LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-641057

First Principles Methods: A Perspective From Quantum Monte Carlo

M. A. Morales, R. Clay, C. Pierleoni, D. M. Ceperley

July 17, 2013

Entropy

Disclaimer

Submitted to *Entropy*. Pages 1 - 38.

OPEN ACCESS

Article

First Principles Methods: A Perspective From Quantum Monte Carlo

Miguel A. Morales [1,], Raymond Clay[2], Carlo Pierleoni[3]* and David M. Ceperley [2]

[1]Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, Ca, 94550

[2] Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street Urbana, IL 61801-3080

[3] Dipartimento di Scienze Fisiche e Chimiche, Università de L'Aquila, Italy

* Author to whom correspondence should be addressed; carlo.pierleoni@aquila.infn.it

Version July 9, 2013 submitted to Entropy. *Typeset by LaTeX using class file mdpi.cls*

Abstract: Quantum Monte Carlo methods are the most accurate algorithms for predicting properties of general quantum systems. We briefly introduce ground state, path integral at finite temperature and coupled-electron Monte Carlo methods, their merits and limitations. We then discuss recent calculations using these methods for dense liquid hydrogen as in undergoes a molecular atomic (metal/insulator) transition. We then discuss a procedure that can be used to assess electronic density functionals, which in turn can be used on a larger scale for first principles calculations and apply this technique to dense hydrogen and liquid water.

Keywords: quantum Monte Carlo, first-principles simulations, hydrogen, Coupled Electron-Ion Monte Carlo, high pressure

1. Introduction

With the increasing computational power and the greater access to large clusters seen during the last decade, simulation methods have become an increasingly useful tool for many fields of science, including chemistry, materials science, condensed matter physics, and biophysics. In this article we explore some of the future impact of Quantum Monte Carlo in the field of first principles simulation (FPS). By this we mean reliable simulations that can be performed on condensed matter systems in the absence of detailed experimental information on those systems[1]. Starting with the general Hamiltonian in Eq. (1), and taking as input only the chemical compositions, masses, density, temperature etc, currently there is a hierarchy of methods that are used to perform such a simulation. In this introduction we focus on only three classes of methods: the use of semi-empirical interatomic potentials together with Monte Carlo (MC) or molecular dynamics (MD) simulations , DFT-based simulation methods, and quantum Monte Carlo simulations.

The first member of the hierarchy uses interatomic potentials, the best known of which is the Lennard-Jones potential. Such potentials are routinely used in the vast majority of simulations (soft condensed matter, biophysics, materials science) and are reviewed in a different contribution to this issue [1]. The first question is how do we construct such a potential? The typical approach is to use available experimental data. However, it is well known that those potentials are not very accurate in the vast majority of systems, even if they match experimental data. Hence, though they can be used to say something about generic properties of systems, quantitative predictions for defect energies, energy barriers, melting temperatures, cannot be trusted [2]. Another fundamental limitation of this approach is that it becomes difficult to construct reliable interatomic potentials for complex systems containing several types of atoms, for example a solvent with various solutes, or systems under extreme conditions, since it becomes difficult to get enough reliable experimental data to constrain all of the parameters. For these reasons, it is highly desirable to have methods that can provide reliable predictions without input from experimental measurements.

Density Functional Theory (DFT) in the Kohn-Sham formulation solves a one body problem, and adds electronic correlation through an exchange-correlation functional. A breakthrough in the usefulness and popularity of simulations occurred with the development of the ab initio molecular dynamics (AIMD) approach by Car and Parrinello [2], when they combined molecular dynamics and DFT to perform simulations of complex chemical systems. Due to its favorable ratio between accuracy and computational cost, DFT has become the workhorse in the field of first-principles simulations. In fact, the recent explosion in the popularity of first-principles methods is, to a large part, due to the success of DFT in providing a fairly accurate description of the electronic structure of materials at a reasonable computational cost. DFT also gives access to a large range of observables. While DFT has been very successful in the description of many types of materials, e.g. metals and weakly correlated systems, many of the currently available exchange-correlation functionals in DFT possess well-known limitations [3], including the failure to properly describe strongly correlated materials, self-interaction errors, etc. It is recognized that even for such a fundamental system as water, the AIMD procedure is not accurate

[1] Of course experiment has been used to give us the general physical laws, and to validate the computational approaches.

[2] If the potential has been adjusted to reproduce experimental measurements, then the method is no longer first principles, and the question becomes whether the potential is transferrable, i.e. reliable for properties that are not fitted for.

enough, giving large errors in many basic properties including the melting temperature, the diffusion constant, the compressibility, among others [4].

In the past decade there has been an explosion of new DFT functionals with various characteristics because of the difficulty of making systematic improvements to the functional or judging the accuracy of a functional. If the DFT functional is considered as "variable" then how does the user, in the absence of experimental data, decide on the functional? In the case of finite molecular systems, the availability of high-level quantum chemistry methods offers a possible path towards the improvement of DFT, for example by minimizing errors in a training set between DFT and CCSD(T) results. In fact, many exchange-correlation functionals contain optimizable parameters that are obtained from calculations on finite molecular systems (exceptions to this include LDA, PBE, among others), where results of quantum chemistry methods are routinely used as a references. In solids, accurate calculations using many-body methods are computationally expensive, which has limited their use in the development of density functionals. While there has also been considerable developments in other correlated approaches for bulk systems, such as GW, Bethe-Saltpeter, and DMFT, they are more expensive and still leave questions of accuracy. For reasons of space, we do not discuss these approaches further.

The third approach is the use of quantum Monte Carlo (QMC) methods, which are generalizations of the classical Monte Carlo techniques to quantum statistical physics and fundamentally based on imaginary-time path integrals. For a class of systems (bosons, one dimensional physics) such techniques provide an exact computational method. For general problems, though not exact, they are highly accurate *and* systematically improvable. Although there are a variety of QMC methods (ground state, variational, Path Integral, auxiliary field...) fundamentally they are closely related. QMC are the most accurate general methods but are less developed and require much more computational facilities than DFT methods (although the scaling of computer time versus system size is similar) limiting the systems on which such simulations can and have been done. The largest impact to date of QMC has been in the development and improvement of DFT methods; specifically we mention the correlation energy of the electron gas [5], a fundamental component in almost all exchange-correlation functionals used in DFT. Recent calculations [6] give the corresponding correlation energies for temperatures above zero.

Later in this paper we give an example of work in progress in this direction where QMC is used to directly rank various DFT functionals. We suggest that this benchmark quality data could be used to improve directly the best functionals. One can then envision using the highest ranked functional to develop intermolecular potentials that would then be of higher quality. Ercolessi et al. [7] have developed the force-matching procedure to find the optimal pair potential reproducing the forces appearing in an AIMD simulation. Such an approach is now feasible using QMC calculated forces and energies.

First principles simulation methods entirely based on QMC have also been developed in the last decade. These are the Coupled Electron-Ion Monte Carlo method [8] and the QMC-Molecular Dynamics [9], and have been recently reviewed in [10]. However their application to condensed phases has been limited so far to high pressure hydrogen, and hydrogen-helium mixtures because of the considerable computation cost of those methods. In this paper we will illustrate their use to investigate the dissociation of liquid molecular hydrogen under pressure, a problem which is still unsolved by DFT methods.

The article is organized as follows. We first describe in section 2 the various QMC methods. In the following section 3.1 we then present a few applications of QMC to the high pressure phases of hydrogen.

This is followed in section 3.2 by a description of the use of these methods to provide quantitative information on the accuracy of various DFT functionals. Finally we close with a discussion in section 4.

2. Computational Methods

In this section, we review some of the Quantum Monte Carlo methods used in the first principles modeling of condensed matter systems. Under normal conditions of temperature and pressure, such systems are described to a high degree of accuracy by the non-relativistic Hamiltonian for a collection of electrons and ions:

$$\begin{aligned} \hat{H} &= \hat{T}_n + \hat{H}_{el} = \hat{T}_n + \hat{T}_e + \hat{V}, \\ \hat{T}_n &= -\sum_{I=1}^{N_n} \lambda_I \hat{\nabla}_I^2, \quad \hat{T}_e = -\lambda_e \sum_{i=1}^{N_e} \hat{\nabla}_i^2, \\ \hat{V} &= \sum_{I<J} \frac{z_I z_J}{|\vec{R}_I - \vec{R}_J|} + \sum_{i<j} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{i,I} \frac{z_I}{|\vec{r}_i - \vec{R}_I|}, \end{aligned} \tag{1}$$

where N_n and N_e are the number of ions and electrons, respectively, in atomic units $\lambda_e = 1/2$, $\lambda_I = 1/(2M_I)$, and M_I and z_I are the mass and charge (in units of the electron mass m_e and charge e) and charge of the nucleus I. The system occupies a volume Ω.[3]. Note that $\vec{r}$ with lower case indexes (i, j, ...) is used to denote the position of electrons and $\vec{R}$ with upper case indexes (I, J, ...) is used for the nuclei. When no indices are used, $\vec{r}$ and $\vec{R}$ represent the full $3N_e$ and $3N_n$ dimensional vectors, respectively. The electronic Hamiltonian $\hat{H}_{el}$ corresponds to the solution of the problem in the clamped-nuclei approximation, where the ions produce a fixed external potential for the electrons. Another quantity that will be of interest is the electron number-density given by $\rho = N_e/\Omega$, and parameterized with $r_s = a/a_0$, where $4\pi a^3/3 = \rho^{-1}$. Given Eq. (1), we only need to add the temperature, particle statistics and boundary conditions to completely specify the physical and numerical problem to be solved.

Finding the eigenvalues and eigenfunctions of the Hamiltonian in Eq. (1) is a formidable task, impossible to do analytically except for a few simple systems such as the single hydrogen atom. In practice, numerical or approximate theoretical methods must be used. Two of the most widely applicable methods are based either on imaginary-time path integrals or density functional theory (DFT), as discussed in the following subsections.

[3] In this section, we use atomic units, where Planck's constant $\hbar = m_e = k_B = e = 1$ with k_B being Boltzmann's constant, and the energy is measured in Hartrees $E_\mathrm{h} = 315,775$ K$= 27.2114$ eV. Note that, in these units, the energy of a hydrogen atom is $0.5E_\mathrm{h}$, the binding energy of a hydrogen molecule is $0.17E_\mathrm{h}$, the unit of length is the Bohr Radius $a_0 = 0.0529$ nm, and the equilibrium bond length is 1.4 a0.

2.1. Ground State Methods

The following ground state methods seek to evaluate expectation values of physical observables taken over the ground state wavefunction $\phi_0(R)$:

$$\langle \hat{\mathcal{O}} \rangle = \frac{\int dR\, \phi_0^*(R) \hat{\mathcal{O}} \phi_0(R)}{\int dR\, |\phi_0(R)|^2}. \tag{2}$$

Two problems are evident from this formula. The first is that we almost never know $\phi_0(R)$ exactly. The second is that even if we did, Eq. (2) is a high dimensional integral. The following methods address both these problems. For sake of notation simplicity, throughout the sections 2.1-2.2 we will indicate by R the set of all coordinates of the quantum degrees of freedom without distinction between electrons and nuclei.

2.1.1. Variational Monte Carlo

Variational Monte Carlo (VMC) is conceptually the simplest of the ground-state QMC methods. It works by approximating the true ground-state wavefunction $\phi_0(R)$ with some trial wavefunction $\Psi_T(R)$. Integrals like Eq. (2) are then performed using Metropolis Monte Carlo sampling, with $\Psi_T(R)$ in place of $\phi_0(R)$ [11]. The accuracy of this method depends strongly on how closely $\Psi_T(R)$ approximates $\phi_0(R)$. Fortunately, the variational principle of quantum mechanics gives us a metric by which to improve the quality of trial wavefunctions. Consider the expectation value of the Hamiltonian and its variance:

$$E[\Psi_T] = \frac{\int dR\, \Psi_T^*(R) \hat{H} \Psi_T(R)}{\int |\Psi_T(R)|^2 dR} \tag{3}$$

$$\sigma_E^2[\Psi_T] = \frac{\int dR\, \Psi_T^*(R) (\hat{H}^2 - E[\Psi_T]) \Psi_T(R)}{\int dR\, |\Psi_T(R)|^2} \tag{4}$$

$$\tag{5}$$

The variational theorem states that:

$$E[\Psi_T] \geq E[\phi_0] \tag{6}$$

$$\sigma_E^2[\Psi_T] \geq \sigma_E^2[\phi_0] = 0. \tag{7}$$

Based on this, improvements to the wavefunction can be quickly gauged by whether they lower the energy and variance. A popular approach for fermionic problems is to assume a Slater-Jastrow wavefunction: a Slater determinant of single-particle orbitals, multiplied by a function that is symmetric under particle exchange [12,13]. The single-particle orbitals are typically taken from other quantum-chemistry methods (Hartree-Fock, DFT, etc). Other possible choices include multi-Slater determinant expansions [14], geminals [15], etc.

VMC can be improved if we consider classes of trial wavefunctions $\Psi_T(R, \alpha)$ parameterized by $\alpha = (\alpha_1, ..., \alpha_m)$. We then minimize the energy and/or variance with respect to these parameters. Recent improvements to optimization algorithms allow the optimization of thousands of variational parameters [16,17]. Traditionally, only the Jastrow functions have been parameterized, although work has been done using parameterized single particle orbitals and multi-Slater determinantal expansions.

