New QMC approaches for the simulation of electronic systems: a first application to aromatic molecules and transition metal compounds

Thesis submitted for the degree of

Philosophiæ Doctor

Candidate:
Michele Casula

Supervisor:
Prof. Sandro Sorella

October, 24th 2005

Contents

Introduction

The aim of this thesis is to develop a suitable computational framework that allows the study of realistic strongly correlated compounds. In particular, we would like to focus our attention on iron, which is one of the most important transition metal elements, and one of the most challenging for computational physics. Indeed, from the biochemical point of view, iron compounds, like iron-sulfur clusters, are the basic constituents of various proteins, where they play the role of catalytic centers of many enzymatic reactions. Therefore they are present in all living organisms, and the iron-sulfur proteins are among the oldest known biological catalysts. They are responsible for electron transfer reactions, like nitrogenase and hydrogenase, and thus they are essential compounds in anaerobic bacteria which metabolize hydrogen with high efficiency.

A detailed understanding of the physics underlying these processes is still elusive, although a great improvement has been done by crystallographic studies and the production of synthetic analogues. They clarified the key role of iron active sites, by reproducing in simple structures the same properties verified in much bigger proteins. A recent study [1, 2] proposed to apply the iron based synthetic compounds to the construction of bioinspired hydrogen fuel cells, where an iron-sulfur active site can mimic the hydrogen-producing enzymes to supply electrical power, in place of platinum, which is the currently used electrocatalyst but is expensive and limited in availability.

Numerical studies of the iron compounds are necessary to explain their electronic structure, their geometry and most importantly the mechanisms behind the chemical reactions. Unfortunately density functional theory (DFT), which is a very reliable and convenient method for transition metals having either an almost empty (like Sc, Ti) or an almost full (like Cu, Ag, Zn) d-shell, is not accurate in the case of intermediate elements (like Cr, Mo, Fe, Ni), where different states are

competitive and it is difficult to resolve the right quantum numbers of the ground state. Indeed, the well-known tendency of DFT methods to favor d^n over $d^{n-1}s^1$ configurations might lead to incorrect results. Moreover the iron clusters display peculiar spin properties, since the high spin sites are usually antiferromagnetically coupled to yield a singlet ground state, and an energy spectrum which is a Heisenberg spin ladder [3]. The direct description of the individual pure spin state is difficult within the DFT framework, since the single-determinant Kohn-Sham orbital representation is not able to correctly reproduce the total spin symmetry. This is another drawback of DFT calculations, which has been partially overcome using local spin density functionals, that however can give only an approximate estimate of the Heisenberg coupling. Alternative high level *ab initio* techniques, such as multiconfigurational active space (CASSCF), multireference configuration interaction (MRCI), and coupled cluster (CC) methods, could be used to calibrate DFT results on small clusters, but unfortunately their computational cost scales as N^5 or more, where N is the number of electrons, making calculations for large systems unfeasible. In principle the number of configurations to be included in the wave function grows exponentially with the number of atoms involved in the system, and only proper approximations and truncations can change this scaling from an exponential to a polynomial behaviour.

An alternative to the previous methods is represented by Quantum Monte Carlo (QMC) techniques. Indeed, they can deal with highly correlated wave functions, without any restriction on the functional form used in the variational ansatz. In this way, it is possible to include in a compact and efficient form the correlation effects and evaluate the quantum expectation values of any operators by a stochastic sampling of the configuration space. A related drawback of this technique is the presence of the statistical error, which must be taken under the desired accuracy with a sufficient long sampling of the correlated wave function. As well known, the more accurate is the wave function, the smaller is the error. Therefore it is extremely important to have a good variational wave function, since in this case both the statistical and the systematic errors are reduced. Usually, in QMC calculations a Jastrow-Slater wave function is employed, where the Jastrow factor explicitly correlates the electron motion, while the Slater part includes the single particle orbitals with the proper overall antisymmetry. The Slater part can be extended to a multireference correlated wave function coming from a post Hartree-Fock cal-

culation. Unfortunately, within such schemes the computational cost scales very inefficiently with the system size, since the number of configurations needed in the wave function grows exponentially with the complexity of the system, as previously discussed. From this point of view, a lot of effort has been done in this thesis to develop an accurate and efficient variational ansatz. The final task it to take into account both dynamical correlations and near degeneracy effects, and include all possible bonds among the nuclear sites of a compounds through the computation of only a single Slater determinant, which yields a global scaling of $N^3 - N^4$ for variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC) calculations. We were inspired by the resonating valence bond (RVB) approach[4], widely used in strongly correlated lattice models to describe the exponentially growing singlet pairs in a spin liquid system [5] by means of a simple projected-BCS wave function. Our effort allowed us to find the quantum chemical analogue of the RVB ansatz. It has been called Jastrow correlated antisymmetrized geminal product (JAGP) and we applied it to to a quite wide set of molecules.

