METALLIZATION AND DISSOCIATION IN HIGH PRESSURE LIQUID HYDROGEN BY AN EFFICIENT MOLECULAR DYNAMICS WITH QUANTUM MONTE CARLO

Author: Guglielmo Mazzola

Supervisor: Prof. Sandro Sorella

A thesis submitted for the degree of Doctor of Philosophy

31 October 2014

Contents

Chapter 1

Introduction

Computer simulations of matter are the modern realizations of a very old idea in science, namely that the evolution of a system can be computed exactly given the set of starting conditions and the forces of interaction[1]. Moreover, numerical simulations have a central role in testing models and theoretical predictions as they may be directly compared with real experiments. These numerical experiments may also replace their real-life counterparts if one wants to study materials under extreme conditions.

Molecular dynamics (MD) is a class of computer simulations in which the motion of interacting particles is integrated numerically given the forces that act between them, and therefore is a powerful tool to predict properties of materials. In the so-called *classical* molecular dynamics, the forces between the particles are defined through empirical force fields. This approach is particularly cheap from the computational point of view, and very large systems can be simulated. The electrons are integrated out from the system and they are replaced by effective ionic force fields, which are determined in advance. However, in atomistic simulations, classical MD has a very limited predictive power as the electronic structure, and thus the bonding pattern, changes qualitatively with the ionic configurations generated along the simulation.

On the other hand, the so-called *ab-initio* molecular dynamics considers both ionic and electronic degrees of freedom. The forces between the ions are evaluated at each step during the dynamics, by computing the electronic structure, altough in approximate ways for practical implementations. The Born-Oppenheimer approximation is implicitly understood, where electrons are considered in their instantaneous ground state[2] while the nuclei can be treated as classical particles moving in the field corresponding to the ground state electronic potential energy surface. The foundation of the *ab-initio* MD field can be traced back to the pioneering work of Car & Parrinello[3], in the 1985, where for the first time, an ab-initio

electronic technique such as Density Functional Theory[4] was combined with MD. Clearly, ab-initio MD is a much more computationally demanding technique, and the size of the systems that can be simulated are smaller compared to the classical MD case. Until very recently, the Density Functional Theory (DFT) method has been considered the standard tool for the atomistic simulation of materials, as it allows the simulation of many electrons with a reasonable computational effort. DFT is in principle an exact theory, but in practical calculations can be used only with approximations, namely the exchange-correlation contribution in the functional is approximated[5], and the results are only as good as the approximation used. Conversely, the heavy computational cost of more accurate, such as *post Hartree-Fock* techniques, strongly limits the number of explicitly treated electrons. The quest for an efficient ab-initio molecular dynamics technique is motivated by the need of describing accurately the electronic correlation and, at the same time, treating a large number of atoms, thus allowing a realistic prediction of thermodynamic properties even for complex materials.

Dense hydrogen is certainly one of these cases. Indeed extensive experimental as well as theoretical efforts have been spent for about one century up to the present days, to understand the high pressure phase diagram of hydrogen. In spite of seemingly a simple system to study, hydrogen displays a very rich phase diagram, that in most cases, it is not completely determined from an experimental point of view. At low pressures and temperatures, hydrogen crystallizes as a molecular solid and undergoes a number of phase transitions with increasing density. Up to the present days, this material has been probed experimentally, at low temperatures, up to the pressure of 360 GPa[6], and it always stays in the insulating molecular phase. Monovalent elements like the alkali metals are not normally found in the paired state because the cohesive energy of the metallic state is larger than the binding energy of the diatomic state. In hydrogen, the opposite happens. However, the monoatomic state can be achieved by either high temperatures or high pressures. Its existence is not in doubt, at least from a theoretical standpoint: at some sufficiently high density (and corresponding pressure) a molecular state becomes impossible because the kinetic energy of a localized wave function would exceed the bonding energy. Nevertheless, quantify precisely this dissociation pressure is an extremely hard task.

In 1935, Wigner and Huntington proposed, for the first time, that a pressure of 25 GPa would have been sufficient to break the molecules and create a monoatomic crystal[7]. Moreover, according to the band theory of solid, this compound would be a metal, due to the half-filled conduction band. On the other hand, any indication of metallization or dissociation has not been observed in the room temperature solid phase yet [8, 9, 6]. At high temperatures instead, experimental evidences for a insulator-to-metal (IM) transition have been reported[10] in the range of 100-200 GPa and 2000-3000 K, altough a clear understanding of

its underlying mechanism is still missing. Notice that, apart from its importance as a model system, hydrogen is the most common substance in the universe and the understanding of its equilibrium properties, when compressed, is crucial for a satisfactory description of many astrophysical bodies. Indeed the experimentally observed IM transition occurs at pressures and temperatures that are realized in the interior of giant planets such as Jupiter and Saturn.

Performing experiments under these extreme conditions is really hard, and only few attempts, at high temperatures, have been made[10, 11, 12]. Therefore, numerical simulations are very important, both to better characterize what is observed in the experiment, and to replace real experiments in the vast range of the phase diagram that is still unaccesible with the present experimental techniques. Dense liquid hydrogen has been extensively studied with ab-initio MD driven by forces calculated at the DFT level. Most of these simulations agree on the existence of a first-order transition in the liquid phase between the molecular and the atomic fluid[13, 14, 15, 16, 17]. Nevertheless, its predicted position in the phase diagram strongly depends on the particular approximation used for the exchange-correlation functional, the difference being of more than one hundred GPa's. Moreover a recent systematic study [18] clearly demonstrates the inadequacy of standard approximated DFT functionals calculations for solid molecular hydrogen, strongly limiting the DFT predictive power on this subject.

Among all first principles simulation methods, the quantum Monte Carlo (QMC) method[19] provides a good balance between accuracy and computational cost. QMC methods are the most accurate and efficient statistical methods for treating many body quantum systems. They do not rely on uncontrolled approximations for the electronic correlation as they allow, in principle, the possibility to reach any desired accuracy, by simply considering correlated wave functions of increasing complexity. Moreover, from a computational point of view, QMC algorithms scale almost linearly with the number of processors, thus they are particularly suitable for the present massively parallel supercomputers. Nevertheless, until recently QMC has never been used as a tool to calculate forces within an ab-initio MD framework. For this reason, previous QMC studies on dense liquid hydrogen have been limited to small number of atoms[20, 21, 16], employing a Metropolis sampling technique based on the evaluation of (noisy) energy differences[22]. Indeed, computing forces in QMC is not simple, since finding a finite-variance force estimator is not straightforward[23, 24, 25]. Conversely, evaluating forces using finite energy differences, which correspond to infinitesimal ionic displacements, is computationally expensive[26], as the computational cost increase linearly with the system size[27]. Moreover, it is not possible to use QMC forces, which come with finite and possibly large error bars, in a standard Hamiltonian MD framework, since the simulation would not conserve the total energy. Therefore, altough many efforts have been spent in this

direction, the application of forces from QMC has been limited to few cases[24, 26, 27, 28] and the first attempt by Attaccalite & Sorella (2008) to combine QMC with molecular dynamics is fairly recent[25].

In this Thesis, we improve upon the original Attaccalite & Sorella simulation on dense liquid hydrogen and, for the first time, we perform large scale QMC simulations on this system, namely employing long and equilibrated MD simulations, using a larger number of atoms, on a wide range of temperatures and pressures[29]. We find that the transition from a molecular fluids towards a fully dissociated one appears at higher pressures than previously believed, as we observe that a liquid with mixed atomic-paired character is stable over a wide range of pressures. The connection of this dissociative transition with the insulator to metal transition is also discussed and extensive tests on the accuracy of the method are reported.

Finally, we apply also this ab-initio QMC-MD on a very different system, liquid water at ambient conditions. Understanding completely the structure of water is a recurring theme at the interface of contemporary physics, chemistry, and biophysics. Also in this case, many open questions still exist, which wait to be answered by both experimental and theoretical techniques[30, 31]. From the point of view of the simulations, it has been shown that DFT based molecular dynamics yield in general an overstructured liquid at the 1 g/cm^3 density and ambient conditions[32, 33, 30], namely displaying structural and dynamical properties that are closer to a supercooled liquid, rather than to the one that we encounter in our everyday life. For the first time, we are able to perform a bulk water MD simulation within a QMC approach. We obtain results for static quantities, like the atomic radial pair distribution, which are very encouraging, in view of their good agreement with experiments[34, 35].

1.1 Thesis outline

This thesis is organized as follows

- In Chapter 2 we describe the techniques that are used in ab-initio molecular dynamics, with particular focus on QMC. The first part is devoted to the electronic methods, while the second to the sampling techniques for the ionic configurations. In particular, the QMC-MD methods, which makes use of QMC noisy forces to drive ionic systems to thermal equilibrium is illustrated in Sect. 2.5

- Chapter 3 deals with the dense hydrogen problem. In the introduction we give an overview of the wide literature on hydrogen, from the early speculations to the latest experiments and numerical simulations. Then we introduce the results obtained with

our QMC ab-initio molecular dynamics. These are the main outcome of the thesis and are followed by accuracy tests, discussions and comparison with other existing techniques.

- In Chapter 4 we report preliminary results on liquid water at ambient conditions. Along with the physical results obtained, our purpose is also to show the feasibility of the QMC-MD as an alternative and promising tool to investigate equilibrium properties of bulk materials.

1.2 List of publications

The results contained in Sect. 2.6 and in Sect. 3.3 have been published in the following papers:

- "Finite-temperature electronic simulations without the Born-Oppenheimer constraint", G. Mazzola, A. Zen, and S. Sorella, *J. Chem. Phys.* **137**, 134112 (2012).

- "Unexpectedly high pressure for molecular dissociation in liquid hydrogen by a reliable electronic simulation", G. Mazzola, S. Yunoki, and S. Sorella, *Nature Communications*, **5**, 3487 (2014).

The result concerning water will be published in shortcoming publications

- "Langevin dynamics in curved manifold for an accelerated canonical sampling: application to bulk and water clusters" G. Mazzola, A. Zen, and S. Sorella, *in preparation.*

- "Ab-initio molecular dynamics simulation of liquid water by Quantum Monte Carlo" A. Zen, G. Mazzola, Y. Luo, L. Guidoni and S. Sorella *in preparation.*

The result reported in Sect. 3.4 and Sect. 3.5 about the insulator to metal transition and the validation of the QMC-MD technique will be also reported in a manuscript still in preparation.

Chapter 2

Methods

In this chapter, we review some well-known computational methods and we illustrate new frameworks that are used to model our system. Realistic systems are made up of electrons and ions and can be described with great accuracy by a standard non relativistic hamiltonian. In the first part of the chapter we describe the electronic methods, such as Quantum Monte Carlo (QMC) and Density Functional Theory (DFT), that are used in this work to treat the electronic problem at fixed nuclei. In the second part we illustrate the methods used for the sampling of the nuclei once that the electronic problem is solved. We are going to highlight some issues connected with the ionic sampling of extensive systems and we propose a framework based on a Generalized Langevin equation that can both enhance the efficiency of the sampling and allow the simulation of large systems with QMC noisy forces.

The fact that we can divide our realistic systems in electronic and ionic separable problems is intimately connected with the use of the Born-Oppenheimer approximation. This is a good approximation for the main physical problem of hydrogen under pressure. However, in the end of the chapter, we also introduce a different framework in which the electronic and ionic coordinates are coupled. This scheme allows to do simulations at finite temperature for both ions and electrons and is tested on the simple hydrogen molecule.

2.1 The Born-Oppenheimer approximation

The basis of most finite temperature computational approaches is the Born-Oppenheimer (BO) separation of the system in the electronic and nuclear subsystems, implying that, at zero temperature, the nuclei move in a potential energy surface (PES) provided by the electrons. This approximation is justified by the large difference between the electronic mass m_e and the average nuclear mass M, as demonstrated in the seminal paper of BO[2], by expand-

ing perturbatively the Schrödinger equation in terms of $(m_e/M)^{1/4}$. This separation of the electronic and nuclear degrees of freedom is known as the Born-Oppenheimer approximation. Moreover, when studying systems at non zero temperatures, it is a common practice to consider electrons in their instantaneous ground state while the nuclei can be treated either as classical particles, following an *ab initio* finite temperature molecular dynamics or as quantum particles by solving the nuclear Schrödinger equation with potential given by the ground state electronic PES. These methods implicitly use also an adiabatic approximation, because the BO approximation does not provide only one PES, but several adiabatic potential energy surfaces (PESs), one for each electronic eigenstate. However the nuclei are evolved only according to one adiabatic PES and the non-adiabatic coupling between adiabatic PESs is neglected. This approximation is reliable only if the electronic energy gap is large, namely when the gap between the electronic ground state and the low-lying excited states is much larger than the thermal energy. If the temperature is high enough this is not the case, and it is not correct to assume that the electrons are constrained into one adiabatic PES. In Section2.6 a QMC method that can be used to take in account all the (several) PESs implicitly is described. For the time being let us consider only cases in which the ground state BO approximation is valid and justifies the electron-ions decoupling.

2.2 Density Functional Theory

Density Functional Theory (DFT) is now a standard technique in most branches of chemistry and materials science. Its origin can be traced back to the Thomas-Fermi theory (1927), which is the first approximate method for solving the electronic structure of atoms using just the one-electron ground-state density $\rho(\boldsymbol{r})$. In 1964, Hohenberg and Kohn[4] proved that properties of the ground state of an N-particle system can be expressed as functionals of its density. DFT could simplify a lot the description of the many electron problem since it employ the density (which only depends on 3 spacial coordinates) rather than the full wave function as key variable, but, in practice, the universal energy functional is still unknown. The modern version adopted today is Kohn-Sham (KS) DFT[36], which defines self-consistent equations that must be solved for a set of orbitals whose density, $\rho(\boldsymbol{r})$ is defined to be exactly that of the real system, described by the Hamiltonian

$$\hat{H} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{i,A} \frac{Z_A}{|\boldsymbol{r}_i - \mathbf{R}_A|} + \sum_{i>j} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{2.1}$$

where we use atomic units (energy in Hartree, length scale in Bohr) and the indexes i, j run over the electronic coordinates while A over the nuclei with charges Z_A.

The starting point consist in writing the energy functional as a sum of several contributions[37, 5]:

$$E[\rho] = T[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.2}$$

The electron density of the KS reference system is given by

$$\rho(\mathbf{r}) = \sum_i \phi(\mathbf{r})^2. \tag{2.3}$$

The kinetic energy for the KS single particle orbitals is

$$T[\rho(\boldsymbol{r})] = \sum_i \langle \phi_i | -\frac{1}{2}\nabla^2 | \phi_i \rangle. \tag{2.4}$$

The other two known components are the external potential given by the charged nuclei,

$$V_{ne}[\rho] = \int \rho(\boldsymbol{r}) v(\boldsymbol{r}) d\boldsymbol{r}, \qquad v(\boldsymbol{r}) = -\sum_A \frac{Z_A}{|\boldsymbol{r} - \mathbf{R}_A|} \tag{2.5}$$

and the Hartree component of the electron-electron interaction energy

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} d\boldsymbol{r}' \tag{2.6}$$

The exchange-correlation functional E_{xc} is everything else and no explicit form, given in terms of $\rho(r)$, is available up to now. In practical calculations, the XC contribution is approximated, and the results are only as good as the approximation used. The simplest possible form, known as the LDA (local density approximation), assumes that the functional is a local functional of the density

$$E_{xc}^{LDA} = \int \rho(\boldsymbol{r})\, \epsilon_{xc}(\rho(\boldsymbol{r})) d\boldsymbol{r} \tag{2.7}$$

where ϵ_{xc} is the exchange-correlation energy density of the homogeneous electron gas with density ρ which can calculated using QMC methods. Starting from this very first example, many density functional (DF) approximations have been developed for practical applications and we refer the reader to DFT textbooks or reviews[37, 5] for a more detailed description of all the classes of possible functionals. The Perdew-Burke-Ernzerhof (PBE)[38] exchange-correlation functional is the most used for condensed systems. This functional belong to the GGA family (generalized gradient approximation) of functionals, in which not only the local value of the density at position $\boldsymbol{r}$ enters, but also its gradient. Instead in the early 1990s, hybrids functionals were introduced by Becke[39], by replacing a fraction of GGA exchange with Hartree-Fock (HF) exchange. Among this kind of functionals we find the B3LYP, which is currently the most popular approximation in chemistry. The best known hybrid functionals in the condensed matter community are the PBE0, which uses a mixture of 25% of Hartree-Fock exchange with 75% PBE, and the HSE[40] functional, which uses a combination of range separation and the same mixing fractions as in PBE0 but with the Hartree-Fock

calculation done only on the short-range part of the potential. We briefly introduced LDA, PBE and HSE functionals since they are employed in the dense hydrogen DFT simulations that will be reported in the following chapters. Many other functionals have been developed in these years, however a functional that performs uniformly better than B3LYP is still missing. Some functional may be more accurate (compared to the experiments) for a specific system but may fail for the others. Clearly, when calculating a property to compare with experiment, one could keep trying functionals and tune the parameters until agreement with the measured value is reached but these can not be considered first principle calculations. An interesting discussion on whether or not DFT, with standard approximated functionals, should be called *ab-initio* or empirical can be found in Ref.[5].

2.3 Quantum Monte Carlo methods

The words *Monte Carlo* were used for the first time in the 1940's by scientists who were working in Los Alamos at the Manhattan Project, to indicate a class of methods for the numerical solution of integration problems in a probabilistic way, although the ideas underlying Monte Carlo (MC) methods were already known in the eighteenth century. In many cases, MC methods are the only efficient approach to evaluate many-dimensional integrals of physical interest. For example, in a system of 100 electrons the required integrals to compute expectation values are 300 dimensional, and for such problems MC integration is dramatically more efficient than conventional quadrature methods such as Simpson's rule. Indeed the most successful application of MC methods in physics, is the possibility of solving exactly the quantum many body problem, at least in some cases. We refer to textbooks(see e.g. Ref.[41]) for a complete introduction of MC methods and the probability theory on which they are based. In the following sections we briefly describe two kinds of ground-state Quantum Monte Carlo (QMC) methods that are employed in this thesis:

1. **Variational Monte Carlo (VMC).** This method consists in finding an approximate solution of the Schrödinger equation, depending on some parameters. We make use of the variational principle for obtaining of the best possible value of such parameters.

2. **Diffusion Monte Carlo (DMC).** This technique is a stochastic projector method for solving the imaginary-time many-body Schrödinger equation. This method is more accurate than VMC but it is computationally more expensive. However its success often rely on the use of an optimized trial wave function which can be obtained only by a good VMC calculation. Moreover DMC is not an exact method for fermions as it needs in practice to be used within the fixed node approximation in order to avoid the

infamous *sign-problem*.

QMC results are typically (but not always) an order of magnitude more accurate than density-functional results obtained using the best available approximate exchange-correlation energy functionals[19] but since QMC is much more computationally demanding than DFT it has been applied up to now to much smaller system.

We conclude this section by noting that QMC (at variance with DFT) is not the only many-body approach for solving electronic systems, but can be considered the one in which electronic correlations can be included in the most simple and computationally affordable way. Indeed starting from the simplest single-determinant Hatree-Fock (HF) (1930) theory , which includes the exchange effects arising from the antisymmetry of the many-electron wave function, but neglects the electronic correlations caused by the electron-electron Coulomb repulsion, several post-HF methods have been developed. Within these methods, an almost exact solution is recovered by enlarging the basis and using linear combinations of Slater determinants, e.g. configuration interactions (CI) methods. These methods are often referred as *quantum chemistry* methods since they are computationally very demanding (cost growing as N_{el}^6) Recently, a combined post-HF and QMC approach has been developed by Alavi and coworkers. This is a new promising quantum Monte Carlo method for the simulation of correlated many-electron systems in full configuration-interaction (Slater determinant) spaces[42, 43] and has recently been applied even in simulation of real solids[44].

2.3.1 Variational Monte Carlo

The Variational Monte Carlo method is in principle very simple and it is used to compute quantum expectation values by sampling the electronic coordinates $\vec{r}$ from the squared modulus of a trial function $|\psi_T(\vec{r})|^2$. In this way one can compute the upper bound E_V to the true ground state energy E_{GS} as:

$$E_{GS} \leq E_V = \frac{\int d\vec{r}\, \psi_T^*(\vec{r}) \hat{H} \psi_T(\vec{r})}{\int d\vec{r}\, \psi_T^*(\vec{r}) \psi_T(\vec{r})} \tag{2.8}$$

$$= \frac{\int d\vec{r}\, |\psi_T(\vec{r})|^2 E_L(\vec{r})}{\int d\vec{r}\, |\psi_T(\vec{r})|^2} \tag{2.9}$$

where we have defined the *local energy* operator

$$E_L(\vec{r}) = \frac{\hat{H}\psi_T(\vec{r})}{\psi_T(\vec{r})} \tag{2.10}$$

The expectation value in eq.(2.9) can be calculated by standard Metropolis MC technique[45] by sampling the un-normalized distribution $|\psi_T(\vec{r})|^2$. At the end the total energy is com-

puted as a time average (over the total M simulation steps) of the values of the local energy

$$E_V = \frac{1}{M} \sum_{m=1}^{M} E_L(\vec{r}_m) \tag{2.11}$$

The fluctuations of the local energy during the sampling are connected to the quality of the trial wavefunction as for an exact eigenstate these should be exactly zero and $E_V \to E_{GS}$. For this reason the choice of the trial function is crucial in VMC calculations. Finally we remark that expectation values of operators other than the Hamiltonian, such as forces, may also be expressed as 3N-dimensional integrals (2.9) and evaluated in an analogous way. Care should be taken on the evaluation of the error bars of these statistical quantities. Since the samples generated in the Metropolis walk are correlated a binning technique must be used to compute the variance[46].

2.3.2 Trial wavefunctions

The quality of the trial wave function controls the statistical efficiency and constrains the final accuracy of any VMC (and DMC) simulation. Also, the evaluation of the trial wave function (and its gradient in many cases) is the most demanding part of the computation. Moreover the trial function contains several variational parameters that have to be optimized in order to approach the ground state, i.e. find the lowest energy within a given trial ansatz. The larger is the number of variational parameters the better should be in principle the convergence to the exact ground state. However, it is also true that optimization schemes may fail if the the total number of parameters is too large. One should therefore seek trial wave functions that are both accurate, easy to evaluate, and compact (so that they require the smallest number of parameter to characterize physically the wavefunction). The advantage in adopting QMC consists essentially in using explicitly correlated electronic wavefunction that can substantially improve on the single particle HF ansatz.

In QMC simulations, as in this thesis, trial wave functions of the Slater-Jastrow type (Jastrow, 1955) are used. They consist of a single Slater determinant multiplied by a totally symmetric non-negative Jastrow correlation factor that includes the cusps.

$$\psi_T(\vec{r}) = J(\vec{r})\, D(\vec{r}) \tag{2.12}$$

The Slater determinant D gives the correct fermionic/bosonic symmetry of the wavefunctions while the Jastrow term J takes in account the electronic correlations as it depends also on the interparticle distances. The orbitals in the Slater determinant are usually obtained from LDA or HF calculations. However in our calculations, after initialization given by LDA, the variational parameters contained in the orbitals are optimized in the presence of the Jastrow term. We refer the reader to the seminal works of Casula, Sorella and

coworkers[47, 48, 49] for an exhaustive description of this kind of variational trial wavefunction as well as a possible generalization, namely the *Resonating Valence Bond (RVB)* wavefunction. Indeed in Ref. ([49]) it is shown how the Slater determinant function is a particular case of the more general *antisymmetrized geminal power* ansatz, which is based on the RVB picture of the chemical bond.

2.3.2.1 Atomic orbitals and Slater Determinant

The atomic orbitals (AO) are the key ingredient on which both the determinant and the Jastrow part are constructed. Indeed our QMC code *TurboRVB* employs localized AO instead of plane waves as fundamental basis set and in Ref. ([50]) the reader can find a recent and very useful overview of all the possible types of AO that can be found in this code. In this section we only describe briefly the ones that are used for simulating dense hydrogen.

A generic atomic orbital $\phi^a_\mu(r_{ia})$ of the atom a is written in terms of the radial vector $r_{ia} = \boldsymbol{r}_i - \mathbf{R}_a$ connecting the nucleus of the atom a with the position r_i of the electron i. An (uncontracted orbital) $\phi_{l,m}$, having azimuthal quantum number l and magnetic quantum number m, is the product of an angular part, that is, real spherical harmonic, and a radial part. The latter may have several functional forms; in this application, we have considered the Gaussian type orbitals

$$\phi_{l,m} \propto r^l e^{-\zeta r^2} Z_{l,m}(\Omega) \tag{2.13}$$

where $Z_{l,m}$ is the spherical harmonic. Notice that we can use gaussian functions because the nuclear cusp condition is satisfied by the 1-body Jastrow term. A possible generalization consists in the so-called *hybrid orbitals* that represent linear combinations of different uncontracted AOs.

$$\phi_a(\boldsymbol{r}; \{\zeta_{k,l}, c^k_{l,m}\}) = \sum_{l=0}^{l_{max}} \sum_{k=1}^{K_l} \sum_{m=-l}^{l} c^k_{l,m}\, \phi_{l,m}(\boldsymbol{r}, \zeta_{k,l}) Z_{l,m}(\Omega) \tag{2.14}$$

An atomic hybrid orbital ϕ_a, related to the atom a, is written as the sum of all the uncontracted orbitals, of any azimuthal and magnetic quantum numbers, that we want to use to describe the atom. Therefore, in the language of chemistry, an hybrid orbital thus may contain mixtures of s and p AOs. Both the exponents and the coefficients have to be conveniently optimized, but the advantage of this much more compact definition is that it will reduce a lot the number of variational parameters connecting orbitals localized on different atoms (cf. Eq.(2.16)).

Starting from this AOs we can construct the molecular orbitals (MOs) as linear combinations of AOs. Thus we can obtain our Slater Determinant D using the $N_{el}/2$ lowest energy MOs.

In the Slater determinant $\Psi_{\rm SD}$, the double occupied molecular orbitals ψ_i, with index $i = 1, \ldots, N/2$ (N is the number of electrons), are a linear combination of the localized atomic hybrid orbitals:

$$\psi_i(\mathbf{r}) = \sum_{a=1}^{M} \sum_{\mu_a=1}^{L_a} c_i^{a,\mu_a} \phi_{a,\mu_a}(\mathbf{r}) \tag{2.15}$$

where ϕ_{a,μ_a} is the μ_a-th atomic hybrid orbital of atom a, centered around the position $\mathbf{R}_a$ of nucleus a, L_a is the number of atomic hybrid orbitals used for atom a, for a total of $L = \sum_a L_a$ hybrid orbitals in the overall system, M is the total number of atoms and the $M \times L$ coefficients c_i^{a,μ_a} are variationally optimized. From what follows it may be understood why adopting a localized atomic basis set is crucial when dealing with large systems.

Indeed, the optimization of the Slater determinant is performed by using the correspondence between the single SD written in terms of molecular orbitals, and a truncated antisymmetrized geminal power (AGPn)[49] with $n = N/2$, with a geminal

$$g(\mathbf{r}_1, \mathbf{r}_2) = \sum_{a,b}^{M} \sum_{\mu_a}^{L_a} \sum_{\mu_b}^{L_b} g_{\mu_a,\mu_b}^{a,b} \phi_{a,\mu_a}(\mathbf{r}_1) \phi_{b,\mu_b}(\mathbf{r}_2) \tag{2.16}$$

The $L \times L$ parameters $g_{\mu_a,\mu_b}^{a,b}$ are related to c_i^{a,μ_a} by the relation:

$$g_{\mu_a,\mu_b}^{a,b} = \sum_{i=1}^{N/2} c_i^{a,\mu_a} c_i^{b,\mu_b} \,. \tag{2.17}$$

The present formulation is adopted in the *TurboRVB* code, therefore the parameters, that are actually optimized in this approach, are the $g_{\mu_a,\mu_b}^{a,b}$ ones. They are then used to obtain the molecular orbital coefficients c_i^{a,μ_a} via the diagonalization described in Refs. [49]. This choice gives a very important technical advantage for systems of large sizes. In particular, in order to decrease the total number of variational parameters, we have fixed to zero all the coefficients $g_{\mu_a,\mu_b}^{a,b}$ connecting the atoms a and b that are at a distance $R_{ab} = \|\mathbf{R}_a - \mathbf{R}_b\|$ larger than an appropriately chosen cut-off $r_{\rm max}$. In this way, a straightforward reduction of the number of variational parameters is possible, by exploiting the fact that matrix elements connecting localized orbitals above a threshold distance $r_{\rm max}$ do not contribute significantly to the energy. Indeed, as we have systematically checked in several test cases (see Sect. 3.5.1), it is enough to consider $r_{\rm max}$= 4 a.u. to have essentially converged results for the molecular orbitals, implying a drastic reduction of the variational space (from $\simeq$ 33000 parameters to $\simeq$ 5000 in a 256 hydrogen system, using one hybrid atomic orbital with 2s gaussian basis per atom).

2.3.2.2 Jastrow factor

The bosonic Jastrow term, $J = e^U$, represents a compact way to take into account explicitly the electronic correlations since it depends directly on distances between electrons. There are many different choices for this factor; in this work we have used a Jastrow factor that account up to the 3-body interaction. The 1-body term is used to satisfy the nuclear-electron cusp condition. Hence the total Jastrow reads

$$J = J_1\, J_2\, J_3 \tag{2.18}$$

The term U_1 is a one electron interaction term which improves the electron-nucleus correlation and satisfies the nuclear cusp conditions. The exact functional form is given by

$$U_1 = -\sum_a^M [(2Z_a)^{3/4} \sum_i^N u_1(\sqrt[4]{2Z_a}\, r_{ia})] + \sum_a^M \sum_{\nu_a}^{L_a^J} \sum_i^N f_{\nu_a}^a \phi_{\nu_a}^a(\boldsymbol{r}_{ia}) \tag{2.19}$$

where the vector $r_{ia} = \boldsymbol{r}_i \mathbf{R}_a$ is the difference between the position of the nucleus a and the electron i, Z_a is the electronic charge of the nucleus a, L_a^J is the number of atomic orbitals $\phi_{\nu_a}^a$ that are used to describe the atom a, $f_{\nu_a}^a$ are variational parameters and the function $u_1(x) = (1 - e^{-b_1 x})/2b_1$ depends parametrically on the value of b_1.

The $U_2(r)$ factor is an homogeneous two body interaction term. It depends only on the relative distance r_{ij} between pairs of electrons. The specific functional form reads

$$U_2(\boldsymbol{r}) = \sum_{i<j}^N u_2(r_{ij}) \tag{2.20}$$

where $u_2(x) = (1 - e^{-b_2 x})/2b_2$ and b_2 is a variational parameter. Finally the 3 body term is an inhomogeneous two electron interaction that depends also on the relative position of the electrons and the nucleus, i.e. it's an *e-e-n* interaction. Its functional form is

$$U_3(\boldsymbol{r}, \mathbf{R}) = \sum_{i<j}^N [\sum_a^M \sum_{\mu_a,\nu_a}^{L_a^J} f_{\nu_a,\mu_a}^a \phi_{\nu_a}^a(\boldsymbol{r}_{ia}) \phi_{\mu_a}^a(\boldsymbol{r}_{ja})] \tag{2.21}$$

where $\phi_{\mu_a}^a$ are the uncontracted atomic orbitals centered on atom a. Notice that this is an *on site* interaction which is included as a particular case of the more general 4-body *e-e-n-n* interaction $U_4(\boldsymbol{r}, \mathbf{R}) = \sum_{i<j}^N [\sum_a^M \sum_b^M \sum_{\nu_a}^{L_a^J} \sum_{\nu_b}^{L_b^J} f_{\nu_a,\nu_b}^{a,b} \phi_{\nu_a}^a(\boldsymbol{r}_{ia}) \phi_{\nu_b}^b(\boldsymbol{r}_{jb})]$

2.3.2.3 Optimization method

The integral (2.9) provides the value of the energy within a given trial wavefunction ansatz $\psi_T(\alpha)$ which depends on a (possibly very large) set of parameters $\{\vec{\alpha}\}$. These are for example the parameters $f_{\nu_a,\nu_b}^{a,b}, b_1, b_2$ defined in the Jastrow above as well as the parameters that define

the AOs and the hybrid orbitals $\{\zeta_{k,l}, c^k_{l,m}\}$ and the expansion coefficients of the MOs. According to the variational principle, the exact ground state energy E_{GS} represents the lowest bound for any variational wave function; so the idea is to optimize all these parameters in order to get the lowest possible energy within the given functional ansatz $\psi(\alpha)$.

In order to optimize the wavefunction we employ an iterative method in which the parameters $\vec{\alpha}$ are changed at each step accordingly to the generalized forces $\vec{f}(\vec{\alpha}) = -\partial E[\vec{\alpha}]/\partial\vec{\alpha}$. Advanced method such as *stochastic reconfiguration (SR)*[51, 52] and the linear method[53, 54] also take in account higher order derivatives, like the Hessian matrix.

For the problem of dense hydrogen we tested different optimization procedures with the linear method[53] being the optimal choice. The number of optimization steps clearly depends on the accuracy required for the total energy. This number can be substantially reduced if the trial wavefunction represents already a good starting point, and this is the case for the optimization during the molecular dynamics. Indeed, given that the wavefunction $\psi(\vec{\alpha}[\mathbf{R}])$ is correctly optimized at ionic position $\mathbf{R}$, $\psi(\vec{\alpha}[\mathbf{R}])$ is a good starting point for the new wavefunction at position $\mathbf{R} + \delta\mathbf{R}$. The smaller is $|\delta\mathbf{R}|$, which is controlled by the time step Δ, the closer will be the two optimized wavefunctions. In this case a small number, i.e. 5-6, of SHR steps are sufficient to maintain the wavefunction close to ground state PES that is reachable within the given variational ansatz.

Instead, if an optimization *from scratch* is required, the minimization procedure takes longer. The first step consists in initializing the Slater Determinant by constructing the set of MOs with a preliminary LDA calculation. In the second step we optimize the Jastrow keeping fixed the determinant. We call such wavefunction ansatz as JDFT. Finally we improve on the DFT Slater Determinant (i.e. modifying the nodal surface) by optimizing the determinant parameters in the presence of the Jastrow. In this way we obtain the first JSD ansatz, at fixed ionic positions $\vec{\mathbf{R}}$ that we use as a starting point for the dynamics.

2.3.3 QMC nuclear forces

The majority of QMC applications has been limited to the calculation of the ground-state total energy. In principle, there is no difficulty for computing other observables. However, in practice, the convergence of the Monte Carlo calculations can be either much more slower than the energy or even impossible if *infinite-variance* problems arise. Accurate energy calculations are directly related to good trial wave functions as it can be proven that the error in the energy E_V and variance scales as the error squared in the wavefunction (see. e.g. Ref.[24])

$$\delta E \equiv E_V - E_{GS} \sim \mathcal{O}[\delta\psi^2] \tag{2.22}$$

$$\sigma^2[E_L] = \langle (E_L - E_V)^2 \rangle_{\psi_T^2} \sim \mathcal{O}[\delta\psi^2] \tag{2.23}$$

with the variational and local energy as defined in 2.3.1 and the error in the wavefunction is $\delta\psi \equiv \psi_T - \psi_{GS}$, where ψ_{GS} is the exact ground state wavefunction. Eq.(2.23) tells that energy is accurate at order two in the wavefunction. This very favourable scaling is no longer true for other properties. In this section we will briefly focus on evaluation of nuclear forces that are essential to perform molecular dynamics. The nuclear force acting on nucleus a is defined as

$$\boldsymbol{f}_a = -\nabla_a E_V[\psi] \tag{2.24}$$

Since both the Hamiltonian operator $\hat{H}$ (2.1) and the wavefunction ψ depend parametrically on the atomic positions R, the expression for the force is

$$\boldsymbol{f}_a = -\frac{\int \frac{\partial E_L}{\partial \boldsymbol{R}_a}\psi^2\, d\boldsymbol{r}}{\int \psi^2\, d\boldsymbol{r}} - 2\frac{\int (E_L(\boldsymbol{r}) - E_V[\psi])\frac{\partial \log|\psi|}{\partial \boldsymbol{R}_a}\psi^2\, d\boldsymbol{r}}{\int \psi^2\, d\boldsymbol{r}} + \boldsymbol{f}_a^{bias} \tag{2.25}$$

The first term is referred as *Hellmann-Feynman* term while the second as *Pulay* term. Notice that the name Hellmann-Feynman (HF) is given only on the basis of a similarity with the HF theorem, which in this case can not be applicable since the variational wavefunction is not an eigenstate of $\hat{H}$. Therefore, from the magnitude of the HF term with respect to the Pulay we can estimate the quality of the trial wavefunction. Let us investigate one major difficulty connected with the QMC evaluation of Eq.(2.25), the so-called infinite variance problem. For example, the HF term contains a contribution to the force which is proportional to $1/d^2$, where d is the electron-nucleus distance. This leads to a diverging variance even for the simplest hydrogen atom case (see e.g Ref.[55]). Also, in the Pulay force there is a similar problematic behavior in proximity of the nodal surface, because both E_L and $\partial \log|\psi|$ diverge as the nodal surface is approached. While the the infinite variance problem in the HF term can be solved by employing *renormalized* force estimators, following a scheme proposed by Assaraf and Caffarel[23, 24], the Pulay divergence needs to be removed in a different way, that is using the *reweighting* technique introduced by Attaccalite and Sorella[25].

Finally, the last term in Eq.(2.25) is

$$\boldsymbol{f}_a^{bias} = -\sum_k \frac{\partial E_V[\psi]}{\partial \alpha_k}\,\frac{\partial \alpha_k}{\partial \boldsymbol{R}_a} \tag{2.26}$$

and is exactly zero if the energy is correctly optimized, i.e. $\partial E_V[\psi]/\partial\alpha_k = 0$. For this reason, assuming that the wavefunction is very close to the energy minimum, we neglect this complicated term in our calculation. Nevertheless, we experienced that this bias force is far from being negligible if the optimization has not been correctly achieved. Moreover, we observed that this kind of contribution acts like a dissipative force, i.e. limiting the displacements of the ions during the dynamics. This effect is well known also in the context of DFT within the so-called *second generation Car-Parrinello dynamics* approach, where a dissipative unknown

force arises from a non optimal propagation of the electronic density at each MD step (see Ref.[56, 57] for details).

Despite the apparent complexity of the expression of the force estimator of Eq.(2.25), the computational gain with respect to the simple finite-difference technique is evident, especially for large systems. Indeed the choice

$$\boldsymbol{f}_a = -\frac{E_V(\boldsymbol{R}_a + \Delta\boldsymbol{R}_a) - E_V(\boldsymbol{R}_a)}{\Delta\boldsymbol{R}_a} \tag{2.27}$$

leads to the computation of $3M$ energy differences, with M number of atoms, that is a computational bottleneck if M is very large. This scaling clearly holds even if Eq.(2.27) is evaluated by means of a correlated sampling such as the *Space Warp Coordinate Transformation (SWCT)*[58].