VMC has some advantages that keep it in use. First, it is usually computationally cheaper than more accurate QMC methods, scaling like $\mathcal{O}(N^3)$ for fermions but with a smaller prefactor. VMC can also include several different types of electron correlations. Lastly, it doesn't suffer from a sign problem. However, it is at heart an approximate method, and does depend on the choice of trial wavefunction.

2.1.2. Projector Methods

2.1.3. Formalism

Projector methods attempt to stochastically project out the exact many-body ground state, allowing us to sample this distribution for Monte Carlo integration. The "projector", or imaginary-time Green's function $G(R', R, \beta' - \beta)$, is the operator solution to the imaginary-time Schrödinger equation:

$$\frac{\partial \Psi}{\partial \beta} = -\hat{H}\Psi(R, \beta), \tag{8}$$

subject to the boundary condition that $\lim_{\beta' \to \beta} G(R', R, \beta' - \beta) = \delta(R' - R)$. One can verify that the formal solution is $\hat{G} = \exp(-\beta \hat{H})$. Now consider an arbitrary wavefunction $\Psi(R, \beta = 0)$ that is not orthogonal to the ground state $\phi_0(R)$. Expanding this function in terms of the eigenfunctions of the Hamiltonian, and applying the projector to this, we find:

$$\begin{aligned} \Psi(R, \beta) &= \sum_i a_i \phi_i(R) e^{-\beta \epsilon_i}, \\ &\propto a_0 \phi_0(R) + \sum_i a_i \phi_i(R) e^{-\beta(\epsilon_i - \epsilon_0)}. \end{aligned} \tag{9}$$

This implies that as $\beta \to \infty$, we are left with just the ground state wavefunction.

For efficiency reasons, it is better to use the "importance-sampled" Schrödinger's equation [12,18,19]. We obtain this by writing the original equation in terms of $f(R, \beta) = \Psi_T(R)\Psi(R, \beta)$. After some algebra [12], we find that

$$\begin{aligned} \frac{\partial f(R, \beta)}{\partial \beta} &= \hat{L} f(R, \beta), \\ &= \lambda \nabla \cdot [\nabla - F(R)] f(R, \beta) + [E_T - E_L(R)] f(R, \beta). \end{aligned} \tag{10}$$

$F(R)$ is the quantum force defined by $F(R) = \nabla \ln |\Psi_T(R)|^2$ and $E_L(R)$ is the local energy. If $f(R, \beta) \geq 0$ everywhere, then we can interpret f as a probability distribution. This amounts to demanding a bosonic many-body ground state (fermions will be covered in a later section). Eq. (10) can then be interpreted a generalized Smoluchowski equation for a drift-diffusion process with sources and sinks. The first term represents a drift-diffusion process, whereas the second term represents an exponential growth/decay process. When we get around to simulating this equation, we will use the mapping between a Smoluchowski equation governing probability distributions, and Langevin-like equations, governing the diffusion and growth of *particles*.

The Green's function for this equation is formally $\tilde{G}(R', R, \beta) = \langle R' | \exp(\beta \hat{L}) | R \rangle$, although it is easy to show that this is related to the original projector by the transformation $\tilde{G}(R', R, \beta) =$

$\Psi_T(R')G(R',R,\beta)\Psi_T(R)^{-1}$. In the short-time approximation, we can decouple the drift-diffusion and growth operators by the Trotter formula. The result (for the symmetric decomposition) is:

$$\tilde{G}(R',R,\tau) = G_{DD}(R',R,\tau)G_B(R',R,\tau) \quad (11)$$

$$G_{DD} = \exp\left(-\frac{(R-R'-2\lambda_e\tau_e F)^2}{4\lambda_e\tau_e}\right) \quad (12)$$

$$G_B(R',R,\tau) = \exp(-\frac{\tau}{2}[E_L(R')+E_L(R)-2E_T]) \quad (13)$$

The short-time approximation allows us to deal with the full propagator as a product of short-time propagators, $\hat{G}(\beta) = (\hat{G}(\tau=\beta/M))^M$. The cost is that we have now incurred a time-step error that we must take into account.

2.1.4. Diffusion Monte Carlo

In diffusion Monte Carlo [19–21], we represent the distribution function $f(R,\beta)$ as an ensemble of 3N-dimensional samples $\{R_1,...,R_M\}$, which are known as "walkers". The average density of walkers is proportional to the distribution function f.

As in classical diffusion, we would then simulate Eq. (11) by use of a Langevin-like process acting on the walkers. Assuming that the time step $\tau=\beta/M$ is sufficiently small, we advance from $f(R,\beta)\rightarrow f(R,\beta+\tau)$ by first proposing to move each walker R_i to R_i' by a drift-diffusion step, prescribed by Eq. (12). Then we accumulate a weight associated with walker i, given by $w_i(\beta+\tau)=w_i(\beta)G_B(R_i',R_i,\tau)$. To calculate the expectation value of an operator $\hat{\mathcal{O}}$ over $f(R,\beta)=\Psi_T(R)\Psi(R,\beta)$, we average over the ensemble of walkers, including the appropriate weights:

$$\langle\hat{\mathcal{O}}\rangle = \frac{\sum_{i=1}^M w_i(\beta)\mathcal{O}(R_i)}{\sum_{i=1}^M w_i(\beta)}. \quad (14)$$

If we stopped here, this would be the basis of pure-diffusion Monte Carlo [22]. Because these weights are exponential factors, the variance associated with Eq. (14) will increase exponentially as the simulation progresses: the weights of a few walkers will exponentially grow, whereas the rest will exponentially tend to zero.

Branching diffusion Monte Carlo [20], by far the most used form of DMC, fixes this problem by using the weights to either replicate or kill off walkers. After each drift-diffusion step, the number of walkers associated with the single walker R_i to advance to the next time-step, M_{next}^i is chosen to be $M_{next}^i = \mathrm{INT}(w_i(\beta+\tau)+\xi)$, where ξ is a random number between $[0,1]$. The weights of the replicated walkers are all adjusted to conserve the total weight of walker i as much as possible. Modern methods are typically hybrids, where the weights of walkers are carried until they exceed certain established bounds, at which point they are branched [23].

The simulation is run by initializing the starting ensemble according to $f(R,0) = |\Psi_T(R)|^2$. Assuming β is the projection time required to reach the ground-state, the simulation is incremented $M=\beta/\tau$ steps, at which point our ensemble is distributed according to $f_0(R)=\Psi_T(R)\phi_0(R)$. Samples can then accumulated, and the simulation is run for a long enough time to achieve the desired error bars.

It is important to note that since we are sampling $f_0(R)$, this corresponds to the following type of expectation value, known as a "mixed-estimate":

$$\langle \hat{\mathcal{O}} \rangle_{DMC} = \frac{\langle \Psi_T | \hat{\mathcal{O}} | \phi_0 \rangle}{\langle \Psi_T | \phi_0 \rangle} \tag{15}$$

For observables that commute with the Hamiltonian, this gives us exact, unbiased estimates over the true many-body ground state wavefunction. For those that don't, the estimators will be biased by the quality of the trial wavefunction. This bias is less than that encountered by VMC, but still present. This can be alleviated somewhat by the use of "extrapolated estimators", and by the "forward-walking" method [24].

2.1.5. Reptation Monte Carlo

Reptation Monte Carlo is based on the path-integral representation of the projector. Assuming that β is large enough to guarantee sufficient convergence to the ground state, we begin by partitioning the full projector into M segments of time-interval $\tau = \beta/M$, called "time slices". Inserting a resolution of the identity between each short-time projector, we find the following path-integral expression for the mixed distribution $\langle \Psi_T | \phi_0 \rangle$:

$$\langle \Psi_T | \phi_0 \rangle = \int dR_0 \dots dR_M \Psi_T(R_0) G(R_0, R_1, \tau) \dots G(R_{M-1}, R_M, \tau) \Psi_T(R_M) \tag{16}$$

Using the short-time approximate Green's function at the beginning of this section, we can recast this expectation value in a more traditional path-integral form:

$$\langle \Psi_T | \phi_0 \rangle = \mathcal{Z} = \int \mathcal{D}X e^{S[X]} \tag{17}$$

$$S[X] = \ln \Psi_T(R_0) + \ln \Psi_T(R_M) - \sum_{i=0}^{M-1} L_s(R_i, R_{i+1}) \tag{18}$$

$$L_s(R', R) = \frac{(R'-R)^2}{4D\tau} + \frac{1}{2}(R'-R)\cdot(F'-F) \tag{19}$$

$$+\frac{\tau}{2}\left[E_L(R') + E_L(R) + D(F^2(R') + F^2(R))\right] \tag{20}$$

Here, X is shorthand for the directed path $X = R_0, \dots, R_M$. Eq. (17) plays the role of a partition function in statistical mechanics, where the $\Pi[X] = e^{S[X]}/\mathcal{Z}$ is the probability of a given path X, $-S[X]$ is the ground-state action, which includes the wavefunctions at the ends of the path, as well as a sum over "link-actions" $L_s(R', R)$, which are defined between time slices (see Eqs. (19-20)). The form we used for the link-action comes from symmetrizing the normal Green's function by hand, and writing it in terms of the importance-sampled Green's functions [25].

The versatility of reptation Monte Carlo comes from how $\Pi[X]$ is sampled. In the original method [26], one takes a given path X and chooses a growth direction at random. One then proposes a new path X^* by adding a time slice to the "head" and removing one from the "tail". Acceptance or rejection of this move based on the usual Metropolis acceptance step. This type of move is called "reptation", reminiscient of a "reptile", from which the method derives its name. The proposed head move is done by a drift-diffusion move, as in DMC, and rigorously preserves detailed balance.

Most practical implementations use what's known as the "bounce algorithm" [25]. Rather than choose the growth direction randomly, the direction is set at the beginning of the simulation and is changed only after a rejection step, hence the name "bounce". This method does not satisfy detailed balance, but does satisfy a more general stationarity condition that we require for Markov chain Monte Carlo. This dramatically decreases the autocorrelation time of the method, and also tames ergodicity problems that have been observed to crop up in the method.

RMC is appealing for two main reasons. It gives us the same level of accuracy for the energy as in DMC, but correlated sampling between different configurations can be done without approximation. This is particularly useful in methods like CEIMC. RMC also gives us the ability to sample expectation values over the pure distribution, as seen below:

$$\langle \hat{\mathcal{O}} \rangle_{pure} = \frac{\langle \Psi_T | e^{-\frac{\beta}{2}\hat{H}} \hat{\mathcal{O}} e^{-\frac{\beta}{2}\hat{H}} | \Psi_T \rangle}{\langle \Psi_T | e^{-\beta \hat{H}} | \Psi_T \rangle} \tag{21}$$

$$= \frac{1}{\mathcal{Z}} \int \mathcal{D}X e^{-S[X]} \mathcal{O}(R_{\beta/2}) \tag{22}$$

This shows that the center time slice of the reptile is distributed according to $|\phi_0(R)|^2$, whereas the ends are distributed according to the mixed distribution $f(R)$. This easy access to the pure distribution makes RMC ideal for calculations of unbiased observables and correlation functions, doing so in a more efficient manner than "forward-walking" in DMC.

2.1.6. The Fixed-Node Approximation

The previous projector methods we mentioned are in principle *exact* for bosonic systems, since the mapping to a diffusion process is valid when $\phi_0(R) \geq 0$ everywhere. However, since the wavefunction for a fermion systems must be antisymmetric under exchange, the ground state wavefunction will have as many negative configurations as positive ones[4]. We can restore the probabilistic interpretation of the wavefunction $\Psi(R, \beta)$ if we factor the sign into the weight of the walker, or into the observable itself. It turns out that in doing so, we will have large and almost equal contributions to the expectation value of opposite signs. This leads to an exponentially decaying signal to noise ratio, implying that the computational effort required to treat the fermion problem directly scales exponentially.

By far, the most common means of alleviating the sign-problem in both DMC and RMC is applying the "fixed-node" approximation [20,21]. We assume that the nodes of $\phi_0(R)$ are the same as the nodes for $\Psi_T(R)$. We then propagate our ensemble of walkers or our reptile strictly within restricted space where $\Psi_T(R)$ doesn't change sign. This can be implemented by rejecting moves that carry walkers across a node, or bouncing a reptile whenever a head move is proposed across a nodal surface. Though this is an uncontrolled approximation, it turns out to be an extremely good one in most cases. Fixed-node energies are known to be upper bounds, which allows us to optimize the nodal surfaces and to compare fixed-node DMC and fixed-node RMC energies with other methods. It turns out that both of these methods are among the most accurate computational methods known for electronic systems.

[4] In many cases the wavefunction can be made real

2.2. Finite-Temperature Methods

Next, we summarize path integral methods. These methods are similar to DMC but can treat systems at non-zero temperature: a many-body density matrix replaces the trial wave function. Concerning first principles simulations the path integral method can be used either to simulate the properties of thermal electrons or to simulate the zero point effects of light nuclei or both. For electronic simulations there are two major problems. First, the energy scale of electrons is 1 Hartree or above, thus to reach ambient temperature requires very long paths. Second, since electrons are fermions, antisymmetrization and hence the sign problem is inevitable. For a more complete overview of the method and its application to fermion systems, see Refs. [27] and [28] respectively.