QMC methods include also projection schemes, like the Green function Monte Carlo (GFMC) and diffusion Monte Carlo (DMC) algorithms, the former used for lattice Hamiltonians, the latter for realistic systems. No matter what the initial wave function is, they are able to filter out the high energy components of the starting guess and to project it to the ground state. Therefore these projections techniques are extremely powerful, and they can further improve an already optimized trial wave function. However, the antisymmetric character of the fermionic wave function is a limitation for the precision of the methods, since the fixed node (FN) approximation must be applied. Nevertheless, DMC calculations usually provide results within the chemical accuracy (0.05 eV). Another approximation, necessary to implement the DMC scheme in a feasible way, is the locality approximation (LA), which is used in the presence of non local potentials, and which provides non variational results, whose accuracy is strongly dependent on the quality of the initial trial wave function. Pseudopotentials are non local potentials, and since they are unavoidable to reduce the computational time, the LA has been widely employed in DMC electronic structure calculations of realistic systems. In the second part of this thesis, we provide a solution to overcome the LA, by developing a novel QMC framework, the lattice regularized diffusion Monte Carlo (LRDMC), able to treat non local potentials in a variational way, without

resorting to the locality approximation.

The last part of this work gathers the variational JAGP ansatz and the novel LRDMC method, to study the iron dimer. Despite its simplicity, this molecule is a challenge for traditional numerical methods, as the ground state symmetry of the anion and neutral dimer is still controversial. Moreover, it is the first test for our approach in the field of transition metal compounds.

This thesis is organized as follows:

- In Chapter 1 we introduce the Monte Carlo methods;
- In Chapter 2 we present the JAGP ansatz, with its application to some light atoms, simple dimers and the benzene molecule;
- In Chapter 3 we introduce the new LRDMC scheme;
- In Chapter 4 we will use both the JAGP ansatz and the LRDMC method to study the iron dimer.

Chapter 1

Quantum Monte Carlo methods

1.1 Introduction

The term "quantum Monte Carlo" (QMC) covers several different techniques based on random sampling. The simplest of these, variational Monte Carlo (VMC), uses a stochastic integration method to evaluate expectation values for a chosen trial wave function. For a sufficiently high number of variables in the integrand, QMC methods are much more efficient than a deterministic integration such as Simpson's rule, and the many-body systems are certainly the case. The major advantage of this framework is the possibility to freely choose the analytic form of the trial wave function, which may contain highly sophisticated terms, in such a way that electron correlation is explicitly taken into account in the variational ansatz. This is an important feature valid not only for the VMC method, but in general for all the QMC techniques, which are therefore extremely useful to study physical cases where the electron correlation plays a crucial role. In particular they have been widely applied not only to strongly interacting lattice systems, but also to realistic continuous models, such as electron gas [6], quantum dots[7], nanoclusters [8], solid hydrogen [9] and liquid helium, which was the first attempted application of the VMC method [10].

In the last few years, QMC methods have also been used in quantum chemistry, for their favorable scaling with the number of particles N, as an alternative to post Hartree-Fock theories. Indeed the computational effort scales as $N^3 - N^4$, depending on the QMC method. This property is appealing since in principle it

could allow to treat complex molecular systems, not accessible by the full configuration interaction (CI) or the coupled cluster (CC) approaches, whose scaling is worse already for the case where only single and double excitations are taken into account. On the other hand, since the chemical bond can be very weak, and the correlation can contribute to a large fraction of the binding energy, a highly precise method is required in the most difficult cases. In general a good theory for the proper description of energetics must fulfill the so called "chemical accuracy", defined as 1 kcal per mole ($\approx 0.04\ eV$ per molecule). Usually the post Hartree-Fock methods are able to reach this accuracy, at least for small compounds. Also QMC techniques can afford such a precision, in particular the diffusion Monte Carlo (DMC) method, which filters out the high energy components of a given trial wavefunction, and projects the starting guess to a state much closer to the real ground state of the system.