The great advantage in evaluating the forces without resorting to a finite difference scheme has been proved by Capriotti and Sorella[27]. In that work the evaluation of the force estimator in Eq.(2.25) was done in conjunction with SWCT and *algorithmic differentiation (AD)*; the scope of SWCT is in reducing the statistical error in the evaluation of the force estimator rather than performing the finite-difference evaluation of the total energy. AD is employed instead to compute automatically all the partial derivatives, $(\partial/\partial\boldsymbol{R}_a)E_L$, $(\partial/\partial\boldsymbol{R}_a)\log\psi$, $(\partial/\partial\boldsymbol{r}_i)E_L$, $\cdots$ that appear in Eq.(2.25) and in the modified version in which SWCT are used[1]. We refer the reader to the original paper[27] for a general overview of the AD technique and to Refs.[27, 50] for a self-contained description of the SWCT in the context of QMC. Although the SWCT+AD technique was implemented in our hydrogen calculations this new achievement was not essential in performing the large scale simulation described in Chapter3. Indeed in the seminal paper of Attacalite and Sorella[25] a system of 128 hydrogen atoms was already simulated, without computing forces with the SWCT+AD method. The SWCT+AD technique has proven to be substantially more efficient for heavy atoms instead and when pseudopotentials are used, as in the case of water simulations[27]. For the above reasons we choose to avoid a lengthy description of these techniques in this thesis which is mainly devoted to hydrogen at high pressures.

2.3.4 Diffusion Monte Carlo

This last section about QMC methods deals with the *Diffusion Monte Carlo (DMC)* technique. DMC is an -in principle exact- projection methods, unlike VMC. However, as anticipated, for practical purposes the introduction of a trial wavefunction ψ_T is essential even in this framework, both to enhance the efficiency of the method and to circumvent the *fermionic sign problem* by introducing some approximations. For these reason, a good DMC calculation

[1] see Eq. (15) of Ref.[27].

has always to be done on top of a good VMC trial wavefunction optimization. DMC with ψ_T *importance sampling* was first used in the work of Ceperley and Alder[59] and is a well established technique. Suggested reviews on this topic are Ref.[60] for molecular systems and [19] for solid state systems. In this section we give a short self-contained description of the DMC in the configuration space, following the lines of Foulkes[19] and Pederiva[41].

DMC is a stochastic projector method for solving the imaginary-time many-body Schrodinger equation

$$-\frac{\partial\psi(\boldsymbol{r},\tau)}{\partial\tau} = (H - E_0)\psi(\boldsymbol{r},\tau) \tag{2.28}$$

The propagation in imaginary time of an arbitrary wavefunction of a system, projects from this function the eigenstate of H which has the lowest energy, among those along which the initial state has a non-vanishing component. Indeed the solution of Eq.(2.28) reads

$$\psi(\boldsymbol{r},\tau) = e^{-\tau(H-E_0)} = \sum_n c_n e^{-\tau(E_n-E_0)}\phi_n(\boldsymbol{r}) \tag{2.29}$$

where E_n are the eigenvalues corresponding to the eigenstates of H, ϕ_n. The projections along the eigenstates decay exponentially with different time scales τE_n. By using the normalization $e^{-\tau E_0}$ is possible to filter out every excited eigenstate for a sufficient large time, as in the limit $\tau \to \infty$ we obtain the ground state of H:

$$\lim_{\tau\to\infty} \psi(\boldsymbol{r},\tau) = c_0\phi_0(\boldsymbol{r}) \tag{2.30}$$

In order to introduce how the diffusion algorithm works let us explicitly write Eq.(2.28) with the standard hamiltonian

$$-\frac{\partial\psi(\boldsymbol{r},\tau)}{\partial\tau} = \frac{1}{2}\sum_i \nabla_i^2\psi(\boldsymbol{r},\tau) + V(\boldsymbol{r})\psi(\boldsymbol{r},\tau) \tag{2.31}$$

where $\boldsymbol{r} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N\}$ represents the collection of all the particles. This equation represents a diffusion equation with an absorption term, as can be realized by considering two limiting cases. If we neglect the potential V, then Eq.(2.31) becomes a diffusion equation

$$\frac{\partial\psi(\boldsymbol{r},\tau)}{\partial\tau} = \frac{1}{2}\sum_i \nabla_i^2\psi(\boldsymbol{r},\tau). \tag{2.32}$$

The exact solution can be recast in integral form as

$$\psi(\boldsymbol{r},\tau) = \int G(\boldsymbol{r}-\boldsymbol{r}',\tau)\psi(\boldsymbol{r}',0)\, d\boldsymbol{r}' \tag{2.33}$$

where the Green's function is the free diffusion propagator

$$G(\boldsymbol{r}-\boldsymbol{r}',\tau) = -\frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\boldsymbol{r}-\boldsymbol{r}')^2}{2\tau}} \tag{2.34}$$

Then if we represent the desired distribution $\psi(\boldsymbol{r},\tau)$ by a set of discrete or random walkers, the (pseudo)dynamics of the $3N$-dimensional walker is given by $\boldsymbol{r}'_j = \boldsymbol{r}_j + \boldsymbol{\eta}_j$, with $\boldsymbol{\eta}_j$ being a gaussian distributed random variable with variance τ.

If we neglect the diffusion term, then Eq.(2.32) (with a renormalization energy E_0) reduces to

$$-\frac{\partial\psi(\boldsymbol{r},\tau)}{\partial\tau} = [V(\boldsymbol{r}) - E_0]\psi(\boldsymbol{r},\tau) \tag{2.35}$$

and its solution

$$\psi(\boldsymbol{r},\tau) = e^{[V(\boldsymbol{r})-E_0]\tau} \tag{2.36}$$

represents a survival probability (after a time τ) that depends on the position $\boldsymbol{r}$. From the point of view of the algorithm, this process is called branching or reweighting and can be implemented in several ways. Either we can associate to each walker a weight that depends on this probability, or we can kill or spawn more walkers at each step depending on this weight. The latter technique is more efficient in practice.

In the general intermediate situation an exact expression for the propagator can not be obtained. To this end the common Trotter-Suzuki approximated formula can be used

$$e^{-\tau H} = e^{-\tau(T+V)} \approx e^{-\tau V/2}e^{-\tau T}e^{-\tau V/2} + \mathcal{O}[\tau^3] \tag{2.37}$$

and the algorithm now will have to implement both the diffusive dynamics and the branching process. Since this approximation introduce a time-step error, the algorithm consist in iterating many times diffusion/branching steps, keeping the time step τ as small as possible. After this process has been iterated for a sufficient long time, the distribution of the walker will be stationary and should sample correctly the ground state $\phi_0(\boldsymbol{r})$. The value of E_0, which is an initial guess, should automatically stabilize during the dynamics in order to keep the walker population stationary.

However, this simple DMC algorithm turns out to be "usually spectacularly inefficient"[19] due to the large fluctuations of the branching weight, hence of the walker population. This happens when the (charged) particles become very close and in these cases is convenient to introduce an approximate *guiding function* to drive the walkers far away from the problematic regions in the configuration space. Let us consider, then, a trial wavefunction ψ_T (obtained for instance from a VMC calculation). In the importance sampling DMC the walkers are sampled from a density $f(\boldsymbol{r},\tau) = \psi_T(\boldsymbol{r})\phi(\boldsymbol{r},\tau)$ which evolves in imaginary time, and is proportional in the large time limit to $\psi_T(\boldsymbol{r})\phi_0(\boldsymbol{r})$. By multiplying Eq.(2.31) by ψ_T we obtain the equation which control the evolution of f

$$-\frac{\partial f(\boldsymbol{r},\tau)}{\partial\tau} = -\frac{1}{2}\nabla^2 f(\boldsymbol{r},\tau) + \nabla\cdot[\mathbf{v}_D(\boldsymbol{r})f(\boldsymbol{r},\tau)] + [E_L(\boldsymbol{r}) - E_T]f(\boldsymbol{r},\tau) \tag{2.38}$$

where $\nabla = (\nabla_1, \nabla_2, \cdots)$, $\mathbf{v}_D(\boldsymbol{r})$ is a 3N-dimensional *drift* velocity given by

$$\mathbf{v}_D(\boldsymbol{r}) = \nabla \log |\psi_T(\boldsymbol{r})| \tag{2.39}$$

and E_L is the usual local energy as defined in Eq.(2.10) in the VMC section. The corresponding (approximate short time) propagator is now the product $G_d(\boldsymbol{r}, \boldsymbol{r}', \tau) \approx G_d\, G_B$ of a modified diffusion Green's function

$$G_d(\boldsymbol{r}, \boldsymbol{r}', \tau) = -\frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(\boldsymbol{r} - \boldsymbol{r}' - \tau \mathbf{v}_D(\boldsymbol{r}))^2}{2\tau}\right] \tag{2.40}$$

and a modified branching weight

$$G_b(\boldsymbol{r}, \boldsymbol{r}', \tau) = \exp\left[-\tau/2\left(E_L(\boldsymbol{r}) + E_L(\boldsymbol{r}') - 2E_0\right)\right] \tag{2.41}$$

The advantages connected with the importance sampling are clear both in the diffusion and in the branching part. Since the drift velocity $\mathbf{v}_D(\boldsymbol{r})$ carries the walkers along in the direction of increasing ψ_T, the region in which ψ_T is large are sampled more frequently and the walker may avoid specific regions in which $\psi_T = 0$, i.e. when two electron becomes in contact. Moreover the exponent in the reweighting term Eq.(2.41) now contains the local energy instead of the potential energy. This is important because for a ψ_T the local energy is close to the ground-state energy and roughly constant, so the population fluctuations are strongly reduced. The importance sampling energy estimator can be computed as

$$E_0 = \frac{\langle \phi_0 | H | \psi_T \rangle}{\langle \phi_0 | \psi_T \rangle} = \lim_{\tau \to \infty} \frac{\int f(\boldsymbol{r}, \tau) \frac{H\psi_T(\boldsymbol{r})}{\psi_T(\boldsymbol{r})} d\boldsymbol{r}}{\int f(\boldsymbol{r}, \tau) d\boldsymbol{r}} \tag{2.42}$$

Moreover importance-sampling transformation is also extremely helpful in satisfying the fixed-node constraint that will be introduced in the following. The fixed node (FN) approximation is, up to now, mandatory when applying DMC to fermionic systems.

Indeed, since in DMC the wave function has to be a population density, DMC can only employ the constant sign solutions of Schrödinger equation. This is a serious issue if one is interested in the ground-state of a fermion system where the wave function is antisymmetric i.e. both positive and negative. The standard way of dealing with negative weights of the fermionic system is to sample with respect to the bosonic system by using the absolute values of the weights $|f(\boldsymbol{r}, \tau)|$ and to assign a sign to the quantity being sampled. However, the statistical errors increase exponentially with the particle number N. Many attempts to solve this (negative) sign problem have been made, although it is now believed that an algorithm which scales polinomially with N (this is the definition of "solving the sign problem") can not be obtained for the exact solution of a generic fermionic problem[61].

Fixed-node DMC is an alternative and approximated method for dealing with the fermion antisymmetry[60, 62]. Although not exact, it gives ground-state energies that satisfy a variational principle and is stable for large systems. The basic idea of FN DMC is very simple. A trial many-electron wave function is chosen and used to define a trial many-electron nodal surface. In this way $f(\boldsymbol{r}, \tau) > 0$ by construction. Then a standard DMC algorithm can be applied in the regions enclosed by a nodal surface (nodal pockets). Then the Eq.(2.28) is solved by simulating the diffusion process within the domains bounded by the assumed fixed nodes.

This condition is satisfied if the random walks are constructed not to cross the nodes of the trial wave function. Moreover, as mentioned before, the importance-sampling transformation automatically enforces the fixed-node constraint, since whenever a walker approaches a nodal surface of ψ_T, the drift velocity grows as $1/d$, with d being the distance from the nodal surface, and carries it away (see again Ref.[19] for details).

This FN method has many positive features: If the nodes are correct, then the solution is exact. Then the algorithm scales with a polynomial time: for large systems, the computational bottleneck is represented by the time taken to determine whether the system has crossed a node, and since the nodes are given by the Slater determinant part of the ψ_T, the evaluation of a general determinant requires order N^3 operations[62].

Recently, a new DMC scheme, the Lattice Regularized DMC (LRDMC)[63, 64] was introduced by Casula & Sorella in order to deal with heavy atoms. The main advantages of the lattice regularization are that i) it allows the inclusion of nonlocal potentials (important for the study of heavy atoms) in a consistent variational scheme and ii) can deal with several length scales by using multiple lattice spaces, with great reduction of autocorrelation times for heavy atoms or complex systems such as low density electron gases.

Finally, let us spent few words about DMC mixed estimators, i.e. estimators of observables O different then the energy, like the ionic forces. For these observables one does not get mean values on the ground state, but rather matrix elements between the ground state and the trial variational state:

$$< O >_{MA} = \frac{\langle \phi_0 | O | \psi_T \rangle}{\langle \phi_0 | \psi_T \rangle} \tag{2.43}$$

It is found that a good approximation for the exact mean value on the ground state is given by the so-called *Ceperley's correction*:

$$\frac{\langle \phi_0 | O | \psi_0 \rangle}{\langle \phi_0 | \psi_0 \rangle} \approx 2 < O >_{MA} - < O >_{VMC} + \mathcal{O}[\epsilon^2] \tag{2.44}$$

where $< O >_{VMC}$ is the standard VMC expectation value over ψ_T and the error $\mathcal{O}[\epsilon^2]$ is of order 2 in the distance between the real ground state and the trial function, i.e. $\psi_T - \psi_0 \sim \epsilon\, \psi_{residual}$.

2.4 Molecular dynamics and sampling

In the previous section we described state-of-the-art electronic methods that are used in ab-initio simulations of materials. However this is not the end of the story as realistic materials are composed by electrons and ions, and several properties also depend on the ionic degrees of freedom. In order to investigate thermodynamics properties such as structural relaxations and phase transitions it is therefore necessary to deal with the ions and sample efficiently also their equilibrium distribution.

In the past, performing accurate electronic calculations was so computationally expensive that most of the abovementioned electronic methods could not have been employed in a molecular dynamics (MD) scheme. Indeed in all practical applications, MD calculations employed empirical interatomic potentials and this clearly limited the predictive power of the MD framework. The first *ab-initio* MD technique dates from the pioneering work of Car & Parrinello[3] in the 1985, where for the first time, an ab initio electronic technique such as DFT was combined with MD. Car-Parrinello (CP) method significantly reduced the amount of computer time necessary per time step and opened the field of ab-initio simulations of materials. By contrast, the straightforward approach within DFT consists in solving the Kohn-Sham equations for a new wave function at each time step instead of performing the fictitious CP dynamics on the wave function. This approach is sometimes referred as Born-Oppenheimer dynamics (BOMD)[65] and in the last decade, due to the availability of faster computer machines, re-gained popularity among the DFT community. Indeed within BOMD a greater time step in the MD integration can be used, thus providing less correlated configurations at each step.

Moreover, within BOMD one avoids to deal with hidden parameters such as the electron fictitious mass and the artificial electron thermostat for systems with a small band gap. Indeed, in the context of hydrogen research it was shown that without a full reoptimization of the electronic wave functions in every step (as in CPMD), the degree of dissociation may be artificially enhanced[66]. At present days, for hydrogen simulations and within standard DFT density functionals, there is no computational barrier that prevents BOMD simulations of large system, i.e. with a number of proton in the simulation cells large enough to neglect finite size effects[15, 17].

This situation changes dramatically if one switches from standard DFT towards more accurate methods such as QMC simulations. Let us start by clearing the field from a possible ambiguity about the term *dynamics*. Strictly speaking, molecular dynamics simulations generate trajectories in phase-space by treating the nuclei classically and integrating Newton's or

Hamilton's equations of motion numerically. In most cases MD is employed to obtain equilibrium distributions at finite temperature by means of some appropriate thermostats rather than to investigate real dynamical properties. Thermostats are used to simulate the canonical, constant temperature, NVT ensamble, starting from a NVE microcanonical Newtonian dynamics. For this reason the term *molecular dynamics* is commonly used also in the context of the sampling as in this thesis, provided that some kind of thermostat is used to enforce the correct probability distribution. For example, previous DFT simulations on hydrogen were performed within NVT ensamble by means of a Nose-Hoover thermostat[66, 15].

2.4.1 Ionic Metropolis sampling

A different route for the sampling of ionic property is represented by the standard Metropolis Monte Carlo (MC) algorithm[45]. At variance with the MD, this method employs a Markov walk to sample the canonical distribution and thereby calculate thermodynamic properties of classical many-body systems.

The advantages connected with this scheme are that i) it does not require the evaluation of the force but only of the total energy and ii) it is possible in principle to use efficient global updates to sample the configuration space, generating very uncorrelated samples at each iteration. MC sampling may be superior in cases of complex systems such as macromolecules and polymer physics[67], and generally, in all system in which there is a clear decoupling of length scales. For example, in biophysics simulations, one would like to flip or rotate rigidly a molecule or translate and swap entire groups of atoms, in order to enhance the sampling. On the other hand, an efficient sampling of phase space in MC requires smart and system-dependent trial moves, making the application of the method more intricate than molecular dynamics where the configurational sampling obeys the general principle of following the forces.

At the end, the choice between MD and MC sampling should depend on the system under consideration. In practice, for materials and systems whose equilibration time is much shorter than polymers and biomolecules, this choice only depends on the availability of the ionic forces within the electronic method used. For example, simulations of water at ambient conditions are usually performed with MD if the electronic method is DFT (with forces)[68, 30], while MC sampling is employed if more accurate post-Hartree-Fock methods are employed (such as MP2, without the possibility to calculate forces)[69]. The liquid water simulations reported in Chapter 4 being a recent remarkable exception to this general rule.

Turning back to the hydrogen problem, we believe that a MD-based sampling is more efficient than MC for liquid hydrogen at high pressure since there is no decoupling of length scales at the densities under considerations and there is any clear notion of "rigid molecules"

on which global trial moves, such as block rotations and translations, can be applied.

Historically, the first QMC simulations on hydrogen[20, 70, 71, 72] employed a variant of MC sampling instead, the *Coupled Electron-Ion Monte Carlo (CEIMC)*[20]. This technique is essentially an implementation of the *Penalty method*[22] which extends the standard Metropolis scheme in the case of noisy energy difference estimations, the energy difference being evaluated with QMC electronic calculations (at different levels of accuracy, i.e. VMC or DMC). Let us briefly recall how Metropolis sampling works and the issues connected with this choice. It is well known that the acceptance probability of a trial move in the configuration space $\boldsymbol{R}$ is given by

$$A(\boldsymbol{R} \to \boldsymbol{R}') = \min\left[1, \exp\left(\frac{E(\boldsymbol{R}') - E(\boldsymbol{R})}{kT}\right)\right] \tag{2.45}$$

hence by the energy difference between the new $\boldsymbol{R}'$ and the old $\boldsymbol{R}$ position. This algorithm works if $E(\boldsymbol{R})$ is a function that can be analytically evaluated given the position $\boldsymbol{R}$, like in the original implementation[45] or in the context of VMC Metropolis sampling where the local energy E_L is a well defined function of the electron positions. If the energy difference $E(\boldsymbol{R}') - E(\boldsymbol{R})$ is evaluated with some statistical error δE instead, the acceptance/rejection will be biased by an exponentially large factor $\exp(\delta E/kT)$. Naively one would require to kill the error bar $\delta E \to 0$ to a negligible level, instead, it is still possible to use a finite error $\delta E > 0$ and modify the acceptance threshold. This is the idea behind the Penalty method[22]. Despite this remarkable improvement, the problem is still far from being solved since, in practice, the error bar on the energy difference must be of order $\delta E \sim kT$. Therefore, since the error δE grows with the size of the system while kT is an intensive quantity, the larger is the size the larger must be the QMC computational time spent to reduce the noise level to kT. For this reason, hydrogen simulations within CEIMC technique employ supercells containing up to 54 protons, a number that is considered almost one order of magnitude smaller than state-of-the art DFT simulations. Although the twist averages boundary conditions (TABC)[73], that can be viewed as the generalization of Brillouin zone integration to many-body quantum systems, are used within CEIMC to eliminate shell effects in the kinetic energy of metallic systems. However this number could be too small to reproduce the thermodynamic limit.

For this reason, we believe that a molecular dynamics framework should be more efficient provided that,

- the QMC forces can be feasibly computed and drive the sampling to the relevant configurations reducing the autocorrelation times,
- the noisy QMC forces (which are themselves evaluated with a finite error bar) can still be employed in the sampling,
- we are not forced to reduce the error bar on the (extensive) energy to $\ll k_B T$.

A QMC-MD scheme which is based on a generalized second order Langevin equation and make use of (noisy) QMC forces , was originally proposed by Attaccalite & Sorella[25] in 2008. A very similar Langevin scheme was also proposed in Refs.[56, 57] at the DFT level. In this thesis, we report, for the first time, systematic QMC-MD large scale simulations obtained within this framework. Moreover we describe the setup of the MD parameters in order to obtain ergodic simulations and illustrate in more details the efficiency gain connected with this new generalized Langevin scheme.

2.4.2 Langevin dynamics

Langevin dynamics (LD) was first introduced in molecular simulations to evaluate the properties of mesoscopic systems composed by particles in a solvent[74]. Here a noise and a dissipative force are added to the Hamilton equations to model a bath of lighter particles. Recently, a broad interest on LD has been renewed at the interface of contemporary molecular biology, biochemistry, and biophysics. For example, the protein folding problem, that is a class of thermally activated rare events whose mechanism is central to the understanding of neurodegenerative diseases, have been tackled both via brute-force atomistic LD simulations[75] and by a more sophisticated theoretical framework based on LD[76, 77, 78] that is giving very promising results[79].

However, LD is now also commonly used to obtain canonical ensembles[74, 80] at finite temperature T, even if the system is not surrounded by a "real" thermal bath. Usually, in MD simulations the sampling can be biased, due to the finite time-step integration errors. On the other hand, in a MC simulation the moves are accepted or rejected in such a way that the exact distribution is enforced. Due to its stochastic nature LD, that is a stochastic molecular dynamics, can be viewed as an efficient MC scheme where all the moves are accepted. Indeed, the trial moves are not chosen at random but moving the particles in the direction of the forces acting on them. If the usual Metropolis acceptance/rejection is also performed on top of this proposal scheme, a variance of the MC algorithm, i.e. the *Smart Monte Carlo*[74] is recovered and the equilibrium distribution sampled is exact.

Employing Smart MC sampling if the energetics are given by QMC noisy calculations is simply too expensive and inefficient for the same reason described in Sec.(2.4.1). Therefore, given that we have QMC ionic forces, we may choose to employ the LD sampling scheme and pay the price of having a time step discretization error. Clearly, also the ionic forces given by QMC calculations will be noisy quantities affected by a statistical error. Nevertheless, the naive reason on why it is better to implement QMC noisy forces in a LD scheme rather than in a Newtonian dynamics is quite clear, since the LD already contains a noise term. Let us start with the first order LD, i.e. the LD that mostly resembles a MC update scheme for the

3N-dimensional positions $\boldsymbol{R}$,

$$\dot{\boldsymbol{R}} \quad = f(\boldsymbol{R}) + \boldsymbol{\eta} \tag{2.46}$$

where $\boldsymbol{f}$ represent the force acting on the configuration $\boldsymbol{R}$ and the noise $\boldsymbol{\eta}$ is connected with the temperature T and is given by the fluctuation-dissipation theorem$\langle \eta_i(t)\eta_j(t')\rangle = \delta(t - t')\delta_{i,j}\, 2kT$. A friction γ, usually present in Langevin equations, would only rescales the time scale in the first order Langevin equation, therefore we put $\gamma = 1$. We also consider the unitary mass equation. The above equation can be easily discretized with time step Δ

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n + \Delta\, \boldsymbol{f}_n + \sqrt{2kT\Delta}\, \boldsymbol{z}_n \tag{2.47}$$

This is an iteration scheme that defines the new variables $\boldsymbol{R}_{n+1}$, at time t_{n+1} , in terms of the $\boldsymbol{R}_n$ at time t_n and of a set of normal gaussian distributed variables (3N-vector) $\boldsymbol{z}_n$. Thus, this iteration represents just a Markov process, that can be implemented by a simple iterative algorithm, once the force $\boldsymbol{f}_n$ can be evaluated for a given position $\boldsymbol{R}_n$ of the N classical particles.

It is important to emphasize that, in this iteration scheme, the noise is proportional $\sqrt{\Delta}$ and dominates, for $\Delta \to 0$, over the deterministic force contribution, which is linear in Δ. For this reason, if the force $\boldsymbol{f}_n$ is evaluated with a statistical error $\delta\boldsymbol{f}_n$ at each step, this statistical noise contribution becomes negligible in the $\Delta \to 0$ limit compared to the thermal noise. In practice, a finite time step must be used, therefore the forces must be evaluated with a statistical noise $|\delta\boldsymbol{f}_n|\Delta$ much smaller than $\sqrt{2kT\Delta}$. Nevertheless, also this condition can be computationally too hard to be met. Therefore the strategy is to employ a generalized second order LD in which the statistical error in the forces can be taken in account without biasing the simulations.

2.5 Generalized second order Langevin dynamics

In this section we describe the framework of the second order LD in which one exploit the freedom given by the fluctuation-dissipation theorem of choosing arbitrary set of matrices for the friction and the power spectrum of the noise. Although Generalized Langevin equations (GLE) have been widely employed in the literature and are at the root of successful algorithm such as the *Colored Noise* technique[81, 82, 83], we encountered some skepticism about the choice in which the power spectrum of the noise and -consequently- the friction matrix explicitly depends on the positions $\boldsymbol{R}$. Therefore we are going to prove the correctness of our GLE both formally and empirically with an illustrative example within DFT.

Let us start with the second order Langevin equations

$$\dot{\boldsymbol{v}} = -\boldsymbol{\gamma}(\boldsymbol{R})\boldsymbol{v} + \boldsymbol{f}(\boldsymbol{R}) + \boldsymbol{\eta}(t) \tag{2.48}$$

$$\dot{\boldsymbol{R}} = \boldsymbol{v} \tag{2.49}$$

$$\langle \eta_{i,\nu}(t)\eta_{j,\mu}(t')\rangle = \alpha_{i\nu,j\mu}\delta(t-t') \tag{2.50}$$

In order to derive the fluctuation-dissipation relation between the α and γ matrices we consider the time-discretized version of the above equation

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n - \Delta\boldsymbol{\gamma}(\boldsymbol{R})\ \boldsymbol{v}_n + \Delta\ \boldsymbol{f}_n + \sqrt{\Delta}\boldsymbol{z}_n \tag{2.51}$$

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n + \Delta\ \boldsymbol{v}_n \tag{2.52}$$

where the covariance matrix of the random vector $\boldsymbol{z}_n$ is given (for $\Delta \to 0$) by

$$\langle \boldsymbol{z}_n \boldsymbol{z}_{n'}\rangle = \delta_{n,n'}\ \boldsymbol{\alpha}(\boldsymbol{R}_n) \tag{2.53}$$

The strategy is now to write explicitly the Master equation for this Markov process (defined on the joint variable $(\boldsymbol{R},\boldsymbol{v})$) and read off the corresponding Fokker-Planck equation in the $\Delta \to 0$ limit. The conditional probability $K(\boldsymbol{R}',\boldsymbol{v}'|\boldsymbol{R},\boldsymbol{v})$ is given by

$$K(\boldsymbol{R}',\boldsymbol{v}'|(\boldsymbol{R},\boldsymbol{v}) = \int d\boldsymbol{z}_n \delta(\boldsymbol{R}'-\boldsymbol{R}-\Delta\boldsymbol{v})\delta(\boldsymbol{v}'-\boldsymbol{v}-\Delta(\boldsymbol{f}-\boldsymbol{\gamma}\boldsymbol{v})-\sqrt{\Delta}\boldsymbol{z}_n)\exp\left[-\frac{1}{2}(\boldsymbol{z}_n\boldsymbol{\alpha}^{-1}\boldsymbol{z}_n)\right] \tag{2.54}$$

we obtain therefore, by using that $P_{n+1}(\boldsymbol{R}',\boldsymbol{v}') = \int\int d\boldsymbol{R}\, d\boldsymbol{v}\ K(\boldsymbol{R}',\boldsymbol{v}'|(\boldsymbol{R},\boldsymbol{v})P_n(\boldsymbol{R},\boldsymbol{v})$,

$$P_{n+1}(\boldsymbol{R}',\boldsymbol{v}') = \int d\boldsymbol{z}_n \exp\left[-\frac{1}{2}(\boldsymbol{z}_n\boldsymbol{\alpha}^{-1}\boldsymbol{z}_n)\right]\mu(\boldsymbol{R}')P(\boldsymbol{R}'-\Delta\boldsymbol{v}',\boldsymbol{v}-\Delta(\boldsymbol{f}-\boldsymbol{\gamma}(\boldsymbol{R}')\boldsymbol{v}')-\sqrt{\Delta}\boldsymbol{z}_n) \tag{2.55}$$

where the measure factor is $\mu(\boldsymbol{R}') = [1+\Delta\ Tr\boldsymbol{\gamma}(\boldsymbol{R}')] + \mathcal{O}(\Delta^{3/2})$. We can now expand in small Δ and read off the Fokker-Planck equation

$$\frac{\partial P(\boldsymbol{R},\boldsymbol{v},t)}{\partial t} = \left[-\frac{\partial}{\partial \boldsymbol{R}}\boldsymbol{v} + \frac{\partial}{\partial \boldsymbol{v}}\left[\boldsymbol{\gamma}(\boldsymbol{R})\boldsymbol{v}-\boldsymbol{f}(\boldsymbol{R})\right] + \frac{\partial}{\partial \boldsymbol{v}}\left[\frac{\boldsymbol{\alpha}(\boldsymbol{R})}{2}\frac{\partial}{\partial \boldsymbol{v}}\right]\right]P(\boldsymbol{R},\boldsymbol{v},t) \tag{2.56}$$

The desired equilibrium distribution (which is time independent) is the unnormalized Boltzmann distribution,

$$P_{eq}(\boldsymbol{R},\boldsymbol{v}) = \exp\left[\frac{\frac{1}{2}\boldsymbol{v}^2 + U(\boldsymbol{R})}{kT}\right] \tag{2.57}$$

Therefore in order to make Eq. 2.57 a stationary solution of Eq. 2.56, the following relation must hold:

$$\boldsymbol{\gamma}(\boldsymbol{R}) = \frac{\boldsymbol{\alpha}(\boldsymbol{R})}{2kT} \tag{2.58}$$

In this way we have proven the generalized fluctuation-dissipation relation which prescribes how to set the friction tensor $\gamma(\boldsymbol{R})$, given the temperature T and the noise $\alpha(\boldsymbol{R})$, in order to sample the correct Boltzmann distribution of Eq.(2.57).

2.5.1 An illustrative example for DFT

In this section we describe one possible choice -among many others as we will see- for the position depend $\gamma(\boldsymbol{R})$ tensor. We are going to show that it is possible to enhance the efficiency of the sampling by modifying the standard fluctuation-dissipation relation, adopting a position dependent friction tensor. In this example, we focus on liquid hydrogen at high pressure. The forces in Eq.(2.48) are obtained within standard LDA-DFT, since the purpose of this section is to show the computational gain connected only with a GLE dynamics displaying a non-trivial choice for the friction/noise matrices, i.e. $\gamma \neq \gamma_0 \boldsymbol{I}$.

An ubiquitous problem in MD is the presence of multiple time scale motion. The separation of time scales occurs when both high and low frequency degrees of freedom appears simultaneously in the system[84], for example, in molecular liquids the bond vibrations usually occur on a time scale which is short compared to the translations and rotations. In these cases standard MD integrators are inefficient as the maximum time step Δ that can be used to integrate the equations of motion must be chosen to insure accurate integration of the highest frequency motion present in the system. In particular, if the time step Δ is large, two atoms originally far apart at time t_n may become dangerously close at the next step t_{n+1} due to a random displacement $\boldsymbol{\eta}(t_n)$ given by thermal fluctuations. This causes serious instabilities in the MD and constrains the size of the time step Δ. Ideally one would like to choose a time step which is different for each spacial direction and, for example, perform a smaller displacement along the line that connects two atoms.

We show that a position and velocity dependent friction tensor may alleviate this problem. Let us begin with a simpler system, that is two particle a, b in one dimension and a (2×2) non-diagonal friction matrix given by

$$\bar{\gamma} = \begin{pmatrix} \gamma_0 & -\delta \\ -\delta & \gamma_0 \end{pmatrix} \tag{2.59}$$

If $\delta = 0$ we recover the standard case in which the friction is a diagonal matrix with scalar intensity γ_0. The LD explicitly reads

$$\dot{v}_a = -\left(\gamma_0 v_a - \delta v_b\right) + f_a + \eta_a \tag{2.60}$$

$$\dot{v}_b = -\left(\gamma_0 v_b - \delta v_a\right) + f_b + \eta_b \tag{2.61}$$

and the two equations are now coupled. Let us put now $\delta > 0$. We note that if the two particles are approaching, i.e. $v_a/v_b < 0$ then the effective friction acting on a and b is greater than the original γ_0. It is clear that this choice greatly reduces the possibility that the two particles coincides and therefore gives a much more stable dynamics. If the two particle are moving in the same direction instead, then this motion is less damped and therefore the

collective translation of the two particle in one direction is enhanced and the "molecule" may visit the space more efficiently. This illustrative example may be extended to realistic systems in several way. For example, we can define the $3N \times 3N$ friction matrix in such a way that we enhance the friction along the radial direction between two particle whose relative distance $r \equiv |\boldsymbol{R}_i - \boldsymbol{R}_j|$is under a cutoff r_c,

$$\bar{\gamma}_{i,\mu;j,\nu} = \left(\gamma_0 + \delta \sum_k |u_\mu^{i,k}| \varepsilon_{i,k} \delta_{\mu,\nu} \right) \delta_{i,\mu;j,\nu} \; - \; \delta \, |u_\mu^{i,j}| \delta_{\mu,\nu} \varepsilon_{i,j} \tag{2.62}$$

with

$$\varepsilon_{i,j} = \begin{cases} 1, & \text{if } |\boldsymbol{R}_i - \boldsymbol{R}_j| < r_c \\ 0, & \text{otherwise.} \end{cases} \tag{2.63}$$

where i, j are the particle indexes, $\mu, \nu = \{x, y, z\}$ and $\boldsymbol{u}^{i,j}$ is the unit vector $\boldsymbol{u}^{i,j} = \boldsymbol{R}_i - \boldsymbol{R}_j / |\boldsymbol{R}_i - \boldsymbol{R}_j|$ denoting the direction of the relative $i - j$ "bond". With this choice, only particles whose distance is below r_c are effectively damped along the interparticle separation $\boldsymbol{u}$.

We discovered that this particular choice of the friction matrix leads to equations that are very similar to the *Diffusive particle dynamics* technique, originally proposed in the context of hydrodynamics simulations[85] and the so-called *Peters thermostat*[86, 87] employed in mesoscopic simulations of liquids. Nevertheless, to our knowledge, it is the first time that these equations are explicitly devised with the goal of stabilizing a Langevin dynamics and applied to an atomistic system rather than represent merely a convenient thermostat for hydrodynamics mesoscopic simulations.

We employ this technique for simulating liquid at high pressure in the molecular phase (at a density of r_s=1.8 and target temperature of 600 K, N=64) with DFT[2]. We choose the cutoff $r_c = 1.6$ Bohr, that is, slightly larger than the equilibrium molecular distance. In this way, with $\delta > 0$ we will slow down more the intramolecular vibration compared to the intermolecular relaxations and the molecular rotations. Nevertheless, our choice is general and does not require an a-priori definition of "rigid molecules", as the pairs that feel the additional friction may vary during the simulation. In Fig.(2.1) and (2.2) we show that the GLE with $\delta > 0$ does not introduce any bias in the simulation. Moreover, the $\delta > 0$ simulations always perform better for fixed time-step Δ than the standard diagonal friction, as the finite time-step error is greatly reduced. Overall, these simulations look more stable and a larger integration time step can be used. In Fig.(2.1) is shown the convergence of the pressure P and temperature T as a function of the Δ, while in In Fig.(2.2) we plot the radial pair distribution functions $g(r)$ for the two LD (GLE and standard) at fixed Δ.

[2]We use a modified version of the QuantumEspresso code[88] in which we have implemented this GLE. We use the LDA XC functional.

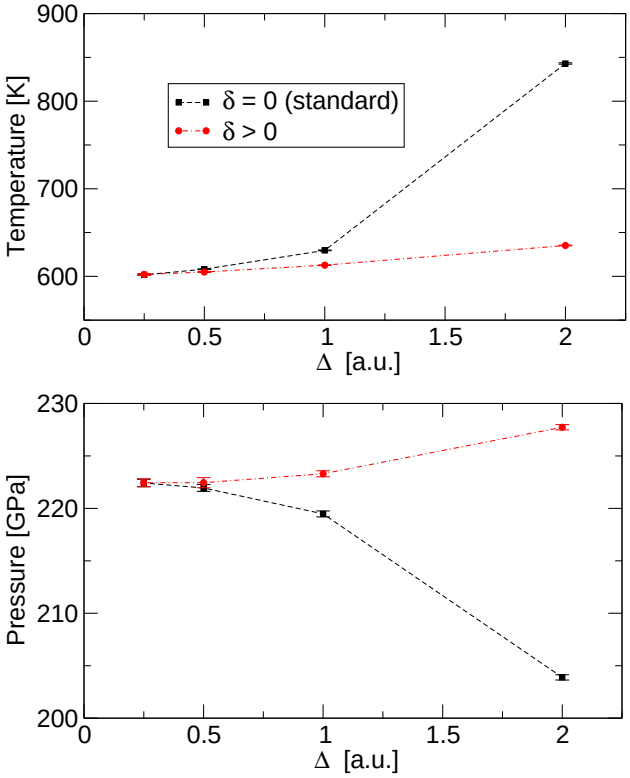


Figure 2.1: Convergence of temperature T and pressure P as a function of the GLE time step Δ. Black points refer to a standard second order LD with diagonal friction $\gamma = \gamma_0 I$ while red point to a GLE with friction given by Eq.(2.62). We put $\gamma_0 = 1.2$ a.u. and $\delta = 1.6$ a.u. The GLE simulations with $\delta > 0$ extrapolate better than the standard LD.

As expected Eq.(2.62) reduces the time step error for all these quantities. Clearly it is not possible to eliminate entirely this discretization error by solely acting on the friction within second order LD. Instead, this should be possible only within the first order Langevin equation framework, as in this case the friction completely rescale the times. However, as shown in Sec(A.2), the first order discretized equation with position dependent friction and noise are much more involved and difficult to implement in atomistic simulation codes.