2.2.1. Path Integrals

To begin, we define the many particle density matrix:

$$\rho(R, R'; \beta) = \langle R \mid e^{-\beta \hat{H}} \mid R' \rangle \tag{23}$$

where $R \equiv (r^{(1)}, \ldots, r^{(N)})$ with $r^{(i)}$ specifying the spacial coordinates of the i^{th} of N particles, and $\beta \equiv 1/k_B T$, the inverse temperature. The diagonal of the density matrix $\rho(R, R'; \beta)/Z(\beta)$ is the quantity that we need to sample in a first principles simulation. The partition function is defined as the trace of the density matrix,

$$Z(\beta) = Tr(\rho) = \int dR \langle R \mid e^{-\beta \hat{H}} \mid R \rangle = \int dR \rho(R, R; \beta). \tag{24}$$

The expectation value of any observable may be computed from this definition as

$$\langle \hat{\mathcal{O}} \rangle = Tr(\hat{\mathcal{O}}\rho)/Z = Tr(\hat{\mathcal{O}}\rho)/Tr(\rho). \tag{25}$$

Using the product property of the density matrix M times, such that $\beta = M\tau$, we write the partition function (or the diagonal density matrix) as an integral over a discrete path:

$$Z(\beta) = \prod_{i=0}^{M-1} \int d^d R_i \rho(R_0, R_1; \tau)\rho(R_1, R_2; \tau) \ldots \rho(R_{M-1}, R_0; \tau). \tag{26}$$

We have reduced the problem of sampling a low temperature density matrix to one of finding a high temperature density matrix and integrating over the path. The action, defined as

$$S(R_i, R_j; \tau) \equiv -ln[\rho(R_i, R_j; \tau)]. \tag{27}$$

can be broken into kinetic and potential parts, using Trotter's formula. The integration over all of the path variables is done using a specialized form of either Metropolis Monte Carlo or Molecular Dynamics.

Finally, in order to account for the particle statistics of the simulated system, we must sum over permutations $\mathcal{P}$, giving

$$Z(\beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \int_{R \to \mathcal{P}R} dR_t e^{-S[R_t]} \tag{28}$$

where R_t represents the generic path starting at R and ending at $\mathcal{P}R$.

250 2.2.2. Restricted Paths

For fermions, negative terms enter in this sum, leading to a sign problem. As was done in the previous discussion of DMC, one way to circumvent this issue is to impose a nodal constraint [29]. We define the *nodal surface* $\Upsilon_{R_\star\beta}$ for a given point $R_\star$ and inverse temperature β to be

$$\Upsilon_{R_\star\beta} = \{R \mid \rho(R, R_\star; \beta) = 0\} \tag{29}$$

which is a $(dN-1)$-dimensional manifold in dN-dimensional configuration space. Here, $R_\star$ is dubbed the *reference point*, as it is needed to define the nodal surfaces. Inside a nodal cell, by definition the sign of the density matrix is uniform. Using Dirichlet boundary conditions, we may solve the Bloch equation within each nodal cell. We define the *reach* $\Gamma_\beta(R)$ as the set of all continuous paths R_τ, for which $\rho(R_\tau, R_\star, \beta) \neq 0$ for all intermediate τ ($0 < \tau \leq \beta$), i.e. a node-avoiding path.

$$\Gamma_\beta(R) = \{\gamma : R \to R' \mid \rho(R, R_\tau; \beta) \neq 0\}. \tag{30}$$

It is clear that all paths contributing to the Bloch equation solution must belong to this reach. For all diagonal contributions, odd permutations must cross a node an odd number of times and thus are not allowed by this constraint and are exactly cancelled by all paths of node-crossing even permutations. This leaves us with the following expression for the density matrix,

$$\rho(R, R; \beta) = \frac{1}{N!} \sum_{\mathcal{P}, even} \int_{\gamma:R\to\mathcal{P}R}^{\gamma\in\Gamma_\beta(R)} \mathcal{D}R_\tau e^{-\mathcal{S}[R_\tau]/\hbar}. \tag{31}$$

We have thus turned the sign-full expression for the density matrix into one which includes only terms of a single sign, allowing efficient computation. However, because ρ appears on both sides of Eq. 31, this requires a priori knowledge of the density matrix nodal structure, which is generally unknown. To escape this self-consistency issue, an ansatz density matrix that approximates the actual nodal structure, is introduced. This will give an exact sampling of the fermi density matrix if its nodes are correct. This method is called *restricted* PIMC (RPIMC). The density matrix for non-interacting fermions, is a Slater determinant of single-particle distinguishable density matrices, $\rho(R, R_\star; \beta) = \frac{1}{N!} \det \rho_{ij_\star}$ where

$$\rho_{ij_\star} = (4\pi\lambda\beta)^{-d/2} \exp(-\frac{(r_i - r_{j\star})^2}{4\lambda\beta}). \tag{32}$$

251 It is a good approximation to use the free particle density matrix at high temperatures (say for
252 temperatures greater than the Fermi energy) and when correlation effects are weak. Furthermore, due
253 to the constraint of translational invariance, free particle nodes are quite reasonable for homogeneous
254 systems.

255 The nodal error, arising from using an approximate restriction is problematic since it is uncontrollable.
256 The finite temperature variational principle is through the free energy, as opposed to the internal energy
257 in the ground state. Thus one possible solution is parameterize the nodal ansatz, and then minimize
258 the free energy by varying the parameters. This will require a thermodynamic integration, in general.
259 Systems analyzed to date suggest that the nodal error arising from the free-particle ansatz is small since
260 the correlation from the potential is fully taken into acccount.

261 2.2.3. Path Integrals for Nuclei

262 Even when quantum particles can be considered distinguishable, as for instance light nuclei in
263 condensed phases, there could be substantial physical effects arising from their quantum behavior, i.e.
264 resulting from the $\hat{T}_n$ in Eq. (1). For example in bulk hydrogen and in water, the zero point motion
265 of the protons must be taken into account for an accurate description. Furthermore, the frequently
266 used harmonic approximation is often inadequate since non-harmonic effects can be as significant as
267 harmonic effects. In contrast to the situation with electrons, our ability to simulate the nuclei with
268 current algorithms and hardware is well controlled; because the nuclei are thousands of time heavier,
269 they are much closer to the classical limit, so that fewer path steps are needed. For hydrogen containing
270 compounds at room temperature, one can often get away with about ten steps. A second consequence is
271 that particle statistics (either fermi or bose) can typically be ignored; a notable exception is the difference
272 between para- and ortho-hydrogen, important for modeling the low-temperature low-pressure crystals of
273 molecular hydrogen and deuterium.

274 A frequent use of path integrals for nuclei occurs when DFT is used to integrate out the electronic
275 degrees of freedom. However, one wants to use the DFT energy surface for the properties of the quantum
276 nuclei in equilibrium, using the path integral method. To perform the path integration, it is advantageous
277 to use molecular dynamics instead of Monte Carlo since that will allow the electronic wave functions to
278 evolve smoothly in time, and thus reduce the time to convergence in solving the DFT self-consistency
279 conditions. M. Ceriotti, et al. [30] have devised an ingenious noise filtering scheme to reduce the number
280 of needed path integral steps. Assuming the density functional description of the electrons is accurate,
281 thermodynamic (static) properties of the simulated system will be accurate. However, the dynamical
282 properties are not to be trusted. In general, the algorithm does not give an accurate representation of
283 quantum dynamics.

2.3. Coupled Electron-Ion Monte Carlo

The QMC methods described so far, when applied to an ion-electron system, treat all particles on the same footing, either both in the ground state [31–33] or both at the same finite temperature [34–36]. However the large nucleon-electron mass ratio implies a wide separation of time and energy scales and it is a common practice to adopt the adiabatic, or Born-Oppenheimer (BO), approximation. Ignoring such an approximation in QMC causes difficulties. The imaginary time step of the path integral representation (both in DMC/RMC and PIMC) is imposed by the light electron mass. In DMC this means that nuclear "dynamics" (the speed of sampling configuration space) is much slower than electron "dynamics" requiring very long (and time consuming) trajectories. In PIMC the separation of time scales presents itself as a separation in the regions where thermal effects are relevant: in high pressure hydrogen for instance nuclear quantum effects becomes relevant below $\sim 2000K$ where electrons are, to a very good approximation, in their ground state. Performing PIMC in this region of temperatures requires very long electronic paths causing a slowing down of the exploration of configuration space and effectively limiting the ability of PIMC to perform accurate calculations at low temperatures.

The Coupled Electron-Ion Monte Carlo method (CEIMC) is a QMC based on the BO approximation [8]. In CEIMC a Monte Carlo calculation for finite temperature nuclei (either classical or quantum represented by path integrals) is performed using the Metropolis method with the BO energy obtained by a separate QMC calculation for ground state electrons. CEIMC has been extensively reviewed in refs. [8,10]. Here, we only briefly report the main technical features of the method.

2.3.1. Penalty Method

In CEIMC the difference of BO energies of two nearby nuclear configurations in a MC attempted step, as obtained by an electronic QMC run, is affected by statistical noise which, if ignored, results in a biased nuclear sampling. To cope with this situation either the statistical noise needs to be reduced to a negligible value by long electronic calculations (very inefficient), or the Metropolis acceptance/rejection scheme needs modifications to cope with noisy energy differences. The latter strategy is implemented in the Penalty Method [37] which enforces detailed balance to hold on average over the noise distribution. The presence of statistical noise causes an extra rejection for a single nuclear move with respect to the noiseless situation. An extra "penalty" defined as the variance of the energy difference over the square of the physical temperature is added to the energy differences. Therefore running at lower temperatures requires a reduced variance to keep an acceptable efficiency of the nuclear sampling. Small variances can be obtained if correlated sampling is used to compute the energy of the two competing nuclear configurations. In an attempted nuclear MC step, a single ground state electronic run is performed with a trial wave function which is a linear combination of the wave functions of the two nuclear configurations considered. The BO energy of the two nuclear configurations is obtained by a reweighting procedure which provides energy differences with a much reduced variance with respect to performing two independent electronic runs if the "distance" between the two nuclear configurations is limited (i.e. the overlap between the trial wave functions of the two configurations is large) [38]. This strategy allows an efficient sampling of nuclear configuration space for high pressure hydrogen and helium down to temperature as low as $\sim$200K.

2.3.2. Nuclear PIMC

When nuclear quantum effects are included using a path integral representation (see §2.2), the relevant inverse temperature in the penalty method is the imaginary time discretization step τ, so that no loss of efficiency is experienced when lowering the temperature (i.e. taking longer paths). For quantum protons in high pressure hydrogen, CEIMC can be used to efficiently study systems at temperatures as low as $\sim$200K. In the present implementation of nuclear quantum effects in CEIMC, we introduce an effective pair potential between nuclei and use the pair density matrix corresponding to the effective potential to factorize the imaginary time propagator. The residual difference between the energy of the effective system and the BO energy of the original system is considered at the primitive approximation level of the Trotter break-up of the proton propagator [8]. In high pressure hydrogen ($r_s = 1.40$) it is found that with this strategy, an inverse time step of $\tau^{-1} \simeq$4800K is enough to reach convergence of the thermodynamics properties, which allows to study systems at low temperature with a limited number of time slices ($\leq$50).

In CEIMC many-body nuclear moves are preferred to single-body moves. The reason is that even if only few nuclei are moved the entire electronic calculation must be repeated, by far the most expensive part of the method. For this reason we sample nuclear configuration by a smart Monte Carlo method [39] in the normal mode space of the path [40] with forces from the effective two body potential. This strategy allows us to simulate systems of $\sim$100 protons (for hydrogen) at temperature as low as 200K with an acceptable efficiency.

2.3.3. VMC vs RQMC

The main ingredient of CEIMC is the electronic QMC engine used to compute the BO energy. As mentioned a very important aspect for the efficiency of CEIMC is the noise level which is related to the variance of the local energy. In ground state QMC (see §2.1) the "zero variance principle" applies: if the trial wave function is an eigenfunction of the Hamiltonian, the local energy is no longer a function of the electronic coordinates and a single calculation provides the exact corresponding eigenvalues. Therefore by improving the trial wave function and approaching the exact ground state, the variance of the local energy decreases to zero. In connection with CEIMC, this is important not only for the accuracy of the BO energy but also for the efficiency of the nuclear sampling since the extra rejection due to the noise is reduced for a more accurate trial wave function.

To go beyond VMC accuracy in CEIMC we have implemented Reptation QMC method (RMC) [8, 26]. RMC is superior to DMC in the CEIMC context since it uses an explicit representation of the statistical weight of each path and therefore the reweighting procedure needed for estimating energy differences is easily applied. Going from VMC to RMC accuracy in CEIMC requires at least one order of magnitude more computer time. This is because it is in general more difficult to properly sample the configuration space of a 3N-dimensional path than of a 3N-dimensional point. It is analogous to the difficulty of sampling the configuration space of a long polymers with respect to point particles. For any proposed nuclear move one has to relax the electronic path to the new equilibrium and perform long enough sampling of the electronic configuration space to compute the energy difference with the required noise level.