The analogous of the DMC approach for lattice systems is the lattice Green function Monte Carlo (GFMC) method, which is based upon a lattice Hamiltonian and uses the discrete configuration space to find out the lowest state of the system. It is an extremely important tool to analyze the zero temperature properties of strongly correlated lattice Hamiltonians, and in this thesis will be exploited in connection with continuous systems. Indeed in Chapter 3 the lattice regularized diffusion Monte Carlo (LRDMC) will be presented, a novel projection method that uses the machinery of the GFMC algorithm to solve an Hamiltonian, which is the lattice regularization of the continuous one. This novel method is efficient and allows to treat non local potentials in a variational way, contrary to the usual DMC approach, as we will see later.

In principle both the accuracy and the favorable size scaling would allow VMC and DMC methods to successfully study a wide range of chemical systems. However, it has to be noticed that the random sampling behind any quantum Monte Carlo (QMC) framework yields results which are affected by a statistical error, whose scaling with the computational time is characteristic of any stochastic technique:

$$\mathrm{statistical\ \ error} \sim \frac{1}{\sqrt{\mathrm{computational\ \ time}}}. \tag{1.1}$$

Therefore the QMC methods are sometimes very time consuming, since the simulation has to produce a sample large enough to get results within a given statistical

error. This drawback is partially alleviated by the "zero variance principle", i.e. if the wave function is an eigenstate of the Hamiltonian, the energy is not affected by fluctuations, and the result is free of statistical errors. Thus, the better is the trial wave function, the lower is the error and the more efficient is the simulation. Unfortunately this property is valid only for the energy and for those operators which commute with the Hamiltonian. Instead the other observables will be affected by statistical fluctuations during the sampling, even if the trial state is exact. However recent progresses have been made toward the extension of the zero variance principle to operators which do not commute with the Hamiltonian [11, 12].

The quality of the trial wave function has another outcome, which is related to the fixed node approximation (FNA) [13–15] of the DMC method. Indeed, the naive DMC evolution is spoiled by the antisymmetry of the fermionic wave function, and in order to keep the simulation stable, the FNA must be applied, i.e. the sampling must be constrained to stay within a nodal pocket of the trial wave function, and the nodal surface crossing must be avoided. Now, if the nodal surface is exact, the FNA is exact. Therefore the FNA depends on the quality of the trial wave function, and in particular on its nodal structure.

Another drawback of the DMC technique is the bad scaling of its efficiency with the atomic number. For DMC simulations involving atoms, the computational cost scales following the relation:

$$\mathrm{computational} \quad \mathrm{cost} \sim \mathrm{Z}^{5.5}, \tag{1.2}$$

which has been estimated by Ceperley[16] and recently verified by Ma *et al.*[17], by carrying out simulations on noble gas atoms with large atomic numbers. This scaling rules out applications to heavy elements, unless pseudopotentials are used, which reduce the effective atomic number and alleviate significantly the computational cost. The need of pseudopotentials leads nevertheless to another problem in the DMC approach. Indeed pseudopotentials are usually non local and it is impossible to deal with non local potentials within the standard DMC framework. For this reason, the locality approximation (LA) has been introduced [18], which is equivalent to replace the Hamiltonian with an effective one containing only local potentials. This approximation fulfills the zero variance property, i.e. it is exact if the trial wave function is exact, like the FNA, but contrary to the latter, its accuracy depends not only on the nodes of the trial state, but also on its amplitude.

Once again, it is crucial to have a good variational wave function, to be used as trial state in the DMC procedure. Therefore, an intense effort has been done to find efficient and robust QMC methods to optimize the parameters of a given wave function.

Optimization schemes are divided into two categories: variance and energy minimization. The former is widely used, since it has proven to be stable and reliable, even for a poor sampling of the variational wave function. Nevertheless, a lot of effort has been put recently into developing new energy minimization methods, which could be as efficient and stable as the variance based optimizations. Indeed the use of the energy minimization is generally assumed to provide "better" variational wave functions, since the aim of either the VMC or the DMC calculations is to deal with the lowest possible energy, rather than the lowest variance. Moreover the latest energy minimization schemes based on the stochastic evaluation of the Hessian matrix (HSR) [19] are shown to be robust, stable and much more efficient than the previously used energy optimization methods, like e.g. the Stochastic Reconfiguration (SR) algorithm [20].