In order to demonstrate that this technique really represents an improvement with respect to the standard LD it is necessary to show that with Eq.(2.62) we are damping only selected modes and not all the degrees of freedom, uniformely, with a friction which is $\approx \gamma_0 + \delta$. Indeed, the eigenvalues spectrum of the friction matrix Eq.(2.62) shows that M modes are always completely untouched (the friction remains γ_0) while the others $3N - M$

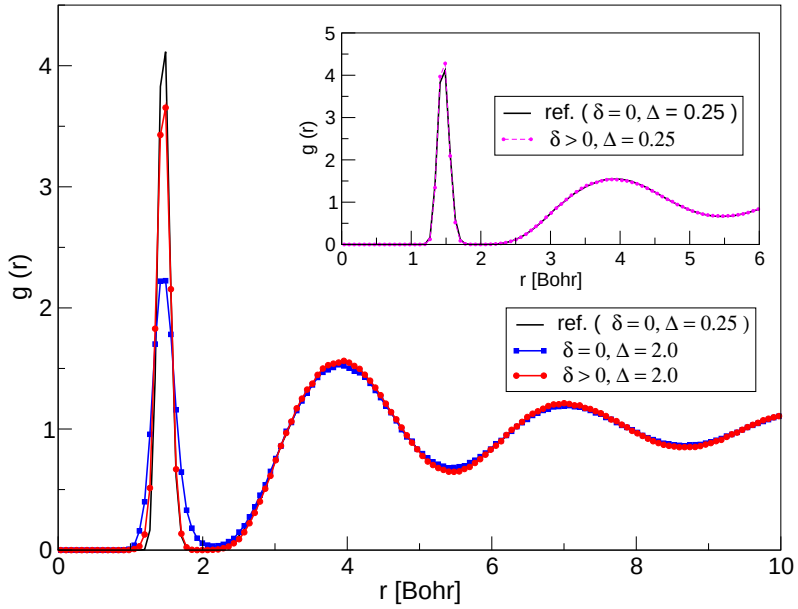
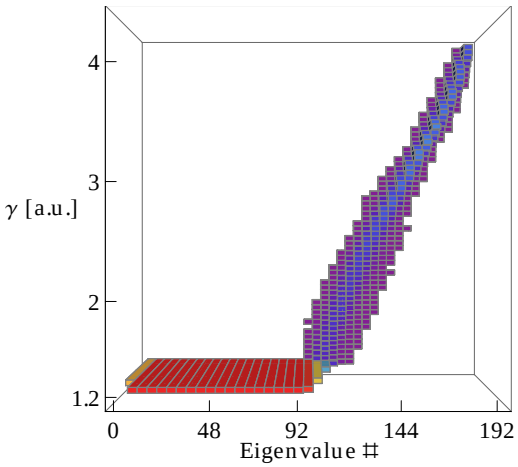


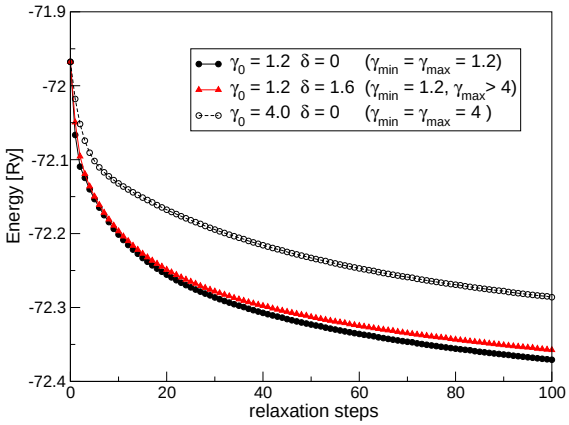
Figure 2.2: Radial pair distribution functions. Blue points refer to a standard second order LD with diagonal friction $\gamma = \gamma_0 I$ and $\Delta = 2$ a.u. while red point to a GLE with friction given by Eq.(2.62) again and $\Delta = 2$ a.u. The red GLE curve displays a profile which is in better agreement with the correct $\Delta \to 0$ function. We put $\gamma_0 = 1.2$ a.u. and $\delta = 1.6$ a.u. As comparison we plot also a converged standard LD profile (black line). Finally in the inset we show that the same limiting curve is obtained for small Δ for both the standard (black) and the GLE (pink).

are damped with a friction $\gamma_{eff} > \gamma_0$. The ratio $M/3N$ depends on the choice of the cutoff r_c. The minimum eigenvalue is $\gamma_{min} = \gamma_0$, while the maximum is $\gamma_{max} > \gamma_0 + \delta$ (see Fig.(2.3(a))). Notice that few modes are effectively damped with a friction that is 3 times larger than $\gamma 0$. In Fig.(2.3(b)) we show that global relaxations, starting from an out-of-equilibrium protonic configuration, are instead not damped by Eq.(2.62). Therefore the larger friction only correspond to the intramolecular radial motion and does not affect the relative intermolecular motion (slow modes).

In conclusion, we have provided a simple example of GLE that improves the $\Delta \to 0$ convergence of the simulations and allows to employ larger time steps, thus providing less autocorrelated samples. This particular choice, that can be eventually improved, can be adopted, without any effort, in every DFT or empirical force field code.



(a) Friction matrix eigenvalues



(b) Structural relaxations for different γ_0 and δ

Figure 2.3: Friction eigenvalues. **a** Histogram of the $3N$ eigenvalues of the friction matrix given by Eq.(2.62) with $\gamma_0 = 1.2$ a.u. and $\delta = 1.6$ a.u. The spectrum goes from a minimum value of $\gamma_{min} = \gamma_0$ to a maximum $\gamma_{max} > 4$. **b** Structural relaxation employing zero-temperature GLE. Black solid circles points refer to a standard second order LD with diagonal friction $\gamma = \gamma_{min} I$ while red triangles to a GLE with friction given by Eq.(2.62). Black open circles refer instead to a standard second order LD with diagonal friction given by $\gamma_{max} = 4$. Notice that these two latter setting are far from being equivalent since, in the GLE case the γ_{max} value only applies for few intramolecular degrees of freedom that do not affect the structural optimization.

2.5.2 QMC noise correction

The basic idea underlining the *QMC noise correction* technique is the following: ideally one would like to follow the Newton's equations

$$\dot{v} = f(R) \tag{2.64}$$

$$\dot{R} = v \tag{2.65}$$

as in DFT. Instead, within QMC, the forces F are given with some statistical error η^{QMC}, which depends on the length of the QMC electronic simulation. This η^{QMC} prevent us from integrate the Newtonian dynamics since the total energy is not conserved. This effect can be seen even in the simple case of H_2 molecule(see Fig.2.4). Therefore, we can not employ QMC noisy forces in real-time MD simulations but still we can use them for Langevin sampling. In order to recover a Langevin equation instead and, at least, sample the Boltzmann distribution, following Eq.(2.48), we should take in account the QMC noise by introducing a suitable damping γ^{QMC},

$$\dot{v} = -\gamma^{\mathrm{QMC}}(R)v + f(R) + \eta^{\mathrm{QMC}} \tag{2.66}$$

$$\dot{R} = v \tag{2.67}$$

$$\langle \eta^{\mathrm{QMC}}_{i,\nu}(t)\eta^{\mathrm{QMC}}_{j,\mu}(t')\rangle = \alpha^{\mathrm{QMC}}_{i\nu,j\mu}\delta(t-t') \tag{2.68}$$

The effective damping γ^{QMC} is connected with the strength of the QMC noise power spectrum α^{QMC} by the fluctuation dissipation theorem. Therefore, the larger is the noise, the larger is the effective friction that will be artificially put in the MD.

Let us now describe the QMC noise correction equations in more detail. For the time being, we are going to illustrate such framework in the case of the simple integrator scheme for the Langevin dynamics. Once again we start from the Langevin equations of motions,

$$\begin{aligned}
\dot{v} &= -\gamma(R)\cdot v + f(R) + \eta(t) && (2.69)\\
\dot{R} &= v && (2.70)\\
\langle \eta(t)\rangle &= 0 &&\\
\langle \eta_i(t)\eta_j(t')\rangle &= \alpha_{ij}(R)\,\delta(t-t') &&
\end{aligned}$$

where R, v, f are the $3N$-dimensional vectors made respectively by the positions, the velocities and the forces of the M nuclei, the indexes i, j run over all the $3N$ nuclear coordinates, and η is a $3N$-dimensional vector representing the noise force, that is determined by the fluctuation-dissipation theorem of Eq.(2.58). Since in Eq.(2.58) one of the two matrices is arbitrary, this time we can choose $\alpha(R)$ in the following form:

$$\alpha(R) = \alpha_0 I + \Delta_0 \alpha^{\mathrm{QMC}}(R) \tag{2.71}$$

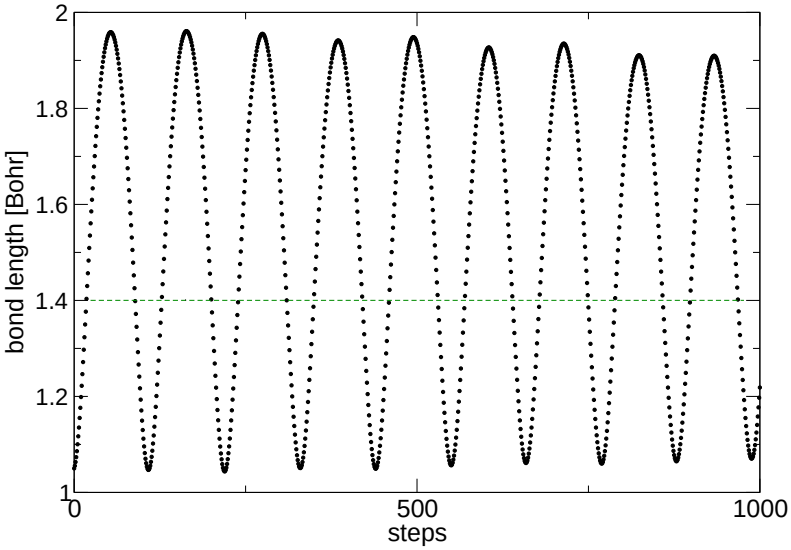


Figure 2.4: Vibration of the H_2 molecule simulated with Newtonian dynamics (Verlet algorithm), using VMC noisy forces. The bin length of each VMC electronic simulation is 2000, that is quite large, given that the variational trial wavefunction contains only 20 parameters. We can see that the oscillations becomes more and more damped as the simulation proceeds, due to the VMC noise in the forces. Notice also that the period of the vibrations is in good agreement with the experimental one that is $1/1.3 \times 10^{14} Hz \approx 320$ a.u. in atomic units of time (Hartree^{-1}). Indeed, from the plot we can estimate the period being in ≈ 333 a.u., since the integration time $\Delta \approx 3.03$a.u. The $\approx 6\%$ deviation from the experimental value is due to the anharmonicity effect, as in our simulation the bond length oscillations extend far away from the equilibrium distance 1.4 Bohr. Nevertheless this is a good quality check for the VMC forces.

where the $3N$-dimensional matrix $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$ defines the correlation of the forces in QMC:

$$\alpha_{ij}^{\mathrm{QMC}}(\boldsymbol{R}) = \left\langle \left(f_i(\boldsymbol{R}) - \langle f_i(\boldsymbol{R})\rangle\right)\left(f_j(\boldsymbol{R}) - \left\langle f_j(\boldsymbol{R})\right\rangle\right)\right\rangle \tag{2.72}$$

where $\langle\cdot\rangle$ indicates the statistical average over the QMC sampling. $\boldsymbol{I}$ is the identity matrix, α_0 is a constant that should be optimized to minimize the autocorrelation time and therefore the efficiency of the sampling. Indeed, it is better for practical purposes (and for the stability of the MD) to add an external and QMC-independent friction to the Langevin dynamics. In the limiting case of very low QMC noise the LD can be eventually underdamped providing

an unoptimal sampling[89, 81]. We may discretize Eqs.(2.69) with the simple

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n - \Delta\boldsymbol{\gamma}(\boldsymbol{R})\, \boldsymbol{v}_n + \Delta\, \boldsymbol{f}_n + \boldsymbol{\eta}_n \tag{2.73}$$

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n + \Delta\, \boldsymbol{v}_n \tag{2.74}$$

with the discrete (time integrated) noise correlator at time t_n is given by

$$\langle \eta_i \eta_j \rangle = \boldsymbol{\alpha} = \frac{2T\gamma_n}{\Delta} \tag{2.75}$$

Since the forces already contains the QMC noise at each time step, the true external noise that we have to put in the LD is

$$\boldsymbol{\eta}^{\text{ext}} = \boldsymbol{\eta} - \boldsymbol{\eta}^{\text{QMC}} \tag{2.76}$$

Since the external noise and the QMC noise are clearly independent the corresponding correlator matrix is

$$\langle \eta_i^{\text{ext}} \eta_j^{\text{ext}} \rangle = \boldsymbol{\alpha}^{\text{ext}} = \boldsymbol{\alpha} - \boldsymbol{\alpha}^{\text{QMC}} \tag{2.77}$$

which is a positive definite matrix provided that $\Delta_0 \geq \Delta$. In this way, the QMC noise rescale the friction matrix but is not "double counted" in the thermal noise $\boldsymbol{\eta}$ that one puts in the Langevin simulation, since it is already present in the forces $\boldsymbol{f}$. In Fig.2.5) we report an illustrative example of this connection between the statistical QMC noise and the damping of the MD. The main advantage of this technique is that the QMC noise on the total energy and forces can be also larger than the target temperature T.

2.5.2.1 Large friction integration scheme

Within this framework, the friction can be quite large due to the QMC contribution. Therefore, we developed a more accurate integration scheme (see Ref.[29]) that allows us to employ a large time step Δ even in presence of large friction γ, i.e. without suffering from the stability constrain $\Delta\gamma < 1$ like in standard integration scheme defined in Eqs.(2.73).

When the friction matrix $\gamma(\boldsymbol{R})$ is large, the velocities can have strong variations in the discrete time integration step Δ. Indeed, we now assume only that in the time interval

$$t_n - \tau/2 < t < t_n + \tau/2,$$

n indexing the time steps, the positions $\boldsymbol{R}$ are changing a little and, within a good approximation, we can neglect the $\boldsymbol{R}$ dependence in the right hand side of Eq.(2.69). Moreover the velocities $\boldsymbol{v}_n$ are computed at half-integer times $t_n - \tau/2$, whereas coordinates $\boldsymbol{R}_n$ are assumed to be defined at integer times:

$$\boldsymbol{v}_n \equiv \boldsymbol{v}(t_n - \tau/2) \tag{2.78}$$

$$\boldsymbol{R}_n \equiv \boldsymbol{R}(t_n) \tag{2.79}$$

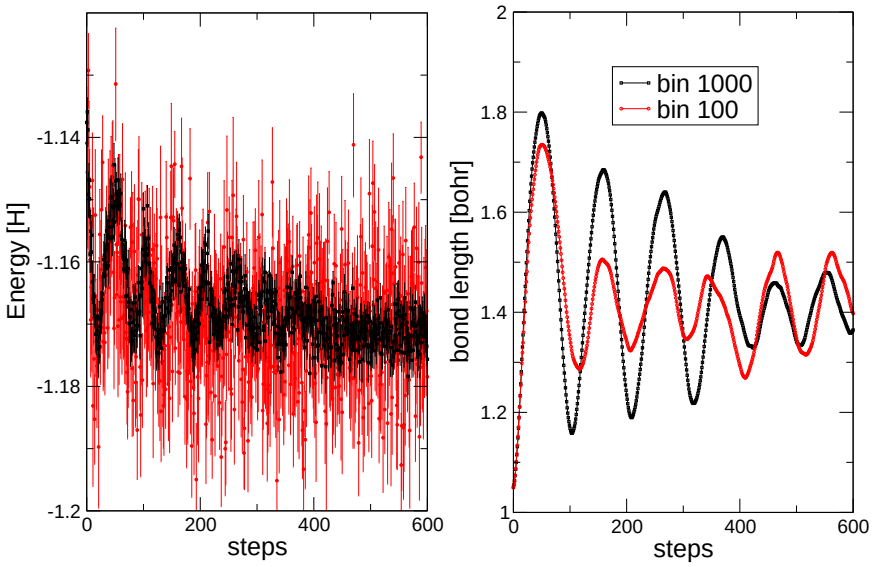


Figure 2.5: H_2 molecule simulated with GLE with noise correction for two different values of the VMC bin length. Left Panel: total energy as a function of the MD steps. The smaller the bin, the greater are the error bars. Right panel: bond length as a function of the MD steps. The smaller the bin, the more damped are the oscillations.

and the quantities that are functions of $\boldsymbol{R}$ in Eq.(2.69) are calculated in $\boldsymbol{R}_n$:

$$\boldsymbol{f}_n \equiv \boldsymbol{f}(\boldsymbol{R}_n) \tag{2.80}$$

$$\boldsymbol{\gamma}_n \equiv \boldsymbol{\gamma}(\boldsymbol{R}_n). \tag{2.81}$$

Then the solution can be given in a closed form:

$$\boldsymbol{v}_{n+1} = e^{-\boldsymbol{\gamma}_n \tau} \boldsymbol{v}_n + \boldsymbol{\Gamma}_n \cdot (\boldsymbol{f}_n + \tilde{\boldsymbol{\eta}}) \tag{2.82}$$

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n + e^{-\boldsymbol{\gamma}_n \tau/2} \cdot \boldsymbol{\Gamma}_n \cdot \boldsymbol{v}_n + \boldsymbol{\Theta}_n \cdot (\boldsymbol{f}_n + \tilde{\tilde{\boldsymbol{\eta}}}) \tag{2.83}$$

where we have introduced the following quantities:

$$\boldsymbol{\Gamma}_n = \gamma_n^{-1}(\boldsymbol{I} - e^{-\gamma_n \tau}) \tag{2.84}$$

$$\boldsymbol{\Theta}_n = \gamma_n^{-1}(\tau \boldsymbol{I} - e^{-\gamma_n \tau/2}\boldsymbol{\Gamma}_n) \tag{2.85}$$

$$\tilde{\boldsymbol{\eta}} = \frac{\gamma_n}{2\sinh(\gamma_n \tau/2)} \int_{t_n-\tau/2}^{t_n+\tau/2} e^{\gamma_n(t-t_n)}\boldsymbol{\eta}(t)\, dt \tag{2.86}$$

$$\tilde{\tilde{\boldsymbol{\eta}}} = \boldsymbol{\Theta}_n^{-1} \int_{t_n}^{t_{n+1}} dt \int_{t_n-\tau/2}^{t} dt' e^{\gamma_n(t'-t)}\boldsymbol{\eta}(t') \tag{2.87}$$

By using that $[\boldsymbol{\alpha}, \gamma] = 0$ and a little algebra, the correlator defining the discrete (time integrated) noise can be computed and gives:

$$\langle \tilde{\eta}_i \tilde{\eta}_j \rangle \equiv \bar{\alpha}_{ij}^{1,1} \tag{2.88}$$

$$\left\langle \tilde{\tilde{\eta}}_i \tilde{\tilde{\eta}}_j \right\rangle \equiv \bar{\alpha}_{ij}^{2,2} \tag{2.89}$$

$$\left\langle \tilde{\eta}_i \tilde{\tilde{\eta}}_j \right\rangle \equiv \bar{\alpha}_{ij}^{1,2} \tag{2.90}$$

$$\langle \tilde{\tilde{\eta}}_i \tilde{\eta}_j \rangle \equiv \bar{\alpha}_{ij}^{2,1} = \bar{\alpha}_{ij}^{1,2} \tag{2.91}$$

where the values of $\bar{\alpha}_{ij}^{1,1}$, $\bar{\alpha}_{ij}^{1,2}$ and $\bar{\alpha}_{ij}^{2,2}$ are defined by the following $3N$-dimensional matrices:

$$\begin{aligned}
\bar{\boldsymbol{\alpha}}^{\mathbf{1,1}} &= 2T\,\gamma_n^2 \frac{\sinh(\gamma_n \tau)}{[2\,\sinh(\gamma_n \tau/2)]^2} \\
\bar{\boldsymbol{\alpha}}^{\mathbf{2,2}} &= 2T\,\gamma_n^{-1} \cdot \boldsymbol{\Theta}_n^{-2} \cdot \\
&\quad \left[\tau \boldsymbol{I} + \gamma_n^{-1}\left(-\boldsymbol{I} + e^{-\gamma_n \tau}/2 + e^{-2\gamma_n \tau} - e^{-3\gamma_n \tau}/2\right)\right] \\
\bar{\boldsymbol{\alpha}}^{\mathbf{1,2}} &= 2T\,\frac{\gamma_n \cdot \boldsymbol{\Theta}_n^{-1} \cdot \left(2e^{\gamma_n \tau/2} - 2\boldsymbol{I} - e^{-\gamma_n \tau} + e^{-2\gamma_n \tau}\right)}{4\gamma_n \sinh(\gamma_n \tau/2)}
\end{aligned}$$

The equations determining the noise correlations are now more complicated, as they involve a 2×2 block matrix $\bar{\boldsymbol{\alpha}}^{a,b}$, where each block is a $3N \times 3N$ submatrix, and $a, b = 1, 2$. Apart for this, the generalization of the noise correction to this case is straightforward, as to each of the four submatrices we have to subtract the $3N \times 3N$ QMC correlation of the forces $\boldsymbol{\alpha}^{\mathrm{QMC}}$, namely

$$\bar{\boldsymbol{\alpha}}_{\mathrm{ext}}^{a,b} = \bar{\boldsymbol{\alpha}}^{a,b} - \boldsymbol{\alpha}^{\mathrm{QMC}} \tag{2.92}$$

is the true external noise we have to add to the system. Now the resulting matrix $\bar{\boldsymbol{\alpha}}_{\mathrm{ext}}$ is indeed positive definite provided $\Delta_0 > \frac{4}{3}\Delta$.

We test the convergence of this improved numerical scheme against the standard discretization scheme of Eqs.(2.73) on an analytically solvable classical toy model (2.6). In this example forces are computed analytically, i.e. $\alpha_{QMC} = 0$, nevertheless the gain obtained by this new discretization scheme is clear.

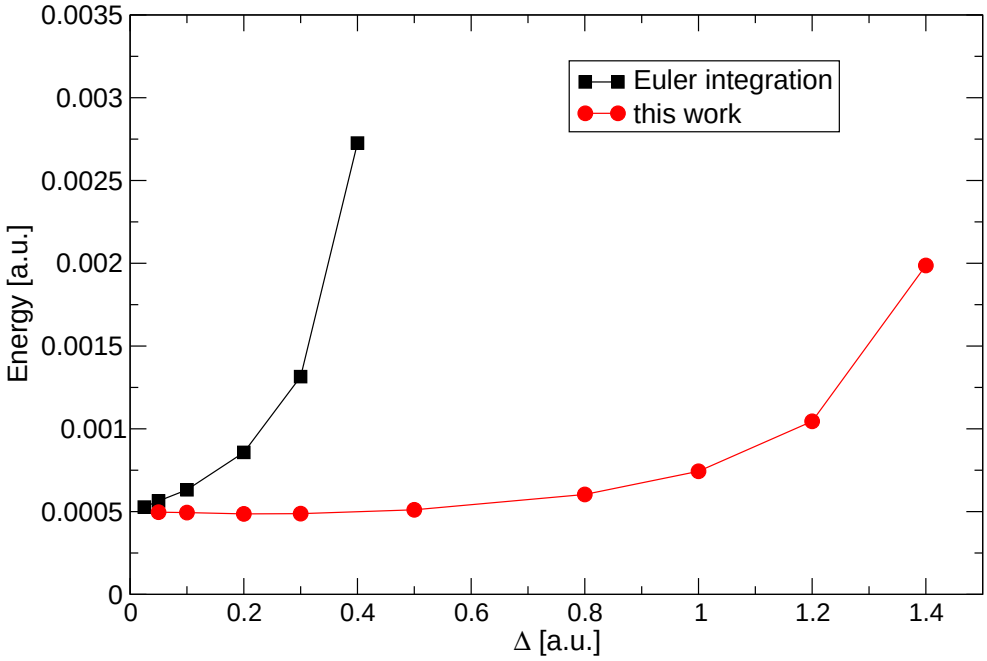


Figure 2.6: Convergence of total energy as a function of the time step Δ for fixed friction and temperature (arbitrary unit). The toy model consists in a 2D particle subject to a radial potential $U(r) = k(r - r_0)^2$. The new sampling scheme (red) is more accurate for larger Δ than the standard so-called *Euler* discretization scheme (black).

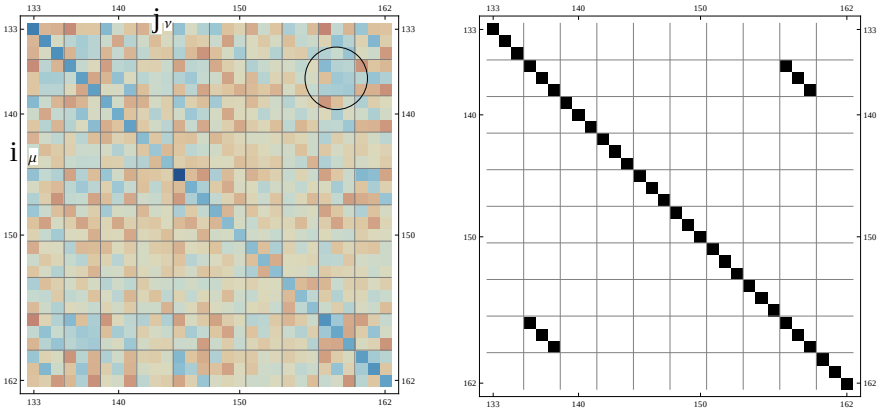
2.5.2.2 Covariance matrix of the QMC forces

The advantages given by the $\boldsymbol{\alpha}^{\mathrm{QMC}}$ as a key ingredient of the MD are essentially two, one is related to the stability of the MD against the QMC fluctuations, the other is an increase in the efficiency of the sampling.

The first one is simple to describe. Let us suppose that the component f_k of the total force $\boldsymbol{f}$ has a large fluctuation due to the QMC statistical errors, i.e. $f_k = f_k^{\mathrm{true}} + \eta_k^{\mathrm{QMC}}$ so that the estimated forces is very different from the value f_k^{true} which one would obtain with a QMC run with infinite bin size. Therefore, the corresponding change in the velocity along the component v_k is great and may eventually lead to a serious instability in the dynamics. If instead the friction γ_k is proportional to such noise fluctuation the velocity is damped by this factor and the corresponding ionic displacement along this direction R_k is reduced. Thus the probability that an unfortunate QMC noise fluctuation may cause an instability in the MD is

also reduced.

The second advantage is less evident. Since the friction matrix $\gamma(\boldsymbol{R})$ is now position dependent, it encodes valuable informations about the system. We already showed the efficiency gain connected with a suitable position dependent friction in Sect. 2.5.1. In that case,$\gamma(\boldsymbol{R})$ was arbitrarily chosen by us, following a reasonable stability consideration. In the noise correction case, the friction is automatically set by $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$. Nevertheless, also the $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$ contains informations on which pairs of atoms are closer, as the matrix $\bar{\gamma}(\boldsymbol{R})$ defined in Sect. 2.5.1, since the statistical QMC noise of the forces is more correlated on the nearest neighbours atoms than on atoms that are placed at large distance. In Fig.(2.7) we show that, for the same configuration $\boldsymbol{R}$, $\bar{\gamma}(\boldsymbol{R})$ and $\gamma^{\mathrm{QMC}}(\boldsymbol{R})$ share some similarities. Therefore, by using the position dependent $\gamma^{\mathrm{QMC}}(\boldsymbol{R})$ one also automatically increases the friction on the direction along two nearest neighbours atoms.



(a) $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$ matrix (b) $\bar{\gamma}(\boldsymbol{R})$ matrix

Figure 2.7: Small portion of the $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$ and $\bar{\gamma}(\boldsymbol{R})$ matrices evaluated on the same 64 atom configuration $\boldsymbol{R}$. The off-diagonal terms of the $\bar{\gamma}(\boldsymbol{R})$ matrix, defined in in Sect. 2.5.1, basically gives the information on which pair of atoms i, j are close (see upper-left/bottom-right corners of these submatrices). The indexes i, j run on the particle, while $\mu, \nu = \{x, y, z\}$. Although the $\boldsymbol{\alpha}^{\mathrm{QMC}}(\boldsymbol{R})$ submatrix is quite noisy, it has (on average) larger entries on the same block i, j which has nonzero elements in $\bar{\gamma}(\boldsymbol{R})$ (highlighted with a circle). Therefore the covariance matrix of the forces has already in itself information on the relative distances of the particles within the configuration $\boldsymbol{R}$. The color legend in the left panel is: red $\rightarrow$ low values; blue $\rightarrow$ high values.

Besides from these qualitative considerations, it was recently discovered that the covariance matrix $\boldsymbol{\alpha}^{\mathrm{QMC}}$ obtained with QMC is empirically proportional to the dynamical matrix (see Ref.[90]). Therefore with a finite and large Δ_0 the high energy modes with high frequency vibrations can be systematically damped, and this clearly allows a faster propagation with larger time step Δ.

2.5.2.3 Efficiency of the MD and MC sampling

The GLE with noise correction is very convenient with respect to the MC sampling with Penalty technique introduced in Sect. 2.4.1. Let us briefly summarize the advantages of this scheme

- The ionic forces provide an efficient way to generate global move.
- The noise correction technique allows to employ a large integration time step Δ, therefore we obtain less autocorrelated samples during the MD iteration scheme.
- Within GLE we are not forced to reduce the error bar on the total energy to order $\sim kT$, therefore the computational saving per MD step is substantial.

These positive features lead to a dramatically improved sampling of the ionic configurational space. Indeed, in Fig.2.8, the efficiency of the MD based sampling with respect to the CEIMC technique -on the very same system- is clear, since we can afford a larger number of global sampling updates. Therefore, within this MD-GLE framework we can obtain ergodic and well equilibrated simulations.

In conclusion, this QMC-MD with noise correction provides an efficient framework for sampling realistic systems and evaluate equilibrium properties. The noise correction technique provides a stable and efficient MD Langevin simulation. In Sect. 3.5.1 a systematic study on the bias given by the finite QMC bin size N_{QMC} is shown. The MD is very robust against the noise fluctuations and quite small QMC bins can be used, i.e $N_{\mathrm{QMC}} \sim \sharp$ of variational parameters. Nevertheless, all the possible sources of bias in this ab-initio MD are systematically controllable, namely

- On the electronic side, the trial wavefunction can be systematically improved, reaching the complete basis set limit. In practice, within VMC, one has to carefully check that the all observables of interest are converged. Moreover, in principle, one can check the accuracy of the result with a fixed node DMC calculation.
- During the dynamics we must stay close to the Born-Oppenheimer potential energy surface $E(\boldsymbol{R})$. We can always increase the number of iteration steps of the optimization procedure in order to meet this requirement. However, within QMC, we cannot precisely fulfill the BO constraint. Instead, we can only stay close to the potential energy within an error that scales as $1/\sqrt{N_{\mathrm{QMC}}}$. This bias can not be eliminated with the noise correction technique. However, also this bias can be removed in the $N_{\mathrm{QMC}} \to \infty$ limit. In all the practical applications we can check that the BO surface is followed within an acceptable error $\delta E_{\mathrm{BO}}/\mathrm{atom} < kT$.

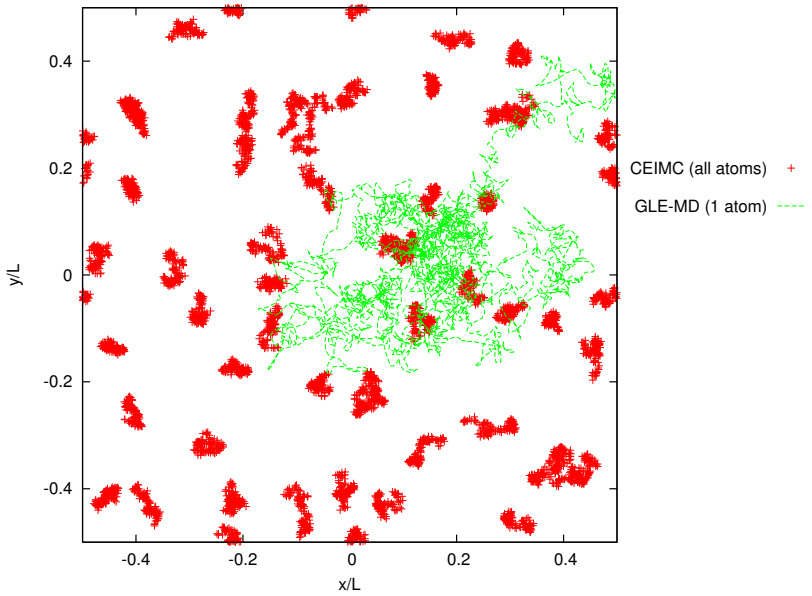


Figure 2.8: Sampling efficiency comparison between GLE-MD and CEIMC. We plot the rescaled ionic positions -projected on the xy plane- sampled during a CEIMC simulation (red points) and a noise-corrected GLE one (green line). For sake of visualization we plot the dynamics of only one ion out of the 54 hydrogen system in the case of the GLE-MD. The different performance of the two schemes is clear. In the CEIMC simulation[91] we observe only small vibrations around the equilibrium positions, therefore much longer simulations are necessary to equilibrate the system. On the contrary, the GLE-MD provides a much more efficient sampling in this case. The liquid hydrogen system is simulated at temperature $T = 600K$, and at density $r_s = 1.35$.

- The finite time step integration error of the MD can be controlled by inspecting the limit $\Delta \to 0$. This error is still present in the noise correction technique, altough can be effectively reduced. Notice that, we cannot employ a very small Δ since we have a finite computational time, i.e. we can afford a finite number M of MD iteration steps, and the total integration time $M\ \Delta$ must be larger than the equilibration time of the system in order to sample correct equilibrium properties. Indeed we experienced that, the lack of equilibration introduces a major source of errors in the MD that can substantially bias the outcomes of the simulation.

A final remark on the unit of measure of the LD. Notice that in the above Langevin equations we have assumed that all the masses of the particles are set to unit values in atomic Rydberg units, namely twice the electronic mass $2m_e$ is implicitly set to unity. In order to sample the canonical distribution the actual values of the masses are immaterial, i.e. the equilibrium properties does not depend on the mass. In order to match the usual atomic units, for instance in the hydrogen case, the time units have to be scaled by the square root

of the ratio between the proton mass and twice the electron mass ($\sqrt{m_p/2m_e} \sim 30.3$).

2.6 A Langevin dynamics for the variational parameters

In this short Section we introduce of a novel framework to include thermally excited states in finite temperature electronic simulations. The method described in this Section is not used to obtain the main physical result of this thesis. Nevertheless this framework, and the class of problems on which it could be applied, may still be interesting and therefore is reported in the thesis. At variance with respect to the previous Sections, the focus is on a generalized first order Langevin equation, whose time discretized form looks much more complicated than its second-order counterpart in the case of coordinate dependent noise correlator. All the details are reported in Appendix A.

2.6.1 Introduction

The adiabatic approximation, typically assumed when performing standard Born-Oppenheimer (BO) molecular dynamics, can become unreliable at finite temperature, and specifically when the temperature is larger than the electronic energy gap between the ground state and the low-lying excited states. In this regime, relevant for many important chemical processes, the non-adiabatic couplings between the electronic energy states can produce finite temperature effects in several molecular properties, such as the geometry, the vibrational frequencies, the binding energy and several chemical reaction. In this Section we introduce a novel finite-temperature non-adiabatic molecular dynamics, based on a novel covariant formulation of the electronic partition function. In this framework the nuclei are not constrained to move in a specific electronic potential energy surface. Then, by using a rigorous variational upper bound to the free energy, we are led to an approximate partition function that can be evaluated numerically. The method can be applied to any technique capable to provide an energy value over a given wave function ansatz depending on several variational parameters and atomic positions.

The calculation of finite temperature properties is one of the most important and challenging aspects of the numerical simulations. The basis of most finite temperature computational approaches is the Born-Oppenheimer (BO) separation of the system in the electronic and nuclear subsystems, implying that at zero temperature the nuclei move in a potential energy surface (PES) provided by the electrons. Moreover, when studying systems at non zero temperatures, it is a common practice to consider electrons in their instantaneous ground state while the nuclei can be treated as classical particles following an *ab initio* finite temperature molecular dynamics. This is the scheme followed in the previous Sections, as the force

, which appears in the MD equations, is the gradient of the ground state total energy with respect to the ionic coordinates (see Eq.2.25). These methods implicitly use also an adiabatic approximation, because the BO approximation does not provide only one PES, but several adiabatic potential energy surfaces (PESs), one for each electronic eigenstate. However the nuclei are evolved only according to one adiabatic PES and the non-adiabatic coupling between adiabatic PESs is neglected. This approximation is reliable only if the electronic energy gap is large, namely when the gap between the electronic ground state and the low-lying excited states is much larger than the thermal energy. If the temperature is high enough this is not the case, and it is not correct to assume that the electrons are constrained into one adiabatic PES. This happens also at room temperature in many chemical processes, and effects can be observed in several molecular properties, such as the geometry, the vibrational frequencies, the binding energy and several chemical reactions. Moreover these effects can be important also in many physical phenomena, such as the occurrence of magnetic or insulating phases below a critical temperature, where the electronic entropy cannot be neglected. We aim to account for the possibility of electronic excitations by removing the adiabatic constraint forcing the electrons to remain, during the dynamics, only in a specific PES. This is particularly important for QMC methods as, due to the statistical noise, it is extremely difficult to satisfy exactly the BO constraint. Following the BO derivation, by using the smallness of the ratio m_e/M, the total partition function Z can be expressed in terms of an electronic partition function $Z[\mathbf{R}]$ at fixed nuclei positions:

$$Z = \int d\mathbf{R}\, Z[\mathbf{R}] \tag{2.93}$$

$$Z[\mathbf{R}] = \mathrm{Tr}\, \exp(-H_{\mathbf{R}}/T) \tag{2.94}$$

where T is the temperature (here and henceforth the Boltzmann constant is assumed to be one and we neglet for simplicity the overall constant coming from integration of the atomic momenta), $H_{\mathbf{R}}$ is the standard electronic Hamiltonian, that includes also the classical ionic contribution, and that depends only parametrically upon the atomic positions $\mathbf{R}$. We are assuming here that the temperature is high enough that quantum effects on heavy nuclei can be neglected, so that it is justified to have taken the classical limit for the nuclei. Observe that, within the standard BO approach, whenever the electronic gap is much larger than T, the electronic partition function $Z[\mathbf{R}]$ in (2.94) can be approximated by $\exp(-E_0(\mathbf{R})/T)$ where $E_0(\mathbf{R})$ is the ground state energy of the hamiltonian $H_{\mathbf{R}}$; in other words a single adiabatic PES is implicitly considered in this case.