In order to improve the efficiency of CEIMC while keeping the RMC accuracy, we have recently developed a method, based on a peculiar thermodynamic integration, to estimate the free energy of the system with RMC based BO energy from the knowledge of the free energy of the system with VMC based BO energy [41]. This allows to extensively use VMC rather than RMC, performing RMC on selected thermodynamic states only.

2.3.4. Hydrogen trail wave function

For high pressure hydrogen we have developed a quite accurate trial function of the Slater-Jastrow, single determinant, form. The Jastrow part has an electron-proton and electron-electron RPA term plus two-body and three-body numerical terms. The Slater determinants (one for each spin state) is built with single electron orbitals, obtained by a self-consistent DFT solution, expressed in terms of quasi-particle coordinates with a backflow transformation. [42,43]. We have recently integrated the PWSCF [44] DFT solver into our CEIMC code to ensure a faster and uniform convergence of the single electron orbitals in different physical conditions. Our trial wave function has a total of 13 variational parameters to be optimized [38,43].

In view of the large variability of DFT results from different exchange-correlation approximations in the dissociation region of high pressure hydrogen (see next section), one interesting question is about the sensitivity of the trial wave function to the particular form of the adopted Kohn-Sham orbitals in the Slater determinant. This is particularly relevant since the form of the orbitals determine the nodal surface of the trial wave function, the ultimate limit in the accuracy of fermionic QMC. On one hand one could hope to further improve the quality of the trial wave function by varying the type of orbitals, on the other hand a large sensitivity to the form of the Kohn-Sham orbitals will signal a too constrained form of the wave function, probably with a large room for improvements. The recent technical advance of the CEIMC code, namely the integration of PWSCF, allowed us to test several different types of orbitals: standard local (LDA) and semilocal (GGA-PBE) approximation, a non-local functional devised to reduce the self-interaction error and improve the description of the electronic correlation in DFT (HSE [45]) and a functional devised to improve the description of the dispersion interactions which are absent in a self-consistent mean-field theory (vdW-DF2 [46–48]) . In the range of coupling parameter $1.22 \leq r_s \leq 1.44$ which corresponds approximatively to the range of pressure between 200GPa and 550GPa according to DFT, we have considered four recently proposed candidate structures for the molecular crystal [49], namely C2/c, Cmca-12, Pbcn and P63m. For each structure we have performed parameter optimizations for the four mentioned forms of the orbitals and at eight different densities. Supercells of 96 atoms were considered for C2/c, Cmca-12 and Pbcn structures, while a supercell of 128 atoms was studied for the P63m structure. Moreover for a single structure, Pbcn, at a single value of $r_s = 1.35$ we have performed a complete RMC study. In figures 1 we report for all densities investigated the variational energies from the different orbitals relative to the energy of the trial function with LDA orbitals. We note that for all structures and at all densities LDA, PBE and HSE orbitals provides trial functions of the same quality (differences are of the order of 0.2 mH/atoms=90K except in a single case, Cmca at $r_s =$ where the difference is three time larger). Instead the trial function with orbitals from vdW-DF2 functional provides higher energies, by roughly 0.4mH/at with values up to 1.4mH/atom ($\simeq 630K$). This first result is quite indicative that our trail function is flexible and general enough to be very little sensitive

Figure 1. Variational energy of four different crystalline molecular structures versus r_s: C_2/c upper-left panel, Cmca-12 upper-right panel, P63m lower-left panel and Pbcn lower-right panel. Energies from wave functions with different orbitals relatives to the energy with LDA orbitals: PBE orbitals (red triangles), HSE orbitals (green closed circles) and vdW-DF2 orbitals (blue closed squares).

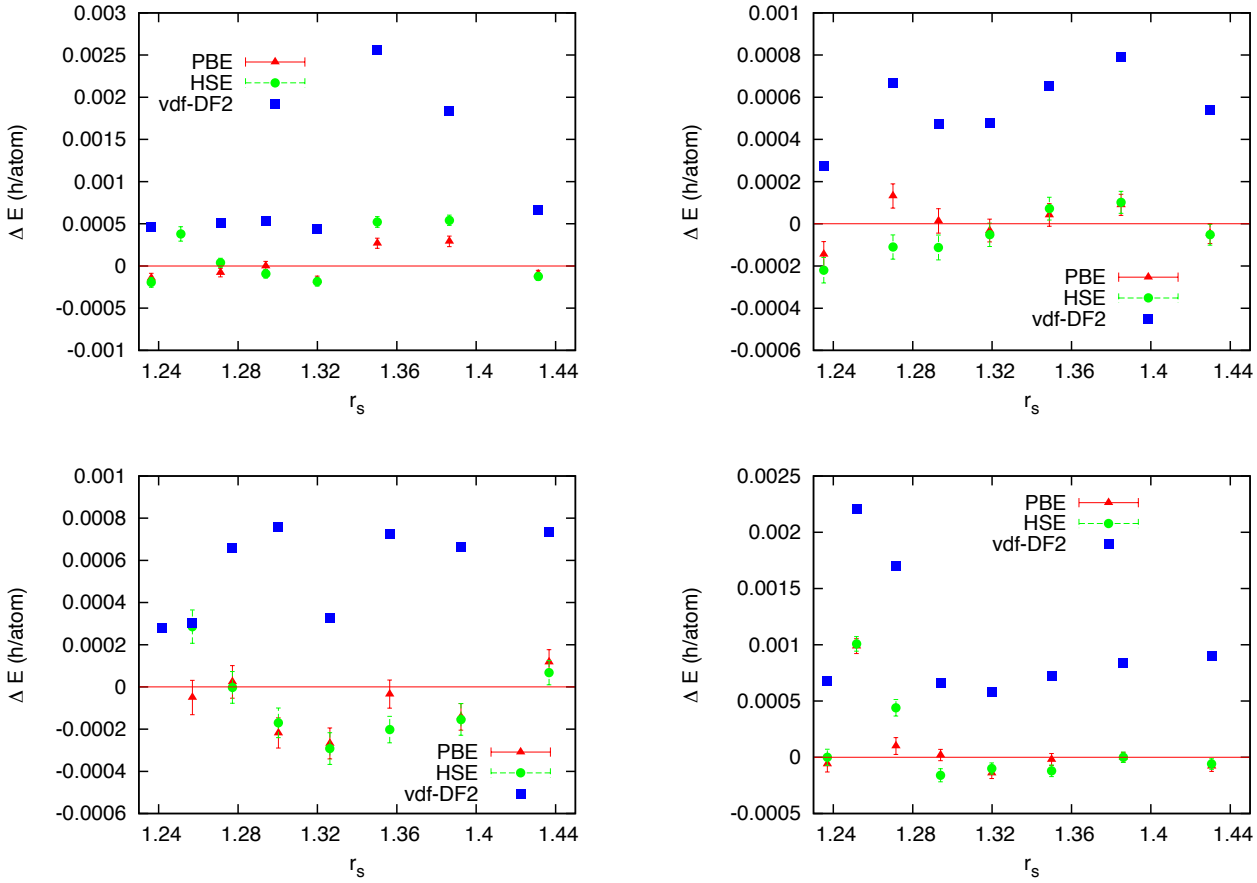
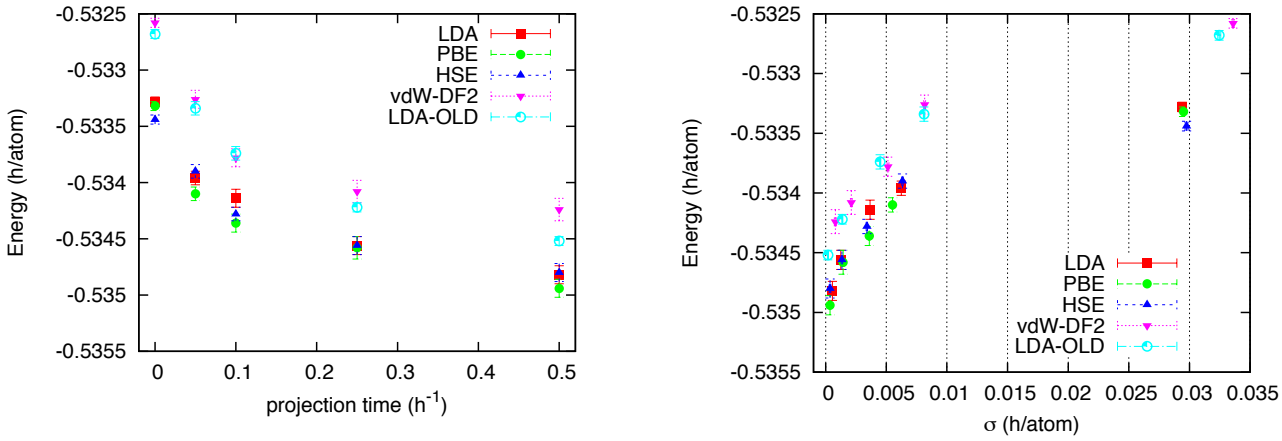


Figure 2. Pbcn structure of molecular hydrogen at $r_s = 1.35$. Left panel: energy per atom versus projection time in RMC from different kind of orbitals: LDA (closed red squares), PBE (green closed circles), HSE (upward blue triangles), vdW-DF2 (downward purple triangles). Also results from the old LDA implementation (cyan open circles) are reported. Right panel: Energy per atom versus variance in RMC from different kind of orbitals.



to the form of the orbitals (more comments please). In order to check whether the observed differences from vdW-DF2 orbitals could be due to optimization problems only, we performed a complete RMC study for a single case, namely the Pbcn structure at $r_s = 1.35$. A time step of $\tau = 0.005h^{-1}$ has been adopted, a value quite small. In figure 2 the energy versus projection time is reported for all kind of orbitals. We also added results from our old DFT solver with LDA orbitals plagued by the truncation error. For all kinds of trial function we observe a very similar relaxation with projection time meaning that the quality of the trial function is similar in all cases. The differences observed at the variational level essentially remain along the projection and therefore in the extrapolated value for the total energy. A quantitative way to estimate the extrapolated ($\beta \to \infty$) value of the total energy is to plot energy versus its variance and use a linear extrapolation at small values of σ^2. This plot for all studied cases is shown in the right panel of figure 2. We see that the three kinds of orbitals, LDA, PBE and HSE all provides extrapolated energies within error bars ($E_0 = -0.5350(2)$), while the the vdW-DF2 orbitals provides a higher value ($E_0 = -0.5342(2)$). The fact that the RMC projection is not able to remove the difference observed at the VMC level means that the nodes from the vdW-DF2 are less accurate than for the other kind of orbitals, which instead, despite their differences, provide essentially the same nodal structure. Finally we note that our old implementation of LDA orbitals provides a less accurate determination of the energy with correspondingly larger variance.

3. Applications

3.1. High-Pressure Hydrogen

Hydrogen is the simplest element of the periodic table and also the most abundant element in the Universe. Because of its simple electronic structure, it has been instrumental in the development of quantum mechanics and remains important for developing ideas and theoretical methods. In the next section we explore its use in developing DFT functionals. Its phase diagram at high pressure has received considerable attention from the first-principles simulation community due to its critical importance in many fields like planetary science, high pressure physics, astrophysics, inertial confinement fusion, among many others [10,50,51]. The phase diagram of hydrogen at high pressure contains many interesting features including: a maximum in the melting line with a subsequent negative slope [52,53], a predicted liquid-liquid transition between an insulating molecular and a conducting atomic phase [54,55], exotic molecular phases at low temperature, and a predicted metal-insulator transition in the solid phase [10,51].

The ground state structure of crystalline hydrogen across the pressure-induced molecular dissociation has been studied by DMC [31–33] which predicted molecular dissociation at density corresponding to $r_s \simeq 1.3$. R-PIMC has been applied to investigate the warm dense matter regime, namely the regime of high pressure and density where thermal and pressure molecular dissociation and ionization occur simultaneously [34,35,56]. Particularly relevant for our current understanding of the phase diagram and the Equation of State (EOS) of compressed hydrogen has been the determination of the primary and secondary Hugoniots lines of deuterium which could be directly compared with experimental data [36,57]. RPIMC predictions for the principal Hugoniot of deuterium were first in disagreement with pulsed laser-produced shock compression experiments [58–60], but were later confirmed by magnetically generated shock compression experiments at the Z-pinch machine [61–66] and by converging explosive-driven shock waves techniques [67,68]. Also relevant for the development and fine tuning of simulation methods for warm dense matter has been the comparison with the less demanding, but also less fundamental methods based on Density Functional Theory (either Kohn-Sham or Orbital-Free functionals). A general agreement between RPIMC and First Principle Molecular Dynamics (FPMD) predictions for the Hugoniot lines was observed [10] except at the lowest temperatures that could be reached by RPIMC ($\sim 10000K$). More recently the synergetic use of BOMD and R-PIMC has allowed to produce first-principle based EOS's in a wide range of physical conditions for hydrogen, helium and hydrogen-helium mixtures [69], instrumental in planetary modeling and crucial ingredients for the hydrodynamic codes used in the large facilities for extreme conditions experiments.

