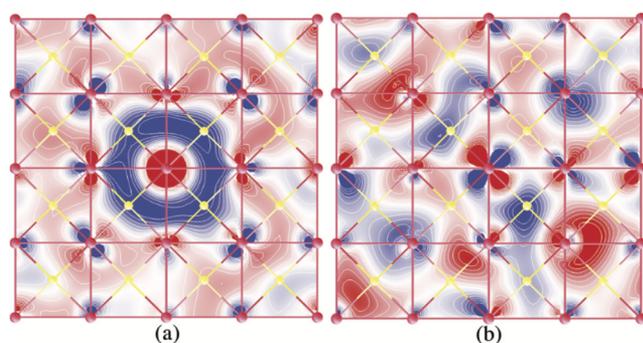


**FIG. 15.** Adsorption of water and methanol on the hydroxyl-terminated and the silicate-terminated faces of kaolinite (side view in the first row and top view in the second row). The adsorbed molecule on kaolinite is depicted in cyan and gray, and the H-bonds are represented by the blue dashed lines. Reprinted with permission from Zen *et al.*, J. Phys. Chem. C **120**, 26402 (2016). Copyright 2016 ACS.

code is expected to describe strongly correlated systems more accurately. In particular, among the various families of correlated WFs that go beyond the simplest JSD form, the JAGP/JPfaffian families are very suitable to describe *extended* systems as they keep a compact and still strongly correlated form even in the thermodynamic limit. Indeed, as detailed in Sec. III, the AGP part is the particle number conserving version of the BCS WF. In the same way, the JAGP variational form can be seen as an efficient formulation of a RVB WF. Thus, superconducting materials including cuprates<sup>161</sup> and iron pnictides and selenides<sup>162</sup> are prominent applications of the JAGP/RVB WFs. By running extensive VMC calculations for  $\text{CaCuO}_2$ , a parent compound of cuprate high-temperature superconductors, Marchi *et al.* validated the JAGP description by correctly finding the expected *d*-wave symmetry of the pairing function.<sup>163</sup> Casula and Sorella applied the RVB WF to FeSe, one of the iron-based high-temperature superconductors,<sup>164</sup> whose electronic structure and pairing mechanism have been unclear and intensively debated. They determined the symmetry of the superconducting order parameter and the size of its gap entirely from first principles, by analyzing the AGP pairing function (Fig. 16). Busemeyer *et al.* applied DMC and LRDMC to study the structural and magnetic properties of the normal state of FeSe under pressure and reported that collinear spin configurations are energetically more favorable than other spin patterns, such as ferromagnetic, checkerboard, and staggered dimer, over a large range of pressures.<sup>165</sup>

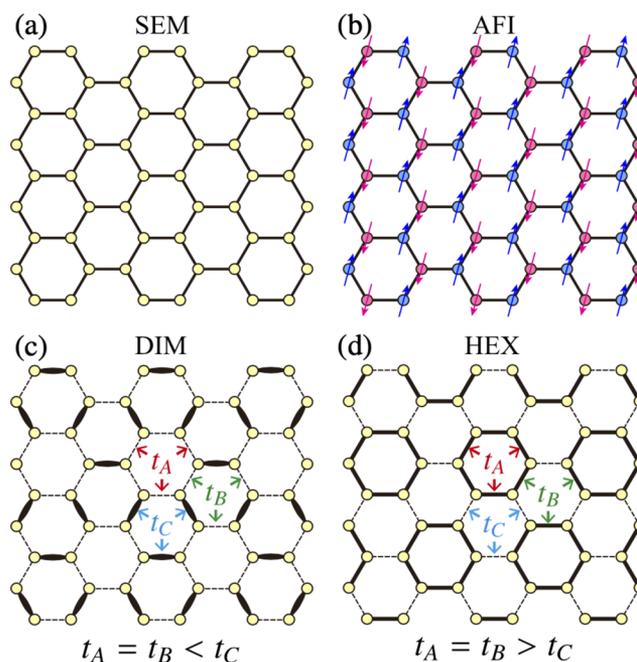
#### D. Aromatic and honeycomb lattice compounds

Aromatic compounds are also successful examples of the RVB theory because they are characterized by resonating C–C bonds that can be efficiently represented by antisymmetrized singlet pairing functions (i.e., antisymmetrized product of geminals or AGP). Casula *et al.* applied the AGP WF combined with a Jastrow factor to the simplest aromatic compound, that is, the benzene molecule ( $\text{C}_6\text{H}_6$ ) for the first time.<sup>127,131</sup> They reported that the inclusion of the resonance between the two possible Kekulé states



**FIG. 16.** (a) Even–even and (b) odd–even components of the AGP pairing function of FeSe at 0 GPa obtained by VMC calculations, where even and odd refer to the parities of the orbitals to the reflection through the Fe plane. The contour plots show  $\Phi(\mathbf{R}_{\text{center}}, \mathbf{r})$  with  $\mathbf{R}_{\text{center}}$  set to be the iron lattice site at the center of the supercell, while  $\mathbf{r}$  spans the plane defined by the  $4 \times 4$  lattice. Red (yellow) balls are iron (selenium) sites. Arbitrary units of blue (red) intensity indicate negative (positive) regions with corresponding magnitude. Reprinted with permission from M. Casula and S. Sorella, Phys. Rev. B **88**, 155125 (2013). Copyright 2013 APS.

significantly lowers the VMC energy compared with the one obtained by a single Kekulé WF. They also showed the importance of adding a three-body Jastrow factor to improve the AGP description, just in the spirit of the RVB framework. By selectively switching on the intersite resonances, they proved that the Dewar contributions (i.e., including the third nearest neighbor carbons) improves the description of the resonating valence bond. Genovese *et al.* have applied a more general AGP WF, i.e., the Pfaffian, to the benzene molecule.<sup>69</sup> They reported that the obtained atomization energies are reasonably consistent with the experimental value at the LRDMC level. On the other hand, when starting from JDFT, i.e., from frozen DFT nodes, the obtained atomization energy is severely underestimated in most cases, implying that the optimization of the nodal surface at the VMC level is essential to obtain a correct value for the atomization energy. As mentioned in Sec. XII B, Sorella *et al.* applied the RVB WF to calculate the binding energy of the face-to-face and displaced parallel benzene dimers<sup>43</sup> and obtained the binding energy very close to the experimental value in the displaced parallel case. Dupuy and Casula applied the same RVB WF to study the ground-state properties of the oligoacene series, from anthracene up to the nonacene.<sup>132</sup> They found that the ground state obtained by the RVB WF has a weak diradical or polyradical instabilities until the nonacene, which is in contrast with the results previously found by lower-level theories, such as Hartree–Fock, hybrid functionals, and CASSCF in a restricted active space. TURBORVB has been applied

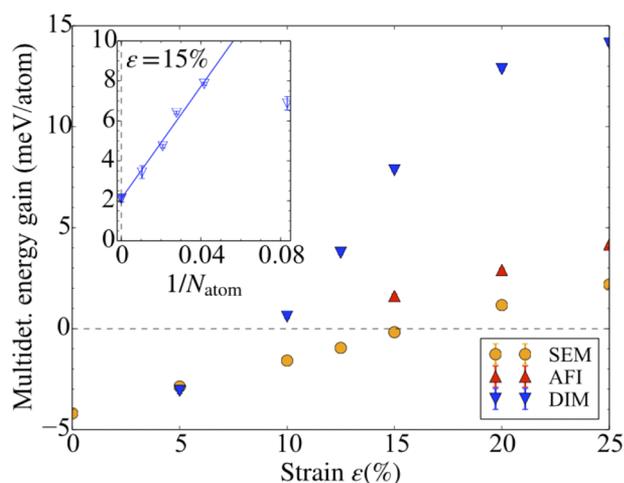


**FIG. 17.** The graphene structures studied in Ref. 166. (a) SEM honeycomb, semimetallic; (b) AFI honeycomb antiferromagnetic insulator; (c) DIM dimerized Kekulé-like insulator; and (d) HEX distorted hexagonal insulator. There are two carbons per unit cell in (a) and (b) and six in (c) and (d).  $t_A$ ,  $t_B$ , and  $t_C$  schematically denote different hopping integrals magnitudes. Reprinted with permission from Sorella *et al.*, Phys. Rev. Lett. **121**, 066402 (2018). Copyright 2018 APS.