In this thesis VMC, SR, HSR, DMC and LRDMC calculations have been carried out to obtain the numerical results presented in the next chapters. In particular, one of the original topics of this thesis is the development of the LRDMC method which overcomes the drawbacks of the LA in the standard DMC method when non local pseudopotentials are included. Therefore an accurate introduction is necessary for each method used, with a careful analysis of the DMC and GFMC algorithms, upon which the new LRDMC approach is based. Of course, other important QMC methods exist, such as auxiliary-field, reptation, and path-integral QMC, but they will not be discussed in this thesis.

The structure of this chapter is organized as follows: in Section 1.2 we present the statistical foundations of the Monte Carlo approach and we describe the Metropolis algorithm, in Section 1.3 we introduce the simple variational Monte Carlo method. The optimization techniques are presented in Section 1.4, while the projection DMC and GFMC methods are analyzed in Section 1.5 and 1.6 respectively. In the last Section we will briefly outline the performances of a parallel implementation of QMC algorithms.

1.2 Statistical foundation and Metropolis algorithm

All QMC methods are based on the sampling of a given probability distribution $\pi(\mathbf{R})$, where $\mathbf{R}$ is a vector of a multi dimensional space S. In order to generate a sequence of points $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ representative of the distribution π, it is necessary to invent a *stochastic process* having π as its unique equilibrium distribution. In practice, we simulate on the computer the evolution of this process, whose fictitious dynamics will converge to the desired distribution, and the expectation values of any observable defined on the space S will be computed as time averages, during this evolution.

The stochastic process is always taken to be a Markov chain, called also *random walk*, where the *walker* is an object defined to have the position $\mathbf{R} \in S$ (usually called *configuration*), and in some algorithms also a real number w (called *weight*). During the fictitious time evolution, both the position and the weight can change, and its random path in the space S will sample the distribution π, with a set of configurations snipped out of the random walk. A Markov chain is fully defined by the initial distribution of the walker, and by the transition probability $p(\mathbf{R}, \mathbf{R}')$, which fulfills the two conditions:

$$p(\mathbf{R}, \mathbf{R}') \geq 0 \quad \text{for all } \mathbf{R} \text{ and } \mathbf{R}' \text{ in } S, \tag{1.3}$$

$$\int d\mathbf{R}' p(\mathbf{R}, \mathbf{R}') = 1. \tag{1.4}$$

The configuration is changed randomly according to p, in such a way that the Markov sequence or sample $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n\}$ of n elements is specified by the joint probabilities:

$$P(\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n\}) \equiv \alpha(\mathbf{R}_1) p(\mathbf{R}_1, \mathbf{R}_2) p(\mathbf{R}_2, \mathbf{R}_3) \cdots p(\mathbf{R}_{n-1}, \mathbf{R}_n), \tag{1.5}$$

where α is the initial probability distribution. As one can see from the above relation, the successive transitions of this random process are statistically independent, i.e. "the future depends on the past only through the present".

An important property of the Markov chain is its *ergodicity*, also called *irreducibility* in the mathematics literature [21, 22]. If the Markov chain is irreducible, for each $(\mathbf{R}, \mathbf{R}')$ there exists an $n \geq 0$ such that $p^n(\mathbf{R}, \mathbf{R}') > 0$, where p^n is the probability to reach $\mathbf{R}'$ from $\mathbf{R}$ in n steps (n-steps transition probability). In other words, a subspace of S can be connected to any other by the Markov process.

Another fundamental topic in the theory of Markov chains is the problem of convergence to equilibrium, and the related definition of *stationary distribution*. A probability distribution π is stationary for the Markov chain p in case:

$$\int d\mathbf{R}'\pi(\mathbf{R}')p(\mathbf{R}',\mathbf{R}) = \pi(\mathbf{R}) \quad \text{for all } \mathbf{R}. \tag{1.6}$$

Now, an important theorem[21] states that if a stationary distribution π exists and the transition probability p is irreducible, then π is unique. In particular, if the process does not come back to the same position with a fixed period (*aperiodicity*), then:

$$lim_{n\to\infty}p^n(\mathbf{R}',\mathbf{R}) = \pi(\mathbf{R}). \tag{1.7}$$

The last statement guarantees that the random walk will converge to the stationary distribution no matter what the initial walker distribution α is. Therefore, under the general properties of ergodicity and stationarity, the Markov chain, after an initial transient, will provide the sampling of a unique stationary state without any dependence on the initial distribution α. The initial transient is usually called *thermalization* or *equilibration*. A Monte Carlo method able to generate a Markov chain based on these properties is therefore suitable to sample a given probability distribution π, and to estimate averages with respect to π. The procedure which allows an arbitrarily complex distribution to be sampled in a straightforward way without knowledge of its normalization is the *Metropolis algorithm*[23].