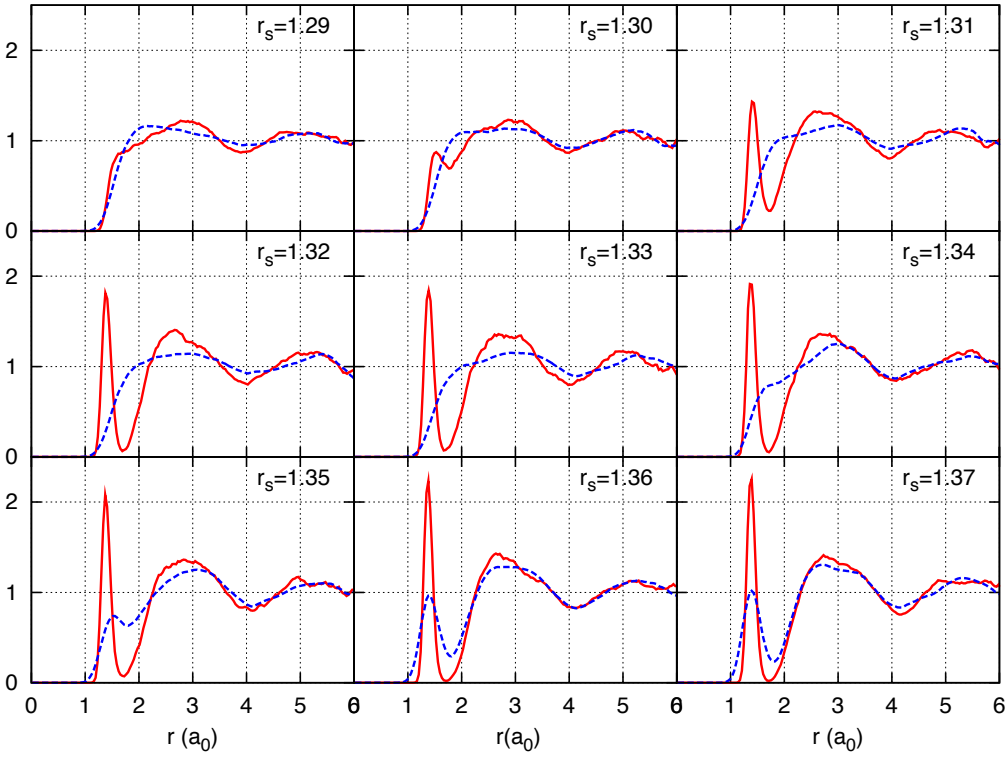
Temperatures lower than $\sim 10000K$ cannot be easily reached by RPIMC without reducing the level of accuracy. However, most of the interesting phenomena in high pressure hydrogen, like molecular dissociation under pressure, metallization, solid-fluid transition, a possible liquid-liquid phase transition and its interplay with melting, the various crystalline phases and the transition to the atomic phases [10], occur at lower temperature out of the reach of RPIMC. Investigating this regime by QMC methods has been the main motivation in developing CEIMC. The other motivation, as mentioned above, is the benchmark of the much more developed (and less demanding) alternative theoretical method, namely

FPMD based on DFT. Indeed the numerical implementation of DFT is based on approximations (the exchange-correlation functional) the accuracy of which can only be established against experiments or, better, against more accurate theories. As mentioned earlier, QMC energy is an upper bound and therefore has an internal measure of accuracy.

CEIMC has been applied to investigate the WDM regime of hydrogen and helium and benchmark FPMD [43,55,70]. In ref [55] an investigation of the fully ionized state of hydrogen in a region of pressure and temperature relevant for Jovian planets found that FPMD based on the GGA-PBE exchange-correlation functional and CEIMC are in very good agreement but both deviates from a widely accepted phenomenological EOS. The agreement between the simulation methods becomes less good when approaching the molecular dissociation regime at slightly lower temperature and pressure. Both CEIMC and FPMD with different approximated functionals has been applied to investigate the Liquid-Liquid phase transition (LLPT) region in hydrogen [41,55,71]. The emerging picture is that a weak first-order phase transition occurs in hydrogen between a molecular-insulating fluid and a metallic-mostly monoatomic fluid. At higher temperature, molecular dissociation and metallization occur continuously. However the precise location of the transition line and the critical point are still matter of debate since several levels of the theory provide different locations. Within FPMD-DFT the location of the transition line depends strongly on the exchange-correlation functional employed and on whether classical or quantum protons are considered [71]. Transition lines from the PBE and vdW-DF2 approximations differ by roughly 200-250GPa, the PBE one being located at lower pressure. The PBE melting line with quantum protons is not in agreement with experiments, which highlights the failure of the PBE approximation when employed together with the quantum description of the nuclei. On the other hand, optical properties for the vdW-DF2 approximation are in agreement with experiments supporting the use of this functional for hydrogen in the WDM regime. The LLPT line from CEIMC lies in between the lines from PBE and vdW-DF2 functionals [41,55]. However, those results were plagued by a truncation error in the calculations of the single electron orbitals which showed up only around the metallization and which resulted in biased estimates. We have now changed the DFT solver in our CEIMC code and checked the convergence. We find a roughly uniform shift of the transition line by ~ 50 GPa to higher pressure and we are performing new calculations with quantum nuclei. Preliminary results, based on VMC electronic energies[5], suggests that, similarly to the DFT scenario, nuclear quantum effects favor molecular dissociation and become increasingly important at lower temperatures. We estimate that the transition pressure is decreased, because of nuclear quantum effects, by $\sim$60GPa at 600K and by $\sim$150GPa at 300K (from 430GPa for classical nuclei to 290GPa for quantum nuclei). The last estimate however is for a metastable liquid state obtained by an instantaneous quenching of the fluid at higher temperature, while it is expected that the equilibrium state at 300K and $\sim$290GPa be crystalline (of unknown structure) [10]. In figure 3 we report CEIMC proton-proton g(r) at various densities along the T=600K isotherm to illustrate the relevance of nuclear quantum effects on the pressure dissociation. The preliminary CEIMC results suggest that, despite the good performance observed on band gap calculation in the crystalline phases [72], the vdW-DF2 exchange-correlation functional has a tendency to over-stabilize molecules.

[5] RMC corrections to the transition lines was previously found to be small and we expect an even smaller effect with the new CEIMC implementation since the VMC variance is roughly half of what it was in the previous code[41]

Figure 3. Proton-proton radial distribution function at various density along the isotherm T=600K. Comparison between classical nuclei (red continuous line) and quantum nuclei (green dashed line) for hydrogen nuclear mass. It is evident the molecular dissociation with increasing density.



Although our results demonstrate the power of CEIMC in predicting the physical properties of hydrogen, its use is still quite demanding in terms of computer time, a fact that limits its applicability. This is particularly true when a much larger exploration of external conditions is needed to clarify the physics. For example, to study the crystalline state of the molecular system and clarify the molecular-atomic transition mechanism in the solid state, it is necessary to consider a large number of candidate structures, some of which have very large unit cells (the recently proposed Pc structure for phase IV of molecular hydrogen contains 192 proton, double what was considered in the LLPT). Moreover, in studying those structure at finite temperature it is important to apply a constant stress algorithm allowing the simulation box to deform and releasing the excess internal stress that otherwise would produce metastable states. While larger systems ($>$ 250 particles) and constant pressure algorithms are routinely used in FP methods based on DFT, their use in conjunction with CEIMC is still problematic. Therefore, it is important to apply CEIMC and other QMC methods to validate DFT predictions and determine the most accurate functional for a given system. The same considerations apply to systems more complex than hydrogen. In the next section we will describe our effort to benchmark functionals for high pressure hydrogen and for water in condensed phase.

3.2. QMC Benchmarks of DFT

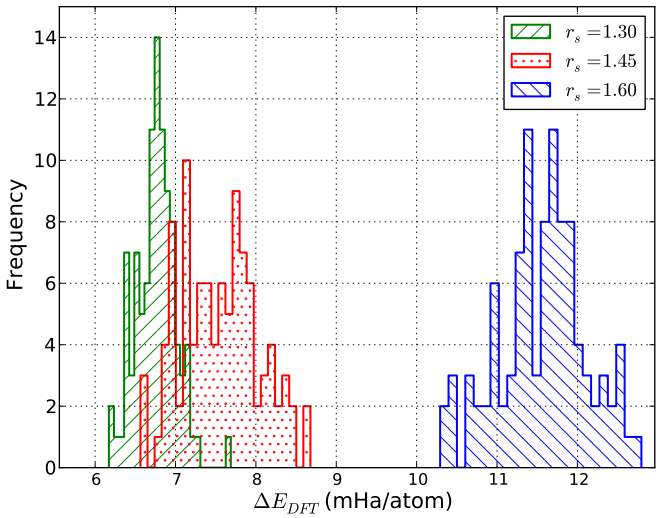
Within the Born-Oppenheimer approximation at low temperatures, the only interaction between ions and electrons comes through the potential energy surface $E_0(\mathbf{R})$, defined as the solution of the electronic hamiltonian for a fixed set of ionic coordinates. $E_0(\mathbf{R})$ is typically approximated by $E_{DFT}(\mathbf{R})$ in first-principles calculations, and obtained from a density functional theory (DFT) calculation. Over the last several years, many-body methods for solids have been developed to the point that the prospect of developing density functionals from accurate reference calculations is now a possibility. In this section, we show how quantum Monte Carlo calculations can be used to benchmark the accuracy of DFT in the description of the potential energy surface. The quality of $E_{DFT}(\mathbf{R})$ defines the predictive capabilities of the resulting first-principles simulation. We use large sets of representative configuration from PIMD simulations, and compare the mean absolute error between accurate QMC calculations and various DFT functionals. We present preliminary calculations on high pressure hydrogen and liquid water at ambient conditions, two materials that are particularly challenging to DFT due to the subtle competition between dispersion interactions, nuclear quantum effects, hydrogen bonding, and anisotropic interactions.

3.2.1. Hydrogen

The phase diagram of hydrogen at high pressure has been extensively explored using first-principles simulations with DFT [54,55,73–76]. In spite of the large number of studies performed on , most of the work so far has employed either the local density (LDA) [77] approximation to the exchange-correlation potential or the Perdew-Burke-Ehrzenhof (PBE) [78] generalized gradient approximation. These are two of the simplest functionals currently available in DFT. In fact, both of them suffer from self-interaction errors and lack a proper treatment of dispersion interactions, making their application in the regime of molecular dissociation questionable. Recently, the use of DFT functionals with an improved description of dispersion interactions has been employed in the study of the liquid and solid molecular phases in the neighborhood of molecular dissociation. It was found that the dissociation density changed when compared to calculations using PBE [71,72]. Since these functionals were not designed for materials at high density, and because dispersion interactions are clearly important in dense molecular hydrogren, there is a crucial need for accurate calculations that can be used to benchmark the different exchange-correlation functionals employed in first-principles simulations.

Since sufficient experimental data does not exist to validate the quality of functionals in the high-pressure high-temperature regime of the phase diagram, we used fixed-node diffusion Monte Carlo (DMC) to benchmark the accuracy of several DFT functionals over a range of densities near the liquid-liquid phase transition (LLPT) at a temperature of T = 1000 K. Henceforth, we will refer to densities using the parameter $r_s = \left(\frac{3}{4\pi}\frac{V}{N}\right)^{1/3}$, where N is the number of hydrogen atoms. First, we ran PIMD simulations with the PBE functional for $N = 54$ hydrogen atoms at three densities: $r_s = 1.30, 1.45, 1.60$. In this range of densities, the liquid goes from an insulating molecular state at $r_s = 1.60$ to a conducting atomic liquid at $r_s = 1.30$. The density $r_s = 1.45$ is intemediate and close to the LLPT for this functional. After equilibration, we sampled 100 ionic configurations from uncorrelated PIMD time slices for each density. For each configuration at each density, we calculated

Figure 4. Histograms of ΔE_{DFT} for the PBE functional for dense hydrogen at densities $r_s = 1.30, 1.45, 1.60$ at $T = 1000K$. ΔE_{DFT} refers to the absolute energy difference per hydrogen atom between the DFT and QMC for a given configuration. There were 54 atoms per configuration.

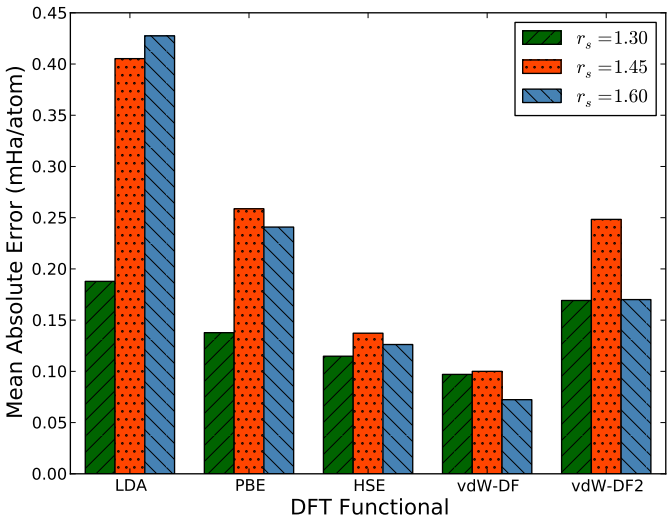


the DMC energy, and then computed $E_{DFT}(\mathbf{R})$ for the following functionals: LDA, PBE, vdW-DF [46], vdW-DF2 [47,48,79], and HSE [45].