not only to simple aromatic molecules but also to honeycomb lattice compounds, in order to study the role of RVB correlations, in other words, to investigate whether the resonance energy is sizable also in extended systems. Marchi *et al.* studied the nature of chemical bonds in graphene using the RVB WF<sup>163</sup> and found that the RVB energy gain becomes extremely small in the thermodynamic limit, i.e., in the infinite lattice. Conversely, Sorella *et al.* applied the RVB WF to isotropically strained graphene<sup>166</sup> (Fig. 17) and found that the RVB effect is crucial in the Kekulé-like dimerized (DIM) phase that has been proposed to become stable under the isotropic strain (Fig. 18). They also reported that the JAGP energy gain with respect to JSD is negligible in the perfect honeycomb structure—as previously found by Marchi *et al.*—but is extremely important in the DIM phase.

### E. Organic compounds: Diradicals and conjugated systems

Diradical and conjugated organic compounds are other interesting applications of the RVB WF because both static and dynamic correlation effects play important roles in these systems. Zen *et al.* found that the RVB WF gives a correct torsion barrier of the ethylene ( $C_2H_4$ ) and the triple-singlet gap in the methylene ( $CH_2$ ).<sup>66</sup> They also investigated the charge-transfer and diradical nature of the electronic states of the retinal minimal model (penta-2,4-dieniminium cation) at the VMC and LRDMC levels.<sup>67</sup> They proved that the dynamical electronic correlation is key to get a reliable ground state energy surface in proximity of the conical intersection between the electronic ground- and first excited-state of the molecule. Their obtained energy landscapes are significantly different from the CASSCF ones, inverting the relative stability of several



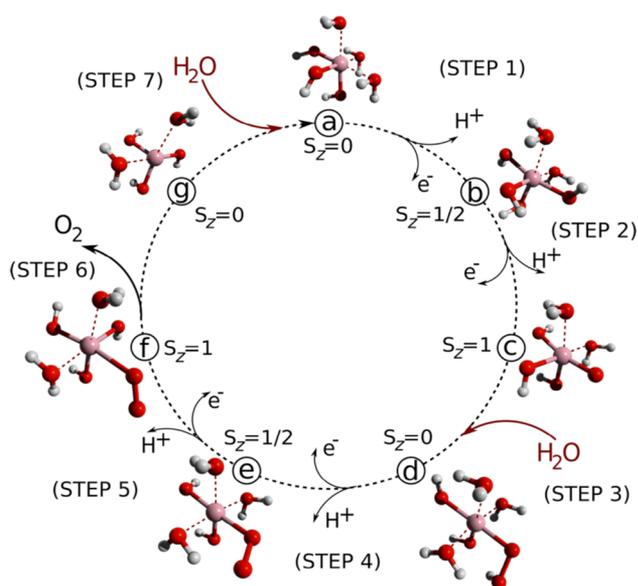
**FIG. 18.** Energy gains by the resonance effect, measured by the energy per atom difference between the single determinant Ansatz (Jastrow-Slater determinant WF) and the corresponding multideterminant JAGP WF. The largest energy gain occurs in the DIM state, underlining its resonating valence-bond nature, actually increasing for large strain  $\epsilon$ . Small negative values at small strain are finite-size effects. Inset: finite-size scaling of this correlation energy gain in the DIM state at  $\epsilon = 15\%$ . Reprinted with permission from Sorella *et al.*, Phys. Rev. Lett. **121**, 066402 (2018). Copyright 2018 APS.

torsional paths, and similar to what was obtained by other correlated approaches. Barborini and Guidoni applied VMC and LRDMC to calculate the torsion barriers of 1,3-butadiene and the ring-opening barrier of cyclobutene,<sup>167</sup> which are in good agreement with the experimental results. Barborini and Coccia investigated the potential energy curves of the two spin states of the tetramethyleneethane (TME) molecule and the two anionic states of TME as a function of the torsion of the central dihedral angle.<sup>168</sup> Through *ab initio* geometrical optimizations at the VMC level, they proposed possible structural interconversions between the states, which are in agreement with the ion photoelectron spectroscopy experiments.

The bond length alternation (BLA), namely, the impact on the geometry of the difference between the single and double carbon bonds, is one of the key structural descriptors in conjugated organic compounds. Barborini *et al.* investigated the effects of the static and dynamical electronic correlations on the BLA of the 1,3-butadiene<sup>169</sup> using a VMC-JAGP Ansatz and other quantum chemistry calculations. They found inconsistency between BLAs obtained by CCSD(T)-CBS (the most accurate) and VMC-JAGP calculations, but the reason for the discrepancy is still unclear. They also investigated the structural properties of polyacetylene chains  $H-(C_2H_2)_n-H$  up to  $N = 12$  acetylene units,<sup>170</sup> in which they revealed that the BLA obtained by the extrapolation to  $n \rightarrow \infty$  is  $0.0910(7)$  Å, which is compatible with the experimental data. Coccia *et al.* applied VMC to study geometries of the retinal model, the chromophore of the rhodopsin involved in the mechanism of vision, and discussed the effects of the electronic correlation on the BLA<sup>171,172</sup> by comparing with DFT and quantum chemistry methods. Wave-function and geometry optimizations of the retinal model were also performed in the presence of the protein field, classically described via electrostatic and mechanical interactions: the complex environment dramatically affects the BLA of the chromophore, which consistently increases with respect to the gas-phase case. They also performed VMC geometry optimization of the peridinin chromophore in order to verify the interplay between the BLA and the optical properties of a large conjugated moiety, involved in the light-harvesting step of photosynthesis in the peridinin-chlorophyll-protein complex.<sup>173</sup> Comparison with DFT and wave-function approaches shows that the combined application of RVB-based VMC and Bethe-Salpeter methods for geometry and excitations, respectively, gives an accurate estimation of the electronic transition energies of peridinin. They recently applied VMC optimization to obtain a ground state structure of keto-1, enol-1, and enol-2 forms of oxyluciferin,<sup>174</sup> which is used to calculate the  $S^1 \leftarrow S^0$  absorption energy based on the Bethe-Salpeter formalism. Coccia *et al.* carried out a systematic basis-set analysis on polarizability, quadrupole moment, and electronic density of the  $C_2H_2$  molecule, with the aim to test VMC and LRDMC calculations in the presence of an external electric field.<sup>175</sup> In particular, a relatively small hybrid (Gaussian + Slater functions) basis set was seen to be able to properly reproduce reference values.

### F. Metal-organic complexes

Recently, TURBORVB has also been used to study metal-organic complexes. Chu *et al.* studied free-energy reaction barriers of water splitting reactions, using a simplified computational model based on the cobalt ion<sup>176</sup> (Fig. 19). They found that the total free-energy differences of the water oxidation, computed at the QMC level,



**FIG. 19.** Reaction cycle involving the cobalt ion model, studied in Ref. 176 by VMC and LRDMC. The pink, red, and white balls represent the cobalt, oxygen, and hydrogen atoms, respectively. Reprinted with permission from Chu *et al.*, *J. Chem. Theory Comput.* **12**, 5803 (2016). Copyright 2016 ACS.

fairly agree with the experimental reference values, in a surprisingly better way than the corresponding CCSD(T) ones. M. Barborini and L. Guidoni applied VMC to tackle the geometry optimization of a  $\text{Fe}_2\text{S}_2(\text{SH})_2^-$  model complex (high-spin and broken symmetry states)<sup>177</sup> based on the extended broken-symmetry (EBS) approach. The number of applications is still limited in this category, but the recent development of sophisticated transition-metal pseudopotentials will make it more feasible in the near future.