In the following derivation we want to include the contribution of all the adiabatic PES corresponding to the ground state and all excited states with an affordable computation, because, as emphasized before, considering only the electronic ground state PES (gsPES) may

fail in several cases, even when we are in the limit of small m_e/M. For instance the occurrence of a broken symmetry phase often implies gapless electronic excitations in $H_{\mathbf{R}}$, and the approximation $Z[\mathbf{R}] \sim \exp(-E_0(\mathbf{R})/T)$ cannot be safely assumed. Other examples are conical intersections[92, 93], when, for some particular ionic positions, $H_{\mathbf{R}}$ becomes gapless and nearby the proximity between different (namely corresponding to low-lying excited states) PES is possible. In these conditions a pure electronic ground state technique fails as the interplay between different adiabatic energy surfaces cannot be taken into account consistently.

2.6.2 Sampling the finite temperature electronic partion function

We consider the problem to estimate the finite temperature partition function of an electronic system with N electrons and M atoms, where we assume in the following that, as discussed in the introduction, the ions are classical particles, whose coordinates $\mathbf{R}$ appear just as simple parameters in the electronic hamiltonian $H_{\mathbf{R}}$ and are confined in a finite volume V. Therefore, once the ion positions are fixed, we need to evaluate the electronic partition function:

$$Z[\mathbf{R}] = \mathrm{Tr} \exp\left(-\beta H_{\mathbf{R}}\right) \tag{2.95}$$

where $\beta = 1/T$. Our derivation applies for an Hamiltonian with a bounded spectrum defined in a finite Hilbert space with dimension D. Generally speaking this is not a relevant restriction as, for instance, in electronic structure calculation one can consider a finite dimensional basis of localized orbitals around each atom. The basis and the dimension D can be increased arbitrarily to reach the so called complete basis set limit, that describes a proper continuous electronic system.

We work within the VMC approach (see Sect. 2.3.1) and with a general variational trial wavefunction

$$\psi(\vec{\boldsymbol{r}}) = J(\vec{x})\, D(\vec{x}) \tag{2.96}$$

The real variational parameters, that define the above wave function, are compactly denoted by $\alpha \equiv \{\alpha_i\}_{i=1,\dots,p}$ and, since all physical quantities do not depend on the norm of the wave function, we consider the $\alpha-$manifold of states:

$$|\alpha\rangle = \frac{|\psi\rangle}{\||\psi\rangle\|} \tag{2.97}$$

The metric in this manifold becomes non trivial as, by a straightforward calculation, the distance between two states $|\alpha\rangle$ and $|\alpha + d\alpha\rangle$ is given by:

$$ds^2 = \||\alpha + d\alpha\rangle - |\alpha\rangle\|^2 = d\alpha^i d\alpha^j S_{i,j} \tag{2.98}$$

where summation over repeated indices is assumed, and S is a $p \times p$ matrix defining the metric tensor of this rather non trivial space, described by p independent variational parameters (e.g. a subset of $v_{i,j}$ and ψ_{ij}). The matrix S can be explicitly evaluated and depends only on average first derivatives of the wave function with respect to the parameters $\alpha's$:

$$S_{i,j} = \frac{\langle \partial_i \psi | \partial_j \psi \rangle}{\langle \psi | \psi \rangle} - \frac{\langle \partial_i \psi | \psi \rangle}{\langle \psi | \psi \rangle} \frac{\langle \psi | \partial_j \psi \rangle}{\langle \psi | \psi \rangle} \tag{2.99}$$

It defines a metric as it is strictly positive definite if all p variational parameters are independent (implying indeed a non vanishing determinant $|S| > 0$). This matrix turns out to be exactly the one used in several optimization techniques[53, 94], and can be computed also for correlated systems by sampling the correlations of the quantities $O_j(x) = \frac{\langle x | \partial_j \psi \rangle}{\langle x | \psi \rangle}$ over the configuration space $\{x\}$ where electrons have definite spin and position, namely:

$$S_{i,j} = \langle O_i O_j \rangle - \langle O_i \rangle \langle O_j \rangle \tag{2.100}$$

where the symbol $\langle \dots \rangle$ denotes average over a distribution $\Pi(x) \propto \langle x | \psi \rangle^2$, that can be sampled by standard variational Monte Carlo. Eq. 2.95 can be also extended in the space α with non trivial metric (see App. A), by using the invariant measure $d\alpha^p \sqrt{|S|}$, corresponding to the metric tensor S:

$$\frac{\int d\alpha^p \sqrt{|S|} \langle \alpha | \exp(-\beta H_{\mathbf{R}}) | \alpha \rangle}{Z_S} = \text{Tr } \exp(-\beta H_{\mathbf{R}}) \tag{2.101}$$

We refer the interested reader to the original paper, Ref.[95], for all the mathematical details and the proof of the above equation.

Then, we can easily bound the exact electronic partition function $Z[\mathbf{R}]$, because, due to the convexity of the exponential function, the expectation value of an exponential operator over a normalized state $|\alpha\rangle$ satisfies:

$$\langle \alpha | \exp(-\beta H_{\mathbf{R}}) | \alpha \rangle \geq \exp(-\beta \langle \alpha | H_{\mathbf{R}} | \alpha \rangle).$$

This immediately provides a rigorous lower bound Z_Q for the partition function Z:

$$Z \geq Z_Q = \frac{\int d\mathbf{R} \int d\alpha^p \sqrt{|S|} \exp(-\beta \langle \alpha | H_{\mathbf{R}} | \alpha \rangle)}{Z_S} \tag{2.102}$$

and a corresponding upper bond F_Q for the free energy $F = -T \ln Z$:

$$F \leq F_Q = -T \ln Z_Q \tag{2.103}$$

In this way it is evident that F_Q represents an improvement to the standard practice to consider only the lowest BO energy surface. In fact in this approximation only one state is

assumed to contribute to the integral in Eq.(2.102), namely the lowest energy state of $H_{\mathbf{R}}$ within the ansatz given by $|\alpha\rangle$:

$$E_{BO}[\mathbf{R}] = \min_{\alpha} \{\langle \alpha | H_{\mathbf{R}} | \alpha \rangle\} \tag{2.104}$$

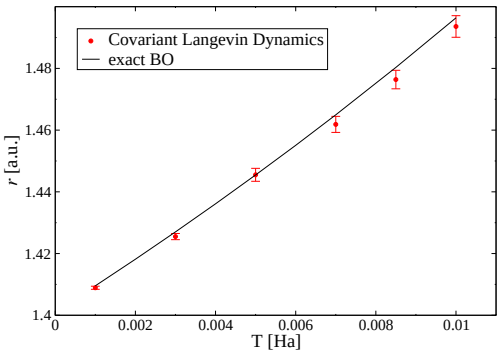
Indeed it is clear that $F = \min_{\mathbf{R}} \{E_{BO}[\mathbf{R}]\}$ only at $T = 0$, and it represents a very bad approximation to F as long as the temperature is raised, whereas the approximate partition function F_Q approaches the correct large temperature limit $-T \ln(DV^M)$ of the exact partition function, while remaining a rigorous upper bound for any T. In Ref.[95] we show in detail a comparison between the approximated partition function Z_Q here introduced, the exact and gsBO ones, showing that our approximation turns out to be better than the gsBO one, above a temperature T^*, that remains meaningful in the thermodynamic limit.

We sample this partition function by means of a first order Langevin equation for the variational parameters α. Such equation of motion turns out to be quite complicated in the first order case, as having a coordinate-dependent diffusion matrix, which is related to the metric tensor S, introduces computationally expensive *drift-diffusion* terms (see App. A).

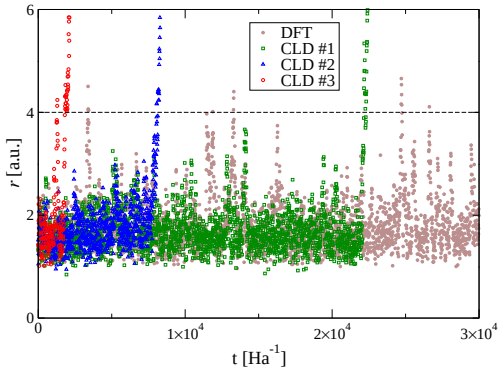
If we couple the Langevin equation for the variational parameters with the one for the nuclei we obtain the set of MD equations which one can integrate numerically in order to simulate realistic systems taking into account electronic entropy. In Fig. 2.9(a) we show that in the $T \to 0$ limit, this MD correctly reproduces the ground state Born-Oppenheimer approximation (see App. A for details).

2.6.3 Results and discussions

This kind of dynamics should behave differently with respect to the standard BO-MD one when the temperature is raised and for $T > T^*$ should be more realistic, because corresponding to a more accurate upper bound of the exact free energy F. In Fig.2.9(a)) we limit the study of the average bond length in a range of temperatures smaller than $3000\ K$ because, above this value, first dissociation events start to appear during the simulations. This temperature is in good qualitative agreement with low pressures experiments[96]. Roughly speaking the dissociation probability depends on the ratio between the thermal energy T and the depth of the free energy well ΔU through the Boltzmann weight[99] $\exp(-\Delta U/T)$ within the assumption that excited electronic eigenstates are well-separated in energy from the ground state. There are instead examples[100] in which BO approximation breaks down, particularly near the transition state of a chemical reaction. In fact, as the reaction coordinate r increases, the energy gap between the ground state and the first (antibonding) excited state becomes smaller[101], for example when $r > 4\ a.u.$ this quantity becomes smaller than $8000K$. Therefore large fluctuations in the bond length, certainly occurring at large tempera-



(a) Average bond length at small T



(b) Dissociation events at $T = 8000$ K.

Figure 2.9: a. Bond length r as a function of temperature. The range of temperatures is well below the electronic gap ~ 0.17 Ha (see Fig.A.1(a)) and therefore the expected exact value is the gsBO one evaluated by eq. (A.16) (black line). Red points are obtained by the MD coupled equations (A.10,A.14). Data are in agreement with the predicted values.
b.Bond length r as a function of simulation time at a temperature of $T = 8000$ K. Coloured points (red, green and blue) correspond to simulations performed with the dynamics presented in this work, while the grey solid ones are obtained with a DFT- first order Langevin gsBOMD (the finite temperature BOMD, defined with the fractional occupation[97, 98] yields qualitatively similar results). The time step used in the integration of the equations is $\Delta_N = 0.1$ Ha^{-1} and points are plotted every 100 iterations. The dashed line indicates the distance r^* such that the energy gap between the ground state PES and the first excited one becomes smaller than T. All the CLD trajectories show escape events while the DFT one describes a stable molecular configuration up to 20×10^4 Ha^{-1} of simulation time (not shown).

tures, are in principle not well described under a BO scheme. Since by definition, an atomic dissociation requires to sample correctly events with large r, we expect to find differences between the standard BO-MD and the dynamics generated by (A.10,A.14), at large enough temperatures. In Fig.2.9(b), we observe that the probability of dissociation is enhanced in our dynamics, which can take implicitly into account also the effective repulsion due to the antibonding state. As expected, this is in sharp contrast with a DFT-BO dynamics obtained using the QUANTUM ESPRESSO package[88, 102]. In the latter dynamics large fluctuations in r do not lead to dissociation, as is partially shown in fig. (2.9(b)), and the H_2 molecule remains stable even when the finite temperature DFT is adopted with the usual practice to work with fractional occupations of the Kohn-Sham energy levels. Indeed no escape events occur within DFT BO-MD, even for long simulations. Moreover in order to compensate the well known overbinding error[37] of the local density approximation (LDA), we have increased the temperature by a factor proportional to the LDA energy barrier ($\sim$ 1.40 times larger than the exact one), and observed no qualitative changes in the trajectories, always confined around the minimum energy value. It is clear therefore that, quite generally, the BO-MD greatly underestimates the evaluation of the reaction rate if, for instance, a *mean first-passage time*[99] analysis is performed. However the time scale used in this first order Langevin dynamics does not have a real physical meaning and accurate transition rates cannot be computed without extending the method to second order dynamics, by taking into account also the mass of the particles.

Chapter 3

High pressure phase diagram of liquid hydrogen

In this Chapter we introduce the main physical results of this thesis. The focus will be on the transition between the molecular and the atomic hydrogen fluid at high pressure. The connection between this liquid-liquid transition (LLT) and the insulator to metal (IM) transition will be also discussed. First we review the state-of-the-art experimental and numerical findings on the hydrogen phase-diagram, with particular reference to the liquid phase. Next we describe the results of our ab-initio molecular dynamics, previously described in Chap.2. We find that the LLT from a molecular fluids towards a fully dissociated one appears at higher pressures than previously believed. These results are very different from the ones obtained in previous simulations within the DFT and QMC framework (see Sect. 2.2 and Sect. 2.4.1). Nevertheless our picture about this elusive transition can be reconciled with some experimental evidences, in which is the conductivity the observable that has been measured and not the molecular dissociation fraction. Indeed, we believe that the molecular-atomic LLT and the IM transition, which may be only a crossover at finite temperature, are two different kinds of phenomena and they may take place far apart in the liquid phase diagram.

3.1 Introduction

Since almost the beginning of the 20th century, extensive experimental as well as theoretical efforts have been devoted to understand the high pressure phase diagram of hydrogen, the simplest possible condensed matter system in nature. Indeed hydrogen can be viewed as a prototypical system showing an IM transition, as a lattice with one atom at each site is the

simplest realization of the half-filled Hubbard model[103], the minimal -but still unsolved[1]- model which takes into account electron correlation. However hydrogen is very far from a simple material, as the study of its high pressure phase diagram remains a fundamental challenge for experimental and theoretical techniques. Apart from its importance as a model system, hydrogen is the most common substance in the universe and the understanding of its equilibrium properties, when compressed, is crucial for a satisfactory description of many astrophysical bodies. Indeed the location of the IM transition in the interior of the giant planets, such as Jupiter and Saturn, clearly has influences on their magnetic fields, convection and even their chemical composition[104]. Materials deep within these planets may behave very differently from their low pressure forms as a pressure of the order of hundreds[2] of GPa's will likely modify the electronic structure of a compound. Indeed the internal energy change ΔE in a small volume of molecular size $V = 10$ Bohr^3, in achieving a pressure $P = 100$ GPa is comparable to typical electronic energies, as $\Delta E = P\,V \sim 1$ eV.

The low pressure and temperature form of hydrogen is molecular due to its strong covalent bond, and the usual phases are described in terms of these molecules, i.e., solid, liquid, and gas molecular phases. In the early days, it was conjectured by Wigner and Huntington[7] that, upon high pressure, this stable entity – the H_2 molecule – can be destabilized, giving rise to an electronic system composed of one electron for each localized atomic center, namely, the condition that, according to the band theory, should lead to metallic behavior. According to them, a pressure of 25 GPa would have been sufficient to break the molecule. Instead, all the attempts to produce metallic hydrogen in the solid have not been successful yet. Indeed at low temperature, this material undergoes a number of phase transitions with increasing density always within the molecular phase and it remains insulating up to 360 GPa, as observed with a recent (2012) experiment[6]. This picture is confirmed by almost all the experiments on the solid phase[8, 9], with the exception of the room-temperature claim of metallization by Eremets & Troyan (2011)[105] at 260 GPa. Nevertheless it is likely[106] that this reported transition is instead the first experimental observation of a new phase boundary between the widely stable phase III and a new phase IV of solid hydrogen[107]. This new phase, which is found to be a mixed molecular and atomic structure, was subsequently better characterized by other experiment[108, 6, 107] in 2012 and 2013.

This phase is the fourth solid phase of molecular hydrogen experimentally discovered, although only the phase I is known in its details. Phase I has a hexagonal closed-packed structure with freely rotating molecules. Phase II possesses some degree of orientational order (the broken symmetry phase) and may differ from the lattice symmetry of phase I and

[1] in dimensionality greater than one.

[2] a typical internal pressure in the core of Jovian planets is of the order of TPa's

exists at temperatures below 135 K and up to 160 GPa. Phase III appears at higher pressure, is separated from phase II by an almost vertical phase line from 0 to 125 K at 155 GPa, and the transition between the two is characterized by a large discontinuity in the vibron frequency. Phases I, II, and III coexist at a triple point at 125 K and 155 GPa. A precise characterization of the crystal structure of phase III is still missing. Numerous structures have been predicted for this dense molecular phases, mainly using a random search technique[109], but no consensus has been reached as within static DFT calculation[109] (i.e. ignoring protonic zero point motion) the candidate structure is the *C2/c* while a fairly recent calculation with DMC suggests the $P6_3/m$ structure to be energetically favourable[110]. Finding theoretically the location of the IM transition depends on the exact determination of the crystal structures, as different lattices lead to different estimated band-gap closure pressures. Moreover, a recent systematic study [18] clearly demonstrates the inadequacy of DFT calculations for solid molecular hydrogen, as the quantitative predictions strongly depends on the choice of the functional used. In conclusion, at zero temperature, any convincing evidence of metallic hydrogen has been found and prediction based both on experimental data and numerical calculations seem to push the metallization pressure above the 400 GPa, i.e. pressures still far from experimental reach.

On the other hand, evidences for a IM transition at high temperatures have been reported[10, 11] altough a clear understanding of its underlying mechanism is still missing. At band theory level, materials can be either perfect metals or perfect insulator at T=0. For metals, any finite resistivity arises from deviations in the crystal lattice structure, such as impurities, defects and phonons, so that one could intuitively expect that in a disordered liquid the onset of the IM transition would be even further delayed. Instead, in Ref.[10] Nellis and coworkers reported a continuous transition from a semiconducting to metallic fluid at 140 GPa, i.e at pressures for which the solid is still in the insulating phase. They also estimate that the metallization occurs still in a partially dissociated but mainly molecular fluid, as the calculated [111] dissociation fraction is around 10 %. The temperature of this shock-compression experiment was not measured, but only estimated to be in the range 2500-3000 K. A similar conductivity behavior is observed also in a similar experiment by Fortov et. al.[11]. In this case, a claimed first-order transition is instead reported based on the observation of a "jump" of the density as a function of the pressure. However, their measurements were sparse in density and a continuous transition can not be ruled out from the data. To complete the experimental data on the liquid phase diagram of hydrogen at large pressures we report also a laser heating experiment, using this time a static pressure technique. In this work, Silvera and coworkers[12] offer an indirect evidence for a first order transition between two "phases" by observing a plateau in a temperature vs. laser power curve. Let us

call these three experimental references with W1996 (Ref. [10]), F2007 (Ref. [11]) and D2013 (Ref. [12]). For the sake of completeness we cite also a forty year old russian experiment[112] on this topic that is not in agreement with F2007 as it shifts the transition to higher pressures. Many other experiments exist on dense hydrogen, but they are mainly on the solid phase and the melting line between the molecular solid and the liquid, i.e. not very relevant on the liquid-liquid transition explored in this thesis. We refer the reader to a recent review[113] for a more complete experimental literature on hydrogen. This P-T phase diagram is reported in Fig.3.1.

Until now, since experiments are not very clear about this transition, numerical simulations would be very welcomed, given also that the range of pressures and temperatures accessible with current experimental setups is limited. One first remarkable success of numerical simulations was in predicting the existence of a re-entranting melting line (Scandolo, 2003[13] with Car-Parrinello MD) and then its location in the phase diagram (Galli and coworkers, 2004[14]). The predicted maximum in the melting line was later confirmed experimentally. The CPMD simulation with DFT-GGA energy functional by Scandolo[13] is also the first on the liquid-liquid transition. He found a first-order phase transition due to large density fluctuations at T = 1500 K and P = 125 GPa with a 6% change of specific volume. He observed a rapid molecular dissociation and metallization occurring at the same point. This result may be qualitatively in agreement with W1996 provided that the first-order line ends in a critical point somewhere below 2500-3000K. Quantitatively, the pressure seems to be underestimated by $\sim$ 50 GPa with respect to W1996. Evidences for a first order transition came from other CPMD simulations[14, 114] while a subsequent BOMD calculation found a crossover instead[66]. The latter result is probably due to finite size effects; indeed a rather small system of 128 atoms has been employed within DFT with Brillouin zone $k-$ integration. A few years later, the possibility for a first order transition was instead reported by the same authors[15] by employing a 256-atom DFT simulation at the Γ point. This time the LLT was characterized by looking at the average lifetime of the molecules.

Ceperley and coworkers have been particularly active in numerical simulations on the LLT, using both CEIMC and DFT (see Sect. 2.2 and Sect. 2.4.1). In a first CEIMC simulation[21] a crossover between the molecular and the atomic liquid was found. They observed also a surprisingly difference between VMC and Reptation[115] QMC methods. Nevertheless it is likely that this issue, as well as a the strong hysteresis effects reported are due to a non-optimal choice of the trial wave function. Later, a clear evidence for a first order transition was found within both CEIMC and DFT-BOMD[16]. LLT lines obtained by CEIMC and PBE-DFT only differs quantitatively by less than 50 GPa. They found that both the density vs pressure curve as well as the conductivity (computed by the Kubo-Greenwood formula

within DFT) undergo a discontinuous jump across the transition. Given that a critical point is found at 2000 K, these results are not qualitatively in disagreement with W1996 but still the pressure appears to be underestimated by 50-100 GPa. Notice that the latest experimental measure reported above as D2013[12] -published in April 2013- is in excellent agreement with these results. Unfortunately, a conclusive word on this topic is still missing as two month before this date, a new study by Ceperley and coworkers appeared[17]. In this 2013 study, by employing both non-local DFT functionals and including quantum proton effects, they shift the LLT by more than 100 GPa (notice that the shift between PBE and these new functionals, within the classical proton framework, is indeed greater than 150 GPa's). This new line is now in very good agreement with W1996 but not with respect to the almost simultaneously published D2013 reference (see Fig.3.1). For these reasons, we believe that our QMC-MD simulations are timely and may give insights on these complicated phenomena.

3.2 Insulator to metal transition in liquid hydrogen

In the 1996 Sir Nevill Mott himself wrote: "I've thought a lot about 'What is a metal' and i think one can only answer the question at T=0. Thus a metal conducts while a non-metal doesn't". At finite temperature, characterizing a metal becomes more difficult, as the raising of temperature thermally excites electrons into the conductive state and the conductivity becomes temperature dependent. Nevertheless, a practical definition of a metal is that of a material with a conductivity in a *metallic range*. This definition applies for the measured metallization at 140 Gpa made by Nellis and coworkers[10, 116]. They observed a change in the slope of the resistivity as a function of the pressure at this value. The change in slope is indicative of the transition of the metallic state. In this region the energy gap is smeared out thermally and by the fluid disorder, and the conductivity remains almost constant, i.e. weakly sensitive to the pressure after the 140 Gpa's. The conductivity measured in this regime is of 2000 $(\Omega\text{cm})^{-1}$, that is a value typical of fluid alkali metals Cs and Rb at 2000 K.

A very important question to be addressed is why the metallization seems to occur in the high temperature fluid phase at lower pressures than in the low temperature solid, for which any clear evidence of metallization has not been found yet. This is probably due to the different protonic structure of the two phases. First of all, the electronic gap is reduced when the molecule is thermally stretched. This is an intramolecular effect that was firstly proposed by Nellis and Aschroft[117, 118] and demonstrated employing a simple $Pa3$ model lattice structure. Moreover, thermal fluctuations in the liquid may cause molecules to approach closer to one another than in the low temperature solid. These structures, that are certainly sampled

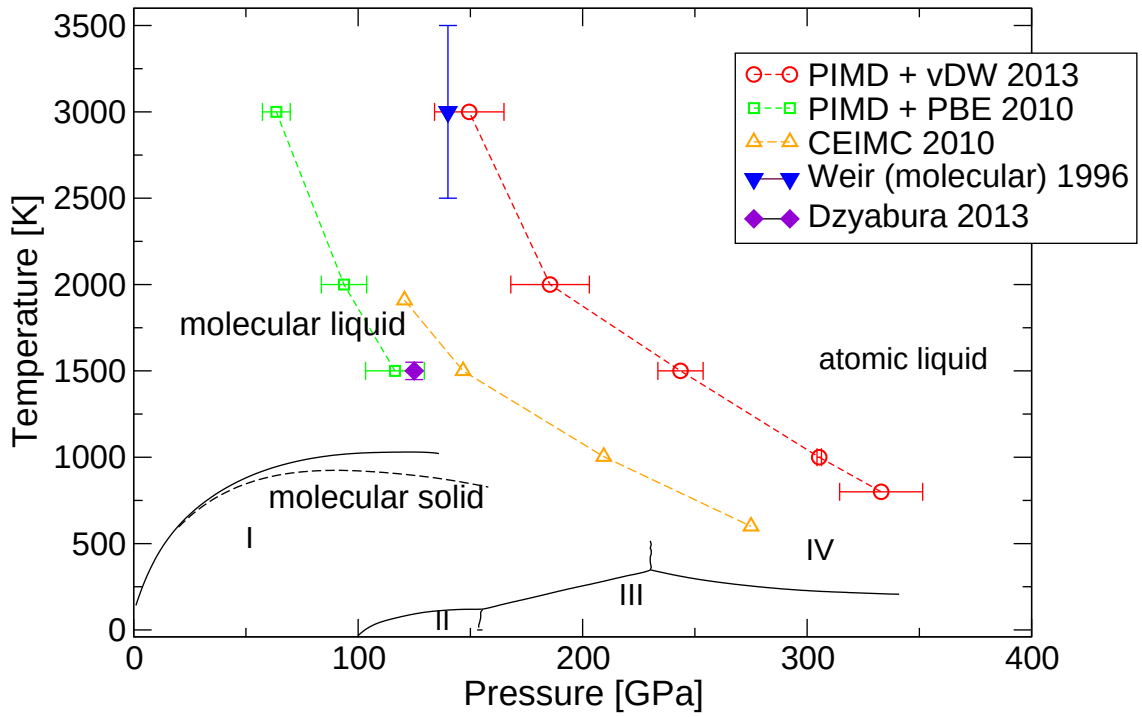


Figure 3.1: Phase diagram of dense hydrogen. All the relevant experimental and numerical references are included up to 2013. Black solid lines indicate experimental boundaries between the molecular liquid and the molecular solid, the latter consisting of four different solid phases denoted by I, II, III, and IV. Solid points are experimental measures on liquid hydrogen[10, 12]. Open symbols with dashed curves correspond to the liquid-liquid transition (LLT) obtained with latest simulations. Red circles and green squares refer to Density Functional Theory (DFT) calculations with different functionals (PBE and vdW-DF2) and quantum protom effects [17] while orange triangles refer to Coupled Electron-Ion Monte Carlo (CEIMC) [16].

during the fluid dynamics, may correspond to lower energy gaps structures with respect to the -almost- fixed lattice structure of the solid. Indeed, at high temperatures, protonic configurations in which the electronic overlap between neighboring atoms is sufficiently large to allow the hopping, are realized more frequently. In the cold solid instead, molecules are more rigid and the electrons are localized around them. This behavior is similar to what happens in fluid iodine. I_2 is a molecular system which experiences the IM transition by bandgap closure due to thermal smearing at lower pressure in the fluid than in the solid. Indeed room temperature solid I_2 becomes metallic at 16 GPa, while the liquid at only 3 GPa

at 1000 K[116].

Apart from these qualitative discussions, a clear picture about the IM transition in hydrogen is still missing. Summarizing, the metallization may take place in the following ways, the first two being the most plausible mechanisms:

- **1. Bandgap closure within the molecular phase.** In this picture, the gap E_{gap} closes continuously and carriers are thermally excited into the conduction band. At a certain density ρ, when $kT \sim E_{gap}(\rho)$ the transition between a semiconducting and a metallic liquid at finite temperature occurs. This mechanism is consistent with Ref.[10] where the conductivity is supposed to be due primarily to electronic excitations, with a small fraction of molecules dissociated. Ashcroft also suggested in Ref.[119] that the conduction may be caused by delocalized electrons around stable H_2^+ ionized pairs.

- **2. Structural phase transition.** Within this mechanism the IM transition coincides with the molecular-atomic dissociation transition. The IM predicted by Wigner and Huntington is within this class, even if appears now irrealistic in the low-temperature solid regime. To date, no evidence of pressure induced dissociation in the solid has been observed. This picture is consistent with Ref.[16]. Nevertheless, within these simulations, the observed jump in the conductivities looks more abrupt than the one regarding the dissociation fraction at the transition. Thus it is unclear why a sharp first-order IM transition arises from a smooth change in the dissociation fraction. Scandolo's CPMD simulations[13] instead show a clear jump both for the structural and the electronic properties. One example of materials that undergo an IM transition driven by a structural change are *Si* and *Ge*, that becomes metallic upon melting[117, 118].

- **3. Mott transition and localization.** This picture involves a IM Mott transition[120] typical of an Hubbard model at half filling, plus additional complication given by the disorder. This mechanism would be very interesting altough it is unlikely to occur. Indeed, the experimental evidence of Ref.[10] suggests that metallization at finite temperature occurs before the complete bandgap closing. A Mott IMT instead would require that exists a structure for which the $E_{gap} = 0$ with the system still being insulating. Moreover, since liquid phases are disordered, an IM transition due to the *Anderson localization*[121] can not be ruled out a priori. Indeed, the fluid motion, generates a sequence of random lattice structures, for which the Anderson IMT may occur, within the Born-Oppenheimer approximation.

The goal of the numerical simulations that will be presented in the next Section is to give insights about these different possibilities, as well as quantitatively locate the LLT in the

phase diagram. Since it is difficult to assess an IM transition without the possibility to calculate the conductivity (we can only obtain the ground state within our QMC technique) we will focus mainly on the LLT between the molecular and the atomic fluids. Electronic properties can be inferred by looking at the variational wavefunction for selected configurations obtained from our QMC-MD. Notice that in the following we will always assume classical protons, this approximation being justified in the range of temperatures studied, that is $>$ 600 K[29].

3.3 Molecular dissociation at large pressure

To identify the LLT, we trace four isotherms in the range 600 - 2300 K, looking for a possible singular behavior of the pressure and the radial pair distribution function $g(r)$, in a wide range of density. Simulations last long enough until thermalization is established. A typical run at fixed density and temperature is about 2 ps long For each of the four isotherm, i.e. 600, 1100, 1700, 2300 K, we perform several simulations varying the density, with a mesh increment of 0.01 (and possibly 0.005 near the transition) for r_s. r_s is the Wigner-Seitz radius defined by $V/N = 4/3\pi(r_s a_0)^3$ where V is the volume, N the number of ions, and a_0 is the Bohr radius. We indeed find a relatively small discontinuity, which appears to be clear also at the highest temperature considered. Since the discontinuity in the pressure appears to be rather small, it is extremely important to identify the LLT by looking also at the radial pair distribution function $g(r)$ for the ions. Notice that, close to the transition a fully molecular phase is not stable, as a large fraction of pairs is found to be already dissociated (Fig. 3.2b). A similar first-order behavior is also found by looking at the pressure as a function of temperature at fixed density (see Fig. 3.3), i.e. by crossing the LLT vertically, along the isochor having density r_s=1.28.

In Ref. [29] all the pressure-vs-density and the $g(r)$ plots for all the four isotherms considered are shown. Here we report the points in the phase diagram at which the LLT occurs (see Table 3.1) The lower the temperature, the larger is the density (hence the pressure) at which the LLT occurs.
Nevertheless, we observe the following common features for all the temperature considered:

- The transition appears to be first order as signaled by a discontinuity of the density as a function of the pressure (Fig. 3.2a). This LLT is always accompanied by a rather clear -altough small- jump of the molecular fraction. This can be traced by looking at the $g(r)$ profiles at different densities (Fig. 3.2b) and computing the average lifetime of the pairs (Fig. 3.2c). The latter observable gives a quantitative information on how much the molecules are stable during the simulation, even if in Langevin dynamics we can

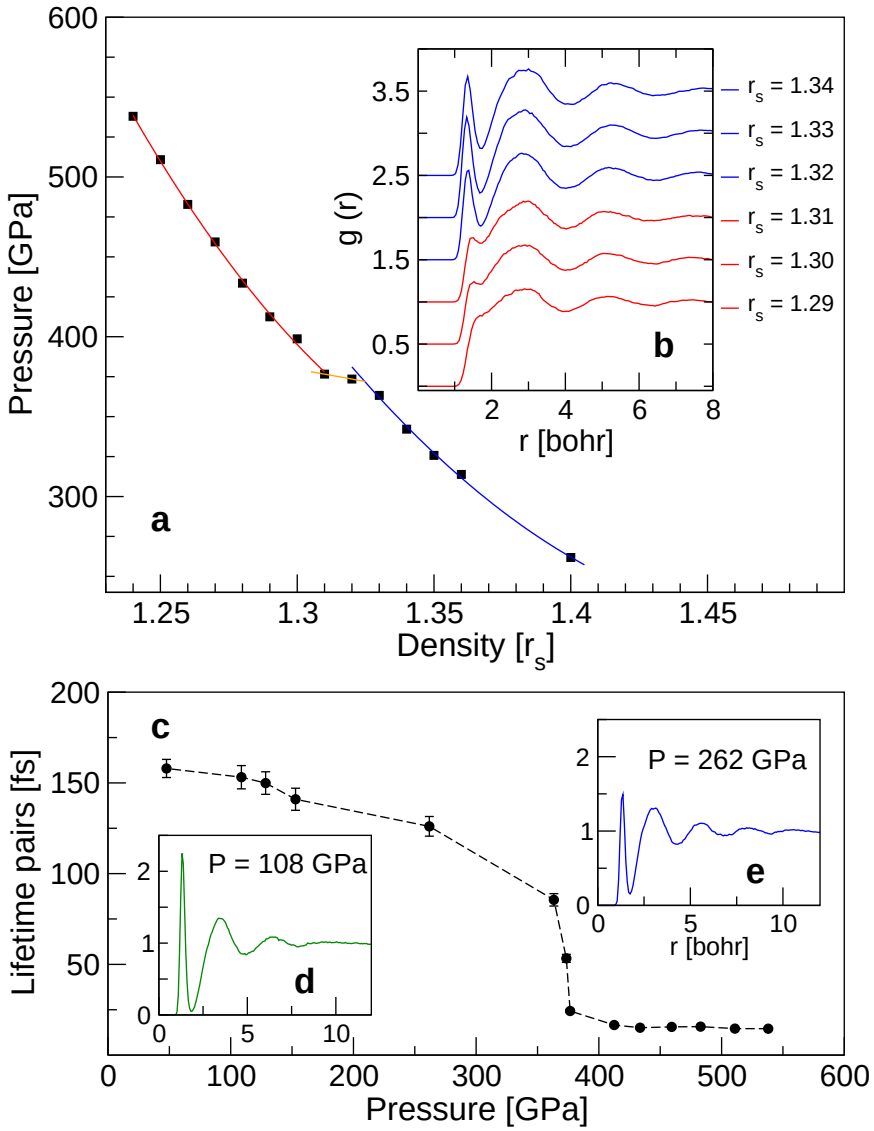


Figure 3.2: Liquid-liquid transition at 2300 K. (**a**) Pressure as a function of the density. A clear plateau is visible around $r_s = 1.31 - 1.32$, indicating the first order transition. Error bars are of the size of the symbols. This evidence is also supported by the discontinuous change with r_s in the radial pair distribution function $g(r)$ (inset **b**). (**c**) Average lifetime of pairs as a function of pressure. A pair is defined here as a couple of ions which are nearest neighbors and whose distance r is smaller than a cutoff $r_c = 1.70$ a.u. The shape of this curve is qualitatively similar for every reasonable choice of r_c although its amplitude may slightly vary. In the insets (**d**) and (**e**), radial pair distribution functions for two different pressures (P) in the molecular fluid. The higher the pressure, the smaller is the molecular peak and more coordination shells appear in the long range tail.

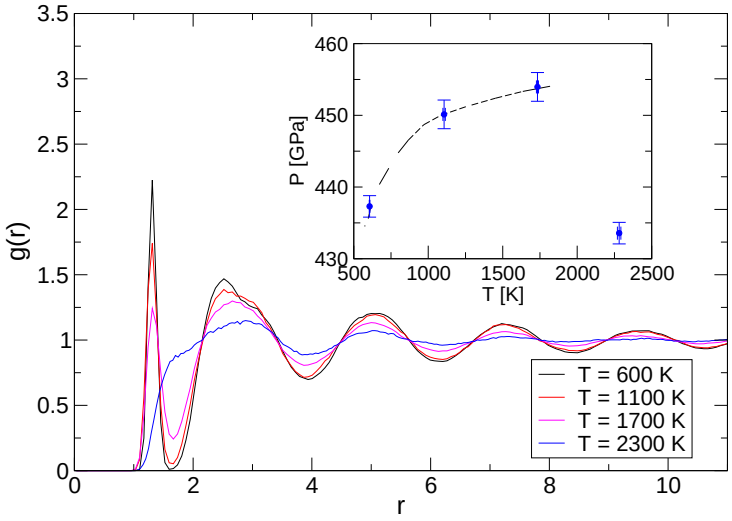


Figure 3.3: Radial pair distribution functions for different temperatures at fixed $r_s = 1.28$ density. The LLT along this isochor occurs between 1700 and 2300 K. Inset. Pressure as a function of the temperature. The pressure increases as long as the fluid remains molecular. A drop in the pressure occurs at the full dissociation.

T [K]	P [GPa]
2315(10)	375(2)
1696(7)	468(7)
1122(30)	639(6)
563(10)	624(8)

Table 3.1: P-T parameters at which the first order LLT is observed. The error bars derive from the fact that, at given density, both P and T are measured quantities. These errors clearly propagates in the estimation of the LLT point in the phase diagram, as the two simulations on the right and on the left of the LLT have clearly different pressures and may have slightly different temperatures. In other words, we are following the isotherm lines along the phase diagram within an error of ~ 10 K. The error in the pressure can be reduced by adopting a finer density grid. These error bars are acceptable for the present accuracy in the phase diagram.

not assign a precise meaning for such time scales.

- The first order transition separates two kind of fluids. A fully atomic liquid at larger pressures and a partially dissociated molecular one at smaller pressures. So we are not

speaking of a sharp first order transition between a fully atomic and a fully molecular liquid.

- This first order LLT occurs at larger pressure than previously believed (see Ref.[13, 16, 17]). In particular we observe an unexpected persistence of the molecules at large pressures, as the molecular fraction slowly disappears with increasing density but the liquid becomes fully atomic only at very large pressures. For instance at 2300 K, the transition is at 375 GPa.

We draw a phase boundary line in the phase diagram (Fig.3.8) putting together these data. The first order character is evident also at the highest temperature of 2300 K, suggesting that even at these temperatures this structural transition is not a crossover. Moreover the shape of the LLT boundary is rather unusual and becomes a vertical line in the P-T phase diagram for T $<$ 1100 K. By assuming that also at lower temperatures no solid phase emerges, the dissociation pressure should remain almost temperature independent. Therefore, even by considering an upper limit of 100 GPa shift to lower pressures, due to proton quantum effects not included here (see Sect. 3.6), we predict that experiments should be done at least above 500 Gpa to realize the Wigner and Huntington dream of hydrogen atomic metallization.