All QMC calculations were performed with the QMCPACK [80–82] software package. We used a Slater-Jastrow trial wavefunction with twist-averaged boundary conditions [83], employing a 3x3x3 grid of boundary conditions. For the Jastrow functions, we used real space b-splines with optimizable knots. We included spin-independent one-body proton-electron terms; a short-ranged term with the appropriate cusp condition, and a long-ranged term. We also included two long-ranged spin-dependent electron-electron functions with appropriate cusp conditions. For each configuration, linear optimization with VMC was performed for all Jastrow parameters at a single twist-angle, these parameters were subsequently used for all twists in the DMC calculations. For the DMC run, a timestep of $\tau = 0.05 Ha^{-1}$ and 6000 walkers were used. The orbitals were obtained from DFT using the Quantum Espresso software package [44], using the PBE functional. We used a plane wave cutoff of 210 Ry. DFT calculations were performed with a Troullier-Martins norm conserving pseudo-potential [84] with a cutoff radius of $r_c = 0.5 a_0$, DMC calculations were performed with the Coulomb potential. Based on the scale of the energy differences, we found a statistical error of 0.02 mHa/particle to be sufficient for present purposes.

An example of the comparison between QMC and DFT is given in Fig. 4. Shown is a histogram of the energy difference between the results of DMC and the PBE functional at the three densities: $\Delta E_{DFT} = E_{DFT} - E_{DMC}$. Given that $r_s = 1.30$ corresponds to the atomic liquid, and $r_s = 1.60$ to the molecular liquid, we immediately see that the errors incurred by using the PBE functional are not consistent across the LLPT. As expected, PBE offers a much better description of the atomic liquid compared to the molecular phase, where self-interaction errors are larger and dispersion interactions are important. This is a well-known failure of most semi-local density functionals, which tend to favor delocalized states.

Figure 5. Mean absolute error of energy/atom vs. functional for dense liquid hydrogen at 1000K. For each functional, we computed the mean absolute error for three different densities, denoted by the different colored bars.



To better quantify and compare the quality of functionals, we have computed the mean absolute error (MAE) from data similar to that shown in Fig. 4. This quantity is defined as $MAE_{func} = \langle|\Delta E_{DFT} - \langle\Delta E_{DFT}\rangle|\rangle$, where the average is taken over all configurations at a particular density. Notice that we subtract the average energy difference in the definition of the MAE, since the zero of energy of each functional is modified by the use of pseudopotentials. Energy differences are more significant since the structure of the liquid is only sensitive to differences. The MAE gives us one measure of the quality, or predictive capability, of a given functional as defined by the reference method, in this case DMC. We have tabulated our results in Fig. 5.

There are several interesting features in Fig. 5 directly related to the expected performance of these functionals in the description of hydrogen near molecular dissociation in the liquid. First, the two semi-local functionals in the comparison, LDA and PBE, have considerably different errors in the molecular and atomic regimes. As described above, the atomic regime is more accurately described in comparison to the molecular phase, leading to a potentially strong underestimation of dissociation transition pressures in both solid and liquid phases. This is consistent with recently reported simulations [71]. On the other hand, both the hybrid HSE and the functionals with improved dispersion vdW-DF and vdW-DF2 offer a more consistent level of description between the two regimes. The mean absolute errors of the HSE and vdW-DF functionals are approximately half that of the PBE functional for all densities, which indicates that these functionals more accurately capture energy *differences* between various liquid configurations.

3.2.2. Liquid Water

Water plays a central role in many scientific fields [85]. It is a critical component to almost all chemical, biological, and geophysical processes. As a result, it is one of the most studied substances in science, both from an experimental and a theoretical point of view. Despite such broad importance,

water's most basic property, its local structure at ambient conditions, characterized by the geometry of its underlying hydrogen-bond (H-bond) network, has remained a matter of debate for over a century [86–88]. Challenges arise because water is only $\approx$ 25 K (at room temperature) from the melting temperature of ice, where a variety of subtle and complex effects become important. While the structure is dominated by H bond between neighboring molecules, both van der Waals (vdW) interactions (which, in this context, refers to dispersion forces resulting from dynamical nonlocal electron correlations) and nuclear quantum effects (NQEs) influence the topology of the H-bond network. In fact, it is precisely these seemingly subtle effects (compared to H bonding) that are key to accurately describing ambient water, but have been (until recently) difficult or impossible to model.

Atomistic simulations have the potential to resolve these issues, particularly using first-principles methods. Providing an accurate theoretical description has been a central topic and open challenge in physical chemistry for many decades. Despite considerable focus over the last decade, to date DFT has proven insufficient for the accurate description of liquid water [4,89]. Nonetheless, much progress has occurred during the last several years. The main advances include the use of functionals that properly describe dispersion interactions in the liquid [46,48,90,91], the use of hybrid functionals [92], and the direct treatment of nuclear quantum effects [93]. The combination of all of these advances in first-principles simulations of liquid water could lead to an accurate description of its interesting properties, including its local structure. At the same time, the choice of exchange-correlation functional in DFT is still a source of complication, mainly due to the large number of possibilities and the inability to test their predictive capabilities without resorting to full first-principles calculations of a large set of observables. As in the case of hydrogen, an accurate first-principles description almost certainly requires the use of path integral methods in order to directly treat nuclear quantum effects, which makes the calculations quite computationally intensive. What is needed is a way to assess the quality of a given functional without having to resort to first-principles calculations of the liquid at the PIMD level, and if possible, a way to systematically improve them using high quality reference calculations from accurate many-body methods.

In this section, we present QMC calculations of configurations of molecules extracted from PIMD simulations of liquid water. QMC has been shown to be a reliable benchmark in the study of small water clusters [94–96], and should provide an accurate reference method to measure the quality of typical density functionals used in simulations of water. All DMC calculations were performed with the QMCPack software package [80–82]. A Troullier-Martins norm-conserving pseudo-potential [84] was used to represent both hydrogen and oxygen. In particular, we used the pseudo-potentials from the CASINO database [97,98], which were recently shown to produce accurate results in the study of small water clusters. A Slater-Jastrow trial wave-function was used. The orbitals in the Slater determinant were obtained from DFT calculations employing the PBE exchange-correlation functional. We do not expect a strong dependence of the resulting comparison on the functional used to generate the orbitals. The Jastrow term contains electron-ion, electron-electron and electron-electron-ion terms, the variational parameters were optimized at the VMC level using a variant of the linear method of Umrigar, *et al.* [99]. A time-step of 0.01 Ha^{-1} was found to be sufficiently small to produce accurate total energies and approximately 4800 walkers were used in the DMC calculations. Casula's T-moves [100] were used to

reduce locality errors, while the Model Coulomb Potential [101] and Chiesa's [102] correction scheme were used to estimate finite-size corrections to the potential and kinetic energies respectively.

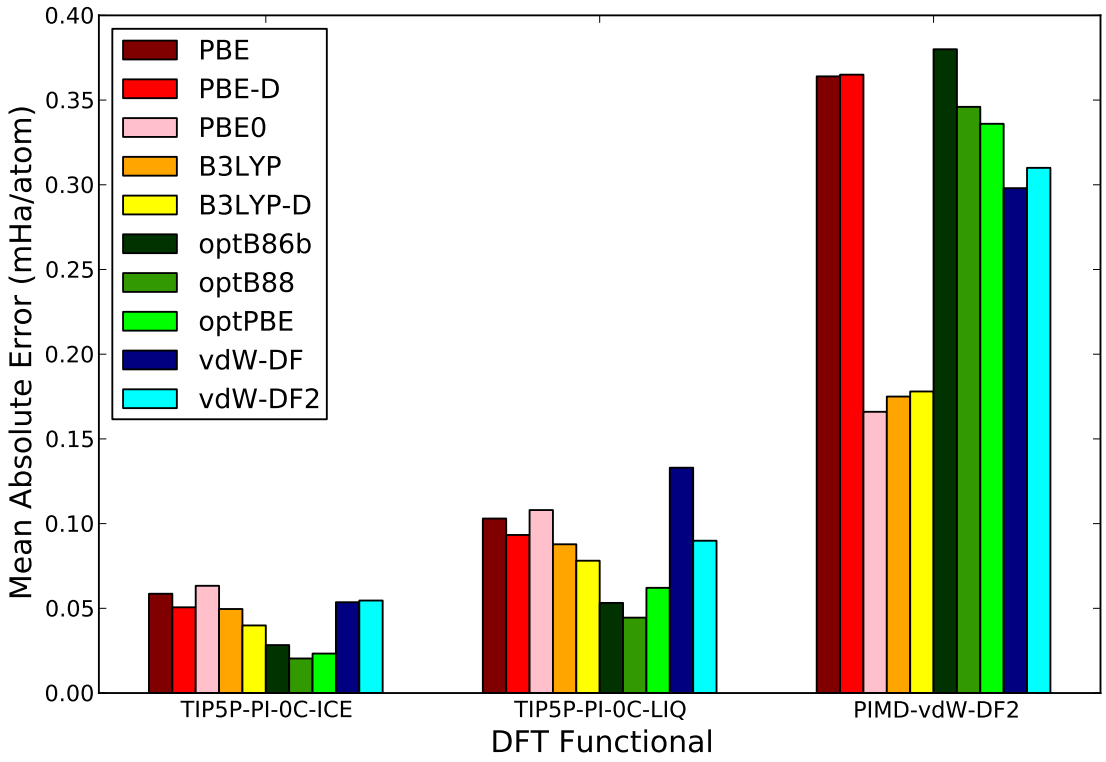
DFT calculations were performed with both Quantum Espresso (QE) [44] and VASP [103–105] simulation packages. In the case of QE calculations we employed norm-conserving Troullier-Martins pseudo-potentials, while in the case of VASP calculations we employed the Projector Augmented Wave method (PAW) [106,107]. A single pseudo-potential (constructed with PBE) was chosen in order to make a homogeneous comparison of all DFT functionals, since some of the functionals employed in this work do not yet allow for the production of pseudo-potentials. All simulations were performed at the Γ point of the supercell in order to be consistent with the corresponding DMC calculations; errors due to the lack of k-point integration were small enough to be safely discarded. We carefully tested the convergence with the plane-wave cutoff in all DFT calculations.

We present calculations for 3 different configuration sets. The first two sets, which we called *TIP5P-PI-0C-ICE* and *TIP5P-PI-0C-LIQ*, were generated with PIMC calculations on simulation cells using the TIP5P water model and 32 molecules [108]. As the name suggests, the PIMC calculations used to generate the configuration set were performed at T = 0 C, from stable solid and liquid phases. The third configuration set was obtained from PIMD calculations of 64 water molecules, at room temperature and a density of 1 g/cm^3, with the vdW-DF2 functional, which has been recently shown to provide an accurate description of the structure of water when combined with a path integral representation [10]. The number of configurations in each set is 20, 47, 50, respectively. The three configuration sets sample different aspects of the potential energy surface of liquid water. While TIP5P is a rigid molecule model, the first-principles simulations with vdW-DF2 are fully flexible, which allows us to emphasize different ranges of the molecular interactions in the liquid. On the other hand, the simulations with TIP5P in both liquid and solid phases at T = 0C sample the configurations that either strongly favor hydrogen bonding in the solid, with those where the hydrogen-bond network has been destabilized in the liquid.

Figure 6 shows the mean absolute difference in the total energy between DMC and DFT calculations, results are separated by configuration sets in order to allow for a more clear comparison between them. Several functionals are considered including the semi-local functionals: PBE [78]; the hybrid functionals: PBE0 [45], B3LYP [109,110]; the non-local van der Waals functionals: optB88 [111], optPBE [111], optB86b [112], vdW-DF [46] and vdW-DF2 [48]; and finally functionals with the empirical van der Waals correction of Grimme, *et al.*, (DFT-D2) [113]. While there are many interesting results in this comparison, the most noticeable feature is the large difference in the scale of the differences between rigid and flexible molecule configurations. This is not unexpected since the larger energy fluctuations in the system are found coupled to the intramolecular degrees of freedom of the molecule. In the case of flexible molecule configurations, hybrid functionals offer a much better agreement with DMC results, producing errors typically a factor if 2 smaller than non-hybrid functionals. This results shows the fact that hybrid functionals do a much better job at describing the intramolecular potential energy surface. This is consistent with the recent calculations of Alfe, *et al.* [95] and with the recent calculations of the absorption spectra of bulk water at ambient conditions of Zhang, *et al.* [92]. On the other hand, the functionals that include an appropriate description of dispersion interactions offer a clearly better comparison with QMC in the rigid-molecule configuration sets. In this case, the intermolecular interactions are the dominant energy contribution and the lack of appropriate dispersion leads to a larger

error. In this case, we can also see a small but finite improvement with the inclusion of empirically corrected vdW functionals (PBE-D, B3LYP-D), but the gain is small and can not compete with non-local vdW functionals. Notice also that the performance of hybrids in the fixed-molecule sets is comparable to the performance of semi-local functionals, due to the fact that neither of these type of functionals can properly describe dispersion interactions. Finally, the configuration set with the smallest overall MAE is the one obtained from the calculations in the solid phase close to melting, showing the fact that most of these functionals can describe hydrogen bonded configurations fairly well.