### G. Crystal polymorphism

*Ab initio* electronic calculations allow one to study the phase stability between crystal polymorphs by comparing the enthalpies or free energies and by computing their equation of states (EOSs). As mentioned in the Introduction, DFT's predictive power strongly depends on the choice of the exchange and correlation functionals, sometimes causing qualitatively different phase stability predictions. On the other hand, VMC and LRDMC calculations do not suffer from this drawback in principle. Sorella *et al.* studied the pressure-induced metal-insulator transition in bulk silicon through EOS calculations.<sup>178</sup> They obtained a reasonable transition pressure at room temperature by VMC, however still a little far from the experimental value. Devaux *et al.* studied the  $\alpha$ - $\gamma$  phase transition in elemental cerium at the VMC and LRDMC levels.<sup>179</sup> They revealed that the volume-collapse transition could be understood as a conventional first-order transition of electronic origin, induced by the  $p$ - $f$  hybridization and by the strong local repulsion affecting the  $f$  states. Hay *et al.* investigated the effect of long-range van der Waals forces in the  $\text{SiO}_2$  polymorphs using DFT and QMC.<sup>180</sup> Their LRDMC calculations showed that the obtained

energy differences in quartz-cristobalite and quartz-stishovite agree well with the experimental values within sub-chemical accuracy, implying that QMC is a promising method in describing both conventional and high-pressure  $\text{SiO}_2$  polymorphs. Note that, at present, only the zero-temperature energy (without the entropic term) is available at the VMC and LRDMC levels. Therefore, the entropic effects are estimated by DFT-PBE calculations.<sup>178</sup> In the near future, it will be possible to estimate the entropic corrections directly within the VMC framework, via phonon calculations under the quasi-harmonic approximation. As described in Sec. IV, considering twist boundary conditions is essential to correct the one-body finite-size errors for periodic systems, especially in a metallic case. Although the twisted average approach is commonly used to correct the error, one can also use the exact special twist (EST) method<sup>87</sup> developed by Dagrada *et al.* to save some computational cost. EST works even in realistic systems, namely, the bcc-hydrogen, the bcc-lithium, and the silicon in the high-pressure  $\beta$ -tin phase.<sup>87</sup>

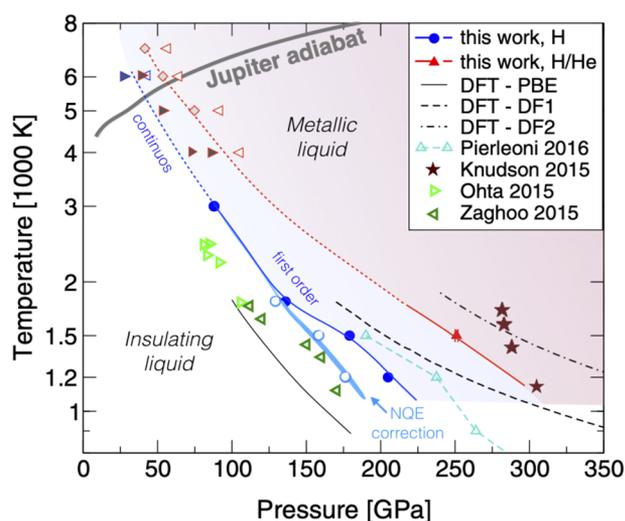
### H. One-dimensional chains

One-dimensional chains of atoms have been intensively studied based on both effective and *ab initio* Hamiltonians because they embody many important open questions in modern condensed matter physics, despite their simplicity. The finite or infinite hydrogen chains are ideal systems for this purpose because it is not necessary to use a pseudopotential or to consider the relativistic effects. They are therefore benchmark systems. Stella *et al.* investigated the metal-insulator transition in the hydrogen chain by explicitly calculating the complex polarization,<sup>125</sup> as shown in Sec. XI. They found that the model has no metal-insulator transition and always behaves as an insulating 1D Hubbard model at half-filling, at least for a large enough interatomic distance ( $R > 1$ ). Motta *et al.* (Simons Collaboration on the Many-Electron Problem)<sup>181</sup> undertook a comprehensive benchmark study in the finite and infinite linear hydrogen chains using state-of-the-art many-body methods, including VMC and LRDMC with extrapolations to the thermodynamic and the complete-basis-set limits. They provided accurate potential energy curves online, which will be very useful for benchmarking other methods. They have recently investigated the ground states of the hydrogen-chain more in detail, namely, the insulator-to-metal transition, dimerization, and magnetic phases.<sup>182</sup> They revealed a fascinating phase diagram, with several emergent quantum phases, depending on the interproton distance.

### I. Liquid and solid hydrogen/helium

Hydrogen behavior at high pressures is still not fully understood and is subject of intense research. In 1935, Wigner and Huntington predicted its metallization upon compression,<sup>183</sup> while Ashcroft also proposed its high- $T_c$  superconductivity.<sup>184</sup> Besides its fundamental interest as the simplest realistic condensed matter system in Nature, calculating its equilibrium properties remains of fundamental importance for planetary science applications.<sup>185</sup> This system is particularly challenging for DFT-based simulations due to the presence of strong correlation effects and a subtle interplay between structural and electronic phase transitions.

A straightforward way to study a phase diagram under pressure and at finite temperature in QMC is to apply the Langevin molecular dynamics described in Sec. VIII. First-order phase transitions can be identified from a discontinuous behavior in the calculated equation of state and of pair correlation functions. Attaccalite and Sorella have formulated a second-order LD suitable for QMC simulation for the first time, as described in Sec. VIII. They applied it to assess the relative stability of solid and liquid hydrogen at high pressure and room temperature.<sup>106</sup> Years later, in a series of works, Mazzola *et al.* applied this technique to study the proposed liquid–liquid transition (LLT) from a molecular insulating to an atomic conducting phase of hydrogen.<sup>113,186</sup> Mazzola and Sorella have also formulated the accelerated first-order LD,<sup>119</sup> as described in Sec. VIII, to study such an elusive phase transition with properly equilibrated simulations.<sup>119</sup> Finally, under the same framework, the first QMC molecular dynamics simulations of a hydrogen–helium mixture has been performed at Jupiter’s interior conditions.<sup>120</sup> The study revealed that mixing He has a significant influence on the H metallization pressure, as shown in Fig. 20, and provided useful benchmark for the equation of states currently adopted by the planetary science community. Dagrada *et al.* revealed that their developed EST technique also works well in the LD simulation for liquid hydrogen, but a perfect agreement with the twisted-average technique is achieved only when relatively large supercells are used.<sup>87</sup> Therefore, a careful finite size scaling analysis is needed when the EST