So far we have considered properties that depends only on the protons and we label the observed phases as *molecular, mixed molecular* and *fully atomic* liquids. The above picture about the LLT seem to rule out the hypothesis number 2 in Sect. 3.2 concerning the insulator to metal transition. Within this mechanism the IM transition is driven by the structural phase transition and the IMT and the LLT would occurs simultaneously. Indeed our results are not well consistent with this mechanism because

- **a.** The transition between the fully molecular and the fully atomic liquid is not sharp. In fact there is not a clear discontinuity between a fluid made of molecules (that one would consider as insulating objects), and a liquid made of unpaired atoms glued by a sea of conduction electrons.

- **b.** If we identify the IMT with the LLT, then the IM would occur at very large pressures. This would not be consistent with Ref.[10] in which a crossover (within the molecular phase) between the nonmetal and the metallic fluid is found at 140 GPa and around 2000 K. Our simulation instead predicts that the LLT is first order at 2300 K and is located at 375 GPa.

For this reason we believe that the metallization at finite temperature takes place within the mixed molecular phase, well before the first order structural phase boundary.

3.4 A possible IMT within the mixed molecular phase

To investigate the occurrence of a possible IMT we look at the electronic wavefunction. We cannot calculate directly the conductivity and the electronic gap between the ground state and the excited states as it is accustomary in DFT, because we have only the ground state $\psi_0(\boldsymbol{R})$ at each configuration $\boldsymbol{R}$ sampled during the dynamics. Nevertheless the VMC wavefunction contains a large amount of information on the electronic properties.

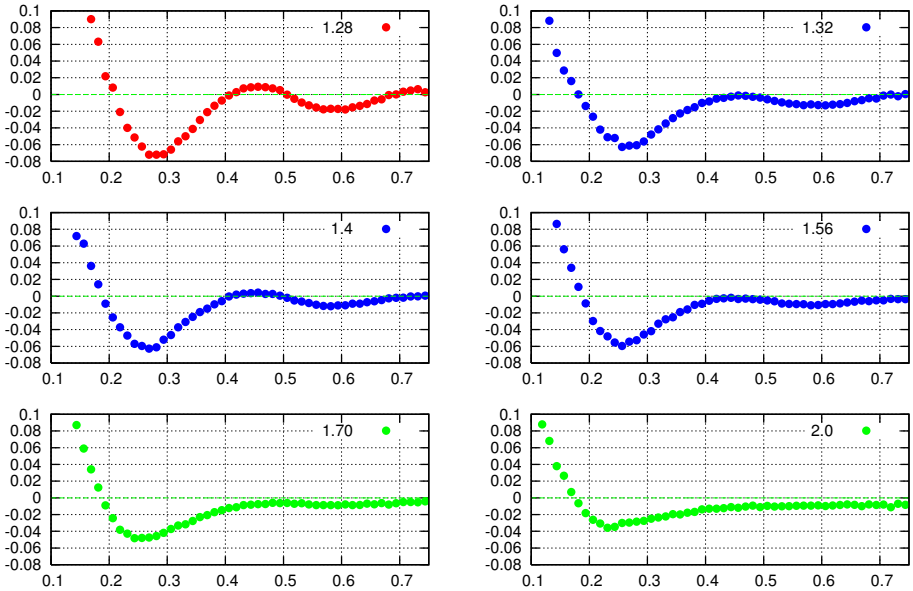


Figure 3.4: Density matrix radial decay $\mathrm{DM}(0,x)$ as a function of the unit-cell rescaled distance $x = r/L$ for different densities $\{r_s$=1.28, 1.32, 1.40, 1.56, 1.70 and 2.0$\}$ at the T=2300 K isotherm. Notice that the smaller is the density -that is inversely proportional to r_s- the faster is the decay and the oscillations are more damped. The LLT occurs at r_s=1.31-1.32. Notice that $\mathrm{DM}(0,x)$ does not change qualitatively at the two end of the first order LLT. For the sake of clarity we plot in green the data which correspond to a fully molecular liquid, in blue the partially dissociated fluid and in red the fully dissociated one. We take averages from 4 equilibrated configurations at each given density.

In 1964 Kohn first claimed that insulators and metals differ in their ground state and not only in the excitation spectrum[122]. Even before the system is excited by any probe, a different organization of the electrons is present in the ground state and this is the key feature discriminating between insulators and metals. He gave evidence that localization of the electronic ground wave function implies zero dc conductivity, and therefore characterizes the insulating state. The locality properties of solids within independent particle theories have been widely studied[123, 124]. It was shown that for insulators the density matrix decays exponentially. The rate of the decay is related to the size of the energy gap in the

band-structure: a larger gaps corresponds to a faster decay. In contrast, for a metal at zero temperature, the decay is much slower, namely power law.

We can observe this property by considering the molecular orbitals ψ_i that define the determinantal part of our wavefunction (see Sec.2.3.2.1). Let us now introduce the density matrix as

$$\mathrm{DM}(\boldsymbol{r}, \boldsymbol{r}') = \sum_{i=1}^{N/2} \psi_i(\boldsymbol{r})\psi_i(\boldsymbol{r}') \tag{3.1}$$

where N is the number of electrons. The abovementioned decay of the density matrix can be seen by looking at the $r \to \infty$ behavior of the function $\mathrm{DM}(0, r)$, where $r = |\boldsymbol{r}|$. The finite size of the system makes the task to quantitatively distinguish between different asymptotic behaviors very hard. Nevertheless a qualitative change of the $\mathrm{DM}(0, r)$ shape as a function of the density can be observed in Fig.3.4.

Notice that, to distinguish between a metal and an insulator, it is sufficient to consider the (uncorrelated) determinantal part of the wavefunction as our short-range Jastrow factor (see Sect. 2.3.2.2) is not able to turn a metallic uncorrelated state into an insulating one[125].

To obtain a more quantitative information about the electronic localization we compute for each density the following quantity

$$|\mathrm{DM}| := \frac{1}{M} \sum_{i=1}^{M} \int d\boldsymbol{r}\, |\mathrm{DM}(\mathbf{0}_i, \boldsymbol{r})| \tag{3.2}$$

namely, the integral of the absolute value of the density matrix $\mathrm{DM}(0, \boldsymbol{r})$, averaging over M different origin points inside the simulation cell. We take average over 4 equilibrated protonic distributions at each density and we observe the behavior of this $|\mathrm{DM}|$ as a function of the pressure for two isotherms: 2300 and 600 K. Due to the different DM decay behaviors, we expect a divergence of this quantity in the metallic phase. Since the system is finite, a dramatic discontinuity is hard to observe, but still, some changes should be visible passing from an insulating to a metallic phase. Notice that the $\mathrm{DM}(0, r)$ decay functions are characterized by Friedel oscillations, i.e. having negative and positive contributions. The wavelength of these oscillation is controlled by the density of the system but the disorder may introduce random shifts. Therefore, we take the sum of the moduli of the $\mathrm{DM}(0, r)$ in order to avoid the zero averaging property, which is likely to happen if one samples over disordered configurations. The results are plotted in Fig.3.5.

As expected the $|\mathrm{DM}|$ values are larger at high pressures, at which the system should be metallic. At 2300 K we note a change in the slope (see Fig.3.5(a)) at $\sim$ 130 GPa, before the real first order transition at 375 GPa. We recall that this is an indirect information on the energy gap. In other words, the gap always decrease with the pressure, but with a different rate beyond the 130 GPa. This value, although its precise connection with the conductivity is still

missing, is surprisingly in agreement with the experimental minimum conductivity value at 140 GPa and $\sim$ 2500 K reported in Ref.[10]. Moreover this is also in agreement that the IM crossover occurs within the molecular phase, as claimed by the authors. In our picture, a real first order transition happens at larger pressure (250 GPa's later) when the complete atomization of the fluid is reached.

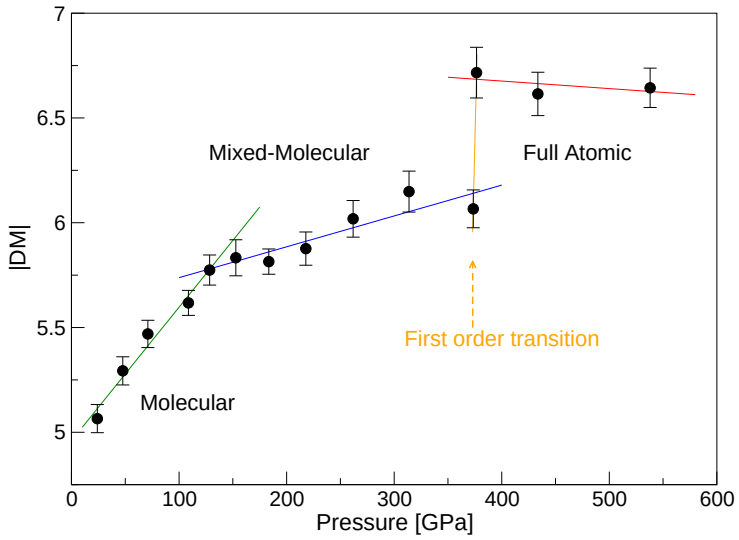
At the lower temperature of 600 K the situation changes slightly (see Fig.3.5(b)), as we observe a more clear change of the |DM| at $\sim$ 440 GPa, namely 180 GPa's before the LLT in which the complete atomization is reached. If we identify this jump with the IMT, then the metallization would occur again in the mixed-molecular liquid at $\sim$ 440 GPa and 600 K. In conclusion, provided that this discontinuity of the density matrix integral is a fingerprint of the IMT, then we observe that

- The IMT occurs before the complete atomization of the fluid, which we label as the LLT
- At 2300 K, this value is in agreement with Ref.[10], both on the pressure value and the mechanism that is supposed to drive the IMT.
- The distance in the phase diagram between the IMT and the LLT is smaller at 600 K than at 2300 K. This is consistent with the fact that at 0 K the two transition should merge, provided that a liquid phase is stable even at this temperature.

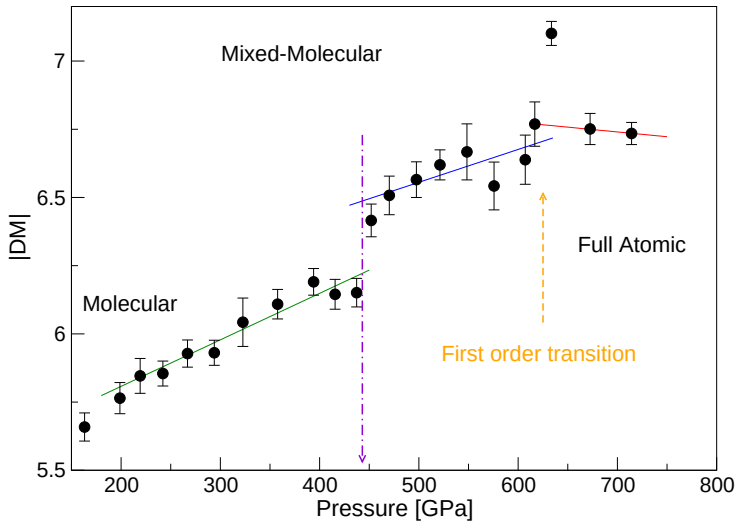
Looking more carefully at the simulations at 600 K, we link the observed jump of the |DM| at 440 GPa with a qualitative change in the $g(r)$'s profiles. Indeed, starting from 440 GPa, the first minimum of the $g(r)$ ceases to be exactly zero and becomes finite. Fluids at densities smaller than r_s=1.28, show a clear separation in the range of the accessible H-H distances, while all the possible interatomic distances are allowed[3] when compressed at more than r_s=1.27 (see Fig.3.6(a)). Exactly between these two endpoint the discontinuity of the |DM| occurs. As the density is increased, the region corresponding to the first minimum becomes more and more populated while the height of the first maximum decreases. At the LLT density, the first molecular peak completely disappears (see Fig.3.6(b)). Combining these two observations, we argue that the electronic distribution is more sensitive to the former type of change in the protonic distribution, rather than to the complete atomization.

In other words, the qualitative difference is between a fluid in which the molecules - on which the electrons are localized- are somewhat defined (the first maximum of the $g(r)$ being clearly separated by the bulk) and a fluid in which there is no more a forbidden region of interatomic distances. These protonic structures may favour the electron hopping between the elongated pairs and the system may become metallic at this densities.

[3] clearly, beyond the equilibrium 1.40 Bohr distance.

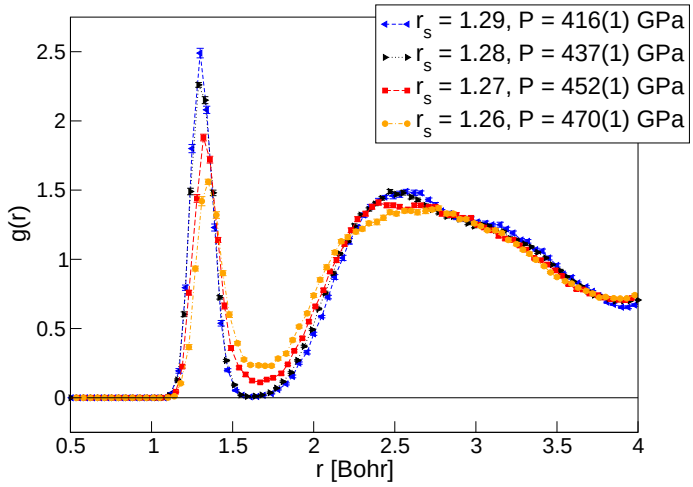


(a) T = 2300 K

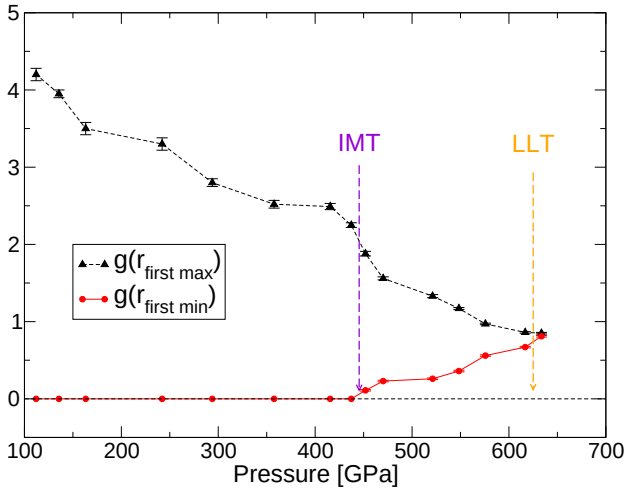


(b) T = 600 K

Figure 3.5: |DM| as a function of the pressure for two different isotherms. In both cases we can observe a change well before the first order LLT, signalled by a discontinuous behavior of the density as a function of the pressure. **a.** T=2300 K. We observe a change in slope of the |DM| at $\sim$ 130 GPa, before the greater discontinuity which occurs at 375 GPa, i.e. at the same pressure of the first order LLT (orange arrow). **b.** T=600 K. This time the discontinuity of |DM| before the LLT is much more clear at $\sim$ 440 GPa (violet arrow). The observed LLT between the mixed molecular and the atomic liquid is instead at $\sim$ 620 GPa (orange arrow).



(a) $g(r)$'s profiles near the IMT



(b) $g(r)$'s first max and min as a function of P

Figure 3.6: a. $g(r)$'s profiles at densities near the |DM|'s jump of Fig.3.5(b). The proposed IMT is between r_s=1.28 and 1.27. Notice the different shapes of the first minimum before and after the IMT. Moreover also the first peak position shifts with the pressure. As the density increases the molecules becomes more elongated as they feel more the attraction of the surrounding atoms. **b.** Evolution of the amplitude of the first molecular peak (black points) and the first $g(r)$'s minimum (red points) as a function of the pressure. At the proposed IMT, all the possible distances ranges beyond the first peak becomes populated. At the same time, the amplitude of the first maximum decrease faster. These two quantities coincides -by definition- at the complete atomization of the fluid.

3.4.1 Molecular lifetime and hopping

The molecular lifetime is the dynamical observable which is directly related to the particles density $g(r)$ around the first minimum. Altough our Langevin dynamics does not, in principle, allows us to compute dynamical quantities we can still infer interesting properties about the nature of the liquids observed. For example, let us compute the number n_r of distin-

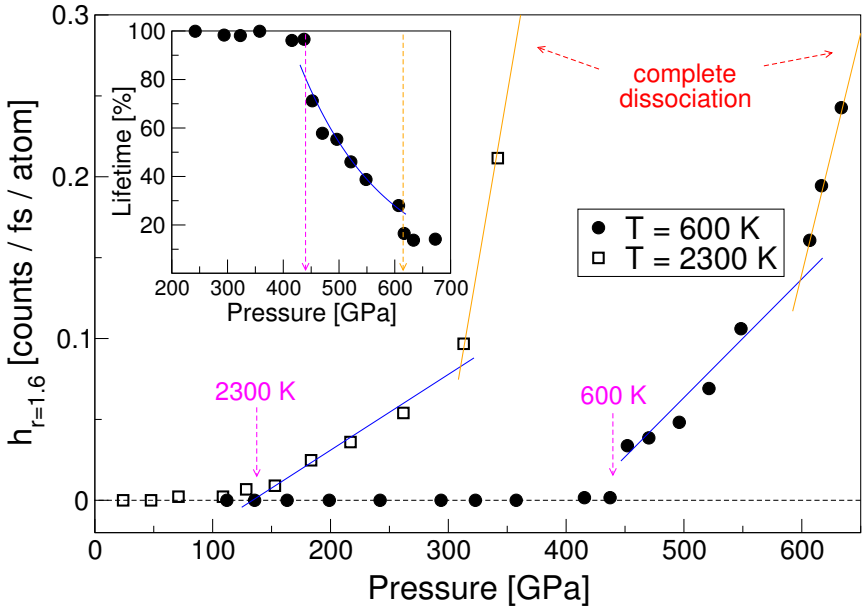


Figure 3.7: Recombination rate $h_{r=1.6}$ as a function of the pressure and for two isotherms, T=2300 K (600 K) with empty squares (full circles) points. Blue and orange lines are guides to the eye to distinguish the two phases, namely the partially dissociated (blue) and the transition towards the complete atomization (orange). In the inset, the average molecular lifetime, with cutoff $r = 1.6$ Bohr (see text) as function of the pressure, at 600 K. The lifespan is measured in % with respect to the total time of observation (400 fs). All these data refer to the 256-atoms system.

guishable particles which establish a nearest neighbor link to a given atom i during a given simulation time Δt. Namely, for each atom i we count how many different hydrogens j form a paired state with i during the simulation. In order to define a molecule we choose that the $i - j$ bond must be the shortest among all the possible $i - j$ distances and shorter than a reasonable cutoff radius r. For a pure dense atomic liquid, the atoms establish transient bonds with a very large number of particles during the simulation. Conversely, in a pure molecular liquid this number should be exactly $n_r = 1$, as all the molecules remain stable along the simulation and there is no proton exchange between the molecules or umpaired atoms in the liquid. Therefore, the quantity h_r, defined as

$$h_r = n_r - 1 \tag{3.3}$$

counts how many new $i-j$ bonds are formed during a given time. This h_r is an *hopping* rate and is directly related to the probability of molecular breaking. We calculate this rate for the T=2300 K and T=600 K isotherms and adopt a cutoff of $r = 1.6$ Bohr (the results does not depends qualitatively by the particular choice of this cutoff). In Fig. 3.7 we observe that $h_{r=1.6}$ becomes nonzero exactly at proposed IMT pressures (pink arrows). Therefore, the insulator to metal transition occurs when the fluids are sufficiently compressed to experience bond breaking and reforming, between the short-lived molecules. These critical pressures depend on the temperature, and, as expected, the larger is the temperature the smaller is the pressure at which the onset of molecular dissociation is observed. The $h_{r=1.6}$ rate increases almost linearly at larger pressures and finally diverges near the LLT pressures, namely at the complete dissociation.

These two critical pressures, namely the IMT and the LLT pressures, define the boundaries of the mixed-molecular liquid phase. In this liquid, although the (static) average number of molecules is constant, the atoms participating to the bonding change throughout the simulation, namely the atom i replaces several times the j molecular partner in a given time but the total molecular fraction remains constant. A typical recombination rate value for this liquid is of the order of $h_{r=1.6} \sim 0.1$ fs^{-1}, which means that the molecular lifetimes are $\approx 10^{-14}$ fs, i.e, of the order of few molecular oscillations. This value is qualitatively in agreement with an old tight-binding MD simulations[126]. We can also directly compute the molecular lifetime as the average lifespan of bonds shorter than a cutoff r. This observable takes into account not only the recombination process but also simple atomization events. We plot the lifetime as a function of the pressure at 600 K in the inset of Fig. 3.7. The qualitative distinction between the three fluids is again evident. The liquid with mixed character, made of short-lived molecules, appears between 440 and 620 GPa.

Finally, we update our high pressure phase diagram in Fig. 3.8 with these data concerning the IMT and the boundaries of the mixed atomic-molecular liquid. Notice that our proposed IMT lines is surprisingly in agreement with the LLT lines obtained by DFT-HSE simulations with quantum protons of Ref.[17], taking also in account that the zero point motion should shift the LLT towards smaller pressures as the temperature is decreased. However, this may be only a coincidence because the authors of Ref.[17] claim that the IMT and the LLT transitions coincide along that line in the phase diagram.

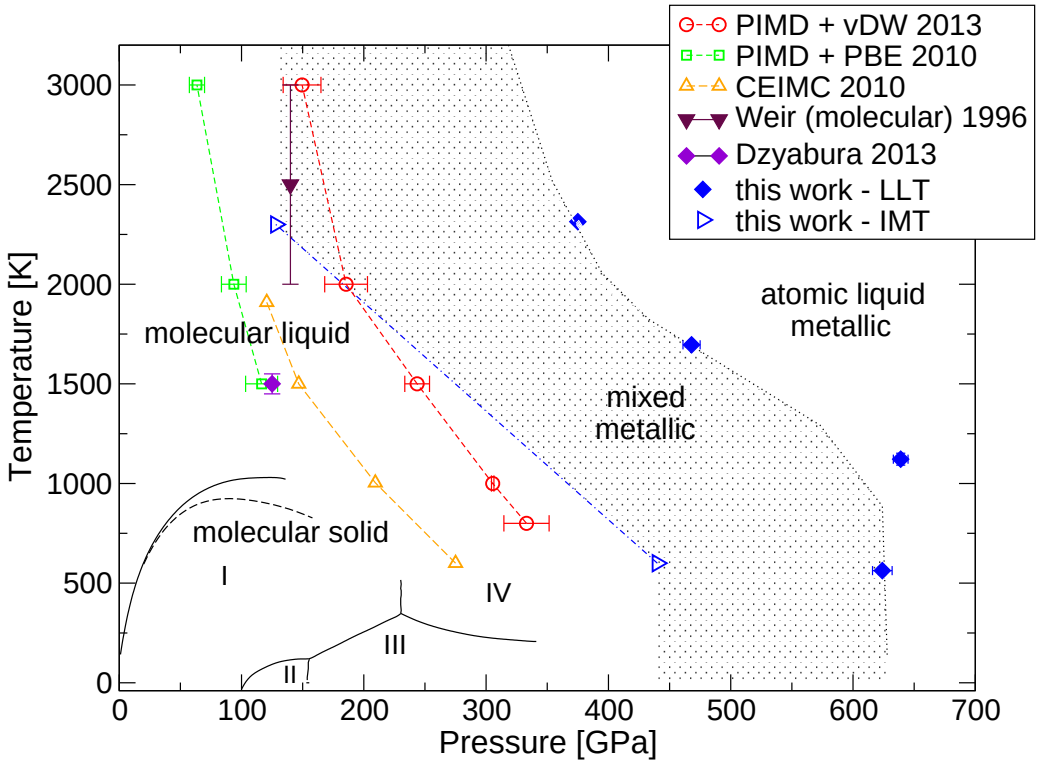


Figure 3.8: *P-T* **phase diagram of hydrogen.** Black solid lines indicate experimental boundaries between the molecular liquid and the molecular solid, the latter consisting of four different solid phases denoted by I, II, III, and IV as in Ref. [6]. Colored symbols with dashed curves correspond to the liquid-liquid transition (LLT) obtained with latest simulations and experimental points. Red circles and green squares refer to Density Functional Theory (DFT) calculations with different functionals (PBE and vdW-DF2) and including proton quantum effects [17] while orange triangles refer to Coupled Electron-Ion Monte Carlo (CEIMC) [71, 72]. Blue diamonds correspond to the LLT estimated in this work. Blue triangles are instead placed at the proposed IMT points, obtained by looking at the density matrix discontinuities. The (black dotted) LLT lines divides the completely dissociated fluid (on the right) from the molecular/partially dissociated fluids (on the left). The (blue dashed) IMT line separates the insulating (on the left) from the conducting liquid (on the right). The shaded area indicates therefore a partially dissociated metallic molecular fluid.

3.4.2 Jastrow factor and localization studies

To investigate further how the electronic properties change with the pressure and their possible relation with the IMT we analyze more the variational wavefunction, looking at the behavior of the Jastrow factors. First of all, we look at the two body Jastrow factor defined in Sect. 2.3.2.2: Its specific functional form reads

$$u_2(r) = (1 - e^{-br})/2b \tag{3.4}$$

Therefore, the electron-electron repulsion becomes important if the range of the interaction $\tau_b := 1/b$ is large. In Fig.3.9 we plot the dimensionless quantity τ_b/L a sa function of the pressure. Interestingly we found that this function is not monotonic with P, rather it shows a maximum when the liquid is mixed molecular - atomic. Nevertheless, such behavior is quite smooth and therefore the two body Jastrow - which is certainly important for the overall accuracy of the calculation - does not play a key role as *reaction coordinate* for the IMT.

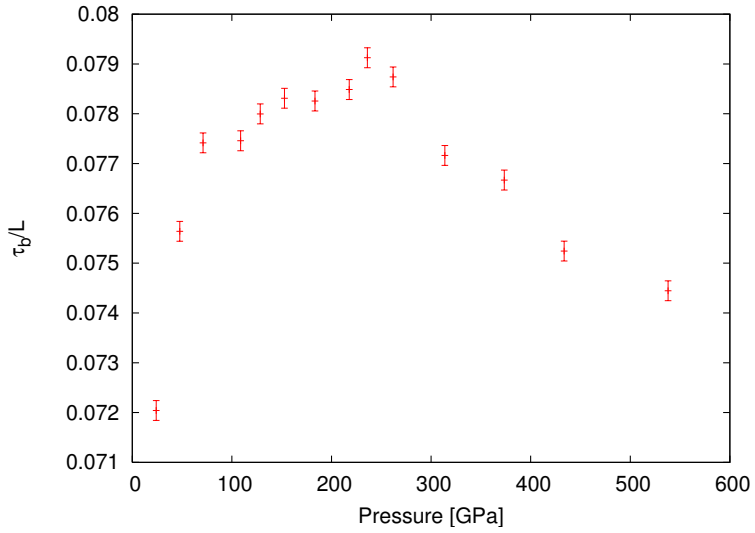


Figure 3.9: τ_b/L as a function of the pressure at T=2300K. This quantity, which tell how much is important the 2-body electron-electron dynamical correlation, is slowly varying in the all range of pressures. We take average over 5 equilibrated configurations at each pressure.

A similar situation occurs for the 3-body jastrow factor. In Fig.3.10 we plot the distribution of the f values of Eq.2.21 for six very different densities. This quantity is directly connected with the *on-site* repulsion of two electrons which are localized on the same ions. We do not observe dramatic changes of these parameters along the transition.

Consider now the highest occupied molecular orbital $\psi = \psi_{HOMO}(\boldsymbol{r})$ obtained with a DFT-LDA calculation performed on our VMC-MD equilibrated protonic configuration. A

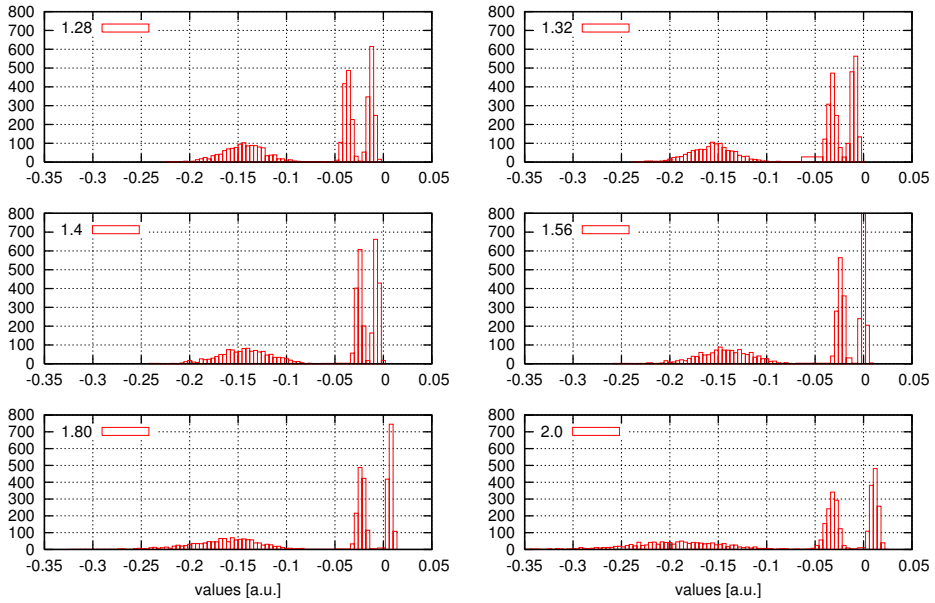


Figure 3.10: Histograms of the f_{ν_a} 3 body Jastrow parameters for six densities at T=2300 K. We take average over 5 equilibrated configurations at each density. Since the 3-body Jastrow is expanded in a localized 2s basis (cf. 2.3.2.2) we observe three peaks. Indeed for each atom there are three variational parameters: $f_{1s,1s}$,$f_{2s,2s}$ and $f_{1s,2s}$. Only the $f_{1s,1s}$ parameters (broad peak) have magnitude larger than 0.

parameter that is indeed changing a lot along the T=2300 K isotherm is the so-called *inverse participation ratio (IPR)*, that we define as

$$IPR = \frac{\int dr^3 \psi^4(\boldsymbol{r})}{\left[\int dr^3 \psi^2(\boldsymbol{r})\right]^2} \tag{3.5}$$

This quantity vanish with the system size if ψ represent an extended state, while remains constant if ψ is a localized state. The IPR function is used in the context of the Anderson IMT. We instead adopt this quantity only to qualitatively observe changes in the wavefunction that we can relate with the IMT. In Fig.3.11 we see that the IPR behavior as a function of the pressure shows a clear change at $\sim$ 100 GPa. At smaller values of the pressure, the IPR is high, therefore the HOMO molecular orbital is localized. Above 100 GPa, the HOMO is extended, and according to this IPR, it should represent a metal. The range of pressure in which such IMT takes place is qualitative agreement with the 130 GPa value at 2300 GPa, given above by the density matrix analysis. Moreover, since the HOMO is obtained with LDA this values is probably underestimated. Notice that, the Fig.3.5(a), is obtained with molecular orbitals optimized with VMC instead.

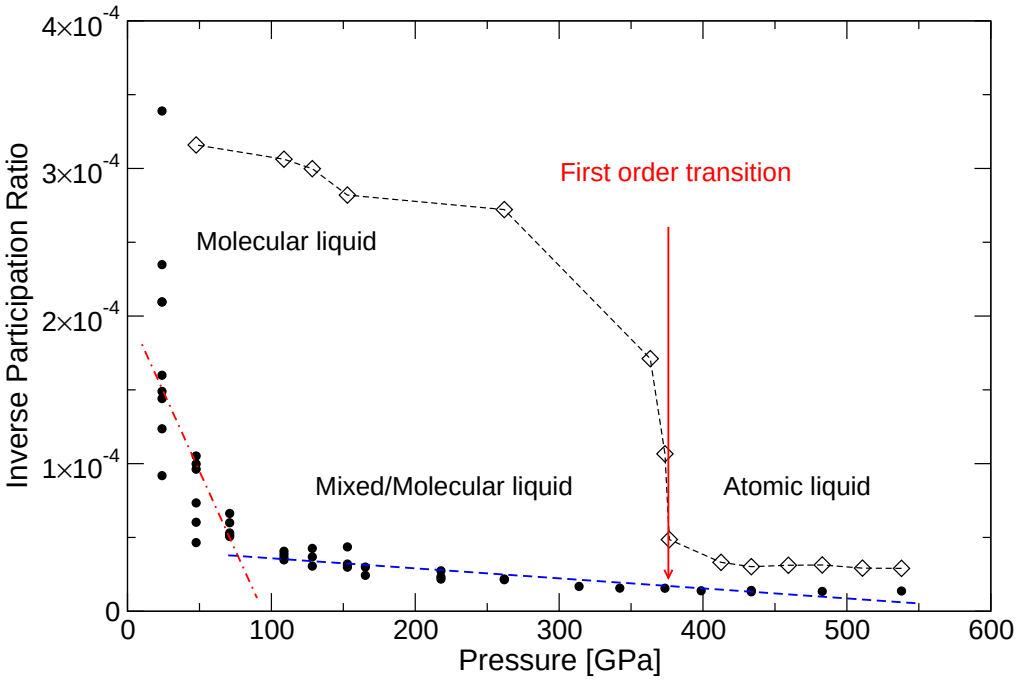


Figure 3.11: Inverse participation ratio, calculated with the LDA HOMO, for several equilibrated protonic configuration at different densities and T=2300 K. Notice that the jump occurs at $\sim$ 100 GPa, that is more than 200 GPa's before the first order LLT (red arrow). The profile of the average lifetime function as a function of the pressure is also plotted (see Fig.3.2.c).

3.5 QMC molecular dynamics: details and tests

We employ simulation cells containing up to 256 hydrogen atoms and use the novel MD scheme with friction in the NVT ensemble described in the previous Chapter (2.5.2), for simulation times of few picoseconds, long enough to have well converged results on the pressure, internal energy, and the radial pair distribution function $g(r)$. Since our conclusion about the LLT are rather unexpected we provide now some important and systematic tests on the accuracy of our method. There are essentially three points to carefully address and they are:

- **Quality of the wavefunction.** Within VMC (and DMC as well) the accuracy of the calculation is related to the accuracy of the trial wavefunctions. In Sect. 3.5.1 we perform systematic tests on the basis set on which our trial wavefunction is expanded.

- **Ergodicity.** In order to locate the first order transition we must ensure that the simulations are equilibrated, within the QMC approach. In Sect. 3.5.2 we prove that the outcome of our simulations are indeed independent from the initial configurations, near the LLT.

- **Finite size effects.** This point is also crucial. Indeed, by using simulation cell containing a finite -and usually small- number of particles, one has to be sure that results are close to the thermodynamic limit. In Sect. 3.5.3 we show that finite size effects are indeed very important even in the liquid phase.

Moreover, we already showed (see Sect. 2.5.1) that the generalized Langevin equation with position dependent friction/noise matrices does not introduce any sort of bias in the simulations.

3.5.1 Basis set convergence

Since QMC, even in its simplest VMC formulation is quite computationally expensive, we must employ a compact wavefunction. Therefore we have to choose a compromise between the accuracy of the calculation and the total number of variational parameters. Our wavefunction is expanded in localized atomic orbital (cf. Sect. 2.3.2). Therefore our way to improve the variational freedom within our ansatz is to enlarge this atomic basis set. The results obtained in this Chapter are obtained with molecular orbitals expanded on a $2s$ Gaussian basis per atom. A $2s$ localized basis set is used also for the 3-body Jastrow We must check that our physical results for the $g(r)$ and the pressure does not significantly depend on these particular choices. To do so, we compare the $2s$ and a $3s1p$ basis on a smaller 54-atoms system.

Total Energy for fixed ionic configurations

	r_s=1.44	r_s = 1.24
VMC - JSD - *2s*	-0.54750(1)	-0.5141(1)
VMC - JSD - *3s1p*	-0.55391(1)	-0.52246(1)
VMC - JDFT - *6s5p1d*	-0.5542(1)	-0.5230(1)
LRDMC - JSD - *2s*	-0.55239(1)	-
LRDMC - JSD - *3s1p*	-0.55678(1)	-0.52535(1)
VMC - "CEIMC"	-0.55605(2)	-0.52564(2)
RQMC - "CEIMC"	-0.5572(1)	-0.52638(3)

Table 3.2: Total energies per atom (in Hatree) obtained with different methods: VMC, DMC or RQMC; different optimization methods JSD (Jastrow and SD optimized together) and JDFT (optimized Jastrow with SD orbitals given by DFT-LDA) and atomic basis sets (*2s, 3s1p, 6s5p1d*) in which the molecular orbitals are expanded. The best energies for both the configuration are obtained with the "CEIMC" ansatz. However, a simple *3s1p* basis set for each atom gives total energy that differ only by few mH/atom. The fact that the LRDMC using such guiding function has energy differences $\lesssim$ 1 mH/atom means that the *3s1p* basis provide an almost optimal nodal surface. LRDMC values are obtained with lattice space extrapolation.

In Table 3.2 we compare the energetics, obtained with different trial wavefunctions, for two different configurations (one molecular and one atomic) at densities r_s=1.44 and r_s=1.24. These two configurations have been provided by Prof. Pierleoni, along with their energies evaluated within VMC and RQMC. We simply label these two energies as "CEIMC" as the precise trial wavefunction in this case is unknown to us. From these data we see that the VMC calculations with $3s1p$ is almost converged, as it gives energies that are 2-3 mH/atom higher than the DMC (or RQMC) values, that we can assume to be the reference values. We can afford to use $3s1p$ basis by adopting the *hybrid orbitals* described in Sect. 2.3.2.1. In this case three hybrids/atom are necessary to recover the energy of the uncontracted $3s1p$ basis per atom[4]. Instead, the $2s$ basis VMC loses about 9 and 11 mH in these configurations, with respect to their large basis counterparts ("CEIMC" calculation).

However, since the size is quite small we can directly run VMC-MD simulations, comparing the $2s$ and the $3s1p$ wavefunctions performance directly for the observables of our interest. In Fig.3.12 and 3.13, we see that the two basis gives almost exactly the same $g(r)$ and the bias in the pressure is only about $\sim$2-4 GPa's. From this fact, we can conclude that

[4]the uncontracted basis means 6 variational parameters for atom. Using hybrid orbitals we save a factor 4 in the total number of variational parameters in the SD.

the difference in the total energy is not particularly relevant. The pressure and forces instead are already evaluated accurately in the small basis case as the $g(r)$'s profile is basically unchanged. Notice also that the QMC noise is totally under control, as different VMC bins length N lead to very similar $g(r)$'s and pressures, and the behavior of the total energy as a function of $1/N$ is smooth and indeed perfectly linear. This property allows, in principle, the possibility of employing a very small VMC bin length, i.e. comparable to the largest number of parameters that enters in the optimization. This fact, combined with the possibility to work with a compact basis set, makes our QMC-MD computationally feasible even for very large sizes.