Figure 6. Mean absolute difference in the total energy between DMC and DFT with various exchange correlation functionals for a supercell containing water molecules. Results presented correspond to calculations using the PAW formulation with VASP. X-D, where X represents a given density functional, designates results using the empirical dispersion corrections of Grimme *et al.*, [113], in particular the DFT-D2 correction scheme as implemented in VASP. Statistical errors on the presented results are on the order of 0.003 mHa and 0.005 mHa for rigid and flexible molecule configurations respectively. They are not shown on the figure for clarity.



4. Discussion

Direct first-principles simulations with QMC accuracy of condensed phases systems are nowadays possible but restricted so far to the simplest first few elements of the periodic table, namely hydrogen, helium and their mixtures. Even for those simple systems challenges are present and the computational demand is large. Nonetheless CEIMC predictions for the liquid-liquid phase transition in hydrogen remains today the target for less accurate but faster DFT-based FP methods. While much work remains to be done in developing QMC-based FP methods, the calculations presented here show one possible use of accurate many-body calculations, using QMC to benchmark the accuracy of DFT functionals. Not only does this allows us to make a judgment of the quality of a functional before its use in first-principles simulations, but it also shows us a path for the systematic improvement of the functionals by adjusting free parameters to minimize the MAE. DFT users will often point to experimental data to validate the quality of a chosen functional. What we have shown is that we can use highly-accurate QMC methods to benchmark functionals around the LLPT of hydrogen, from first-principles. In addition, this set of reference energies for the bulk system can be used to optimize the free parameters in the DFT functional to minimize the MAE, and in the limit of a large data set, reproduce the quality of the more accurate many-body method in first-principles calculations using DFT. This approach will be increasingly necessary as we continue to explore matter under extreme pressures, since experimental data is often insufficient or nonexistent at geophysical/planetary scales. It will also be necessary for other situations where DFT functionals have difficult, such as near metal-insulator transitions.

Let us consider a more general point. We suggest that in general it is superior to use total energies to find an interatomic potential (force field) The traditional approach is to fit experimental data, for example the melting temperature of ice, the density of water versus temperature etc.. Clearly it was necessary to do this in the past since experimental data was all that was available. However, using this approach requires very extensive calculations including free energy or equivalent computations and ultimately only gives a single constraint. We can invoke ”The Allegory of the Cave” from Plato’s *The Republic*. We should not look to fit the atomic potentials using the projections of the energy surface onto thermodynamic properties but to fit directly the energy surface. Thus we will obtain an interatomic potential good for all properties. The situation has changed with mature QMC methods and much more computational power available. (Note that scanning a PES is a task very well suited to massively parallel computers.) Including total energy QMC benchmarks into the fitting procedure in addition to experimental data, can allow for much more systematic improvements. QMC thus can provide a unique role in giving total energies and applicable to large enough systems to approximate condensed matter.

Water and hydrogen show an additional complication of using experimental data: namely because of the importance of quantum zero-point effects of the protons, fitting of the experimental data becomes particularly problematical. A common approach is to do a simulation of the classical system and assume the effective classical system includes effects of ZPE (clearly this then becomes quite approximate). A complication, is that the interatomic potential that results can become temperature and density dependent with all known pathologies related to the use of state dependent potentials [114]. Or one has to do full PIMD simulations of the system in order to determine the best empirical potential, thus increasing the, already large, computational requirements considerably.

One aspect in determining good force fields is to find an appropriate basis set to parameterize the force field. Traditionally, these have contained few functions with very few parameters, e.g. the Lennard Jones potential with ϵ and σ. It is feasible today to calculate millions of points on a PES each of which will be the energy and $3N$ components of the derivative with respect to the ionic coordinates, the forces. Using QMC techniques, each would come with an error estimate. Hence, we can envision fitting this data set to a force field with potential millions of independent parameters. It is now feasible to fit a completely general pair potential (say with a spline basis), three body potential etc. The investigation into effective basis sets is now very important.

We can thus imagine an integrated set of tools: QMC simulations of systems with thousands of electrons produce data sets of energies and forces. These can be used either to tailor a DFT to a particular system, or to determine a force field. The DFT simulations and the force field simulations can then be used to model much larger systems. Thus simulations can thereby become much more predictive, and not just produce universal properties but details important to applications and experiment.

Acknowledgements

M. A. M. was supported by the U.S. Department of Energy at the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and by LDRD Grant No. 13-LW-004. C. P. was supported by the Italian Institute of Technology (IIT) under the SEED project grant number 259 SIMBEDD Advanced Computational Methods for Biophysics, Drug Design and Energy Research.

References

1. Ballone, P. Modelling potential energy surfaces: from first-principle approaches to empirical force fields. *Entropy* **2013**.
2. Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
3. Cohen, A.J.; Mori-Sánchez, P.; Yang, W. Challenges for density functional theory. *Chemical Reviews* **2012**, *112*, 289–320.
4. Schwegler, E.; Grossman, J.C.; Gygi, F.; Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II. *The Journal of chemical physics* **2004**, *121*, 5400–9.
5. Ceperley, D.M.; Alder, B.J. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **1980**, *45*, 566–569.
6. Brown, E.W.; Clark, B.K.; DuBois, J.L.; Ceperley, D.M. Path-Integral Monte Carlo Simulation of the Warm Dense Homogeneous Electron Gas. *Physical Review Letters* **2013**, *110*, 146405.
7. Ercolessi, F.; Adams, J.B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhysics Letters (EPL)* **1994**, *26*, 583–588.
8. Pierleoni, C.; Ceperley, D.M. The Coupled Electron-Ion Monte Carlo Method. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1 SE - 18*; Ferrario, M.; Ciccotti, G.; Binder, K., Eds.; Springer Berlin Heidelberg, 2006; Vol. 703, *Lecture Notes in Physics*, pp. 641–683.
9. Attaccalite, C.; Sorella, S. Stable Liquid Hydrogen at High Pressure by a Novel AbInitio Molecular-Dynamics Calculation. *Physical Review Letters* **2008**, *100*, 114501.
10. McMahon, J.; Morales, M.A.; Pierleoni, C.; Ceperley, D.M. The properties of hydrogen and helium under extreme conditions. *Reviews of Modern Physics* **2012**, *84*, 1607–1653.
11. W.L.McMillan. Ground State of Liquid He4. *Physical Review* **1965**, *138*.
12. Hammond, B.L.; Lester, W.A.; Reynolds, P.J. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; World Scientific Publishing Co. Pte. Ltd.: Singapore, 1994; p. 304.
13. Ceperley, D.; Chester, G.V. Monte Carlo simulation of a many-fermion study. *Physical Review B* **1977**, *16*, 3081–3099.
14. Morales, M.A.; McMinis, J.; Clark, B.K.; Kim, J.; Scuseria, G.E. Multideterminant Wave Functions in Quantum Monte Carlo. *Journal of Chemical Theory and Computation* **2012**, *8*, 2181–2188.
15. Casula, M.; Attaccalite, C.; Sorella, S. Correlated geminal wave function for molecules: an efficient resonating valence bond approach. *The Journal of chemical physics* **2004**, *121*, 7110–26.

16. Neuscamman, E.; Umrigar, C.; Chan, G. Optimizing large parameter sets in variational quantum Monte Carlo. *Physical Review B* **2012**, *85*, 1–6.
17. Clark, B.K.; Morales, M.A.; McMinis, J.B.; Kim, J.; Scuseria, G.E. Computing the energy of a water molecule using multideterminants: a simple, efficient algorithm. *The Journal of chemical physics* **2011**, *135*, 244105.
18. Grimm, R.; Storer, R. Monte-Carlo solution of Schrödinger's equation. *Journal of Computational Physics* **1971**, *7*, 134–156.
19. Ceperley, DM and Kalos, M. Monte Carlo methods in statistical physics. *Topics in current physics* **1979**, *7*, 145.
20. Reynolds, P.; Ceperley, D. Fixed-node quantum Monte Carlo for molecules. *The Journal of Chemical ...* **1982**, *77*, 5593–5603.
21. Anderson, J.B. A random-walk simulation of the Schrodinger equation: H+3. *The Journal of Chemical Physics* **1975**, *63*, 1499.
22. Caffarel, M.; Claverie, P. Development of a pure diffusion quantum Monte Carlo method using a full generalized FeynmanKac formula. I. Formalism. *The Journal of Chemical Physics* **1988**, *88*, 1088.
23. Assaraf, R.; Caffarel, M.; Khelif, A. Diffusion Monte Carlo methods with a fixed number of walkers. *Physical Review E* **2000**, *61*, 4566–4575.
24. Reynolds, P.J.; Barnett, R.N.; Hammond, B.L.; Lester, W.a. Molecular physics and chemistry applications of quantum Monte Carlo. *Journal of Statistical Physics* **1986**, *43*, 1017–1026.
25. Pierleoni, C.; Ceperley, D.M. Computational methods in coupled electron-ion Monte Carlo simulations. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2005**, *6*, 1872–8.
26. Baroni, S.; Moroni, S. Reptation Quantum Monte Carlo: A Method for Unbiased Ground-State Averages and Imaginary-Time Correlations. *Physical Review Letters* **1999**, *82*, 4745–4748.
27. Ceperley, D.M. Path integrals in the theory of condensed helium. *Reviews of Modern Physics* **1995**, *67*, 279–355.
28. Ceperley, D.M. Path Integral Monte Carlo Methods for Fermions. In *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*; Binder, K.; Ciccotti, G., Eds.; Editrice Compositori: Bologna, 1996.
29. Ceperley, D.M. Fermion nodes. *Journal of Statistical Physics* **1991**, *63*, 1237–1267.
30. Ceriotti, M.; Manolopoulos, D.E.; Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *The Journal of Chemical Physics* **2011**, *134*, 084104.
31. Ceperley, D.M.; Alder, B.J. Ground state of solid hydrogen at high pressures. *Physical Review B* **1987**, *36*, 2092–2106.
32. Natoli, V.; Martin, R.M.; Ceperley, D.M. Crystal structure of atomic hydrogen. *Physical Review Letters* **1993**, *70*, 1952–1955.
33. Natoli, V.; Martin, R.M.; Ceperley, D.M. Crystal Structure of Molecular Hydrogen at High Pressure. *Physical Review Letters* **1995**, *74*, 1601–1604.

34. Pierleoni, C.; Ceperley, D.M.; Bernu, B.; Magro, W.R. Equation of State of the Hydrogen Plasma by Path Integral Monte Carlo Simulation. *Physical Review Letters* **1994**, *73*, 2145–2149.
35. Magro, W.R.; Ceperley, D.M.; Pierleoni, C.; Bernu, B. Molecular Dissociation in Hot, Dense Hydrogen. *Physical Review Letters* **1996**, *76*, 1240–1243.
36. Militzer, B.; Kress, J.D.; Johnson, J.D.; Collins, L.A.; Mazevet, S. Calculation of a Deuterium Double Shock Hugoniot from Ab Initio Simulations. *Physical Review Letters* **2001**, *87*, 275502.
37. Ceperley, D.M.; Dewing, M. The penalty method for random walks with uncertain energies. *The Journal of Chemical Physics* **1999**, *110*, 9812.
38. Ceperley, D.M.; Dewing, M.; Pierleoni, C. The Coupled Electronic-Ionic Monte Carlo Simulation Method. In *Bridging Time Scales: Molecular Simulations for the Next Decade SE - 17*; Nielaba, P.; Mareschal, M.; Ciccotti, G., Eds.; Springer Berlin Heidelberg, 2002; Vol. 605, *Lecture Notes in Physics*, pp. 473–500.
39. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Oxford Science Publ, Oxford University Press, 1989.
40. Tuckerman, M. *Statistical Mechanics and Molecular Simulations*; Oxford Graduate Texts, Oxford University Press, UK, 2008.
41. Liberatore, E.; Morales, M.A.; Ceperley, D.M.; Pierleoni, C. Free energy methods in coupled electron ion Monte Carlo. *Molecular Physics* **2011**, *109*, 3029–3036.
42. Holzmann, M.; Ceperley, D.M.; Pierleoni, C.; Esler, K. Backflow correlations for the electron gas and metallic hydrogen. *Physical Review E* **2003**, *68*, 046707.
43. Pierleoni, C.; Delaney, K.T.; Morales, M.A.; Ceperley, D.M.; Holzmann, M. Trial wave functions for high-pressure metallic hydrogen. *Computer Physics Communications* **2008**, *179*, 89–97.
44. Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G.L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A.P.; Smogunov, A.; Umari, P.; Wentzcovitch, R.M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics. Condensed matter : an Institute of Physics journal* **2009**, *21*, 395502.
45. Heyd, J.; Scuseria, G.E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics* **2003**, *118*, 8207.
46. Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D.C.; Lundqvist, B.I. Van der Waals Density Functional for General Geometries. *Physical Review Letters* **2004**, *92*, 246401.
47. Román-Pérez, G.; Soler, J. Efficient Implementation of a van der Waals Density Functional: Application to Double-Wall Carbon Nanotubes. *Physical Review Letters* **2009**, *103*, 096102.
48. Lee, K.; Murray, E.D.; Kong, L.; Lundqvist, B.I.; Langreth, D.C. Higher-accuracy van der Waals density functional. *Physical Review B* **2010**, *82*, 081101.
49. Pickard, C.J.; Needs, R.J. Structure of phase III of solid hydrogen. *Nature Physics* **2007**, *3*, 473–476.