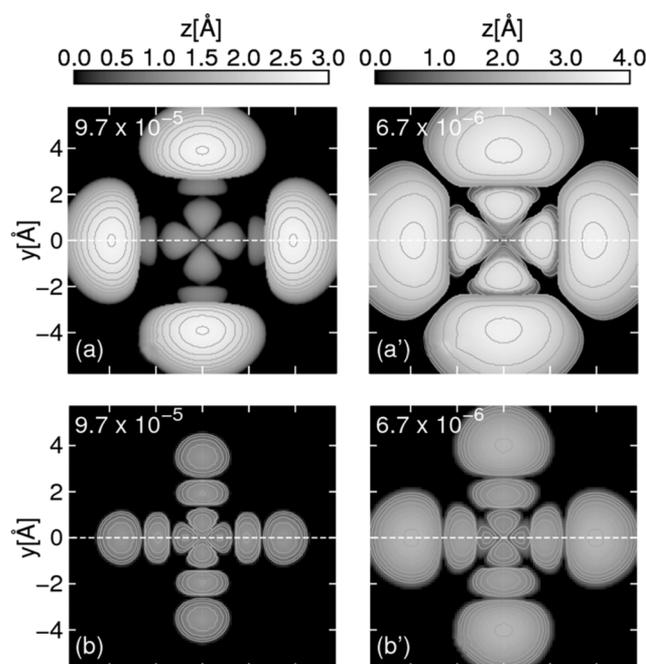


**FIG. 20.** Phase diagram of dense hydrogen and a hydrogen–helium mixture. “This work” refers to the results obtained by VMC-Langevin dynamics simulations. The shaded areas represent the liquid–liquid transitions (LLTs) between the insulating-molecular and the metallic-atomic fluids. The solid symbols refer to their QMC LLTs for the pure hydrogen (blue circles) and for the hydrogen–helium mixture (red triangles), and the blue and red solid lines indicate first order LLTs. The empty (solid) left (right) triangles correspond to simulations displaying a clear atomic (molecular) behavior; on the other hand, the red diamonds represent an intermediate behavior. A calculated Jupiter’s adiabat is shown as a gray line. Pure hydrogen first-order LLT predicted by QMC (cyan) and DFT (solid and dashed lines) are also shown. The light green and dark green triangles and the brown stars represent experimental results. Reprinted with permission from G. Mazzola, R. Helled, and S. Sorella, *Phys. Rev. Lett.* **120**, 025701 (2018). Copyright 2018 APS.

technique is employed, although the thermodynamic limit extrapolation is much smoother at the special twist, saving some computational cost. Moreover, EST allows for the determinant/AGP optimization.

## J. Excited states

TURBORVB allows studying not only the ground-state electronic properties but also excited states. It is trivial to obtain excitation properties when they are characterized by different particles or different spin projection because these quantities are clearly conserved. For example, Barborini *et al.* calculated the  ${}^3B_{1u} \leftarrow {}^1A_g$  vertical triplet excitation of the ethylene molecule and its adiabatic excitation  ${}^3A_1 \leftarrow {}^1A_g$ .<sup>187</sup> One can also deal with nontrivial cases based on a flexible symmetry-adapted RVB WF within its MO representation, which was proposed by Dupuy *et al.* in 2015.<sup>188</sup> Instead of dealing with a linear combination of symmetry-adapted configuration state functions, they construct a symmetry-adapted RVB WF that accurately describes the targeted symmetry. They reported that their developed constrained (fixed-rank) minimization of the geminal expansion is stable without symmetry contamination of the starting excited state symmetry; as a result, the vertical ionization energies and the electron affinities of anthracene obtained by VMC and LRDMC calculations agree with the experimental values.



**FIG. 21.** The STM images of in-plane  $[\text{CuCl}_4]^{2-}$  simulated using UHF and QMC calculations. The left column shows the isosurfaces of the square modulus of the eQPWF for a fixed value of probability density of  $9.7 \times 10^{-5} \text{ \AA}^{-3}$ , while in the right column the isosurfaces take the value of  $6.7 \times 10^{-6} \text{ \AA}^{-3}$ . Panels (a) and (a') display the LUMO+1 UHF molecular orbitals, while panels (b) and (b') display the QMC images. Reprinted with permission from Barborini *et al.*, *J. Chem. Theory Comput.* **12**, 5339 (2016). Copyright 2016 ACS.

### K. Connection to experimental spectroscopies

Barborini *et al.* proposed a new *ab initio* approach to calculate quasiparticle wave functions (QPWFs) of isolated systems and resolved in momentum space, based on the correlated sampling technique. They applied the new approach to simulate Scanning Tunneling Microscopy (STM) images<sup>189</sup> and Angle-Resolved Photoemission Spectroscopy (ARPES).<sup>190</sup> The STM image of  $[\text{CuCl}_4]^{2-}$  obtained by this approach is shown in Fig. 21. It reveals that the electronic correlation has a significant effect on the QPWF in this system. Indeed, the resulting STM image clearly differs from that obtained by uncorrelated methods (i.e., the Hartree–Fock counterpart). In general, the electronic calculation should be useful to interpret the experimental results theoretically. We believe therefore that TURBORVB can become useful also for state-of-the-art research in this field.

### XIII. PYTHON MODULE: TURBO-GENIUS

In general, a computational study involves many complicated operations, such as preparing input files, searching for optimal parameters, performing benchmark studies, transferring information coming from previous calculations to a subsequent simulation's input file, and analyzing output files. Automating these operations can significantly save researcher's burden, avoid trivial human errors, and is beneficial for accelerating the distribution of a computational package to a much wider community. We are now developing a Python-based workflow system suitable for TURBORVB named TURBO-GENIUS, such as Nexus<sup>191</sup> and QMC-SW.<sup>192</sup> The workflow system has been implemented in Python 3 in an object-oriented fashion, so the modules are highly extensible. Note that, since TURBORVB is an all-in-one package, any other commercial code is not necessary to run the code. There are two main modules, turbo-genius-serial.py and turbo-genius-sequential.py, which manage other modules/classes. The former allows one to perform a single job specified by an option (e.g., `turbo-genius-serial.py -j makefort10`), and the latter does a sequential job (e.g., `makefort10.x → convertfort10mol.x → prep.x → turborvb.x`). By using TURBO-GENIUS, one can also analyze a simulation output. For example, one can plot an optimization history, average the variational parameters after optimization, and search for an optimal number of equilibration steps and a reblocking length.<sup>40</sup> TURBO-GENIUS is still at an early stage of development. However, its main features are already there. In the near future, the TURBO-GENIUS modules will be capable of sophisticated TURBORVB job automatizations, as well as setting optimal parameters depending on target systems.

### XIV. CONCLUSIONS

In this paper, we have reviewed TURBORVB, an *ab initio* quantum Monte Carlo package featuring two well established QMC algorithms for electronic structure: variational Monte Carlo and diffusion Monte Carlo. The beginning of the TURBORVB development dates back to 2003, with the Ph.D. thesis project of Casula supervised by Sorella. Since then, TURBORVB has rapidly grown up, and it has been applied to the study of several molecular and condensed matter systems, from benzene to high-temperature superconducting materials. Central to this package are variational parameterizations of

correlated electronic wave functions in the product form  $\Psi = \Phi_{\text{AS}} * \exp J$ , made of an antisymmetric part ( $\Phi_{\text{AS}}$ ) and a Jastrow factor ( $J$ ).

Concerning the antisymmetric part, TURBORVB implements five *Ansätze* (in order of decreasing variational flexibility): (i) the Pfaffian (Pf), (ii) the Pfaffian with a constrained number of molecular orbitals (Pfn), (iii) the Antisymmetrized Geminal Power (AGP), (iv) the Antisymmetrized Geminal Power with a constrained number of molecular orbitals (AGPn), and (v) the single Slater determinant (SD). Notably, the user can freely navigate between these *Ansätze*. Indeed, the package includes conversion modules, which allow the user to choose the most suitable wave function form for the target system.

All the above *Ansätze* can be optimized by state-of-the-art algorithms, namely, the stochastic reconfiguration and the linear method, thus achieving highly accurate variational and fixed-node energies, with a computational cost remaining at the single-determinant level, thanks to efficient algorithmic developments.