A sufficient condition to obtain π as stationary distribution is to choose the transition probability p to satisfy:

$$\pi(\mathbf{R})p(\mathbf{R},\mathbf{R}') = \pi(\mathbf{R}')p(\mathbf{R}',\mathbf{R}), \tag{1.8}$$

which is the *detailed balance condition*. In fact, by integrating the above over $\mathbf{R}$, one gets:

$$\int d\mathbf{R}\pi(\mathbf{R})p(\mathbf{R},\mathbf{R}') = \pi(\mathbf{R}')\int d\mathbf{R}p(\mathbf{R}',\mathbf{R}) = \pi(\mathbf{R}'), \tag{1.9}$$

where the first equality follows directly from the detailed balance condition and the second from the normalization of the transition probability p in Eq. 1.4. The transition probability may be conveniently decomposed into the product of an irreducible proposal or sampling matrix $T(\mathbf{R},\mathbf{R}')$ and an acceptance matrix $A(\mathbf{R},\mathbf{R}')$:

$$p(\mathbf{R},\mathbf{R}') = T(\mathbf{R},\mathbf{R}')A(\mathbf{R},\mathbf{R}'). \tag{1.10}$$

Imposing the detailed balance yields

$$\frac{A(\mathbf{R},\mathbf{R}')}{A(\mathbf{R}',\mathbf{R})} = \frac{\pi(\mathbf{R}')T(\mathbf{R}',\mathbf{R})}{\pi(\mathbf{R})T(\mathbf{R},\mathbf{R}')} \equiv q(\mathbf{R},\mathbf{R}'), \tag{1.11}$$

which can be satisfied quite generally by choosing

$$A(\mathbf{R},\mathbf{R}') = F\left[q(\mathbf{R},\mathbf{R}')\right], \tag{1.12}$$

where the function $F : [0,\infty] \rightarrow [0,1]$ satisfies:

$$\frac{F[z]}{F[1/z]} = z, \quad \text{for all } z. \tag{1.13}$$

In the Metropolis algorithm [23], the function is chosen to be $F[z] = min[1,z]$, but other choices are possible, such as the so called "thermal bath" $F[z] = 1/(1+z)$. The algorithm which exploits the decomposition of the transition probability into the acceptance and proposal probability is the *generalized Metropolis algorithm* [24]. In the simplest version, the acceptance is just the transition probability.

A common task of QMC methods is the estimate of the expectation value of a quantum operator, which turns to be the evaluation of the multidimensional integral $\int d\mathbf{R}\pi(\mathbf{R})\mathcal{O}(\mathbf{R})$. Its value is by definition the mean of the observable $\mathcal{O}$, indicated with the symbol $\langle\mathcal{O}\rangle$. In any Monte Carlo integration, a sample of the distribution π will be generated during the simulation, and the estimate for the true mean value $\langle\mathcal{O}\rangle$ is:

$$\overline{\mathcal{O}} = \frac{1}{M}\sum_{i=1}^{M}\mathcal{O}_i, \tag{1.14}$$

where $\mathcal{O}_i = \mathcal{O}(\mathbf{R}_i)$, and M is the length of the sampling. The quantity in Eq. 1.14 is an unbiased estimator of the true mean, since for the law of large numbers $lim_{M\rightarrow\infty}\overline{\mathcal{O}} = \langle\mathcal{O}\rangle$. Moreover, the central limit theorem states that $\overline{\mathcal{O}}$ is normally distributed around $\langle\mathcal{O}\rangle$. Here, it is apparent that the Monte Carlo approach is affected both by statistical and systematic errors, and thus shares the same features of ordinary experimental work. Indeed, this computational tool can be thought, more than any other numerical method, as the third way of doing science, lying between the theoretical and the experimental sides. In order to compute the statistical error on $\overline{\mathcal{O}}$, we need to evaluate the variance:

$$\sigma^2(\overline{\mathcal{O}}) = \langle(\overline{\mathcal{O}} - \langle\mathcal{O}\rangle)^2\rangle, \tag{1.15}$$

the square root of which will be the estimate of the error bar. One can show that the variance in Eq. 1.15 can be written in an approximate way as