50. Silvera, I. The solid molecular hydrogens in the condensed phase: Fundamentals and static properties. *Reviews of Modern Physics* **1980**, *52*, 393–452.
51. Mao, H.k.; Hemley, R. Ultrahigh-pressure transitions in solid hydrogen. *Reviews of Modern Physics* **1994**, *66*, 671–692.
52. Deemyad, S.; Silvera, I. Melting Line of Hydrogen at High Pressures. *Physical Review Letters* **2008**, *100*, 155701.
53. Bonev, S.A.; Schwegler, E.; Ogitsu, T.; Galli, G. A quantum fluid of metallic hydrogen suggested by first-principles calculations. *Nature* **2004**, *431*, 669–72.
54. Scandolo, S. Liquid-liquid phase transition in compressed hydrogen from first-principles simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 3051–3.
55. Morales, M.A.; Pierleoni, C.; Schwegler, E.; Ceperley, D.M. Evidence for a first-order liquid-liquid transition in high-pressure hydrogen from ab initio simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 12799–803.
56. Pierleoni, C.; Magro, W.R.; Ceperley, D.M.; Berne, B.J. Path Integral Monte Carlo Simulation of Hydrogen Plasma.pdf. Proceedings of the (Binz Germany) International Conference on the Physics of Strongly Coupled Plasmas; Kraeft, W.D.; Schlanges, M., Eds. World Scientific Publishing Co. Pte. Ltd., 1996.
57. Militzer, B.; Ceperley, D.M. Path Integral Monte Carlo Calculation of the Deuterium Hugoniot. *Physical Review Letters* **2000**, *85*, 1890–1893.
58. Da Silva, L.B.; Celliers, P.; Collins, G.W.; Budil, K.S.; Holmes, N.C.; Barbee Jr., T.W.; Hammel, B.A.; Kilkenny, J.D.; Wallace, R.J.; Ross, M.; Cauble, R. Absolute Equation of State Measurements on Shocked Liquid Deuterium up to 200 GPa (2 Mbar). *Physical Review Letters* **1997**, *78*, 483–486.
59. Collins, G.W. Measurements of the Equation of State of Deuterium at the Fluid Insulator-Metal Transition. *Science* **1998**, *281*, 1178–1181.
60. Celliers, P.M.; Collins, G.W.; Da Silva, L.; Gold, D.M.; Cauble, R.; Wallace, R.J.; Foord, M.E.; Hammel, B.A. Shock-Induced Transformation of Liquid Deuterium into a Metallic Fluid. *Physical Review Letters* **2000**, *84*, 5564–5567.
61. Knudson, M.D.; Hanson, D.L.; Bailey, J.E.; Hall, C.A.; Asay, J.R. Equation of State Measurements in Liquid Deuterium to 70 GPa. *Physical Review Letters* **2001**, *87*, 225501.
62. Knudson, M.; Hanson, D.; Bailey, J.; Hall, C.; Asay, J. Use of a Wave Reverberation Technique to Infer the Density Compression of Shocked Liquid Deuterium to 75 GPa. *Physical Review Letters* **2003**, *90*, 035505.
63. Knudson, M.D.; Hanson, D.L.; Bailey, J.E.; Hall, C.A.; Asay, J.R.; Deeney, C. Principal Hugoniot, reverberating wave, and mechanical reshock measurements of liquid deuterium to 400 GPa using plate impact techniques. *Physical Review B* **2004**, *69*, 144209.
64. Bailey, J.E.; Knudson, M.D.; Carlson, A.L.; Dunham, G.S.; Desjarlais, M.P.; Hanson, D.L.; Asay, J.R. Time-resolved optical spectroscopy measurements of shocked liquid deuterium. *Physical Review B* **2008**, *78*, 144107.

65. Hicks, D.G.; Boehly, T.R.; Celliers, P.M.; Eggert, J.H.; Moon, S.J.; Meyerhofer, D.D.; Collins, G.W. Laser-driven single shock compression of fluid deuterium from 45 to 220 GPa. *Physical Review B* **2009**, *79*, 014112.
66. Knudson, M.D.; Desjarlais, M.P. Shock Compression of Quartz to 1.6 TPa: Redefining a Pressure Standard. *Physical Review Letters* **2009**, *103*, 225501.
67. Boriskov, G.; Bykov, a.; Ilkaev, R.; Selemir, V.; Simakov, G.; Trunin, R.; Urlin, V.; Shuikin, a.; Nellis, W. Shock compression of liquid deuterium up to 109GPa. *Physical Review B* **2005**, *71*, 092104.
68. Grishechkin, S.K.; Gruzdev, S.K.; Gryaznov, V.K.; Zhernokletov, M.V.; Ilkaev, R.I.; Iosilevskii, I.L.; Kashintseva, G.N.; Kirshanov, S.I.; Manachkin, S.F.; Mintsev, V.B.; Mikhailov, A.L.; Mezhevov, A.B.; Mochalov, M.A.; Fortov, V.E.; Khrustalev, V.V.; Shuikin, A.N.; Yukhimchuk, A.A. Experimental measurements of the compressibility, temperature, and light absorption in dense shock-compressed gaseous deuterium. *Journal of Experimental and Theoretical Physics Letters* **2004**, *80*, 398–404.
69. Hu, S.; Militzer, B.; Goncharov, V.N.; Skupsky, S. FPEOS: A First-Principles Equation of State Table of Deuterium for Inertial Confinement Fusion Applications. *Physical Review B* **2011**, *84*, 224109.
70. Pierleoni, C.; Ceperley, D.M.; Holzmann, M. Coupled Electron-Ion Monte Carlo Calculations of Dense Metallic Hydrogen. *Physical Review Letters* **2004**, *93*, 146402.
71. Morales, M.A.; McMahon, J.M.; Pierleoni, C.; Ceperley, D.M. Nuclear Quantum Effects and Nonlocal Exchange-Correlation Functionals Applied to Liquid Hydrogen at High Pressure. *Physical Review Letters* **2013**, *110*, 065702.
72. Morales, M.A.; McMahon, J.M.; Pierleoni, C.; Ceperley, D.M. Towards a predictive first-principles description of solid molecular hydrogen with density functional theory. *Physical Review B* **2013**, *87*, 184107.
73. Desjarlais, M.P. Density-functional calculations of the liquid deuterium Hugoniot, reshock, and reverberation timing. *Physical Review B* **2003**, *68*, 064204.
74. Vorberger, J.; Tamblyn, I.; Militzer, B.; Bonev, S. Hydrogen-helium mixtures in the interiors of giant planets. *Physical Review B* **2007**, *75*, 024206.
75. Morales, M.A.; Schwegler, E.; Ceperley, D.; Pierleoni, C.; Hamel, S.; Caspersen, K. Phase separation in hydrogen-helium mixtures at Mbar pressures. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 1324–9.
76. McMahon, J.M.; Ceperley, D.M. Ground-State Structures of Atomic Metallic Hydrogen. *Physical Review Letters* **2011**, *106*, 165302.
77. Perdew, J.P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B* **1981**, *23*, 5048–5079.
78. Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical review letters* **1996**, *77*, 3865–3868.
79. Thonhauser, T.; Cooper, V.R.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D.C. Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Physical Review B* **2007**, *76*, 125112.

80. Kim, J.; Esler, K.P.; McMinis, J.; Morales, M.A.; Clark, B.K.; Shulenburger, L.; Ceperley, D.M. Hybrid algorithms in quantum Monte Carlo. *Journal of Physics: Conference Series* **2012**, *402*, 012008.
81. Esler, K.; Kim, J.; Ceperley, D.; Shulenburger, L. Accelerating Quantum Monte Carlo Simulations of Real Materials on GPU Clusters. *Computing in Science & Engineering* **2012**, *14*, 40–51.
82. Esler, K.; Kim, J.; McMinis, J. QMCPACK.
83. Lin, C.; Zong, F.; Ceperley, D. Twist-averaged boundary conditions in continuum quantum Monte Carlo algorithms. *Physical Review E* **2001**, *64*, 016702.
84. Troullier, N.; Martins, J.L. Efficient pseudopotentials for plane-wave calculations. *Physical Review B* **1991**, *43*, 1993–2006.
85. Franks, F. *Water: A Matrix of Life*, 2 ed.; Royal Society of Chemistry paperbacks, Royal Society of Chemistry: Cambridge, 2000.
86. Ball, P. Water: water–an enduring mystery. *Nature* **2008**, *452*, 291–2.
87. Clark, G.N.; Cappa, C.D.; Smith, J.D.; Saykally, R.J.; Head-Gordon, T. The structure of ambient water. *Molecular Physics* **2010**, *108*, 1415–1433.
88. Nilsson, A.; Pettersson, L. Perspective on the structure of liquid water. *Chemical Physics* **2011**, *389*, 1–34.
89. Grossman, J.C.; Schwegler, E.; Draeger, E.W.; Gygi, F.; Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *The Journal of chemical physics* **2004**, *120*, 300–11.
90. Zhang, C.; Wu, J.; Galli, G.; Gygi, F. Structural and Vibrational Properties of Liquid Water from van der Waals Density Functionals. *Journal of Chemical Theory and Computation* **2011**, *7*, 3054–3061.
91. Mø gelhøj, A.; Kelkkanen, A.K.; Wikfeldt, K.T.; Schiø tz, J.; Mortensen, J.J.r.; Pettersson, L.G.M.; Lundqvist, B.I.; Jacobsen, K.W.; Nilsson, A.; Nø rskov, J.K. Ab initio van der waals interactions in simulations of water alter structure from mainly tetrahedral to high-density-like. *The journal of physical chemistry. B* **2011**, *115*, 14149–60.
92. Zhang, C.; Donadio, D.; Gygi, F.; Galli, G. First Principles Simulations of the Infrared Spectrum of Liquid Water Using Hybrid Density Functionals. *Journal of Chemical Theory and Computation* **2011**, *7*, 1443–1449.
93. Morrone, J.; Car, R. Nuclear Quantum Effects in Water. *Physical Review Letters* **2008**, *101*, 017801.
94. Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. *The Journal of chemical physics* **2008**, *129*, 194111.
95. Gillan, M.J.; Manby, F.R.; Towler, M.D.; Alfè, D. Assessing the accuracy of quantum Monte Carlo and density functional theory for energetics of small water clusters. *The Journal of chemical physics* **2012**, *136*, 244105.

96. Alfe, D.; Bartok, a.P.; Csanyi, G.; Gillan, M.J. Communication: Energy benchmarking with quantum Monte Carlo for water nano-droplets and bulk liquid water. *The Journal of Chemical Physics* **2013**, *138*, 221102.
97. Trail, J.R.; Needs, R.J. Smooth relativistic Hartree-Fock pseudopotentials for H to Ba and Lu to Hg. *The Journal of chemical physics* **2005**, *122*, 174109.
98. Trail, J.R.; Needs, R.J. Norm-conserving Hartree-Fock pseudopotentials and their asymptotic behavior. *The Journal of chemical physics* **2005**, *122*, 14112.
99. Umrigar, C.J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R.G. Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions. *Physical Review Letters* **2007**, *98*, 110201.
100. Casula, M. Beyond the locality approximation in the standard diffusion Monte Carlo method. *Physical Review B* **2006**, *74*, 161102.
101. Fraser, L.; Foulkes, W.; Rajagopal, G.; Needs, R.; Kenny, S.; Williamson, A. Finite-size effects and Coulomb interactions in quantum Monte Carlo calculations for homogeneous systems with periodic boundary conditions. *Physical Review B* **1996**, *53*, 1814–1832.
102. Chiesa, S.; Ceperley, D.; Martin, R.; Holzmann, M. Finite-Size Error in Many-Body Simulations with Long-Range Interactions. *Physical Review Letters* **2006**, *97*, 6–9.
103. Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metalamorphous-semiconductor transition in germanium. *Physical Review B* **1994**, *49*, 14251–14269.
104. Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **1993**, *47*, 558–561.
105. Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **1996**, *6*, 15–50.
106. Blöchl, P.E. Projector augmented-wave method. *Physical Review B* **1994**, *50*, 17953–17979.
107. Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **1999**, *59*, 1758–1775.
108. Mahoney, M.W.; Jorgensen, W.L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **2000**, *112*, 8910.
109. Kim, K.; Jordan, K.D. Comparison of Density Functional and MP2 Calculations on the Water Monomer and Dimer. *The Journal of Physical Chemistry* **1994**, *98*, 10089–10094.
110. Stephens, P.J.; Devlin, F.J.; Chabalowski, C.F.; Frisch, M.J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.
111. Klimeš, J.; Bowler, D.R.; Michaelides, A. Chemical accuracy for the van der Waals density functional. *Journal of physics. Condensed matter : an Institute of Physics journal* **2010**, *22*, 022201.
112. Klimeš, J.; Bowler, D.R.; Michaelides, A. Van der Waals density functionals applied to solids. *Physical Review B* **2011**, *83*, 195131.
113. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry* **2006**, *27*, 1787–99.

114. D'Adamo, G.; Pelissetto, A.; Pierleoni, C. Predicting the thermodynamics by using state-dependent interactions. *The Journal of Chemical Physics* **2013**, *138*, 234107.