The stochastic optimization of many variational parameters (so far up to the order of  $10^5$ ) is feasible in TURBORVB, thanks to an efficient evaluation of energy derivatives using the AAD. This algorithmic scheme, which gives the code an original architecture, drastically decreases the computational cost of ionic force calculations, keeping them of order  $N^3$ . Thus, it paves the way for efficient structural optimizations and molecular dynamics simulations at the VMC level of theory. Indeed, by means of the TURBORVB package, it is at present possible to perform QMC molecular dynamics simulations of dense hydrogen (with up to 256 ions) and liquid water (with up to 64 water molecules), with a remarkably large number of atoms, not far from the DFT state of the art. Thanks to TURBORVB, nuclear quantum effects are also accessible in a correlated electronic environment provided by the VMC wave function.

Large scale calculations are possible with TURBORVB not only because of cutting-edge algorithmic developments but also by virtue of an efficient parallelization based on a hybrid MPI and OpenMP protocol, which is ideal for employing GPU accelerators as well, recently available in the most advanced HPC infrastructures.

A Python wrapper, named TURBO-GENIUS, has recently been developed to make the code user-friendly and to allow both beginners and professional users to handle it more efficiently. We will continue developing TURBORVB and TURBO-GENIUS to implement new QMC algorithms and to extend their range of applications for the current developers and expected new upcoming contributors and users.

### ACKNOWLEDGMENTS

The authors acknowledge all contributors for the development of the code TURBORVB over the last 20 years. We would especially like to acknowledge very fruitful collaborations with F. Becca, G. Carleo, C. Cavazzoni, N. Devaux, N. Dupuy, L. Guidoni, H. Hay, R. Hlubina, Y. Iqbal, R. Maezono, F. Mouhat, A. Parola, K. Seki, T. Shirakawa, L. Spanu, L. Stella, L. F. Tocchio, and S. Yunoki.

K.N. is grateful for computational resources from PRACE, Project No. 2019204934, and those from the facilities of the Research Center for Advanced Computing Infrastructure at Japan Advanced Institute of Science and Technology (JAIST). K.N. also acknowledges financial support from the Simons Foundation and that from Grant-in-Aid for Scientific Research on Innovative Areas

(No. 16H06439). C.A. acknowledges funding from the European Union Seventh Framework Program under Grant Agreement No. 785219 Graphene Core 2 and COST Action TUMIEE CA17126. M.C. is grateful to the French grand équipement national de calcul intensif (GENCI) for the computational time provided through these years under Project No. 0906493; for the funded PRACE Project Nos. 2012061116, 2015133179, and 2016163936; and for the access to the Hokusai and K-computer granted by the Institute of Physical and Chemical Research (RIKEN). Y.L. was supported by the Argonne Leadership Computing Facility, which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-06CH11357. A.Z.'s work was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, U.S. Air Force, under Grant No. FA9550-19-1-7007. S.S. acknowledges financial support from PRIN 2017BZPKSZ and computational resources from PRACE, Project No. 2019204934. S.S. is also grateful to his wife L. Urgias for bearing the too many weekends spent on the development of TURBORVB.

We dedicate this paper to the memory of P. W. Anderson, remembering him as one of the most influential contributors to physics and chemistry of the past century, and in particular for deeply inspiring this work with his RVB theory of matter.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. <http://energy.gov/downloads/doe-public-access-plan>.

## DATA AVAILABILITY

The data that support the findings of this study are available within this article and the cited references. TURBORVB is available from S.S.'s web site (<https://people.sissa.it/~sorella/>) upon request.

## REFERENCES

- R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, *npj Comput. Mater.* **3**, 54 (2017).
- K. Rajan, *Mater. Today* **8**, 38 (2005).
- W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2004).
- M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, *Science* **355**, 49 (2017).
- J. P. Perdew and K. Schmidt, *AIP Conf. Proc.* **577**, 1–20 (2001).
- K. Burke, *J. Chem. Phys.* **136**, 150901 (2012).
- A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Chem. Rev.* **112**, 289 (2012).
- A. Szabo and N. Ostlund, *Modern Quantum Chemistry* (Macmillan Publishing Co., Inc., 1982).
- C. Møller and M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).
- P. J. Knowles and N. C. Handy, *Chem. Phys. Lett.* **111**, 315 (1984).
- B. O. Roos, P. R. Taylor, and P. E. M. Sigbahn, *Chem. Phys.* **48**, 157 (1980).
- J. Čížek, *J. Chem. Phys.* **45**, 4256 (1966).
- W. Foulkes, L. Mitás, R. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017).
- N. Metropolis and S. Ulam, *J. Am. Stat. Assoc.* **44**, 335 (1949).
- J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, M. Casula, D. M. Ceperley, S. Chiesa, B. K. Clark, R. C. Clay, K. T. Delaney, M. Dewing, K. P. Esler, H. Hao, O. Heinonen, P. R. C. Kent, J. T. Krogel, I. Kylänpää, Y. W. Li, M. G. Lopez, Y. Luo, F. D. Malone, R. M. Martin, A. Mathuriya, J. McMinis, C. A. Melton, L. Mitás, M. A. Morales, E. Neuscamman, W. D. Parker, S. D. P. Flores, N. A. Romero, B. M. Rubenstein, J. A. R. Shea, H. Shin, L. Shulenburger, A. F. Tillack, J. P. Townsend, N. M. Tubman, B. Van Der Goetz, J. E. Vincent, D. C. Yang, Y. Yang, S. Zhang, and L. Zhao, *J. Phys.: Condens. Matter* **30**, 195901 (2018).
- R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos, *J. Phys.: Condens. Matter* **22**, 023201 (2009).
- L. K. Wagner, M. Bajdich, and L. Mitás, *J. Comput. Phys.* **228**, 3390 (2009).
- C. J. Umrigar and C. Filippi, Cornell-Holland Ab-initio Materials Package (CHAMP), <http://pages.physics.cornell.edu/~cyrus/champ.html>; accessed 27 October 2019.
- J. S. Spencer, N. S. Blunt, S. Choi, J. Etrych, M.-A. Filip, W. M. C. Foulkes, R. S. Franklin, W. J. Handley, F. D. Malone, V. A. Neufeld *et al.*, *J. Chem. Theory Comput.* **15**, 1728 (2019).
- S. Zhang, J. Carlson, and J. E. Gubernatis, *Phys. Rev. B* **55**, 7464 (1997).
- M. Motta and S. Zhang, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1364 (2018).
- G. H. Booth, A. J. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013).
- A. J. W. Thom, *Phys. Rev. Lett.* **105**, 263004 (2010).
- J. S. Spencer and A. J. Thom, *J. Chem. Phys.* **144**, 084108 (2016).
- N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes, *Phys. Rev. B* **89**, 245124 (2014).
- F. D. Malone, N. Blunt, J. J. Shepherd, D. Lee, J. Spencer, and W. Foulkes, *J. Chem. Phys.* **143**, 044116 (2015).
- S. Ten-no, *J. Chem. Phys.* **138**, 164126 (2013).
- Y. Ohtsuka and S. Ten-no, *J. Chem. Phys.* **143**, 214107 (2015).
- S. Ten-no, *J. Chem. Phys.* **147**, 244107 (2017).
- J. R. McClean and A. Aspuru-Guzik, *Phys. Rev. A* **91**, 012311 (2015).
- A. Nagy and V. Savona, *Phys. Rev. A* **97**, 052129 (2018).
- L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, 1960).
- P. W. Anderson, *Mat. Res. Bull.* **8**, 153 (1973).
- K. Nakano, R. Maezono, and S. Sorella, *Phys. Rev. B* **101**, 155106 (2020).
- C. Umrigar, *Phys. Rev. Lett.* **71**, 408 (1993).
- M. Stedman, W. Foulkes, and M. Nekovee, *J. Chem. Phys.* **109**, 2630 (1998).
- H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- S. Sorella and L. Capriotti, *J. Chem. Phys.* **133**, 234111 (2010).
- S. Sorella, *Phys. Rev. Lett.* **80**, 4558 (1998).
- S. Sorella, M. Casula, and D. Rocca, *J. Chem. Phys.* **127**, 014105 (2007).
- S. Sorella, *Phys. Rev. B* **71**, 241103 (2005).
- C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, *Phys. Rev. Lett.* **98**, 110201 (2007).
- J. Toulouse and C. J. Umrigar, *J. Chem. Phys.* **126**, 084102 (2007).
- M. Casula, C. Filippi, and S. Sorella, *Phys. Rev. Lett.* **95**, 100201 (2005).
- D. F. Ten Haaf, H. J. Van Bommel, J. M. Van Leeuwen, W. Van Saarloos, and D. M. Ceperley, *Phys. Rev. B* **51**, 13039 (1995).
- M. Calandra Buonauro and S. Sorella, *Phys. Rev. B* **57**, 11446 (1998).
- S. Sorella and L. Capriotti, *Phys. Rev. B* **61**, 2599 (2000).
- M. Casula, "New QMC approaches for the simulation of electronic systems: A first application to aromatic molecules and transition metal compounds," Ph.D. thesis, International School for Advanced Studies (SISSA), 2005.
- M. Casula, S. Moroni, S. Sorella, and C. Filippi, *J. Chem. Phys.* **132**, 154113 (2010).
- M. Casula, *Phys. Rev. B* **74**, 161102 (2006).