Analyzing the results also from a physical point of view, we remark here that our $g(r)$'s are not too far from the DFT-HSE predictions for the same system. The HSE functional is one of the most accurate and used xc functional for condensed system (and indeed high pressure hydrogen[17]) while the LDA is the simplest one. This fact validates the approach and, at the same time, shows another evidence on the gained molecular stability within QMC since HSE simulations displays a reduced molecular peak at 120 GPa. The LDA instead shows explicitly its bias towards the atomic (metallic) phase as it destroys all the molecules.

It is unlikely that this enhanced stability of the molecular phase is a bias due to the localized basis set as we showed that two different ansatz -that differ in energy of ~ 5 mH/atom- provided the same $g(r)$, and the $3s1p$ is only ~ 2.6 mH/atom higher in energy with respect to the LRDMC (see Table 3.2). As expected, since the quality of the DFT improves by increasing the density, DFT and VMC give very similar $g(r)$'s at 420 GPa[5].

r_{max}	# parameters	Energy error
0	1542	0.022(6)
2	1813	0.0044(11)
3	3003	0.001(1)
4	5108	0.004(1)
5	8443	-

Table 3.3: Optimized energy for a 256 atom equilibrated configuration as a function of r_{max}. Energies are given as a difference with respect to the $r_{max} = 5$ reference value $E(r_{max} = 5) = -128.236(4)H$. It is noteworthy that even the $r_{max} = 0$ setting improve upon the J+DFT total energy -128.157(4), in which molecular orbitals are not optimized in the presence of the Jastrow factor.

The next test involves the use of the r_{max} cutoff in the optimization of the molecular orbitals (see Sect. 2.3.2.1). For hydrogen it is enough to consider r_{max}= 4 a.u. to have essentially

[5] for the 54 atom system.

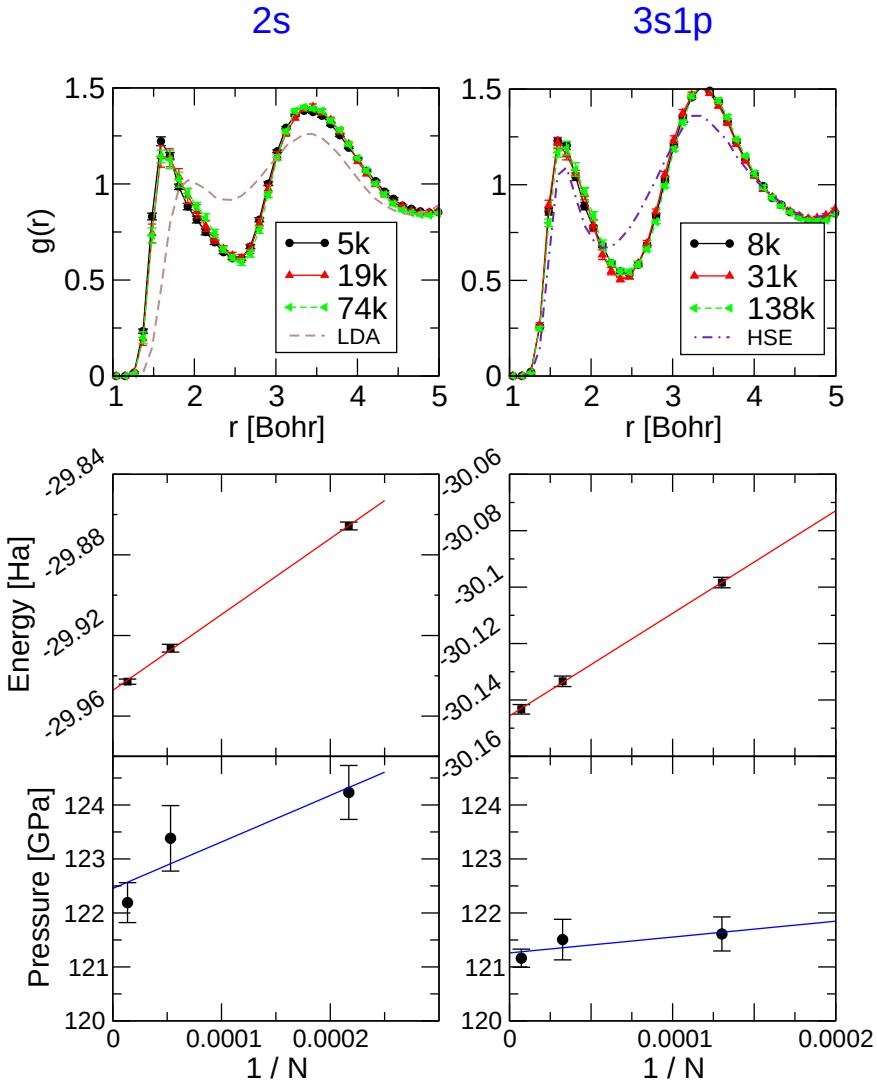


Figure 3.12: Accuracy of the wavefunction: tests at density r_s=1.44 and T=600 K. The panels on the left refer to the 2s basis set, while the panels on the right to the much more accurate 3s1p. We also test the convergence of the observables as a function of the VMC bin length. *Top panels:* Radial distribution function obtained with different VMC bins. Notice that the 2s and the 3s1p results are almost indistinguishable. We also plot the DFT predictions for LDA and HSE xc functional. All the calculation are performed for 54 atoms at the Γ point. *Middle panels:* Total energy (extensive) as a function of the inverse VMC bin size. The two basis extrapolate to two different energies. The extrapolation is linear, suggesting that the bias due to the finite VMC bin length is systematically improvable. *Bottom panels:* Pressure as a function of the inverse VMC bin size. This value is rather insensitive to the bin length and the basis.

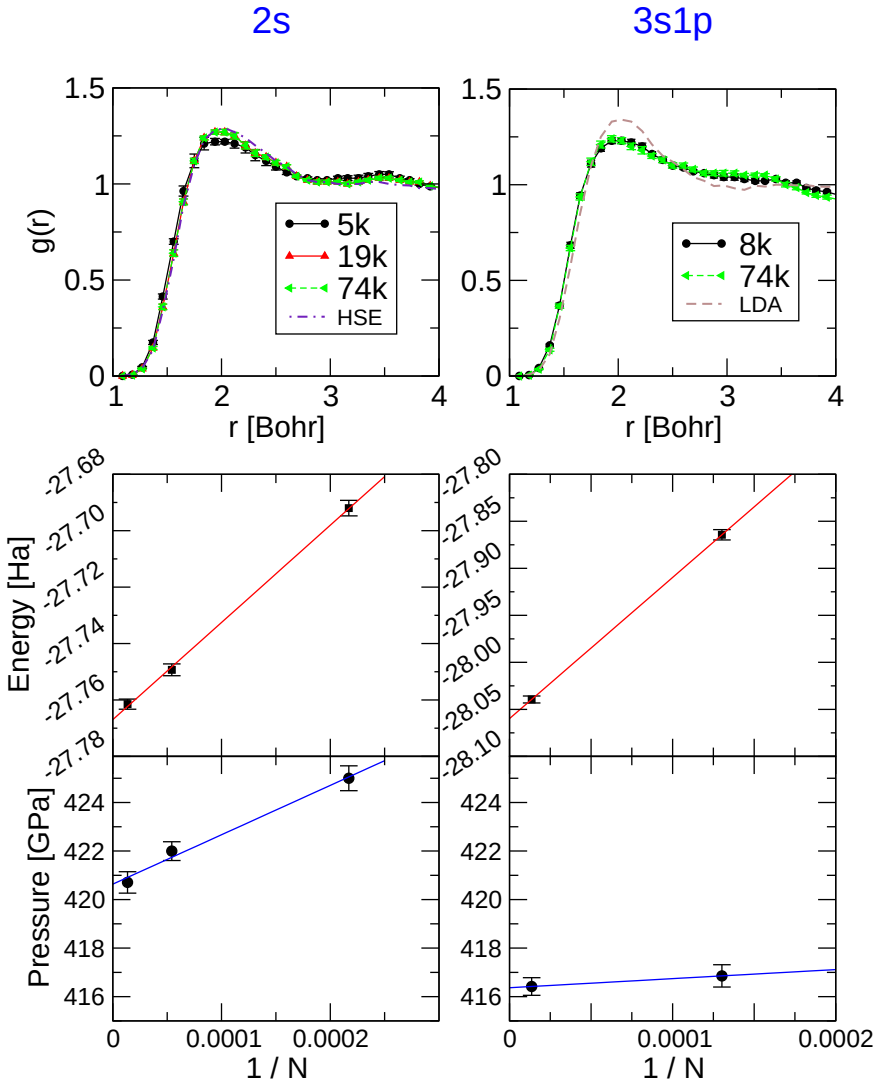


Figure 3.13: Accuracy of the wavefunction: tests at density r_s=1.24 and T=1700 K. The panels on the left refer to the 2s basis set, while the panels on the right to the much more accurate 3s1p. We also test the convergence of the observables as a function of the VMC bin length. *Top panels:* Radial distribution function obtained with different VMC bins. Notice that the 2s and the 3s1p results are almost indistinguishable. We also plot the DFT predictions for LDA and HSE xc functional. All the calculation are performed for 54 atoms at the Γ point. *Middle panels:* Total energy (extensive) as a function of the inverse VMC bin size. The two basis extrapolate to two different energies. Also in this case the extrapolation is linear. *Bottom panels:* Pressure as a function of the inverse VMC bin size. This value is rather insensitive to the bin length and the basis.

converged results for the molecular orbitals (see Table3.3). This possibility implies a drastic reduction of the variational space (from $\simeq$ 33000 parameters to $\simeq$ 5000 in a 256 hydrogen

system, using one 2*s* hybrid atomic orbitals per atom)

In Table3.2 we see that LRDMC energy depends on the choice of the guiding function, i.e. on the nodal surface defined by the trial wavefunction. Nevertheless, DMC improves the total energy of ∼3-4 mH/atom within the same variational ansatz, with respect to the VMC. We perform a DMC molecular dynamics simulation[6] to check whether this energy gain is directly connected with the forces, that is, if DMC may quantitatively change the outcome of the simulations compared to VMC.

The DMC-MD simulation scheme is very similar to its VMC counterpart, with the exception that the ionic forces and the generalized forces acting on the parameters $\{f_k\}$ are computed as mixed averages (see Sect. 2.3.4). At each step, not only the ions are moved accordingly to the ionic forces, but also the guiding function is updated, following the $\{f_k\}$'s. Notice that the equilibrium condition $\{f_k\}$=0 now does not correspond exactly to the minimum DMC energy state, but still defines an improved variational ansatz compared to the simpler VMC optimization. Moreover, this bias vanishes if the VMC trial guiding function is exact.

We perform a DMC-MD at T=600 K and density r_s=1.35 for a 54 proton system. The DMC algorithm is about an order of magnitude slower than the VMC, and larger number of atoms are not possible with the available computational resources. Nevertheless, we do not observe qualitative changes for the radial pair distribution functions and for the pressure, apart from an overall ∼5 mH/atom shift of the total energy (see Fig.3.14).

In conclusion, the DMC-MD is more accurate because the energy - is lower than the one obtained by VMC, but this is achieved with a much more expensive computational cost. Nevertheless, no quantitative changes to the physical results (for the present accuracy in the phase diagram) are found by employing DMC on this system.

[6] within the LRDMC scheme.

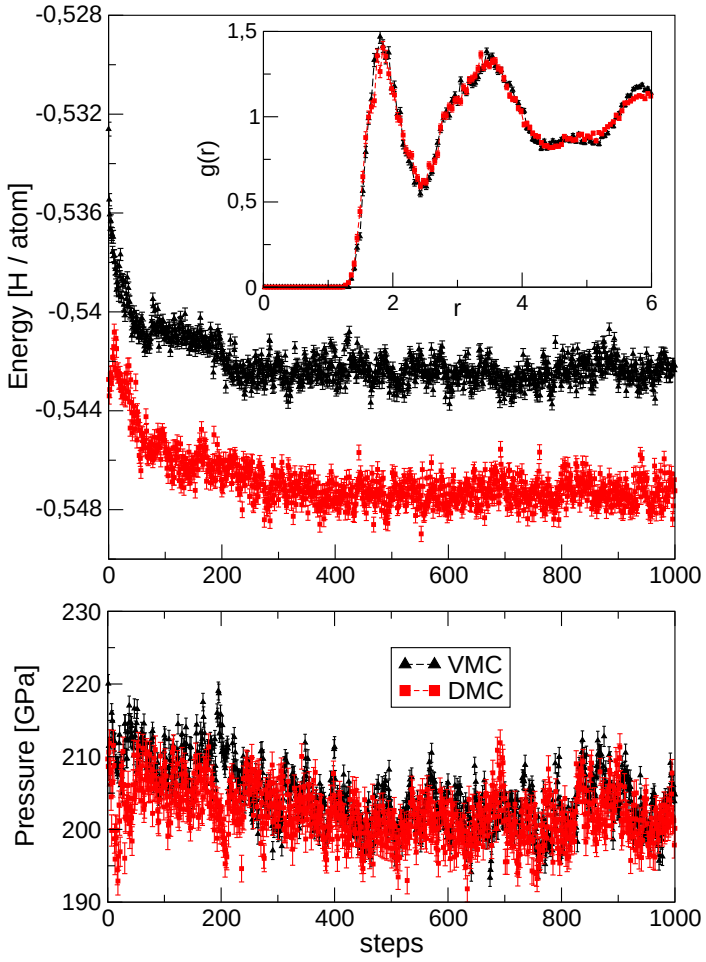


Figure 3.14: Comparison between VMC (black-triangle up points) and DMC (square-red points) dynamics at T=600 K and density r_s=1.35 for a 54 proton system. In the top panel we plot the total electronic energy as a function of the (first 1000) MD steps; in the inset the radial pair distribution function. In the lower panel we plot the pressure. The average values for the pressures (evaluated after that equilibration has been achieved) are P_{VMC} = 204(1) GPa and P_{DMC} = 201(1) GPa. This shift in pressure at fixed density is not relevant for the present accuracy in the phase diagram.

3.5.2 Ergodicity of the MD simulations

The use of a compact basis set is of fundamental importance in order to reduce the computational time required to evaluate the energy and the forces at each MD step. Indeed, it is essential to provide correctly equilibrated MD runs in order to obtain meaningful predictions. One subtle point connected with *ab-initio* MD simulations is the possible lack of ergodicity. If the computational time per step is huge, one may only afford a few MD steps and the total simulation time can be smaller than the equilibration time. In these cases, the *ab-initio* accurate methods is almost useless as it will not provide any structural change, with respect to the starting configuration. In other words, we must be sure that the ab-initio QMC-MD runs can achieve equilibration themselves, and that the outcomes are independent of the particular choice of the initial configuration.

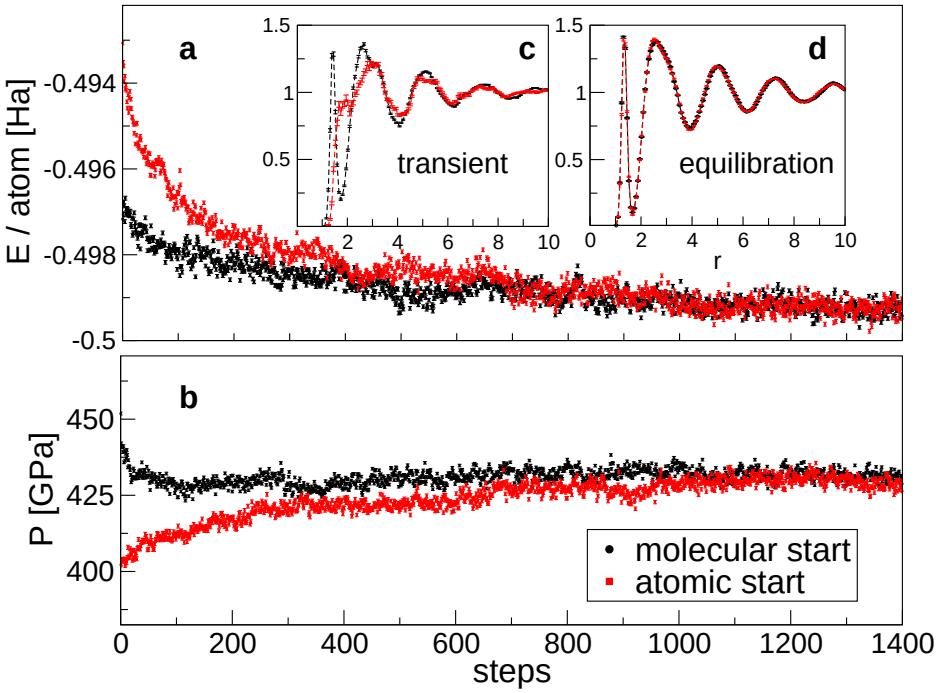


Figure 3.15: Ergodicity of the simulations. (**a**) Energy E and (**b**) pressure P as a function of the ionic steps during a Langevin dynamics simulation at $r_s = 1.28$ and $T = 600$ K. Red squares refer to a simulation whose starting configuration is an atomic fluid, while black circles correspond to a molecular initial configuration (see inset **c** for the radial pair distribution functions of these two configurations). After a short equilibration, energy, pressure, and radial pair distribution function (inset **d**) converge to the same values. The time step used in integrating the SLD is ~ 0.8 fs. 256 hydrogen atoms are considered.

As well known, hysteresis is usually found by using local updates in simple Monte Carlo

schemes, that can not be therefore reliable to determine the phase boundary of a first order transition. The MD-driven sampling is instead powerful enough that different phases can be reached during the simulation, with time scales that remain accessible for feasible computations. In Fig.3.15 and Fig.3.16 we show typical equilibration profiles displayed by the MD. Indeed the equilibration is independent from the starting point both at low temperature (see Fig.3.15) and at high temperature, very close to the LLT (see Fig.3.16). Notice that the pressure typically increases as a function of the density with a rate of $\sim$ 20 GPa/0.01 r_s (see Fig.3.2.a). Therefore, if one repeatedly misses the transition due to hysteresis effects by performing constant density simulations along a given isotherm, the LLT predicted pressure may easily be shifted by tens of GPa's. In Fig.2.8 the efficiency of the VMC-MD in exploring the configuration space with respect to the CEIMC technique -on the very same system- is clear, since we can afford a larger number of global sampling updates. Therefore we believe that the major source of possible errors in the phase diagram is connected to a poorly equilibrated ionic sampling. For this reason we chose to employ a small basis set and perform long and equilibrated MD runs rather than to improve the energy by a few mH's but having short MD simulations.

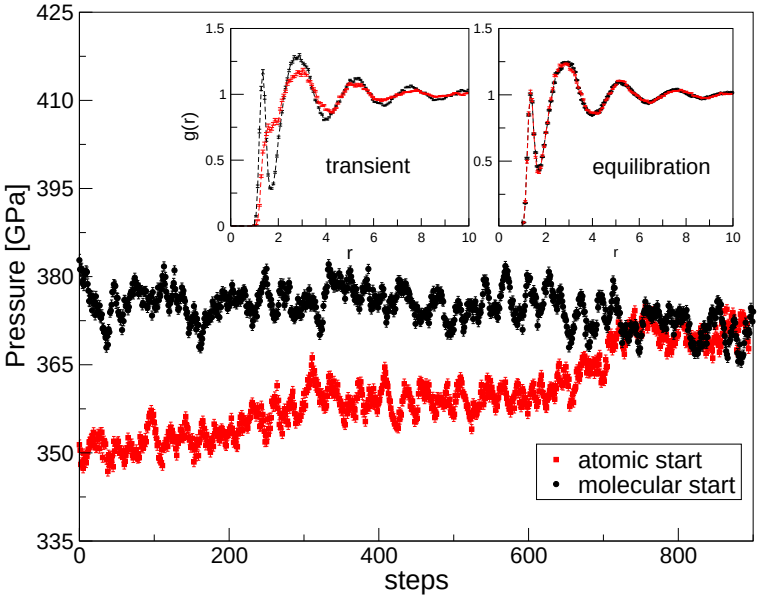


Figure 3.16: Pressure as a function of simulation steps for two different starting configuration at T=2300 K and density r_s=1.32 (the LLT being between r_s=1.31 and 1.32). Black points correspond to a mainly molecular initial distribution while the red ones to an atomic fluid (left inset). The two simulations thermalize halfway between the two possibilities (right inset). The time step used in integrating the SLD is $\sim$ 0.4 fs. 256 hydrogen atoms are considered.

3.5.3 Finite size effects

All the results for the LLT presented here refer to a cubic supercell at the $\Gamma-$point with the largest affordable number of atoms (256) in order to be as close as possible to the thermodynamic limit. Indeed, even though the pressure seems to converge with the size of the simulation cell, the molecular (atomic) nature of the liquid is very sensitive to the number of atoms N.

In Fig.3.17 we show how the finite size effects may bias the result, as the prediction obtained with 256 H is different from the 128 H case. This bias affects both VMC and DFT MD simulation.

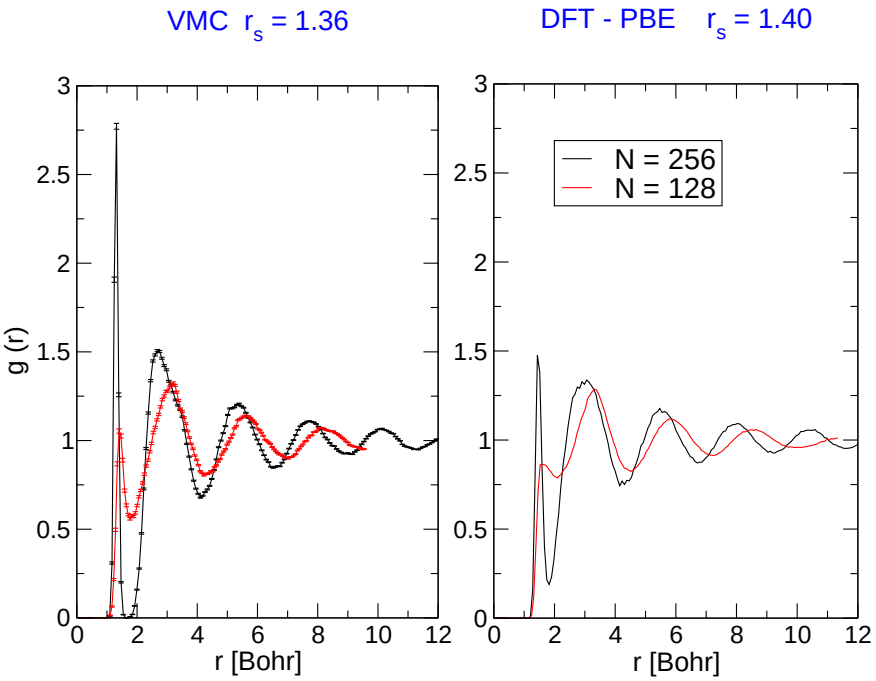


Figure 3.17: Radial pair distribution function calculated with different system sizes (N=256 atoms, black line and N=128 atoms, red line). *Left panel:* Comparison within VMC-MD at r_s=1.36 and T=600 K. The underestimation of the molecular peak is evident in the 128 H case. *Right panel:* Comparison within PBE-MD at r_s=1.40 and T=600 K.

This issue was previously reported in Refs. [15, 127] and cannot be removed with a better k-point sampling, because this will be equivalent to enforce a fictitious periodicity to a liquid phase. Therefore the critical LLT density is severely underestimated by employing supercells smaller than N=256, which is now considered a standard size in DFT simulations of liquids. The main reason of this effect may be structural frustrations, requiring the use of much larger

supercells, whose dimension L has to exceed the correlation length of the liquid. A possible rule of thumb consists in checking that the $g(r)$ is smoothly approaching its asymptotic value 1 at $r = L$. Our evidences support the conclusion that a failure in dealing with the finite size effects will result in a severe underestimation of the LLT critical densities.

Within VMC we did not perform a systematic check on the finite size employing system larger than 256. This may be indeed possible in the near future. For the time being, we can assume that the 256 system provides almost converged result in the liquid phase by comparing with DFT finite size scaling[15] at Γ point. Again, the error in the estimated LLT pressure is more than 100 GPa, if small simulation cells are used.

3.6 Nuclear quantum effects

In this work classical protons are assumed. This may be certainly an issue for low temperature simulations, however our conclusion still holds because

- There is experimental evidence[10] that, at high temperatures $\sim$2000 K, there is no isotope effect. This means that the LLT's shift obtained at high temperature will not be affected by the inclusion of the zero point motion.
- At lower temperatures, path integral MD (PIMD) simulations[17], shows that proton quantum effects only shifts the position of the LLT by 40 GPa's at $\sim$ 1000 K, with respect to the classical case.

Therefore, even by considering an upper limit of 100 GPa shift to lower pressures, still the expected complete atomization will occur at 500 Gpa and 600 K, that is a pressure substantially larger than the DFT prediction.

We spent some effort in order to include quantum proton effects in our framework. To this end, we explored two possible routes, namely the *colored noise* method[81, 82, 83] and the standard *path integral* formulation[128].

3.6.1 Colored Noise

The colored noise thermostat is a particular Langevin thermostat that one put on top of a Newtonian MD equation for the classical nuclei in order to include quantum fluctuations up to the harmonic level. The final equations of motions (EOMs) are Generalized Langevin Equations (cf. Sect. 2.5). In our case, the EOMs for the classical nuclei are already of the GLE type, so we should in principle combine the two noises. Therefore, a correct implementation of the colored noise equations on our Langevin MD is quite an hard task. A possible workaround is to reduce as much as possible the QMC noise and consider the EOM close to

the Newtonian limit (i.e. zero noise and zero friction, cf. Fig2.4) and put the colored thermostat on top of them. However, this strategy provided unsatisfactorily results even for the simple H_2 molecule (see Fig.3.18). Moreover, for the larger 54 and 256 H systems this scheme is extremely inefficient and therefore unpractical.

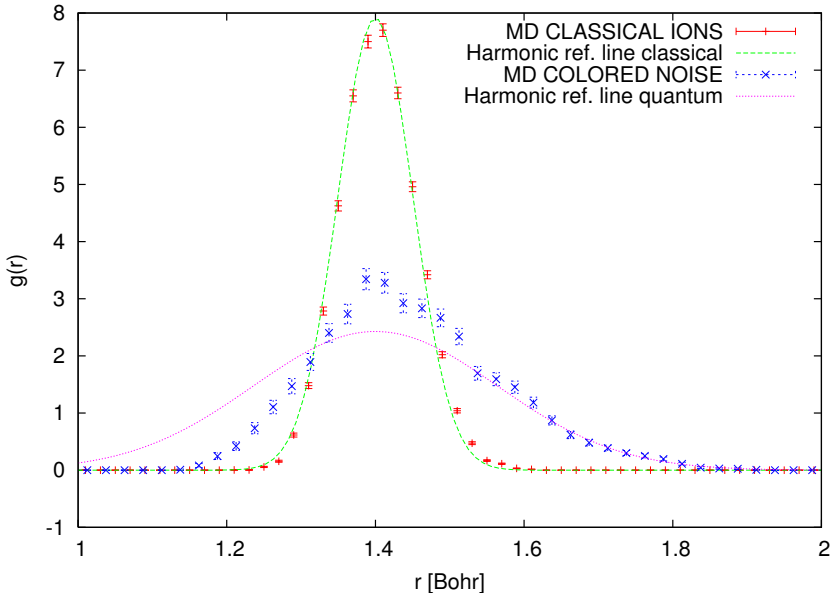


Figure 3.18: VMC-MD for H_2 molecule at 300 K, obtained with classical GLE dynamics (red) and with our implementation of the colored noise (blue). The classical case is in agreement with the theoretical Boltzmann distribution $\rho_{cl} = \sqrt{\frac{\omega_0^2\beta}{2\pi}} \exp\left(-\frac{\beta\omega_0^2(x-x_0)^2}{2}\right)$ with $x_0 = 1.4$ Bohr and $\omega_0 \simeq 0.611$ a.u. given by the harmonic approximation of the realistic potential (green line). The "quantum case" shows a more broad distribution of lengths that is only qualitatively in agreement with the expected distribution of the harmonic model $\rho_{QM} = \sqrt{\frac{\omega}{\pi}\tanh\left(\frac{\beta\omega}{2}\right)} \exp\left(-\frac{\omega}{2}\tanh\left(\frac{\beta\omega}{2}(x-x_0)^2\right)\right)$ with $\omega = \omega_0\sqrt{m/2}$.

3.6.2 Path integral molecular dynamics

Within the Born-Oppenheimer approximation, nuclear quantum effects can be simulated within the imaginary time path integral formalism[128, 129]. This formalism maps the quantum mechanical partition function for a set of distinguishable nuclei onto the classical partition function of a so-called ring polymer, composed of N_B replicas (beads) of the physical system connected by harmonic springs. We refer the reader to textbooks for the details of this technique. Let us just recall that, (i) the spring constant is proportional to the physical temperature T, therefore quantum fluctuation are suppressed at high temperature as the ring polymer becomes extremely localized. (ii) The configurations of the ring polymer can

be sampled from a canonical ensamble with an effective temperature $T^* = T\ N_B$. For this reason, the Langevin dynamics appears to be very suitable to generate the polymer configurations, as the accuracy needed for evaluating the forces decreases with the number of the beads (see Sect. 2.4.2). We implemented the path integral molecular dynamics (PIMD) technique within out generalized Langevin framework using the so-called *primitive* approximation for the action, namely the simplest Trotter discretization for the path integral. The actual quantum correction are recovered in the $N_B \to \infty$ limit. In practice a finite number of beads is sufficient. This number depends on the maximum frequency which appears in the system and does not grow with the system size. Usually a $N_B \sim 40$ is sufficient to obtain converged result for real molecules. We test the algorithm simulating proton quantum corrections on the H_2 molecule at 300 K. We compare our results with a reference quantum harmonic oscillator model, which is analytically solvable. The radial pair distribution (see Fig.3.19) and the total energy (see Fig.3.20) are in good agreement with the predicted values. Notice how the classical behavior turns into the quantum one as the number of replicas is increased.

Finally, we apply the PIMD on the 54 hydrogen bulk system. For this preliminary test, we choose a molecular liquid at density r_s=1.44 and temperature T=600 K, namely at parameters extensively probed in Sect. 3.5.1. In Fig. 3.21 we see that the quantum corrections melt the molecular peak only partially. The estimated pressure is P_{PIMD}=125,1(4) GPa, whereas the "classical" result reported in Sect. 3.5.1 is $P = 122{,}1(4)$ GPa. The total energy now reads E_{PIMD}=-0.54624(5) H/atom. Therefore, the zero point energy is 8.4 mH/atom $\approx$ 0.23 eV/atom (see left panel of Fig. 3.12), namely slightly smaller than the ZPE contribution for the isolated molecule at the same temperature $\approx$ 9.1 mH.

Notice that, at this temperature, quantum corrections are relevant but not dramatically as long as the pressure and the $g(r)$ are concerned. This fact justifies our classical nuclei approximation, at least considering the present accuracy on the hydrogen phase diagram. Nevertheless, the quantum corrections are expected to shift the transition pressure by some tens of GPa's to lower values as already reported with PIMD using DFT[17]. The quantitative determination of this shift using systematic PIMD simulations is left for future work.

3.7 QMC vs DFT: Analyzing the differences

Until very recently, the Density Functional Theory (DFT) method has been considered the standard tool for the simulation of electronic phases, because it allows the simulation of many electrons with a reasonable computational effort. However, there are several drawbacks in this technique especially for the study of the dissociation of hydrogen:

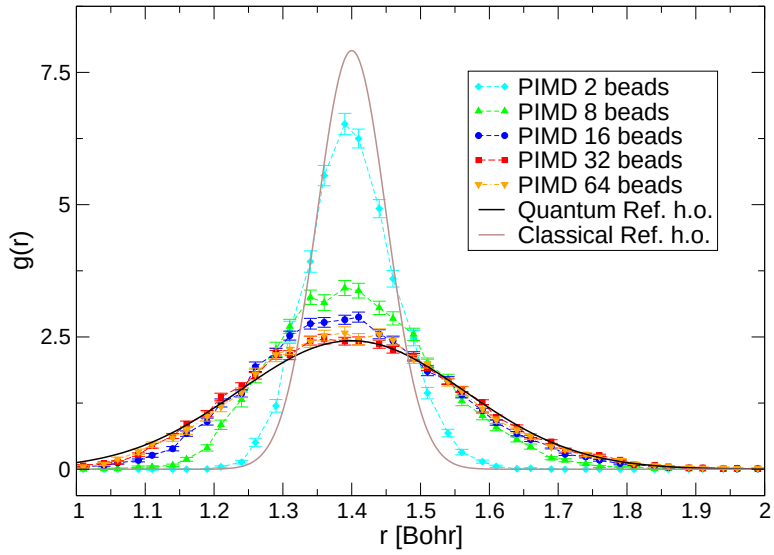


Figure 3.19: Radial pair distributions for H_2 molecule at 300 K obtained with PIMD simulations with different numbers of replicas. The converged result is recovered for N_B=32. The classical case distribution is given by $\rho_{cl} = \sqrt{\frac{\omega_0^2\beta}{2\pi}} \exp\left(-\frac{\beta\omega_0^2(x-x_0)^2}{2}\right)$ with $x_0 = 1.4$ Bohr and $\omega_0 \simeq 0.611$ a.u. given by the harmonic approximation of the realistic potential. The converged PIMD simulations (red, yellow points) show a more broad distributions of lengths which are in agreement with the expected distribution of the harmonic model $\rho_{QM} = \sqrt{\frac{\omega}{\pi}\tanh\left(\frac{\beta\omega}{2}\right)} \exp\left(-\frac{\omega}{2}\tanh\left(\frac{\beta\omega}{2}(x-x_0)^2\right)\right)$ with $\omega = \omega_0\sqrt{m/2}$ and m is the proton mass in atomic unit.

- the single molecule is not accurately described at equilibrium and especially in the dissociation limit [130, 37] (see Fig.3.22).

- Electronic gaps are substantially underestimated [5] within DFT, implying that possible molecular phases are more easily destabilized within standard density functionals (DFs).

For all the above reasons, DFT seems not adequate for the hydrogen problem under high pressure, especially in a range of pressures inaccessible by experiments, where the quality of a particular DF cannot be validated.

Recently, it has also been shown that DFT solid stable phases strongly depend on the DF used [18, 110, 131], suggesting quite clearly that the predictive power of DFT is limited for hydrogen. This issue remains also within the liquid phase, since it has been shown (see Fig. 3.8) that LLT's predictions made with PBE and HSE xc functional differ by more than 100

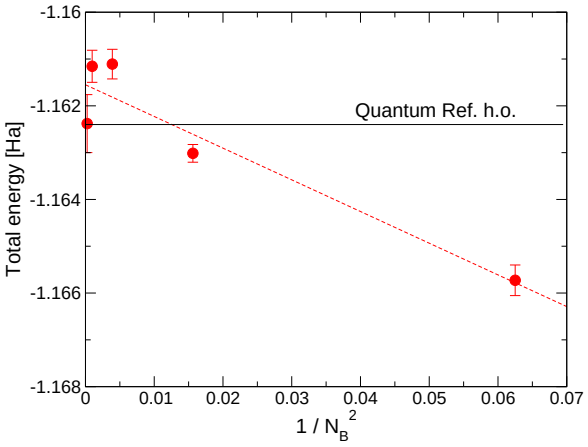


Figure 3.20: Total energy with quantum corrections given by PIMD for the H_2 molecule a T=300 K as a function of the number of replicas. The energy is plotted against $1/N_B^2$ as the Trotter error is expected to scale in this way for the primitive approximation of the action. The reference value for the quantum harmonic oscillator is given by $E_{H.O.}(T) = \frac{1}{2}\omega \coth \frac{\omega}{2T}$ with $\omega = \omega_0\sqrt{m/2}$.

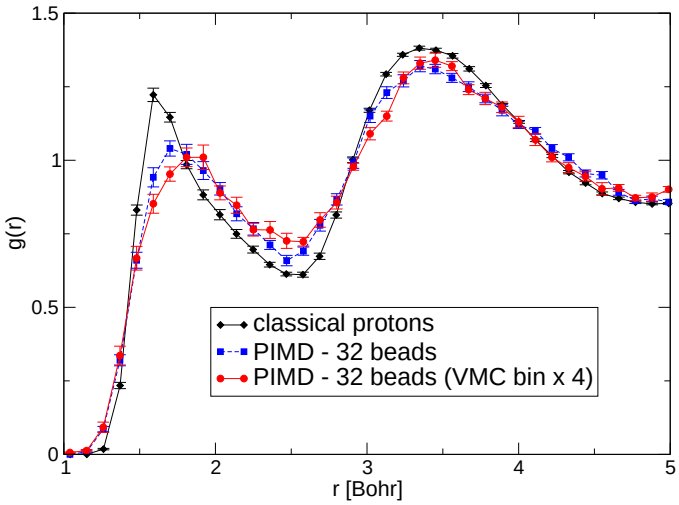


Figure 3.21: Radial pair distribution function with proton quantum corrections (blue and red points). We employ PIMD using 32 beads. We test the convergence of the PIMD run as a function of the VMC bin length. The converged simulation is displayed with red points. The distribution obtained with the VMC-MD using classical protons as in Sect. 3.5.1 is displayed in black. The density is r_s=1.44 and temperature T=600 K.

GPa's. While the HSE[40] functional, gets semi-conductor gaps right to about 0.2 eV error, this does not necessarily imply that HSE can estimate the transition pressure with similar accuracy. Moreover, it is suggested in Ref.[17] that the large difference in the transition pressure estimated by PBE and by HSE is explained, most likely, by the different description of the H_2 bond length by these two methods.

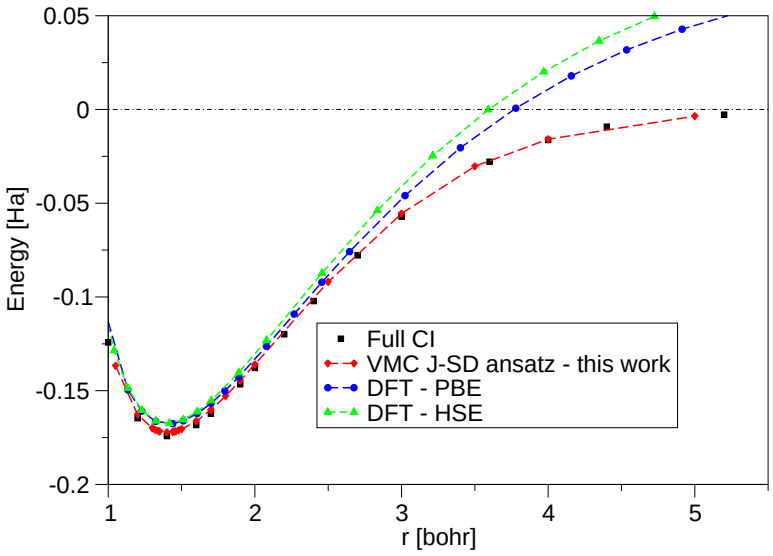


Figure 3.22: Dissociation energy curves for the H_2 molecule for different methods, QMC at the VMC level and with the same 2*s* wavefunction variational ansatz employed in the dynamics, DFT with PBE or HSE DFs, and the exact curve obtained with full configuration interaction (CI) method[101].