$$\sigma^2(\overline{\mathcal{O}}) \simeq \frac{\tau}{M}\sigma^2(\mathcal{O}), \tag{1.16}$$

where $\sigma^2(\mathcal{O}) = \langle \mathcal{O}^2 \rangle - \langle \mathcal{O} \rangle^2$ is the variance of the true average (the latter depending only on the observable $\mathcal{O}$ and the probability distribution π and *not* on the sampling), and τ is the *autocorrelation* time, i.e. the time in the fictitious Markov evolution between two uncorrelated points of the sample. From Eq. 1.16 it is clear that the error on $\overline{\mathcal{O}}$ goes as $1/\sqrt{M}$, as already stated in Eq. 1.1. Instead, if one uses a deterministic d-dimensional cubic mesh to evaluate a d-dimensional integral, like for example the Simpson's rule, the error scales as $M^{-4/d}$, where M is the total number of mesh points. Therefore, if the number of degrees of freedom (dimension of the configuration space) is $d > 8$, it is more convenient to use a Monte Carlo integration scheme based on random walk and Metropolis algorithm.

The Metropolis algorithm is able to compute the averages over a sequence of sampling points $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ generated by moving a single walker, according to the following rules:

- Initialize the system in the state $\mathbf{R}_1$;
- To advance from $\mathbf{R}_n$ to $\mathbf{R}_{n+1}$:
  - sample $\mathbf{R}'$ from $T(\mathbf{R}_n, \mathbf{R}')$,
  - calculate
    $$q(\mathbf{R}_n, \mathbf{R}') = \frac{\pi(\mathbf{R}')T(\mathbf{R}', \mathbf{R}_n)}{\pi(\mathbf{R}_n)T(\mathbf{R}_n, \mathbf{R}')},$$
  - generate a random number r_n and compare it with $q(\mathbf{R}_n, \mathbf{R}')$: if $q(\mathbf{R}_n, \mathbf{R}') > r_n$ accept the move, otherwise reject it;

  If the move is accepted, $\mathbf{R}_{n+1} = \mathbf{R}'$; otherwise, $\mathbf{R}_{n+1} = \mathbf{R}_n$.
- Trow away the first k states as being out of equilibrium;
- Collect averages using the configurations with $n > k$;
- Calculate error bars (using for instance blocking techniques, see Ref.[25]).

1.3 Variational Monte Carlo

Variational Monte Carlo is the simplest of the QMC algorithms, and it is used to compute quantum expectation values of an operator with a given *trial wave function* Ψ_T. In particular, if the operator is the Hamiltonian, its expectation value is the variational energy E_V, which provides a rigorous upper bound on the exact ground state energy E_0:

$$E_V = \frac{\int d\mathbf{R}\Psi_T^*(\mathbf{R})H\Psi_T(\mathbf{R})}{\int d\mathbf{R}\Psi_T^*(\mathbf{R})\Psi_T(\mathbf{R})} \geq E_0. \tag{1.17}$$

The trial wave function is supposed to be as close as possible to the true ground state of the system, or more generally to an exact eigenstate of the Hamiltonian, if one is interested in studying also the properties of excited states. The trial wave function must satisfy some basic conditions. Both Ψ_T and $\vec{\nabla}\Psi_T$ must be continuous wherever the potential is finite, and the integrals $\int d\mathbf{R}\Psi_T^*(\mathbf{R})H\Psi_T(\mathbf{R})$ and $\int \Psi_T^*(\mathbf{R})\Psi_T(\mathbf{R})$ present in Eq. 1.17 must exist. To keep the variance of the energy finite we also require $\int d\mathbf{R}\Psi_T^*(\mathbf{R})H^2\Psi_T(\mathbf{R})$ to exist.