- <sup>54</sup>A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, *J. Chem. Phys.* **151**, 134105 (2019).
- <sup>55</sup>The correlation energy is typically defined as the difference between the exact energy and the Hartree–Fock energy, which is the variational minimum for a SD Ansatz.
- <sup>56</sup>However, the Jastrow factor makes  $\Psi$  not factorizable when expectation values of quantum operators are evaluated. For this reason, it is not a feasible route to traditional quantum chemistry approaches as it requires stochastic approaches to evaluate efficiently the corresponding multidimensional integrals.
- <sup>57</sup>P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- <sup>58</sup>Indeed, with backflow each time an electron is moved, all the entries in  $\Phi_{AS}$  (or several of them, if cut-offs are used) can be changed, resulting in a much more expensive algorithm.
- <sup>59</sup>T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>60</sup>M. Bajdich, L. Mitas, G. Drobný, L. K. Wagner, and K. E. Schmidt, *Phys. Rev. Lett.* **96**, 130201 (2006).
- <sup>61</sup>M. Bajdich, L. Mitas, L. K. Wagner, and K. E. Schmidt, *Phys. Rev. B* **77**, 115112 (2008).
- <sup>62</sup>C. Genovese, T. Shirakawa, and S. Sorella, [arXiv:1911.09748](https://arxiv.org/abs/1911.09748) (2019).
- <sup>63</sup>P. Kasteleyn, *J. Math. Phys.* **4**, 287 (1963).
- <sup>64</sup>Some properties of the Pfaffian operator and their proofs are given in the work of Bajdich *et al.*<sup>60</sup>
- <sup>65</sup>M. Casula and S. Sorella, *J. Chem. Phys.* **119**, 6500 (2003).
- <sup>66</sup>A. Zen, E. Coccia, Y. Luo, S. Sorella, and L. Guidoni, *J. Chem. Theory Comput.* **10**, 1048 (2014).
- <sup>67</sup>A. Zen, E. Coccia, S. Gozem, M. Olivucci, and L. Guidoni, *J. Chem. Theory Comput.* **11**, 992 (2015).
- <sup>68</sup> $S^{-1/2}$  can be computed after a standard block diagonalization of the matrix  $S = UDU^\dagger$ ,  $U$  being a unitary matrix and  $D = \text{diag}\{d_1, \dots, d_{2L}\}$  being a diagonal matrix, such that  $S^{-1/2} = UD^{-1/2}U^\dagger$ , where  $D^{-1/2}$  is the diagonal matrix obtained by taking the inverse square root of each diagonal element  $d_i$  of  $D$ . At this stage, we carefully remove from the basis the elements corresponding to the smallest eigenvalues  $d_i$  in order to work with a sufficiently large condition number that guarantees stable finite precision numerical calculations.
- <sup>69</sup>C. Genovese, T. Shirakawa, K. Nakano, and S. Sorella, [arXiv:2002.03347](https://arxiv.org/abs/2002.03347) (2020).
- <sup>70</sup>The nonzero eigenvalues of  $\tilde{A}$  are  $\pm i\lambda_k$ .
- <sup>71</sup>M. Marchi, S. Azadi, M. Casula, and S. Sorella, *J. Chem. Phys.* **131**, 154116 (2009).
- <sup>72</sup>The output of AAD are matrices  $D_{\mu,\nu} = \frac{\partial F}{\partial A_{\mu,\nu}}$ , where  $F$  is either the log of the WF or the corresponding local energy computed on a given configuration. Then, the projected derivatives corresponding to  $\delta A_{\mu,\nu}^n$  easily follows from Eq. (67), by applying the chain rule.
- <sup>73</sup>It shall be noticed that there is not necessarily one leading term in the expansion in Eq. (71). For instance, when  $|\tilde{\lambda}_p| = |\tilde{\lambda}_{p+1}|$ , the term with  $(i_1, \dots, i_p) = (1, \dots, p)$  has equal weight to the term  $(i_1, \dots, i_{p-1}, i_p) = (1, \dots, p-1, p+1)$ .
- <sup>74</sup>The seniority number is the number of unpaired electrons in a determinant.
- <sup>75</sup>B. P. Pritchard, D. Altarawy, B. T. Didier, T. D. Gibson, and T. L. Windus, *J. Chem. Inf. Model.* **59**, 4814 (2019).
- <sup>76</sup>J. R. Trail and R. J. Needs, *J. Chem. Phys.* **142**, 064110 (2015).
- <sup>77</sup>J. T. Krogel, J. A. Santana, and F. A. Reboredo, *Phys. Rev. B* **93**, 75143 (2016).
- <sup>78</sup>J. R. Trail and R. J. Needs, *J. Chem. Phys.* **146**, 204107 (2017).
- <sup>79</sup>M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, *J. Chem. Phys.* **147**, 224106 (2017).
- <sup>80</sup>M. C. Bennett, G. Wang, A. Annaberdiyev, C. A. Melton, L. Shulenburger, and L. Mitas, *J. Chem. Phys.* **149**, 104108 (2018).
- <sup>81</sup>A. Annaberdiyev, G. Wang, C. A. Melton, M. Chandler Bennett, L. Shulenburger, and L. Mitas, *J. Chem. Phys.* **149**, 134108 (2018).
- <sup>82</sup>M. Burkatzki, C. Filippi, and M. Dolg, *J. Chem. Phys.* **126**, 234105 (2007).
- <sup>83</sup>M. Burkatzki, C. Filippi, and M. Dolg, *J. Chem. Phys.* **129**, 164115 (2008).
- <sup>84</sup>S. Fahy, X. W. Wang, and S. G. Louie, *Phys. Rev. B* **42**, 3503 (1990).
- <sup>85</sup>S. Sorella, N. Devaux, M. Dagrada, G. Mazzola, and M. Casula, *J. Chem. Phys.* **143**, 244112 (2015).
- <sup>86</sup>C. Lin, F. H. Zong, and D. M. Ceperley, *Phys. Rev. E* **64**, 016702 (2001).
- <sup>87</sup>M. Dagrada, S. Karakuzu, V. L. Vildosola, M. Casula, and S. Sorella, *Phys. Rev. B* **94**, 245108 (2016).
- <sup>88</sup>G. Rajagopal, R. J. Needs, S. Kenny, W. M. C. Foulkes, and A. James, *Phys. Rev. Lett.* **73**, 1959 (1994).
- <sup>89</sup>G. Rajagopal, R. Needs, A. James, S. Kenny, and W. Foulkes, *Phys. Rev. B* **51**, 10591 (1995).
- <sup>90</sup>C. Pisani, R. Dovesi, and C. Roetti, *Hartree-Fock Ab Initio Treatment of Crystalline Systems* (Springer Berlin, 1988), Vol. 48.
- <sup>91</sup>R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro *et al.*, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1360 (2018).
- <sup>92</sup>S. Karakuzu, L. F. Tocchio, S. Sorella, and F. Becca, *Phys. Rev. B* **96**, 205145 (2017).
- <sup>93</sup>L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, *Phys. Rev. B* **53**, 1814 (1996).
- <sup>94</sup>A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, and M.-Y. Chou, *Phys. Rev. B* **55**, R4851 (1997).
- <sup>95</sup>P. R. C. Kent, R. Q. Hood, A. J. Williamson, R. J. Needs, W. M. C. Foulkes, and G. Rajagopal, *Phys. Rev. B* **59**, 1917 (1999).
- <sup>96</sup>H. Kwee, S. Zhang, and H. Krakauer, *Phys. Rev. Lett.* **100**, 126404 (2008).
- <sup>97</sup>S. Chiesa, D. M. Ceperley, R. M. Martin, and M. Holzmann, *Phys. Rev. Lett.* **97**, 076404 (2006).
- <sup>98</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 16 Revision C.01, Gaussian, Inc., Wallingford, CT, 2016.
- <sup>99</sup>P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- <sup>100</sup>J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- <sup>101</sup>S. Azadi, C. Cavazzoni, and S. Sorella, *Phys. Rev. B* **82**, 125112 (2010).
- <sup>102</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>103</sup>K. Nakano, R. Maezono, and S. Sorella, *J. Chem. Theory Comput.* **15**, 4044 (2019).
- <sup>104</sup>R. Assaraf, S. Moroni, and C. Filippi, *J. Chem. Theory Comput.* **13**, 5273 (2017).
- <sup>105</sup>C. Umrigar, *Int. J. Quantum Chem.* **36**, 217 (2009).
- <sup>106</sup>C. Attaccalite and S. Sorella, *Phys. Rev. Lett.* **100**, 114501 (2008).
- <sup>107</sup>A. Zen, Y. Luo, S. Sorella, and L. Guidoni, *J. Chem. Theory Comput.* **9**, 4332 (2013).
- <sup>108</sup>G. Mazzola, A. Zen, and S. Sorella, *J. Chem. Phys.* **137**, 134112 (2012).
- <sup>109</sup>S.-I. Amari, *Neural Comput.* **10**, 251 (1998).
- <sup>110</sup>E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, *Phys. Rev. B* **85**, 045103 (2012).
- <sup>111</sup>Y. Luo, A. Zen, and S. Sorella, *J. Chem. Phys.* **141**, 194112 (2014).
- <sup>112</sup>G. E. P. Box and M. E. Muller, *Ann. Math. Stat.* **29**, 610 (1958).
- <sup>113</sup>G. Mazzola, S. Yunoki, and S. Sorella, *Nat. Commun.* **5**, 3487 (2014).
- <sup>114</sup>F. Mouhat, S. Sorella, R. Vuilleumier, A. M. Saitta, and M. Casula, *J. Chem. Theory Comput.* **13**, 2400 (2017).
- <sup>115</sup>H. F. Trotter, *Proc. Amer. Math. Soc.* **10**, 545 (1959).
- <sup>116</sup>M. Tuckerman, B. J. Berne, and G. J. Martyna, *J. Chem. Phys.* **97**, 1990 (1992).