However, as shown in Fig.3.22 we find that the difference between PBE and HSE for the H_2 potential energy surface is rather small when compared to the difference from the exact result, which instead is reproduced correctly by our VMC method, with the small 2*s* basis set.

We tried also to better understand the differences between VMC and DFT in the bulk system, as the errors in the long range part of the dissociation curve are not directly related to the bulk dense liquid. Nevertheless, it is interesting the fact that standard functionals fail for such simple system. Therefore we analyzed a molecular dissociation process in the bulk. To this end, we take a single 54 H molecular configuration and slowly act on the intramolecular distance of only one among all the possible pairs. Notice how the overall shape of the dissociation curve looks very different from its isolated counterpart (see inset of Fig.3.23). Indeed both the equilibrium length and the long range behavior strongly differ

due to the presence of the surrounding atoms.

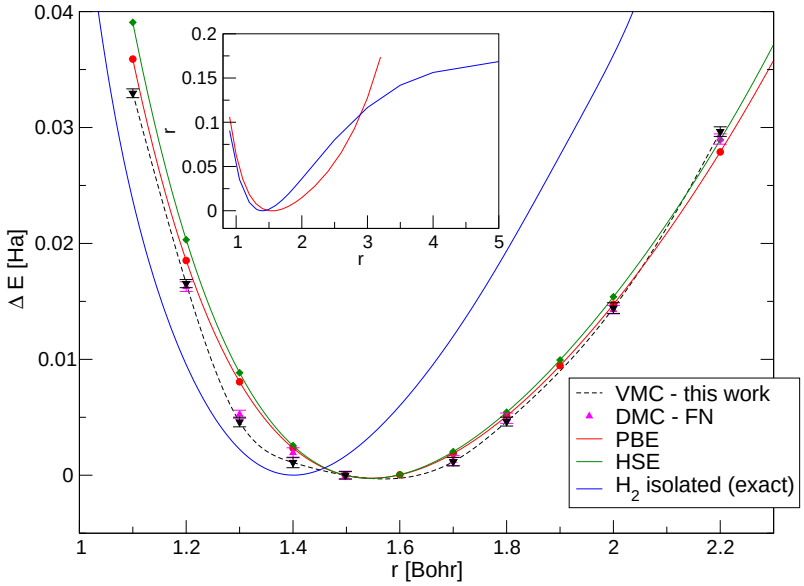


Figure 3.23: Dissociation energy curves for a H_2 molecule in the bulk for different methods, QMC at the VMC (black points) and LRDMC (pink) level, with the same 2*s* wavefunction variational ansatz employed in the dynamics , DFT with PBE (red) or HSE (green) DFs. The precise shape of this functions depend on the particular configuration chosen (i.e. the position of the other 52 atoms), nevertheless, the differences between the methods are quite general. The exact dissociation curve for the isolated hydrogen molecule is also plotted (blue). The dashed black line, which connects the VMC points, is a guide for the eye. Notice in the inset the qualitative difference between the dissociation in the vacuum and in the bulk. The curves offsets are chosen in order to have the lower energy points set to zero.

The equilibrium length may differ from the isolated case even by 0.2 Bohr. In Fig.3.23 we observe that

- The difference between DFT and VMC are rather tiny, and this somewhat validates again our VMC approach.
- The difference between PBE and HSE are even smaller. This fact confirms the sensitivity of the LLT to the accuracy of the method.
- The PES obtained with VMC shows a flatter minimum than DFT. Therefore the molecules in the bulk described by VMC are less *rigid* than their DFT counterparts.

The picture coming from Fig.3.23 is in agreement from the radial pair distribution of the forces, calculated on the same configuration (see Fig.3.24). This time, such target quantity is

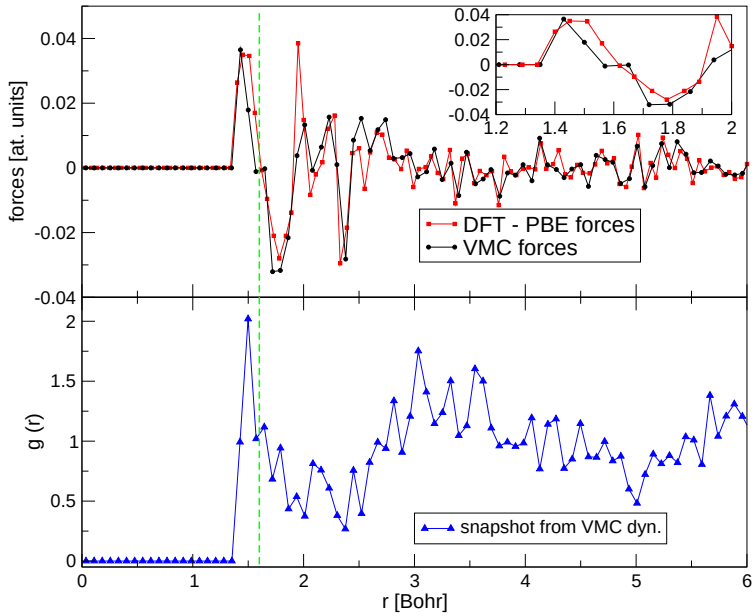


Figure 3.24: *Top panel:* Radial pair distribution of the forces. This is the mean force that 2 ions at distance r feel, averaged on a single configuration. Forces evaluated by VMC and DFT-PBE are overall consistent and they differ mainly (see inset) around 1.6 Bohr, i.e. near the molecular bulk equilibrium distance (dashed green line). *Bottom panel:* Radial pair distribution corresponding to the protonic configuration (single snapshot) taken as a test case.

averaged over all the possible pairs, so it contains independent informations compared to the previous -single molecule- PES calculation. Again, we observe that the molecular stiffness is strongly reduced in the case of VMC forces. Therefore, we believe that QMC provides a more flexible description of the molecular bond in the bulk, that energetically penalizes less intramolecular bond fluctuations (due to the thermal fluctuation in the dense environment). In this way we can explain the enhanced molecular stability at large densities, compared to DFT.

It also interesting to compare energies and forces corresponding to the J-SD and the J-DFT variational ansatz. Let us recall, that the J-SD trial function is obtained after a full Jastrow and determinant optimization whereas in the J-DFT ansatz we simply take the determinantal part from a DFT-LDA calculation and optimize only the Jastrow term. In Fig.3.25 we show that J-DFT forces are substantially less accurate with respect to the ones coming from a J-SD calculation. Therefore we prove that a full wavefunction optimization, is needed to obtain accurate forces. Indeed the J-SD forces, i.e. the ones employed in our molecular dynamics, are almost indistinguishable from the LRDMC (fixed nodes) forces.

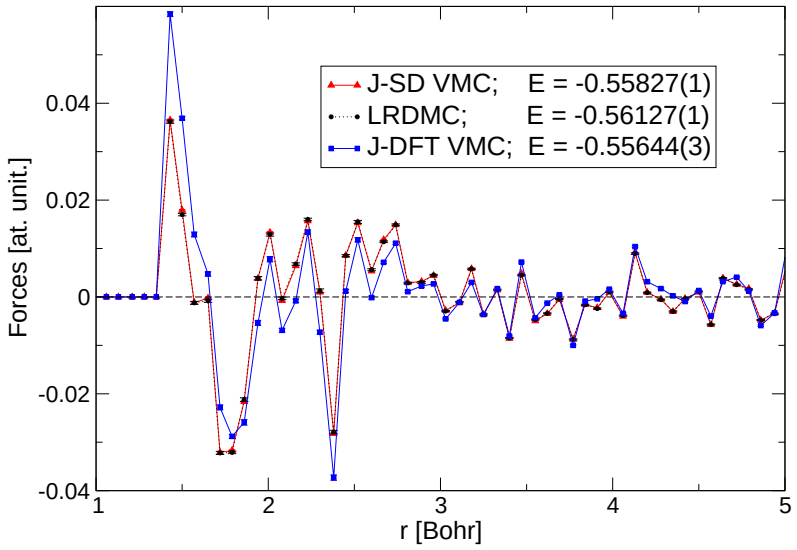


Figure 3.25: Radial pair distribution of the forces and total energy calculated on the same protonic configuration as in Fig.3.24 with different methods. Red line correspond to a VMC calculation using a J-SD ansatz like in our MD. These forces are very close to the LRDMC ones (black dotted line). The J-DFT ansatz provides instead much less accurate forces. Notice also that the J-DFT forces are not zero at the LRDMC molecular equilibrium length. The LRDMC calculation is performed using an a=0.2 a.u. lattice space[63]. A 2s/atom localized basis set is used in all the calculations for the determinantal part of the trial/guiding wavefunction.

To conclude the analysis, we perform also a comparison between DFT, with several xc functionals, and VMC, for different types of liquids, from a fully molecular to an atomic one. In Fig.3.26 we compare the prediction given by the different electronic methods on a 54 H system at Γ-point: we clearly see that DFT always enhances the molecular dissociation and therefore underestimates the LLT transition pressure. As expected, all the methods coincide in the large pressure liquid, since correlations become less important in the high density limit. Qualitatively, VMC and DFT-HSE or PBE differ only at the transition, as VMC still maintains a small molecular fraction. Notice how the HSE/PBE curves are closer to the VMC rather than to LDA predictions. These data confirm the trend for which, the greater is the complexity/accuracy of the xc functional used in DFT simulations, the more molecular the fluids will result.

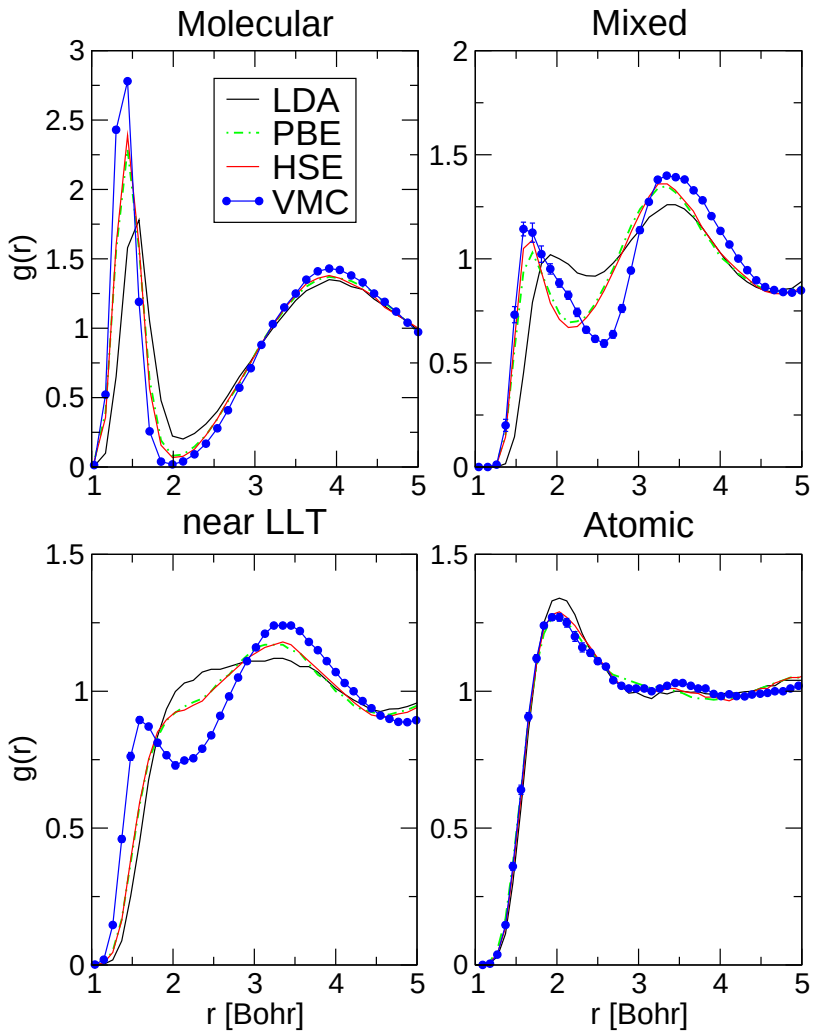


Figure 3.26: Radial pair distribution comparison between DFT with LDA (black), PBE (red), HSE (green) xc functionals and VMC (blue), at different phase space points. *Top left:* fully molecular configuration obtained at parameters r_s=1.80 and T=1800 K. *Top right:* mixed atomic-molecular configuration, at parameters r_s=1.44 and T=600 K. *Bottom left:* simulations near the LLT at r_s=1.44 and T=1800 K. Notice that VMC still predicts a fraction of stable molecules, while the DFT is already in the atomic regime. *Bottom right:* fully atomic configuration, at r_s=1.24 and T=1800 K.

3.8 Conclusions

In this Chapter we have explored the liquid phases of dense hydrogen with an ab-initio QMC molecular dynamics. We focused in particular on the transition between the -low pressure- molecular liquid and the -large pressure- atomic liquid. We observe that the molecules are exceptionally stable, as we found a nonzero molecular fraction up to 375 GPa at 2300 K and 600 GPa at 600 K. Nevertheless, our simulation predict the existence of a large region in the phase diagram that is a mixed molecular-atomic liquid (see Fig.3.8).

Therefore the first order LLT is found at large pressure between this mixed-molecular liquid and the completely dissociated phase (Sect. 3.3). Conversely we collect some clues that the transition between the fully molecular and partially dissociated liquid may be related to the IMT. Thus the metallization at finite temperature occurs within the (partially) molecular phase. At high temperature this IM transition mechanism appears to be smooth, since it is related to the interatomic distances distribution and the onset of molecular dissociation (Sect. 3.4). Our picture is in agreement with the widely discussed experimental Ref.[10], but it is not consistent with previous DFT simulations on this topic[13, 16, 17] as well as with CEIMC simulations[16].

Possible explanation of such differences are provided in Sect. 3.7 , with systematic tests on the accuracy of our QMC framework (Sect. 3.5). We observe that the major sources of bias are given by finite size effects and possible lacks of thermalization. Therefore it is essential to have long MD simulations, on very large sizes.

Moreover, from the point of view of the method, in this work we improved with respect to the original work by Attaccalite and Sorella[25] in the following aspects:

- Using *locality* in the molecular orbital optimization (cf. Sect. 3.5.1) reduces the computational time and offers the possibility to simulate larger sizes (256 electrons) and for a larger number of MD steps.
- We employ a different MD integration scheme, which allows to use a longer time step even in the large friction regime. This helps in providing more uncorrelated samples during the dynamics.
- We understood the role of the covariance matrix of the QMC noisy forces and the most efficient setting of MD/wavefunction/optimization parameters in order to reduce the autocorrelation time in the MD.

Nevertheless, in order to exactly quantify the bias given by the VMC trial wavefunction used in this work, we plan to redo the calculations with a larger basis, both on the smaller 54 proton system and -if the computational resources will be available- on the larger 256 H size.

The main results are obtained employing a strong approximation, namely considering the protons as classical objects. Indeed we have recently explored the possibility of performing PIMD simulations within our framework. Preliminary tests suggest that, even at the lowest temperature considered of 600 K, proton quantum corrections are not substantial, thus validating our general conclusions on the stability of the molecular phase. Moreover, the inclusion of the zero point motion using path-integral techniques may allow us to explore the low temperature region of the hydrogen phase diagram and address many interesting open questions such as the possible existence of a ground state liquid phase.

Chapter 4

Liquid water at room conditions

In this short Chapter we report some preliminary results on liquid water at ambient temperature and density. If dense liquid hydrogen is the most common substance in our universe, liquid water is, by far, the most important one for the life in our planet. Water is also probably the most studied system, both computationally and experimentally. Nevertheless, there are still many open questions on liquid water: one of these concerns the structure of bulk water at room temperature. We address this problem by using the very same QMC molecular dynamics technique which was previously described and employed for the dense hydrogen problem.

In Sect. 4.1, we first provide the motivation of this research, then, in Sect. 4.2 we report some of the latest results obtained with these first QMC-MD simulations on liquid water. We obtain encouraging results, which are very well in agreement with recent experimental references. This fact also provides an independent evidence of the accuracy of this technique.

4.1 Introduction

The structure of water is a recurring theme at the interface of contemporary physics, chemistry, and biophysics. Even if the liquid does not have a precise static *structure* in a lattice sense, the characteristic forces between the water molecules result to a well defined radial and angular molecular distributions at given temperature.

Since it is not possible to directly measure the intramolecular forces, the only way to understand this delicate balance between covalent, hydrogen bond and dispersion forces, is by looking at the structure that they produce. The precise understanding of these many body interactions are also important for developing simplified and computationally cheap classical model which can be used in the ever growing field of biophysical simulations.

From the experimental point of view, even if the experimental conditions are easy to realize (compared to the dense hydrogen problem), the difficulties arise from the interpretation of the spectroscopic measurements[35]. Therefore, there are still many open questions about the structure of liquid water (see e.g. Ref.[31] for discussions on two experimentally proposed structures of water: "tetrahedral" or "chained").

From the point of view of simulations, clusters and bulk water have been extensively studied by ab-initio MD within DFT, employing a large variety of approximations for the exchange correlation density functionals. Among semi-local functionals, the most widely adopted are the gradient corrected functionals BLYP and PBE. The first BLYP-DFT simulation appeared in 1993[132] and, after that, there have been several efforts to obtain structural information for water at ambient conditions from DFT-based molecular dynamics simulations[32, 33, 30]. From the comparison with experiments, it has been shown that DFT yields in general an overstructured liquid at the 1 g/cm^3 density and ambient conditions[32, 33, 30, 133]. The fact that the melting temperature of ice was found to be around 410-420 K for DFT simulations with PBE and BLYP functionals (see Ref.[133] for details), suggests that ambient condition DFT simulations reproduce a supercooled state instead of "real" liquid water.

In our approach we can only provide results on the static quantities, like the Oxigen-Oxigen radial pair distribution function $g_{OO}(r)$. DFT simulations using standard xc functionals, usually predict an overstructured liquid, that is, having an higher first peak and a smaller first minimum, compared with the experimental references. The first peak position also depends on the xc approximation (see Ref.[30] for a systematic comparison between DFT functionals), and its predicted position may vary in a range of $\sim$ 0.05 Å . The inclusion of empirical corrections clearly improve the agreement with the experiment, but at the expense of missing a true *ab-initio* approach.

Moreover, a recent accurate experiment of X-ray diffraction[34] has raised again the issue of the reliability of present ab-initio molecular dynamics schemes, as it was found that, surprisingly, the position of the first peak was shifted towards larger distances. This observation is in excellent agreement with a recent extensive and independent review on the experimental results on the structure of bulk water[35]. Indeed in Ref. [35] a new methodology to interpret the experimental data is employed and a shift also of the intermolecular O-O, O-H and H-H peaks positions with respect to the old (2000) experimental reference[134] are reported. In particular, in 2000, the first O-O peak had a height of 2.75 at a position of 2.73 Å , whereas with the new analysis refinements it has a height of 2.49 at a position of 2.82 Å (see Fig. 4.2)

4.2 Results

We apply the molecular dynamics driven by quantum Monte Carlo forces as in the case of liquid hydrogen at high pressures. We have employed a simulation box containing 32 water molecules in the NVT ensamble at experimental density. Finite size effects studies within DFT[30] showed that this size is already sufficient to obtain quite converged radial pair distribution function. Since equilibrium properties does not depend on the mass, we have used the Oxygen mass equal to the Hydrogen one: this choice clearly improves the autocorrelation times. We have done about 5000 MD iterations, where at each iteration we optimize about 12000 variational parameters. The localized basis set is optimized using the water dimer system as a benchmark. We employ a *5s5p1d* basis set, expanded over 5 hybrid orbital per Oxigen atom and a *3s1p* over 3 hybrid orbitals per Hydrogen atom, for the determinantal part[1] of the wavefunction. We expand the 3-body Jastrow on a *3s2p* basis set per Oxigen atom and a *2s2p* set per Hydrogen atom. Extensive discussions on the trial wavefunction choice and performance comparison between other techniques for the water dimer and examers, will be reported in a forthcoming publication[135].

Our starting configuration is generated by a DFT-BLYP dynamics. The MD Langevin simulation is sufficiently long to observe a complete relaxation for the local quantities, like the short range sector of the $g(r)$'s. We see in fig.(4.1) that these first results are very encouraging. Despite the noise, the outcome is quite clear, because the O-O $g(r)$ is much closer to experiments than the corresponding DFT calculations. Not only the radial distribution function is much less overstructured, but also the position of the first peak is almost indistinguishable from the most recent experiments. In order to avoid possible size effects we study in Fig.(4.2) the position of the first peak with a much shorter simulation (600 steps) with 64 water molecules. Since this position equilibrates rather smoothly with the length of the simulation $\simeq N =$ #steps, we are able to extrapolate the first peak position also for the largest size. Notice that the $1/N$ extrapolation scheme derives from the fact that we include in the analysis also the equilibration time, therefore the out-of-equilibrium value (computed using the first N steps) contributes to the total average with a 1/N weight. We find that the position of the first peak in $g(r)$ remains very close to the smaller size simulation, supporting the validity of our finding.

[1] with a cutoff distance of r_{max}=4.5 Bohr in the molecular orbital optimization.

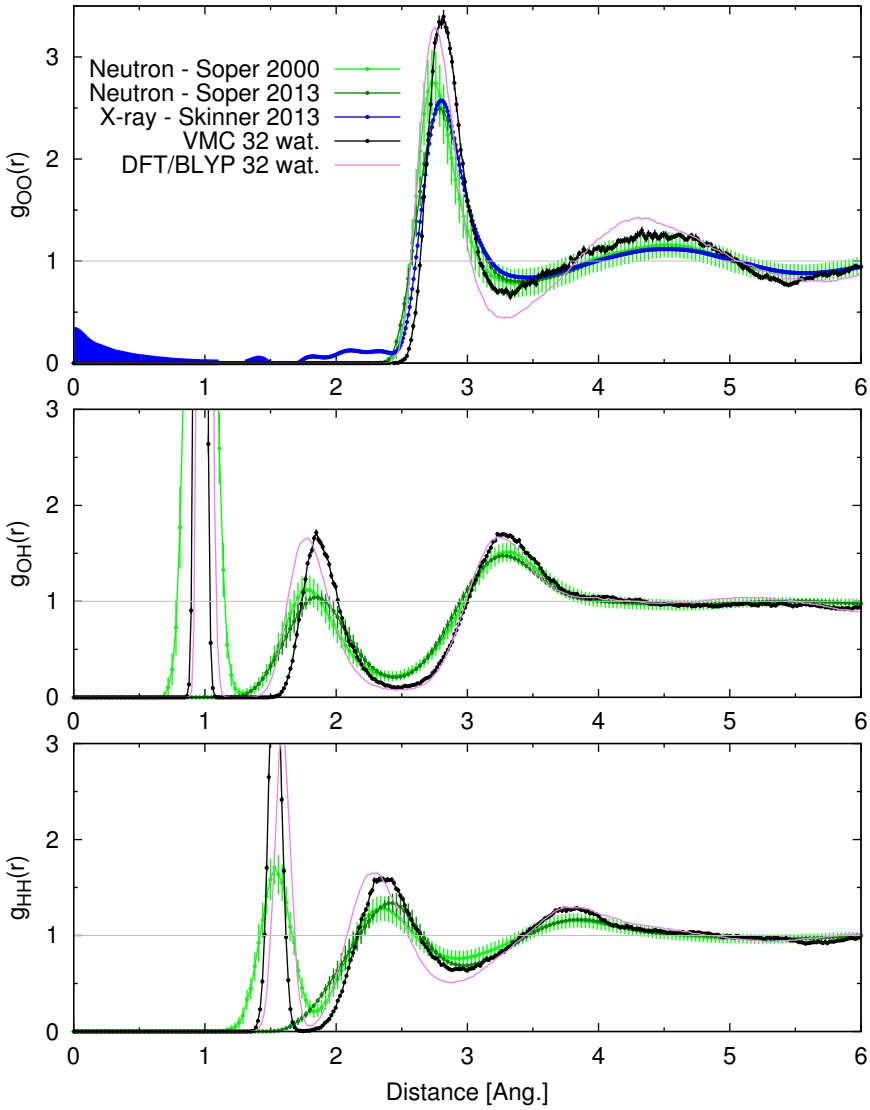


Figure 4.1: Radial distribution function $g(r)$ obtained with 32 waters by VMC as compared with DFT-BLYP and experiments (X-ray[34] or neutron diffraction[134, 35]) (a): Oxygen-Oxygen, (b) Oxygen-Hydrogen, (c) Hydrogen-Hydrogen. A blow-up of the Oxygen-Oxygen first peak is reported in Fig. 4.2.

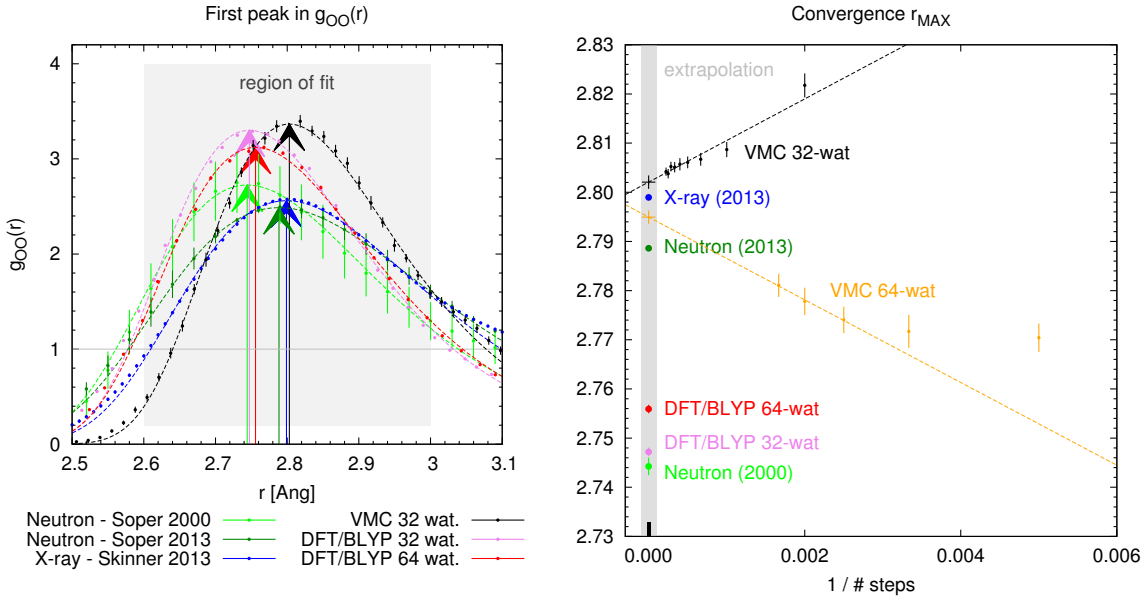


Figure 4.2: Left panel: First peak in the Oxygen-Oxygen radial distribution function $g_{OO}(r)$ obtained from experiments (X-ray[34] in blue, neutron diffraction[134, 35] in green), from the QMC-MD simulations with 32 waters (in black), and a DFT/BLYP-based MD simulation with a PBC box of 32 or 64 waters (in violet and red, respectively). The $g_{OO}(r)$ data have been fitted with a function $f(r) = a\exp(-\frac{r-\mu}{b} - \exp-\frac{r-\mu}{b})$ for r in the range 2.6 – 3.0 Å (gray square). The values of $r_{\mathrm{MAX}} \sim \mu$ have been highlighted with colored coded arrows. Right panel: The plot reports the values of r_{MAX}, for the QMC-MD trajectories that start from a given configuration and are of increasing length, over the inverse of the length of the simulation (*i*.e. the number of steps). The linear fit of the leftmost values provides the extrapolation to zero, that is the equilibrated value (coinciding in the 32 water case with the average over the last-2500 equilibrated steps of dynamics). These values are compared with other methods, reported in the shadowed area.

4.3 Conclusions

We have reported the first ab-initio molecular dynamics simulations within a QMC approach on bulk liquid water. Notice that, before this work, QMC has been recently used only as a tool for benchmarking and analyzing DFT errors on fixed configurations[136, 137], as the current belief is that QMC remains too expensive for regular use in first-principles MD simulations of bulk systems[137]. Indeed, employing the CEIMC technique to this system is particularly expensive because of the rather inefficient Metropolis sampling, based on energy differences, at low temperatures[138].

We have shown the feasibility of combining QMC, at least in its simplest variational formulation, within a MD framework in order to obtain configuration samples of bulk liquid

water in thermal equilibrium, containing up to 64 molecules. The good comparison with recent experimental references shows that this approach is very promising. Systematic comparisons between QMC and DFT, as done in Sect. 3.7 for hydrogen, will be necessary in order to understand the differences between the two techniques. Indeed, the dissociation curve of the isolated water dimer is correctly described by BLYP at equilibrium distance, so it is likely that the DFT shift in the $g_{OO}(r)$ is due to a nonoptimal description of the interaction beyond the two body (two water molecules). This effect may be similar to what happens in the dense hydrogen case, where the interaction between a pair of atoms is strongly affected by the presence of other particles.

Chapter 5

Conclusions

In this thesis we have presented large scale molecular dynamics simulations driven by QMC forces. The realistic system under consideration were dense liquid hydrogen and liquid water at room conditions. These are bulk systems which require simulations of many atoms, in order to reproduce correctly the thermodynamic properties, and are computationally very expensive. Nevertheless, we showed that our approach, based on a novel Langevin molecular dynamics driven by QMC noisy forces (see Sect. 2.4) , is computationally feasible even considering large systems. Indeed, within this scheme we are not forced to decrease to noise level of the extensive total energy to the order of $k_B T$. Therefore, we can perform long and equilibrated QMC simulations on system sizes which can usually be investigated only within a DFT approach.

The main physical result of this thesis is reported in Chapter 3 and concerns the transition between the molecular and the atomic hydrogen fluid at high pressure. From the simulations we can distinguish between three, qualitatively different, types of liquids. The fully molecular liquid at low pressure is characterized by a sharp molecular peak of the radial pair distribution function $g(r)$. We observe that $g(r) \approx 0$ around the first minimum, showing that there is a well-defined covalent bond between the two hydrogen atom, with a well-defined equilibrium molecular distance. A smooth crossover toward a mixed molecular-atomic liquid is observed with increasing pressure. In this liquid, the coordination shells are not well separated but the persistence of a broad molecular peak is evident. We observed that this phase is exceptionally stable, as we found a nonzero molecular fraction up to 375 GPa at 2300 K and 600 GPa at 600 K. Finally, the fluid completely dissociates into an atomic liquid phase. The transition between these two latter kinds of fluids appears to be weakly first order, as signalled also by a discontinuous behavior of the density as a function of the pressure (Sect. 3.3). This first order liquid-liquid transition is found to be at higher pressures

with respect to previous DFT-MD simulations. Conversely we collect some clues that the transition between the fully molecular and the partially dissociated liquid may be related to the insulator to metal transition. Thus the metallization at finite temperature occurs within the (partially) molecular phase. At 2300 K the metallization takes place at a pressure $\sim$130 GPa, a value which is very well in agreement with the experiment, while at 600 K the IMT is predicted to be at $\sim$440 GPa (Sect. 3.4). Notice that these outcomes are obtained with the classical nuclei approximation, therefore the quantum proton corrections are expected to shift the two transitions towards smaller pressures, especially at lower temperatures. Nevertheless, even by considering an upper limit of 100 GPa shift to lower pressures, due to proton quantum effects, we predict that experiments should be done at least above 500 GPa to observe the atomic liquid at room temperature. We plan to perform systematically the path integral molecular dynamics calculations introduced in Sect. 3.6 to better quantify this rough estimation. On the electronic side we performed MD simulations with DFT, Variational and Diffusion Monte Carlo in order to evaluate how the different approximations for the electronic correlation affect the physical results on hydrogen and we ascribe the different predictions given by DFT and QMC to the different descriptions of the molecular bond in the bulk (Sect. 3.7). Indeed the well known large errors of DFT in the long distance part of the molecular dissociation curve in the vacuum may not be directly related to the bulk dense liquid as the presence of the surrounding atoms strongly influences the molecular bond. Moreover, all the possible technical sources of error have been discussed. In fact, our results raised a lively discussion among the QMC community, as they substantially differ from previous investigations within the CEIMC technique, namely a technique which employs QMC energetics to drive the sampling and, therefore, is expected to have a similar degree of accuracy compared to our method. Since the all the systematic tests reported in Sect. 3.5.1 support the quality of our variational wavefunction ansatz, we believe that the major sources of error, that we proved to have solved within our approach, come from finite size and lack of equilibration effects (see Sect. 3.5).

Finally, we performed QMC molecular dynamics, at the VMC level of accuracy, on liquid water at ambient conditions (Chapter 4). Also in this case we observed significative differences with respect to previous DFT based simulations, which are very encouraging, in view of their good agreement with experiments.

Together, these results on water and hydrogen confirm the feasibility of the QMC molecular dynamics as an alternative and promising technique to investigate equilibrium properties of bulk materials.

Appendix A

First order covariant Langevin dynamics

A.1 Partition function

In Eq.(2.95), we use a simple relation for recasting the trace in a finite dimensional Hilbert space as an integral of normalized wave functions $|c\rangle = \sum_{i=1}^{D} x_i |i\rangle$, namely:

$$D \int dx^D \delta(\|x\| - 1) \langle c| \exp(-\beta H) |c\rangle = S_D \mathrm{Tr}\, \exp(-\beta H) \tag{A.1}$$

where $S_D = 2\pi^{D/2}/\Gamma(D/2)$ is the area of the $D-$dimensional unit sphere. The above relation can be immediately proved by substituting $|c\rangle = \sum_{i=1}^{D} x_i |i\rangle$ in the LHS of the above equation and noticing that $\int dx^D \delta(\|x\| - 1) x_i x_j = \delta_{i,j} \int dx^D \delta(\|x\| - 1) x_i^2$, namely it is non zero only for $i = j$. Then, we can sum the integrand over the dummy index i and divide by D, and obtain: $\int dx^D \delta(\|x\| - 1) x_i x_j = \frac{\delta_{i,j}}{D} S_D$, that easily proves Eq.(A.1), as previously stated.

The simple relation (A.1) can be also extended in the space α with non trivial metric, by using the invariant measure $d\alpha^p \sqrt{|S|}$, corresponding to the metric tensor S:

$$\frac{\int d\alpha^p \sqrt{|S|} \langle \alpha| \exp(-\beta H_{\mathbf{R}}) |\alpha\rangle}{Z_S} = \mathrm{Tr}\, \exp(-\beta H_{\mathbf{R}}) \tag{A.2}$$

A.2 Monte Carlo sampling of the partition function

In principle the partition function Z_Q can be sampled by almost standard Monte Carlo methods, whenever the metric S and the expectation value of the energy H over the ansatz $|\alpha\rangle$

are known, for instance within the Hartree-Fock theory, namely when $|\alpha\rangle$ represents just a simple Slater determinant.

However in the truly correlated case, namely when the ansatz $|\alpha\rangle$ differs from a Slater determinant, there are extra complications because both the matrix S and $\langle\alpha|H_{\mathbf{R}}|\alpha\rangle$ are known only within statistical accuracy. In this case a possible way to sample the partition function Z_Q and the corresponding thermodynamic quantities is to use the penalty method[22], introduced some years ago, by using a cost function

$$V_P(\alpha, \mathbf{R}) = \langle\alpha|H_{\mathbf{R}}|\alpha\rangle - \frac{1}{2\beta}\ln|S| \tag{A.3}$$

that can be computed statistically with corresponding error bars.

In the following we have chosen a different route, by employing a finite temperature molecular dynamics rather than Monte Carlo sampling, for the same reason described in the above Sections, namely that the QMC noise represents a bias that vanish linearly with the time step.

Our goal is to sample points in the electronic parameter space α distributed according to the probability distribution defined in Eq.(2.102), by using first order derivatives of the cost function. In the standard Cartesian metric it is possible to use a Langevin dynamics for the variables $\{\alpha\}$ and $\{\mathbf{R}\}$, by means of the standard first order equation of motions:

$$\dot{\vec{x}} \;\; = -\partial_{\vec{x}} V + \vec{\eta} \tag{A.4}$$

where $\vec{x}$ is a covariant vector in a finite dimensional euclidean space, whereas $\partial_{\vec{x}} V(x)$ is the derivative (force) of a potential V. By means of this equation it is well known that it is possible to sample the equilibrium distribution $W_{eq}(x) = \exp(-\beta V(x))$ provided we satisfy the fluctuation dissipation theorem, implying that:

$$\langle\eta_i(t)\eta_j(t')\rangle = \delta(t-t')\delta_{i,j}\frac{2}{\beta} \tag{A.5}$$

Now we suppose to change the reference coordinate system by means of a generic transformation of variables $x \to \alpha$ (a p-dimensional non linear mapping as in general relativity). In the following we have to take into account the Jacobian of such mapping, denoted in the following by the matrix L:

$$L_{i,j} = \partial_{x_j}\alpha_i(\vec{x}) \tag{A.6}$$

Then, the Langevin equation in this new reference can be easily obtained:

$$\dot{\vec{\alpha}} = -S^{-1}\frac{\partial V}{\partial\vec{\alpha}} + L\vec{\eta} \tag{A.7}$$

where $S^{-1} = LL^{\dagger}$, and the equation (A.5) that defines the fluctuation dissipation theorem remains unchanged.

The Eq.(A.7) is covariant if we just replace the matrix S with the matrix defining the metric in a generic curved space:

$$ds^2 = S_{i,j} d\alpha_i d\alpha_j \tag{A.8}$$

where in this formalism sums over repeated indices are conventionally assumed. Indeed after the given transformation the above metric tensor transforms as:

$$S \to (L^{\dagger})^{-1} S L^{-1} \tag{A.9}$$

that leaves unchanged the covariant first order Langevin equation (A.7), as expected.