In order to evaluate E_V with the VMC method, we rewrite the integral in Eq. 1.17 and obtain:

$$E_V = \int d\mathbf{R}\pi(\mathbf{R})E_L(\mathbf{R}), \tag{1.18}$$

where $\pi(\mathbf{R}) = |\Psi_T(\mathbf{R})|^2 / \int d\mathbf{R}'|\Psi_T(\mathbf{R}')|^2$ is positive everywhere and interpreted as a probability distribution, and $E_L(\mathbf{R}) = \frac{H\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}$ is the *local energy*. Using the Monte Carlo integration scheme shown in the previous section, we sample a set of points $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ from the distribution π and at each of these points we calculate the local energy. Then, the QMC estimate of E_V will be:

$$E_V \approx \frac{1}{M}\sum_{m=1}^{M} E_L(\mathbf{R}_m). \tag{1.19}$$

Some effort has been done to find an optimal proposal matrix (Eq. 1.10), in order to improve the overall efficiency and reduce the autocorrelation time. The best proposal probability tries to maximize the diffusion of the Markov process within a given computational time frame. Thus in general it is not true that the most efficient acceptance corresponds to 0.5 (50% of the total moves accepted). In our implementation of the VMC algorithm, we preferred to use the simple Metropolis algorithm with the constant proposal probability and with the single-electron

move, since the evaluation of a complex proposal matrix could become too expensive in the case of simulation involving a large number of electrons. Moreover, we found that the use of a proposal move which sometimes jumps between two nuclear sites is effective in reducing the autocorrelation time, in particular when two nuclei are far apart.

Notice that one does not need to compute the normalization of Ψ_T, which can be extremely difficult to calculate for a wave function with a complex functional form. Indeed, in the Metropolis algorithm Ψ_T appears only in terms that are independent of its overall normalization factor. Moreover, as the trial wave function approaches an eigenstate of the Hamiltonian, the VMC estimate of the energy converges more rapidly with the number M of steps in the random walk, and therefore the algorithm is more efficient in computing energy expectation values. In particular, if Ψ_T coincides with an eigenstate, the variance of the local energy will go to zero since the local energy will be a constant and therefore the statistical error will vanish (see Eq. 1.16). This property is the *zero variance principle*, and suggests that it is very important to deal with a well optimized trial wave function. Although only the operators which commute with the Hamiltonian benefit from this principle, Assaraf and Caffarel [11, 12] have demonstrated that a significant variance reduction can be obtained also by other appropriate estimators, constructed to satisfy the zero variance principle. The methods to optimize Ψ_T will be the subject of the next section.

1.4 Wave function optimization

As already mentioned in the introduction, there has been an extensive effort to find an efficient and robust optimization method with the aim to improve the variational wave function. Indeed, a good wave function yields results with greater statistical accuracy both in VMC and in DMC simulations. Moreover, within the DMC framework, the FNA and the LA (used in the case of non local potentials), benefit from an optimized wave function since all these approximations become exact as the trial state approaches an exact eigenstate of the Hamiltonian. Therefore a well optimized variational ansatz is crucial to obtain reliable and accurate results. The usual trial wave function used in QMC calculation is the product of an antisymmetric part and a Jastrow factor, extensively described in Chapter 2.

The antisymmetric part can be either a single Slater determinant or a multi configuration state, while the Jastrow factor is a bosonic many body function which accounts for the dynamical correlations in the system.

Two different approaches exist for the wave function optimization: the variance and the energy minimization. The former has been presented by Umrigar *et al.*[26] in 1988 and widely used in the last two decades. Let $\{\alpha_i\}$ be the variational parameters contained in the trial wave function. These are obtained by minimizing the variance of the local energy over a set of M configurations $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ sampled from the square of the initial guess $\Psi_T(\mathbf{R}, \alpha^0)$:

$$\sigma^2(\alpha) = \sum_i^M \left[\frac{H\Psi_T(\mathbf{R}_i, \alpha)}{\Psi_T(\mathbf{R}_i, \alpha)} - \bar{E}\right]^2 w(\mathbf{R}_i, \alpha) / \sum_i^M w(\mathbf{R}_i, \alpha), \tag{1.20}$$

where

$$\bar{E} = \sum_i^M \frac{H\Psi_T(\mathbf{R}_i, \alpha)}{\Psi_T(\mathbf{R}_i, \alpha)} w(\mathbf{R}_i, \alpha) / \sum_i^M w(\mathbf{R}_i, \alpha), \tag{1.21}$$

is the average energy over the sample of configurations. The weights $w(\mathbf{R}_i, \alpha) = |\Psi_T(\mathbf{R}_i, \alpha)/\Psi_T(\mathbf{R}_i, \alpha^0)|^2$ take into account the change of the variational wave function due to the change of the parameters, while the set of configurations remains the same. In this way, it is enough to generate about 2000 points from the starting guessed distribution in order to find the minimum of $\sigma^2(\alpha)$ and to iterate few times the procedure until the starting set of parameters is close to the optimal one. An improved version of the algorithm is the unreweighted variance minimization[27, 28], i.e. with all unitary weights, which is more stable since it avoids weights fluctuations. The advantage of this method is that $\sigma^2(\alpha)$ is the sum of all positive terms, therefore the optimization iterated over a restrict and fixed sample leads to a *real* minimization of the variance, once it is calculated over a wider sample based on the new wave function. Instead, for a naive minimization of energy over a limited sample, it is not guaranteed that the new energy will be really lower than the starting one, and often the minimum does not even exist.