- <sup>117</sup>J. C. Sexton and D. H. Weingarten, *Nucl. Phys. B* **380**, 667 (1992).
- <sup>118</sup>M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, *J. Chem. Phys.* **133**, 124104 (2010).
- <sup>119</sup>G. Mazzola and S. Sorella, *Phys. Rev. Lett.* **118**, 015703 (2017).
- <sup>120</sup>G. Mazzola, R. Helled, and S. Sorella, *Phys. Rev. Lett.* **120**, 025701 (2018).
- <sup>121</sup>We assume, without the loss of generality, that  $N^\uparrow \geq N^\downarrow$ .
- <sup>122</sup>R. Assaraf, M. Caffarel, and A. Scemama, *Phys. Rev. E* **75**, 035701 (2007).
- <sup>123</sup>L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, *Phys. Rev. B* **84**, 245117 (2011).
- <sup>124</sup>C. Genovese, A. Meninno, and S. Sorella, *J. Chem. Phys.* **150**, 084102 (2019).
- <sup>125</sup>R. Resta and S. Sorella, *Phys. Rev. Lett.* **82**, 370 (1999).
- <sup>126</sup>M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti, *Phys. Rev. Lett.* **94**, 026406 (2005).
- <sup>127</sup>M. Casula, C. Attaccalite, and S. Sorella, *J. Chem. Phys.* **121**, 7110 (2004).
- <sup>128</sup>S. Sorella and A. Zen, "The new resonating valence bond method for ab initio electronic simulations," in *Many-Electron Approaches in Physics, Chemistry and Mathematics: A Multidisciplinary View*, edited by V. Bach and L. Delle Site (Springer International Publishing, Cham, 2014), pp. 377–392.
- <sup>129</sup>A. Zen, B. L. Trout, and L. Guidoni, *J. Chem. Phys.* **141**, 014305 (2014).
- <sup>130</sup>M. Casula, M. Marchi, S. Azadi, and S. Sorella, *Chem. Phys. Lett.* **477**, 255 (2009).
- <sup>131</sup>M. Casula, S. Yunoki, C. Attaccalite, and S. Sorella, *Comput. Phys. Commun.* **169**, 386 (2005).
- <sup>132</sup>N. Dupuy and M. Casula, *J. Chem. Phys.* **148**, 134112 (2018).
- <sup>133</sup>F. Sterpone, L. Spanu, L. Ferraro, S. Sorella, and L. Guidoni, *J. Chem. Theory Comput.* **4**, 1428 (2008).
- <sup>134</sup>A. Zen, Y. Luo, G. Mazzola, L. Guidoni, and S. Sorella, *J. Chem. Phys.* **142**, 144111 (2015).
- <sup>135</sup>M. Dagrada, M. Casula, A. M. Saitta, S. Sorella, and F. Mauri, *J. Chem. Theory Comput.* **10**, 1980 (2014).
- <sup>136</sup>A theory is size-consistent if the energy  $E_{A+B}$  of the supersystem made of two fragments A and B at a sufficiently large distance is equal to the sum  $E_A + E_B$  of the energy of the fragments. When a theory is size consistent, it is equivalent to take as reference energy for interaction energy evaluation the sum of the fragments or the supersystem with the fragments far apart, whereas if the theory is not, there is an arbitrary shift as large as  $E_A + E_B - E_{A+B}$  in the interaction energy.
- <sup>137</sup>JSD and JAGP could describe the O–O system if the number of unpaired electrons is changed, for instance, if four unpaired electrons are used. However, this is not the right parametrization of the O<sub>2</sub> molecule.
- <sup>138</sup>J. Toulouse and C. J. Umrigar, *J. Chem. Phys.* **128**, 174101 (2008).
- <sup>139</sup>R. A. Aziz and M. Slaman, *Chem. Phys.* **130**, 187 (1989).
- <sup>140</sup>J. M. Merritt, V. E. Bondybey, and M. C. Heaven, *Science* **324**, 1548 (2009).
- <sup>141</sup>S. R. Langhoff and C. W. Bauschlicher, *J. Chem. Phys.* **95**, 5882 (1991).
- <sup>142</sup>L. Bytautas and K. Ruedenberg, *J. Chem. Phys.* **122**, 154110 (2005).
- <sup>143</sup>A. Zen, D. Zhelyazov, and L. Guidoni, *J. Chem. Theory Comput.* **8**, 4204 (2012).
- <sup>144</sup>Indeed, in order to improve the precision by an order of magnitude, the sampling (so, the computational cost) has to be increased by two orders of magnitude. Considering that the evaluation of the energy and all the force components in TURBORVB costs only 4 times the evaluation of only the energy with the same sampling (see Sec. VI), the speedup is  $100/4 = 25$ .
- <sup>145</sup>Y. Luo and S. Sorella, *Front. Mater.* **2**, 29 (2015).
- <sup>146</sup>S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
- <sup>147</sup>B. V. D. Goetz and E. Neuscamman, *J. Chem. Theory Comput.* **13**, 2035 (2017).
- <sup>148</sup>C. Umrigar, M. Nightingale, and K. Runge, *J. Chem. Phys.* **99**, 2865 (1993).
- <sup>149</sup>A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè, *Phys. Rev. B* **93**, 241118 (2016).
- <sup>150</sup>N. Nemec, M. D. Towler, and R. Needs, *J. Chem. Phys.* **132**, 034111 (2010).