Thus, from the above equation, we obtain the desired result with the matrix L given by any solution of the matrix equation:

$$S^{-1} = LL^{\dagger}.$$

Unfortunately Eq. (A.7) looks a bit complicated when it is discretized in times $t_n = \Delta n$, because the integral of the random noise depends explicitly on the curvature of the non linear space by means of the matrix L, and the resulting integration is not univocally defined, simply because the solution $S^{-1} = LL^{\dagger}$ is not unique. Indeed S^{-1} remains unchanged under the substitution $L \to LU$, where U is an arbitrary unitary matrix. In order to remove this arbitrariness, according to Risken[139], we can work out the integral of the equation of motion in a small time interval of length Δ, by requiring also that the corresponding Markov process:

$$\begin{aligned} \alpha(t_{n+1})^i &= \alpha(t_n)^i - \Delta \left[S^{-1}(t_n) \partial_{\vec{\alpha}} \left(V - \frac{1}{2\beta} \ln \mathrm{Det} S \right)(t_n) \right]^i \\ &+ \frac{1}{2} \sum_k \partial_{\alpha_k} D_{i,k} + y_n^i \\ \langle y_n^i y_n^j \rangle &= D_{i,j} = \frac{2\Delta}{\beta} S^{-1}_{i,j}(t_n) \end{aligned} \tag{A.10}$$

has the correct equilibrium distribution for $\Delta \to 0$:

$$W_{eq}(\alpha) \propto \sqrt{\mathrm{Det} S} \exp(-\beta V(\alpha)) \tag{A.11}$$

In fact it is possible to show that, only with the above definition of the drift term, the associated and univocally defined Fokker-Planck equation for the probability distribution $W(\alpha, t)$ reads for $\Delta \to 0$:

$$\begin{aligned} \partial_t W(\alpha, t) &= \sum_j \partial_j \left\{ \sum_i \frac{1}{\beta} S^{-1}_{j,i} \partial_i W(\alpha, t) \right. \\ &+ \left. W(\alpha, t) \left[S^{-1} \partial_{\vec{\alpha}} \left(V(\alpha) - \frac{1}{2\beta} \ln \mathrm{Det} S \right) \right]^j \right\} \end{aligned} \tag{A.12}$$

which has the equilibrium distribution $W_{eq}(\alpha)$ satisfying:

$$\sum_i \frac{1}{\beta} S^{-1}_{j,i} \partial_i W_{eq}(\alpha) + W_{eq}(\alpha) \sum_i S^{-1}_{j,i} \partial_i \left(V - \frac{1}{2\beta} \ln \mathrm{Det} S \right) = 0 \tag{A.13}$$

Indeed, by multiplying both sides of the equations by $S_{k,j}$ and summing over j, we obtain the standard equation for the equilibrium distribution $\sqrt{|S|}\exp(-\beta V)$.

A.3 Covariant Langevin dynamics for ions and electrons

We want to implement the above formalism in an ab-initio molecular dynamics (MD) at finite temperature dealing with electrons and ions within the same formalism, similarly to what was done in the pioneer work by R. Car and M. Parrinello[3]. In the following we will show how the ionic motion can be quite naturally included in the above scheme. In fact what we obtained before does not hold only for the electronic parameters, but for a generic set of parameters which appear in a variational wave function. The ionic positions **R** can thus be thought as complementary parameters. The inclusion of this kind of parameters in the above formalism is straightforward: if M is the number of atoms, then S becomes a $(p+3M)\times(p+3M)$ block-diagonal matrix. The mixed elements $S_{\{\alpha\},\{\mathbf{R}\}}$ are always zero since total wave functions characterized by different sets of atomic positions are orthogonal. Moreover, since the ionic positions **R** belong to the real space, the corresponding metric is the Cartesian one, and is defined by a diagonal matrix $S(R_l, R_r) = S_N \delta_{l,r}$ among all the ion components. Thus, we can esplicitly write down the complete set of equations for both the atomic and electronic parameters. For the ionic positions we use

$$\begin{aligned} R(t_{n+1})^l &= R(t_n)^l + \Delta_N \, F^l(t_n, \{\alpha(t_n\}) + \chi^l_n \\ \langle \chi^l_n \chi^r_n \rangle &= \frac{2\Delta_N}{\beta} \, \delta_{l,r} \end{aligned} \tag{A.14}$$

with $l, r = 1, \cdots, 3M$ and F^l being the force acting on the l-th ionic cartesian coordinate, while for the electronic variables Eq. (A.10) holds with $i, j = 1, \cdots, p$ and where $-\partial_{\vec{\alpha}} V$ is the force acting on the parameters α, i.e, the gradient of the total electronic energy V evaluated at fixed **R** with respect to these parameters.

Notice also that the time discretization corresponding to the ionic dynamics is defined by the arbitrary constant S_N appearing in the extended metric tensor defined before, namely $\Delta_N = \Delta S_N^{-1}$. It is clear therefore that the relative speed between electron and ion dynamics can be tuned to optimize efficiency, exactly as in Car-Parrinello ab-initio molecular dynamics. We emphasize here that in the limit $\Delta, \Delta_N \to 0$ consistent results are obtained because the equilibrium distribution (A.11) remains unaffected by the choice of S_N.

A.4 Result: Born-Oppenheimer limit

Once we set up the discretized equations (A.10,A.14) we can test the above formalism in a simple but realistic case. We are going to study the H_2 molecule, looking at the temperature behaviour of the total energy E and the bond distance r between the two hydrogen atoms, assumed classical. We start with this simple system because the above quantities can be easily computed, providing therefore useful benchmarks for our technique.

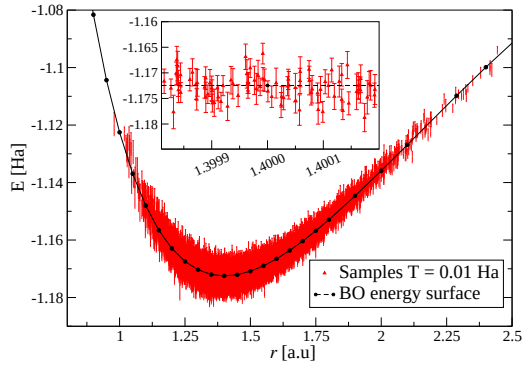
The distribution sampled by means of this covariant Langevin dynamics (CLD) represents an improvement on the gsBO only above a temperature T^*. At $T = 0$ our approximate free energy F_Q coincides with the gsBO one F_{BO}, but as soon as $T > 0$ the F_{BO} becomes better for $T \leq T^*$ (see Ref.[95] for details). If the temperature is much lower than the electronic gap the gsBO approximation should be essentially exact and can be easily obtained from the potential energy surface (PES) $v(r)$ of the H_2 molecule. In the following we are going to show that, in this simple system, we cannot distinguish the correct BO low temperature behavior and the one implied by our approximate technique, clearly indicating that T^* should be almost negligible for this system.

To proceed further we need now to specify what type of correlated variational wave function will be used in all the following calculations, and its dependence on the two electronic positions $\vec{r}_1$ and $\vec{r}_2$. In the singlet state the orbital function $f(\vec{r}_1, \vec{r}_2)$ is symmetric and positive and is parametrized here as a product of two factors $f(\vec{r}_1, \vec{r}_2) = f_0(\vec{r}_1, \vec{r}_2) \times \exp(J(\vec{r}_1, \vec{r}_2))$, where f_0 is taken fixed and allows to satisfy the electron-electron and electron-ion cusp conditions, whereas

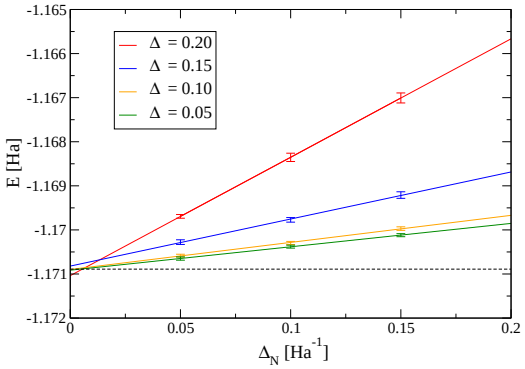
$$J = \sum_{i,j} \lambda_{i,j} \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) \tag{A.15}$$

is cusp free and is expanded systematically in a basis of atomic orbitals centered on each atom containing up to 3s and 1p gaussian functions and a constant one $\phi_0 = 1$. This amounts to $p = 65$ independent variational parameters for the symmetric matrix $\lambda_{i,j}$. The exponent of the gaussians are kept fixed during our simulations. The chosen variational ansatz is particularly useful for evaluating the complicated terms in (A.10), i.e. the *drift-diffusion* ones which depend linearly on the temperature and require the knowledge of the derivative of the matrix S. This is indeed simpler for the parameters $\lambda_{i,j}$ appearing in a linear fashion in the exponential factor J of Eq.(A.15). The first step is thus to construct the PES of the molecule (Fig.A.1(a)). In this way we not only acquire the key information for the numerically exact evaluation of the gsBO observables, but we also check that our choice of the free variational parameters in the wave function allows us to recover the well known PES for this molecule[101].

Canonical averages of an observable $O(r)$ can be obtained by computing numerically the



(a) Potential energy surface



(b) Time step extrapolation

Figure A.1: a. Black line: Total energy E as a function of the bond length r obtained by minimizing the energy of our variational wave function for fixed r; in doing this we act only on those Jastrow parameters $\{\alpha\}$, for which we have been able to compute the covariant forces in Eq.(A.10). Red points: Energy with error bars of configurations sampled in the dynamics (A.10,A.14) with $T = 0.01\ Ha$. The PES is correctly followed during the simulation. In the inset a region around the minimum at $r = 1.40\ a.u.$ is enlarged.
b. Time averages of the total energy E at $T = 0.003\ Ha$ as a function of Δ_N for four values of Δ. All the series converge roughly to the same value which is also the expected one (horizontal dashed line) obtained with eq. (A.16). Thus the second extrapolation $\Delta \to 0$ is not necessary.

one dimensional (conditionally convergent) integral

$$\hat{O} = \frac{\int dr\, r^2\, O(r) \exp\left(-\beta v(r)\right)}{\int dr\, r^2 \exp\left(-\beta v(r)\right)} \tag{A.16}$$

On the other hand we can compute $\hat{O}$ as a time average on the Langevin dynamics (A.10,A.14) for sufficient low T. The extrapolation $\Delta \to 0$ involving the discretized time steps is per-

formed in the order $\Delta_N \to 0, \Delta \to 0$. It is observed (see Fig.A.1(b)) that the Δ_N dependence of the time averages of the quantities is linear for fixed Δ, a property useful in the extrapolation. Finally we show our results for the total energy and the bond distance at various temperatures in the range between $0.001 \div 0.01$ Ha, i.e. from room temperature to ~ 3000 K. The forces acting on the parameters and on the ions, as well as the matrix S are evaluated by a short QMC run at each iteration of the dynamics. We see that our Langevin dynamics gives results in very good agreement with the expected gsBO values. We stress once again that this dynamics does not require an electronic minimization at each ionic move, realizing an impressive gain from the point of view of the computational cost.

Bibliography

[1] P. S. m. d. Laplace, *A Philosophical Essay on Probabilities (1814)*. Wiley, English translation by F.W. Prescott and F.L. Emory (1902). 1

[2] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," *Annalen der Physik*, vol. 389, pp. 457–484, Jan. 1927. 1, 7

[3] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical Review Letters*, vol. 55, pp. 2471–2474, Nov. 1985. 1, 23, 108

[4] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review*, vol. 136, pp. B864–B871, Nov. 1964. 2, 8

[5] K. Burke, "Perspective on density functional theory," *The Journal of Chemical Physics*, vol. 136, p. 150901, Apr. 2012. 2, 9, 10, 87

[6] C.-S. Zha, Z. Liu, and R. J. Hemley, "Synchrotron infrared measurements of dense hydrogen to 360Â GPa," *Physical Review Letters*, vol. 108, p. 146402, Apr. 2012. 2, 52, 69

[7] E. Wigner and H. B. Huntington, "On the possibility of a metallic modification of hydrogen," *The Journal of Chemical Physics*, vol. 3, pp. 764–770, Dec. 1935. 2, 52

[8] A. F. Goncharov, E. Gregoryanz, R. J. Hemley, and H.-k. Mao, "Spectroscopic studies of the vibrational and electronic properties of solid hydrogen to 285 GPa," *Proceedings of the National Academy of Sciences*, vol. 98, pp. 14234–14237, Apr. 2001. PMID: 11717391. 2, 52

[9] P. Loubeyre, F. Occelli, and R. LeToullec, "Optical studies of solid hydrogen to 320âGPa and evidence for black hydrogen," *Nature*, vol. 416, pp. 613–617, Apr. 2002. 2, 52

[10] S. T. Weir, A. C. Mitchell, and W. J. Nellis, "Metallization of fluid molecular hydrogen at 140 GPa (1.4 mbar)," *Physical Review Letters*, vol. 76, pp. 1860–1863, Mar. 1996. 2, 3, 53, 54, 55, 56, 57, 61, 64, 84, 94

[11] V. E. Fortov, R. I. Ilkaev, V. A. Arinin, V. V. Burtzev, V. A. Golubev, I. L. Iosilevskiy, V. V. Khrustalev, A. L. Mikhailov, M. A. Mochalov, V. Y. Ternovoi, and M. V. Zhernokletov, "Phase transition in a strongly nonideal deuterium plasma generated by quasi-isentropical compression at megabar pressures," *Physical Review Letters*, vol. 99, p. 185001, Oct. 2007. 3, 53, 54

[12] V. Dzyabura, M. Zaghoo, and I. F. Silvera, "Evidence of a liquidâliquid phase transition in hot dense hydrogen," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 8040–8044, May 2013. PMID: 23630287. 3, 53, 54, 55, 56

[13] S. Scandolo, "Liquid-liquid phase transition in compressed hydrogen from first-principles simulations," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 3051–3053, Mar. 2003. PMID: 12626753. 3, 54, 57, 61, 94

[14] S. A. Bonev, E. Schwegler, T. Ogitsu, and G. Galli, "A quantum fluid of metallic hydrogen suggested by first-principles calculations," *Nature*, vol. 431, no. 7009, pp. 669–672, 2004. 3, 54

[15] I. Tamblyn and S. A. Bonev, "Structure and phase boundaries of compressed liquid hydrogen," *Physical Review Letters*, vol. 104, p. 065702, Feb. 2010. 3, 23, 24, 54, 83, 84

[16] M. A. Morales, C. Pierleoni, E. Schwegler, and D. M. Ceperley, "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*, vol. 107, no. 29, pp. 12799–12803, 2010. 3, 54, 56, 57, 61, 94

[17] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, "Nuclear quantum effects and nonlocal exchange-correlation functionals applied to liquid hydrogen at high pressure," *Physical Review Letters*, vol. 110, p. 065702, Feb. 2013. 3, 23, 55, 56, 61, 68, 69, 75, 84, 86, 89, 94

[18] S. Azadi and W. M. C. Foulkes, "Fate of density functional theory in the study of high-pressure solid hydrogen," *Physical Review B*, vol. 88, p. 014115, July 2013. 3, 53, 87

[19] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum monte carlo simulations of solids," *Reviews of Modern Physics*, vol. 73, pp. 33–83, Jan. 2001. 3, 11, 19, 20, 22

[20] C. Pierleoni, D. M. Ceperley, and M. Holzmann, "Coupled electron-ion monte carlo calculations of dense metallic hydrogen," *Physical Review Letters*, vol. 93, p. 146402, Sept. 2004. 3, 25

[21] K. T. Delaney, C. Pierleoni, and D. M. Ceperley, "Quantum monte carlo simulation of the high-pressure molecular-atomic crossover in fluid hydrogen," *Phys. Rev. Lett.*, vol. 97, p. 235702, Dec 2006. 3, 54

[22] D. M. Ceperley and M. Dewing, "The penalty method for random walks with uncertain energies," *The Journal of Chemical Physics*, vol. 110, pp. 9812–9820, May 1999. 3, 25, 106

[23] R. Assaraf and M. Caffarel, "Computing forces with quantum monte carlo," *The Journal of Chemical Physics*, vol. 113, pp. 4028–4034, Sept. 2000. 3, 17

[24] R. Assaraf and M. Caffarel, "Zero-variance zero-bias principle for observables in quantum monte carlo: Application to forces," *The Journal of Chemical Physics*, vol. 119, pp. 10536–10552, Nov. 2003. 3, 4, 16, 17

[25] C. Attaccalite and S. Sorella, "Stable liquid hydrogen at high pressure by a novel abÂ initio molecular-dynamics calculation," *Physical Review Letters*, vol. 100, p. 114501, Mar. 2008. 3, 4, 17, 18, 26, 94

[26] C. Filippi and C. J. Umrigar, "Correlated sampling in quantum monte carlo: A route to forces," *Phys. Rev. B*, vol. 61, pp. R16291–R16294, Jun 2000. 3, 4

[27] S. Sorella and L. Capriotti, "Algorithmic differentiation and the calculation of forces by quantum monte carlo," *The Journal of Chemical Physics*, vol. 133, p. 234111, Dec. 2010. 3, 4, 18

[28] S. Saccani, C. Filippi, and S. Moroni, "Minimum energy pathways via quantum monte carlo," *The Journal of Chemical Physics*, vol. 138, no. 8, pp. –, 2013. 4

[29] G. Mazzola, S. Yunoki, and S. Sorella, "Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation," *Nature Communications*, vol. 5:3487, Mar. 2014. 4, 36, 58

[30] T. D. Kühne, M. Krack, and M. Parrinello, "Static and dynamical properties of liquid water from first principles by a novel carâparrinello-like approach," *Journal of Chemical Theory and Computation*, vol. 5, no. 2, pp. 235–241, 2009. 4, 24, 98, 99

[31] T. D. Kühne and R. Z. Khaliullin, "Electronic signature of the instantaneous asymmetry in the first coordination shell of liquid water," *Nature communications*, vol. 4, p. 1450, 2013. 4, 98

[32] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water," *The Journal of Chemical Physics*, vol. 120, no. 1, pp. 300–311, 2004. 4, 98

[33] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water," *The Journal of Chemical Physics*, vol. 122, no. 1, pp. –, 2005. 4, 98

[34] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore, "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range," *J. Chem. Phys.*, vol. 138, no. 7, p. 074506, 2013. 4, 98, 100, 101

[35] A. K. Soper, "The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure?," *ISRN Phys. Chem.*, vol. 2013, no. 11, pp. 1–67, 2013. 4, 98, 100, 101

[36] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, pp. A1133–A1138, Nov. 1965. 8

[37] A. J. Cohen, P. Mori-Sanchez, and W. Yang, "Challenges for density functional theory," *Chemical Reviews*, vol. 112, pp. 289–320, Jan. 2012. 9, 49, 87

[38] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, Oct. 1996. 9

[39] A. D. Becke, "Density-functional thermochemistry. III. the role of exact exchange," *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, Apr. 1993. 9

[40] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened coulomb potential," *The Journal of Chemical Physics*, vol. 118, pp. 8207–8215, May 2003. 9, 89

[41] E. Lipparini, *Modern Many-Particle Physics: Atomic Gases, Nanostructures and Quantum Liquids*. Singapore; Hackensack, NJ; London: World Scientific Pub Co Inc, 2 edizione ed., 2008. 10, 19

[42] G. H. Booth, A. J. W. Thom, and A. Alavi, "Fermion monte carlo without fixed nodes: A game of life, death, and annihilation in slater determinant space," *The Journal of Chemical Physics*, vol. 131, p. 054106, Aug. 2009. 11

[43] D. Cleland, G. H. Booth, and A. Alavi, "Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum monte carlo," *The Journal of Chemical Physics*, vol. 132, p. 041103, Jan. 2010. 11

[44] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, "Towards an exact description of electronic wavefunctions in real solids," *Nature*, vol. 493, pp. 365–370, Jan. 2013. 11

[45] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092, June 1953. 11, 24, 25

[46] V. Ambegaokar and M. Troyer, "Estimating errors reliably in monte carlo simulations of the ehrenfest model," *American Journal of Physics*, vol. 78, no. 2, 2010. 12

[47] M. Casula and S. Sorella, "Geminal wave functions with jastrow correlation: A first application to atoms," *The Journal of Chemical Physics*, vol. 119, pp. 6500–6511, Oct. 2003. 13

[48] M. Casula, C. Attaccalite, and S. Sorella, "Correlated geminal wave function for molecules: An efficient resonating valence bond approach," *The Journal of Chemical Physics*, vol. 121, pp. 7110–7126, Oct. 2004. 13

[49] M. Marchi, S. Azadi, M. Casula, and S. Sorella, "Resonating valence bond wave function with molecular orbitals: Application to first-row molecules," *The Journal of Chemical Physics*, vol. 131, no. 15, pp. –, 2009. 13, 14

[50] A. Zen, Y. Luo, S. Sorella, and L. Guidoni, "Molecular properties by quantum monte carlo: An investigation on the role of the wave function ansatz and the basis set in the water molecule," *Journal of Chemical Theory and Computation*, vol. 9, pp. 4332–4350, Oct. 2013. 13, 18

[51] S. Sorella, "Wave function optimization in the variational monte carlo method," *Physical Review B*, vol. 71, p. 241103, June 2005. 16

[52] S. Sorella, M. Casula, and D. Rocca, "Weak binding between two aromatic rings: Feeling the van der waals attraction by quantum monte carlo methods," *The Journal of Chemical Physics*, vol. 127, p. 014105, July 2007. 16

[53] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, "Alleviation of the fermion-sign problem by optimization of many-body wave functions," *Physical Review Letters*, vol. 98, p. 110201, Mar. 2007. 16, 46

[54] J. Toulouse and C. J. Umrigar, "Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules," *The Journal of Chemical Physics*, vol. 128, p. 174101, May 2008. 16

[55] B. L. Hammond, W. A. J. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*. World Scientific, 1994. 17

[56] F. R. Krajewski and M. Parrinello, "Linear scaling for quasi-one-dimensional systems," *Physical Review B*, vol. 74, p. 125107, Sept. 2006. 18, 26

[57] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, "Efficient and accurate car-parrinello-like approach to born-oppenheimer molecular dynamics," *Physical Review Letters*, vol. 98, p. 066401, Feb. 2007. 18, 26

[58] C. J. Umrigar, "Two aspects of quantum monte carlo: Determination of accurate wave-functions and determination of potential energy surfaces of molecules," *International Journal of Quantum Chemistry*, vol. 36, pp. 217–230, Apr. 1989. 18

[59] D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," *Physical Review Letters*, vol. 45, pp. 566–569, Aug. 1980. 19

[60] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. J. Lester, "Fixed-node quantum monte carlo for molecules," *The Journal of Chemical Physics*, vol. 77, pp. 5593–5603, Dec. 1982. 19, 22

[61] M. Troyer and U.-J. Wiese, "Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations," *Physical Review Letters*, vol. 94, p. 170201, May 2005. 21

[62] D. M. Ceperley, "Fermion nodes," *Journal of Statistical Physics*, vol. 63, pp. 1237–1267, June 1991. 22

[63] M. Casula, C. Filippi, and S. Sorella, "Diffusion monteÂ carlo method with lattice regularization," *Physical Review Letters*, vol. 95, p. 100201, Sept. 2005. 22, 92

[64] M. Casula, S. Moroni, S. Sorella, and C. Filippi, "Size-consistent variational approaches to nonlocal pseudopotentials: Standard and lattice regularized diffusion monte carlo methods revisited," *The Journal of Chemical Physics*, vol. 132, p. 154113, Apr. 2010. 22

[65] R. M. Wentzcovitch and J. Martins, "First principles molecular dynamics of li: Test of a new algorithm," *Solid State Communications*, vol. 78, pp. 831–834, June 1991. 23

[66] J. Vorberger, I. Tamblyn, B. Militzer, and S. A. Bonev, "Hydrogen-helium mixtures in the interiors of giant planets," *Physical Review B*, vol. 75, p. 024206, Jan. 2007. 23, 24, 54

[67] K. B. I. f. P. J. G.-U. Mainz, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford University Press, July 1995. 24

[68] M. Sprik, J. Hutter, and M. Parrinello, "Abinitio molecular dynamics simulation of liquid water: Comparison of three gradientâcorrected density functionals," *The Journal of Chemical Physics*, vol. 105, pp. 1142–1152, July 1996. 24

[69] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, "Bulk liquid water at ambient temperature and pressure from MP2 theory," *The Journal of Physical Chemistry Letters*, vol. 4, pp. 3753–3759, Nov. 2013. 24

[70] K. T. Delaney, C. Pierleoni, and D. M. Ceperley, "Quantum monteÂ carlo simulation of the high-pressure molecular-atomic crossover in fluid hydrogen," *Physical Review Letters*, vol. 97, p. 235702, Dec. 2006. 25

[71] M. A. Morales, C. Pierleoni, E. Schwegler, and D. M. Ceperley, "Evidence for a first-order liquid-liquid transition in high-pressure hydrogen from ab initio simulations," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 12799–12803, July 2010. 25, 69

[72] E. Liberatore, M. A. Morales, D. M. Ceperley, and C. Pierleoni, "Free energy methods in coupled electron ion monte carlo," *Molecular Physics*, vol. 109, no. 23-24, pp. 3029–3036, 2011. 25, 69

[73] C. Lin, F. H. Zong, and D. M. Ceperley, "Twist-averaged boundary conditions in continuum quantum monte carlo algorithms," *Physical Review E*, vol. 64, p. 016702, June 2001. 25

[74] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Clarendon Press, 1987. 26

[75] D. L. Ensign and V. S. Pande, "The fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations," *Biophysical Journal*, vol. 96, pp. L53–L55, Apr. 2009. 26

[76] P. Faccioli, M. Sega, F. Pederiva, and H. Orland, "Dominant pathways in protein folding," *Physical Review Letters*, vol. 97, p. 108101, Sept. 2006. 26

[77] M. Sega, P. Faccioli, F. Pederiva, G. Garberoglio, and H. Orland, "Quantitative protein dynamics from dominant folding pathways," *Physical Review Letters*, vol. 99, p. 118102, Sept. 2007. 26

[78] G. Mazzola, S. a. Beccara, P. Faccioli, and H. Orland, "Fluctuations in the ensemble of reaction pathways," *The Journal of Chemical Physics*, vol. 134, p. 164109, Apr. 2011. 26

[79] S. a. Beccara, T. Skrbic, R. Covino, and P. Faccioli, "Dominant folding pathways of a WW domain," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 2330–2335, Feb. 2012. 26

[80] G. Bussi and M. Parrinello, "Accurate sampling using langevin dynamics," *Physical Review E*, vol. 75, p. 056707, May 2007. 26

[81] M. Ceriotti, G. Bussi, and M. Parrinello, "Langevin equation with colored noise for constant-temperature molecular dynamics simulations," *Physical Review Letters*, vol. 102, p. 020601, Jan. 2009. 27, 36, 84

[82] M. Ceriotti, G. Bussi, and M. Parrinello, "Nuclear quantum effects in solids using a colored-noise thermostat," *Physical Review Letters*, vol. 103, p. 030603, July 2009. 27, 84

[83] M. Ceriotti, G. Bussi, and M. Parrinello, "Colored-noise thermostats à la carte," *Journal of Chemical Theory and Computation*, vol. 6, pp. 1170–1180, Apr. 2010. 27, 84

[84] M. E. Tuckerman, G. G. Martyna, and B. J. Berne, "Molecular dynamics algorithm for condensed systems with multiple time scales," *J. Chem. Phys.*, vol. 93, no. 1287, p. 1287, 1990. 29

[85] P. J. Hoogerbrugge and J. M. V. A. Koelman, "Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics," *EPL (Europhysics Letters)*, vol. 19, p. 155, June 1992. 30

[86] E. a. J. F. Peters, "Elimination of time step effects in DPD," *EPL (Europhysics Letters)*, vol. 66, p. 311, May 2004. 30

[87] L. Pastewka, D. Kauzlarić, A. Greiner, and J. G. Korvink, "Thermostat with a local heat-bath coupling for exact energy conservation in dissipative particle dynamics," *Physical Review E*, vol. 73, p. 037701, Mar. 2006. 30

[88] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. d. Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos,

N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," *Journal of Physics: Condensed Matter*, vol. 21, p. 395502, Sept. 2009. 30, 49

[89] F. Tassone, F. Mauri, and R. Car, "Acceleration schemes for ab initio molecular-dynamics simulations and electronic-structure calculations," *Phys. Rev. B*, vol. 50, pp. 10561–10573, Oct 1994. 36

[90] Y. Luo, A. Zen, and S. Sorella, "Ab-initio molecular dynamics with noisy forces, an affordable quantum monte carlo scheme for accurate vibrational properties," *in preparation.*, 2014. 40

[91] E. Liberatore and C. Pierleoni *private communications*, 2011. 42

[92] G.A. Worth, L.S. Cederbaum, "BEYOND BORN-OPPENHEIMER: Molecular Dynamics Through a Conical Intersection," *Annu. Rev. Phys. Chem*, vol. 55, p. 127, 2004. 45

[93] C. A. Mead and D. G. Truhlar, "On the determination of born–oppenheimer nuclear motion wave functions including complications due to conical intersections and identical nuclei," *The Journal of Chemical Physics*, vol. 70, no. 5, pp. 2284–2296, 1979. 45

[94] S. Sorella, M. Casula, and D. Rocca, "Weak binding betweem two aromatic rings: feeling the van det waals attraction by quantum monte carlo methods," *J. Chem Phys.*, vol. 127, p. 14105, 2007. 46

[95] G. Mazzola, A. Zen, and S. Sorella, "Finite-temperature electronic simulations without the born-oppenheimer constraint," *The Journal of Chemical Physics*, vol. 137, p. 134112, Oct. 2012. 46, 47, 109

[96] I. Langmuir, "The dissociation of hydrogen into atoms.," *J. Am. Chem. Soc.*, vol. 34, pp. 860–877, 1912. 47

[97] M. Grumbach, D. Hohl, R. Martin, and R. Car, "AB-INITIO MOLECULAR-DYNAMICS WITH A FINITE-TEMPERATURE DENSITY-FUNCTIONAL," *Journal of Physics Condensed Matter*, vol. 6, no. 10, pp. 1999–2014, 1994. 48

[98] N. D. Mermin, "Thermal properties of the inhomogeneous electron gas," *Phys. Rev.*, vol. 137, pp. A1441–A1443, Mar 1965. 48

[99] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after kramers," *Rev. Mod. Phys.*, vol. 62, pp. 251–341, Apr 1990. 47, 49

[100] L. Butler, "CHEMICAL REACTION DYNAMICS BEYOND THE BORN-OPPENHEIMER APPROXIMATION ," *Annu. Rev. Phys. Chem*, vol. 49, p. 125, 1998. 47

[101] G. Corongiu and E. Clementi, "Energy and density analyses of the h2 molecule from the united atom to dissociation: The 1g+ states," *The Journal of Chemical Physics*, vol. 131, p. 034301, July 2009. 47, 89, 109

[102] *we used the pseudopotential* H.pz-kjpaf.UPF *from* www.quantum-espresso.org. 49

[103] J. Hubbard, "Electron correlations in narrow energy bands," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 276, no. 1365, pp. 238–257, 1963. 52

[104] T. Guillot, "THE INTERIORS OF GIANT PLANETS: models and outstanding questions," *Annual Review of Earth and Planetary Sciences*, vol. 33, no. 1, pp. 493–530, 2005. 52

[105] M. I. Eremets and I. A. Troyan, "Conductive dense hydrogen," *Nature Materials*, vol. 10, pp. 927–931, Dec. 2011. 52

[106] W. J. Nellis, A. L. Ruoff, and I. F. Silvera, "Has metallic hydrogen been made in a diamond anvil cell?," *arXiv:1201.0407 [cond-mat]*, Jan. 2012. 52

[107] C.-s. Zha, Z. Liu, M. Ahart, R. Boehler, and R. J. Hemley, "High-pressure measurements of hydrogen phase iv using synchrotron infrared spectroscopy," *Phys. Rev. Lett.*, vol. 110, p. 217402, May 2013. 52

[108] R. T. Howie, C. L. Guillaume, T. Scheler, A. F. Goncharov, and E. Gregoryanz, "Mixed molecular and atomic phase of dense hydrogen," *Phys. Rev. Lett.*, vol. 108, p. 125501, Mar 2012. 52

[109] C. J. Pickard and R. J. Needs, "Structure of phase III of solid hydrogen," *Nature Physics*, vol. 3, pp. 473–476, July 2007. 53

[110] S. Azadi, W. M. C. Foulkes, and T. D. Kühne, "Quantum monte carlo study of high pressure solid molecular hydrogen," *New Journal of Physics*, vol. 15, p. 113005, Nov. 2013. 53, 87

[111] N. C. Holmes, M. Ross, and W. J. Nellis, "Temperature measurements and dissociation of shock-compressed liquid deuterium and hydrogen," *Phys. Rev. B*, vol. 52, pp. 15835–15845, Dec 1995. 53

[112] S. B. K. F. V. Grigorev *JETP Lett.*, vol. 16, p. 201, 1972. 54

[113] J. M. McMahon, M. A. Morales, C. Pierleoni, and D. M. Ceperley, "The properties of hydrogen and helium under extreme conditions," *Reviews of Modern Physics*, vol. 84, pp. 1607–1653, Nov. 2012. 54

[114] S. A. Bonev, B. Militzer, and G. Galli, "Ab initio simulations of dense liquid deuterium: Comparison with gas-gun shock-wave experiments," *Physical Review B*, vol. 69, p. 014101, Jan. 2004. 54

[115] S. Baroni and S. Moroni, "Reptation quantum monte carlo: A method for unbiased ground-state averages and imaginary-time correlations," *Phys. Rev. Lett.*, vol. 82, pp. 4745–4748, Jun 1999. 54

[116] W. J. Nellis, S. T. Weir, and A. C. Mitchell, "Minimum metallic conductivity of fluid hydrogen at 140 gpa (1.4 mbar)," *Phys. Rev. B*, vol. 59, pp. 3434–3449, Feb 1999. 55, 57

[117] W. Nellis, A. Louis, and N. Ashcroft, "Metallization of fluid hydrogen," *arXiv preprint cond-mat/9708144*, 1997. 55, 57

[118] W. Nellis, A. Louis, and N. Ashcroft, "Metallization of fluid hydrogen," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 356, no. 1735, pp. 119–138, 1998. 55, 57

[119] N. W. Ashcroft, "Metallic hydrogen: A high-temperature superconductor?," *Phys. Rev. Lett.*, vol. 21, pp. 1748–1749, Dec 1968. 57

[120] N. F. Mott, "The basis of the electron theory of metals, with special reference to the transition metals," *Proceedings of the Physical Society. Section A*, vol. 62, no. 7, p. 416, 1949. 57

[121] P. W. Anderson, "Absence of diffusion in certain random lattices," *Phys. Rev.*, vol. 109, pp. 1492–1505, Mar 1958. 57

[122] W. Kohn, "Theory of the insulating state," *Phys. Rev.*, vol. 133, pp. A171–A181, Jan 1964. 62

[123] E. Koch and S. Goedecker, "Locality properties and wannier functions for interacting systems," *Solid State Communications*, vol. 119, no. 2, pp. 105 – 109, 2001. 62

[124] R. Resta, "The insulating state of matter: a geometrical theory," *The European Physical Journal B*, vol. 79, no. 2, pp. 121–137, 2011. 62

[125] M. Capello, F. Becca, S. Yunoki, and S. Sorella, "Unconventional metal-insulator transition in two dimensions," *Phys. Rev. B*, vol. 73, p. 245116, Jun 2006. 63

[126] T. J. Lenosky, J. D. Kress, L. A. Collins, and I. Kwon, "Molecular-dynamics modeling of shock-compressed liquid hydrogen," *Phys. Rev. B*, vol. 55, pp. R11907–R11910, May 1997. 68

[127] M. P. Desjarlais, "Density-functional calculations of the liquid deuterium hugoniot, reshock, and reverberation timing," *Physical Review B*, vol. 68, p. 064204, Aug. 2003. 83

[128] D. M. Ceperley, "Path integrals in the theory of condensed helium," *Rev. Mod. Phys.*, vol. 67, pp. 279–355, Apr 1995. 84, 85

[129] D. Marx and M. Parrinello, "Ab initio path integral molecular dynamics: Basic ideas," *The Journal of Chemical Physics*, vol. 104, no. 11, 1996. 85

[130] M. Fuchs, Y.-M. Niquet, X. Gonze, and K. Burke, "Describing static correlation in bond dissociation by KohnâSham density functional theory," *The Journal of Chemical Physics*, vol. 122, p. 094116, Mar. 2005. 87

[131] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, "Towards a predictive first-principles description of solid molecular hydrogen with density functional theory," *Physical Review B*, vol. 87, p. 184107, May 2013. 87

[132] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, "ãâabinitioâã liquid water," *The Journal of Chemical Physics*, vol. 99, no. 11, pp. 9080–9089, 1993. 98

[133] S. Yoo, X. C. Zeng, and S. S. Xantheas, "On the phase diagram of water with density functional theory potentials: The melting temperature of ice ih with the perdew–burke–ernzerhof and becke–lee–yang–parr functionals," *The Journal of Chemical Physics*, vol. 130, no. 22, pp. –, 2009. 98

[134] A. K. Soper, "The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa," *Chem. Phys.*, vol. 258, pp. 121–137, Aug. 2000. 98, 100, 101

[135] A. Zen, G. Mazzola, Y. Luo, L. Guidoni, and S. Sorella, "Ab-initio molecular dynamics simulation of liquid water by Quantum Monte Carlo," *in preparation.* 99

[136] D. Alfè, A. P. Bartók, G. Csányi, and M. J. Gillan, "Communication: Energy benchmarking with quantum monte carlo for water nano-droplets and bulk liquid water," *The Journal of Chemical Physics*, vol. 138, no. 22, pp. –, 2013. 101

[137] M. A. Morales, J. R. Gergely, J. McMinis, J. M. McMahon, J. Kim, and D. M. Ceperley, "Quantum monte carlo benchmark of exchange-correlation functionals for bulk water," *Journal of Chemical Theory and Computation*, vol. 10, no. 6, pp. 2355–2362, 2014. 101

[138] D. M. Ceperley *private communication*, 2014. 101

[139] H. Risken, *The Fokker-Planck equation*. methods of solution and applications, Springer Verlag, 1996. 107

RINGRAZIAMENTI

Un rigraziamento veramente sentito al mio *supervisor* Sandro Sorella. Non solamente per le *useful discussions* ma sopratutto per la costanza e l'entusiasmo con cui abbiamo lavorato insieme e con cui mi ha sempre aiutato, dai piu' banali problemi di compilazione, in su.

Acknowledgments agli amici Nicola e Matteo per la piacevole convivenza nello stabile signorile sito in Via Bonomea, 265, interno 414, ai quali aggiungo Cristiano, Ivan e Cristina. Grazie anche Ye Luo, per la sua disponibilita' e ad Andrea Zen per aver lavorato insieme nonostante la distanza.
Un grazie anche ai miei amici a Trento (denominazione generalizzata), Luca Matteo, Davide e il Biondo, oltre che a Martina, Davide e Nonno Nicola.

Grazie ad Aurora. E grazie alla catena di eventi che me l'hanno fatta incontrare.

Grazie ai miei genitori. Se ora sono qui a scrivere questi ringraziamenti, e' per merito loro.