- <sup>151</sup>T. D. Beaudet, M. Casula, J. Kim, S. Sorella, and R. M. Martin, *J. Chem. Phys.* **129**, 164711 (2008).
- <sup>152</sup>L. Spanu, S. Sorella, and G. Galli, *Phys. Rev. Lett.* **103**, 196401 (2009).
- <sup>153</sup>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).
- <sup>154</sup>D. Nissenbaum, L. Spanu, C. Attaccalite, B. Barbiellini, and A. Bansil, *Phys. Rev. B* **79**, 035416 (2009).
- <sup>155</sup>G. Ferlat, M. Hellgren, F.-X. Coudert, H. Hay, F. Mauri, and M. Casula, *Phys. Rev. Mater.* **3**, 063603 (2019).
- <sup>156</sup>M. Calandra, F. Becca, and S. Sorella, *Phys. Rev. Lett.* **81**, 5185 (1998).
- <sup>157</sup>L. Capriotti, A. E. Trumper, and S. Sorella, *Phys. Rev. Lett.* **82**, 3899 (1999).
- <sup>158</sup>F. Becca, M. Capone, and S. Sorella, *Phys. Rev. B* **62**, 12700 (2000).
- <sup>159</sup>L. Capriotti, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. Lett.* **87**, 097201 (2001).
- <sup>160</sup>S. Sorella, G. B. Martins, F. Becca, C. Gazza, L. Capriotti, A. Parola, and E. Dagotto, *Phys. Rev. Lett.* **88**, 117002 (2002).
- <sup>161</sup>J. G. Bednorz and K. A. Müller, *Z. Phys. B* **64**, 189 (1986).
- <sup>162</sup>Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, *J. Am. Chem. Soc.* **130**, 3296 (2008).
- <sup>163</sup>M. Marchi, S. Azadi, and S. Sorella, *Phys. Rev. Lett.* **107**, 086807 (2011).
- <sup>164</sup>M. Casula and S. Sorella, *Phys. Rev. B* **88**, 155125 (2013).
- <sup>165</sup>B. Busemeyer, M. Dagrada, S. Sorella, M. Casula, and L. K. Wagner, *Phys. Rev. B* **94**, 035108 (2016).
- <sup>166</sup>S. Sorella, K. Seki, O. O. Brovko, T. Shirakawa, S. Miyakoshi, S. Yunoki, and E. Tosatti, *Phys. Rev. Lett.* **121**, 066402 (2018).
- <sup>167</sup>M. Barborini and L. Guidoni, *J. Chem. Phys.* **137**, 224309 (2012).
- <sup>168</sup>M. Barborini and E. Coccia, *J. Chem. Theory Comput.* **11**, 5696 (2015).
- <sup>169</sup>M. Barborini and L. Guidoni, *J. Chem. Theory Comput.* **11**, 508 (2015).
- <sup>170</sup>M. Barborini and L. Guidoni, *J. Chem. Theory Comput.* **11**, 4109 (2015).
- <sup>171</sup>E. Coccia and L. Guidoni, *J. Comput. Chem.* **33**, 2332 (2012).
- <sup>172</sup>E. Coccia, D. Varsano, and L. Guidoni, *J. Chem. Theory Comput.* **9**, 8 (2013).
- <sup>173</sup>E. Coccia, D. Varsano, and L. Guidoni, *J. Chem. Theory Comput.* **10**, 501 (2014).
- <sup>174</sup>E. Coccia, D. Varsano, and L. Guidoni, *J. Chem. Theory Comput.* **13**, 4357 (2017).
- <sup>175</sup>E. Coccia, O. Chernomor, M. Barborini, S. Sorella, and L. Guidoni, *J. Chem. Theory Comput.* **8**, 1952 (2012).
- <sup>176</sup>S. Chu, E. Coccia, M. Barborini, and L. Guidoni, *J. Chem. Theory Comput.* **12**, 5803 (2016).
- <sup>177</sup>M. Barborini and L. Guidoni, *J. Chem. Phys.* **145**, 124107 (2016).
- <sup>178</sup>S. Sorella, M. Casula, L. Spanu, and A. Dal Corso, *Phys. Rev. B* **83**, 075119 (2011).
- <sup>179</sup>N. Devaux, M. Casula, F. Decremps, and S. Sorella, *Phys. Rev. B* **91**, 081101 (2015).
- <sup>180</sup>H. Hay, G. Ferlat, M. Casula, A. P. Seitsonen, and F. Mauri, *Phys. Rev. B* **92**, 144111 (2015).
- <sup>181</sup>M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokofev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang (Simons Collaboration on the Many-Electron Problem), *Phys. Rev. X* **7**, 031059 (2017).
- <sup>182</sup>M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K. Chan, N. Chepiga, P. Helms, C. Jimenez-Hoyos, A. J. Millis *et al.*, [arXiv:1911.01618](https://arxiv.org/abs/1911.01618) (2019).
- <sup>183</sup>E. Wigner and H. Huntington, *J. Chem. Phys.* **3**, 764 (1935).
- <sup>184</sup>N. W. Ashcroft, *Phys. Rev. Lett.* **21**, 1748 (1968).
- <sup>185</sup>T. Guillot, *Annu. Rev. Earth Planet. Sci.* **33**, 493 (2005).
- <sup>186</sup>G. Mazzola and S. Sorella, *Phys. Rev. Lett.* **114**, 105701 (2015).
- <sup>187</sup>M. Barborini, S. Sorella, and L. Guidoni, *J. Chem. Theory Comput.* **8**, 1260 (2012).
- <sup>188</sup>N. Dupuy, S. Bouaouli, F. Mauri, S. Sorella, and M. Casula, *J. Chem. Phys.* **142**, 214109 (2015).
- <sup>189</sup>M. Barborini, S. Sorella, M. Rontani, and S. Corni, *J. Chem. Theory Comput.* **12**, 5339 (2016).
- <sup>190</sup>M. Barborini, S. Sorella, M. Rontani, and S. Corni, *J. Chem. Phys.* **149**, 154102 (2018).
- <sup>191</sup>J. T. Krogel, *Comput. Phys. Commun.* **198**, 154 (2016).
- <sup>192</sup>V. Konkov and R. Peverati, *SoftwareX* **9**, 7 (2019).