Despite the efficiency and robustness of the existing variance minimization, the possibility to develop an energy minimization method is still appealing, since the structural optimization of a compound is feasible only within an energy based approach, and also because it has been observed[29] that an energy optimized wave function gives better expectation values for operators which do not commute

with the Hamiltonian. Therefore a lot of energy based optimization methods for QMC calculations have been proposed during these last few years, ranging from the simplest steepest descent (SD) approach [30] to the more sophisticated Newton method [31–33]. The goal is always to design a scheme which is stable even in the presence of the statistical noise of QMC sampling, and which converges quickly to the global minimum of the estimator. In the two next subsections we will present the Stochastic Reconfiguration (SR) method and the Stochastic Reconfiguration method with Hessian accelerator (SRH). Both of them are energy minimization procedures largely used in the present study, the latter is an evolution of the former after the introduction of a reliable and efficient scheme to estimate the Hessian matrix.

1.4.1 Stochastic reconfiguration method

We introduce the *stochastic minimization* of the total energy based upon the SR technique, already exploited for lattice systems [20]. Let $\Psi_T(\alpha^0)$ be the wavefunction depending on an initial set of p variational parameters $\{\alpha_k^0\}_{k=1,\dots,p}$. Consider now a small variation of the parameters $\alpha_k = \alpha_k^0 + \delta\alpha_k$. The corresponding wavefunction $\Psi_T(\alpha)$ is equal, within the validity of the linear expansion, to the following one:

$$\Psi'_T(\alpha) = \left(\Psi_T(\alpha^0) + \sum_{k=1}^{p} \delta\alpha_k \frac{\partial}{\partial\alpha_k}\Psi_T(\alpha^0)\right) \tag{1.22}$$

Therefore, by introducing local operators defined on each configuration $x = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ as the logarithmic derivatives with respect to the variational parameters:

$$O^k(x) = \frac{\partial}{\partial\alpha_k} \ln \Psi_T(x) \tag{1.23}$$

and for convenience the identity operator $O^0 = 1$, we can write Ψ'_T in a more compact form:

$$|\Psi'_T(\alpha)\rangle = \sum_{k=0}^{p} \delta\alpha_k O^k |\Psi_T\rangle, \tag{1.24}$$

where $|\Psi_T\rangle = |\Psi_T(\alpha_0)\rangle$ and $\delta\alpha_0 = 1$. However, as a result of the iterative minimization scheme we are going to present, $\delta\alpha_0 \neq 1$, and in that case the

variation of the parameters will be obviously scaled

$$\delta\alpha_k \rightarrow \frac{\delta\alpha_k}{\delta\alpha_0} \tag{1.25}$$

and Ψ'_T will be proportional to $\Psi_T(\alpha)$ for small $\frac{\delta\alpha_k}{\delta\alpha_0}$.

Our purpose is to set up an iterative scheme to reach the minimum possible energy for the parameters α, exploiting the linear approximation for $\Psi_T(\alpha)$, which will become more and more accurate close to the convergence, when the variation of the parameters is smaller and smaller. We follow the stochastic reconfiguration method and define

$$|\Psi'_T\rangle = P_{SR}(\Lambda - H)|\Psi_T\rangle \tag{1.26}$$

where Λ is a suitable large shift, allowing Ψ'_T to have a lower energy than Ψ_T [20], and P_{SR} is a projection operator over the $(p+1)$–dimensional subspace, spanned by the basis $\{O_k|\Psi_T\rangle\}_{k=0,\ldots,p}$, over which the function $|\Psi'_T\rangle$ has been expanded (Eq. 1.24). In a continuous system, if its energy is unbounded from above, Λ should be infinite. However, in this case, the optimal Λ is finite, since the basis is finite, and the spectrum of the Hamiltonian diagonalized in this basis is bounded from above as in a lattice system. In order to determine the coefficients $\{\delta\alpha_k\}_{k=1,\ldots,p}$ corresponding to Ψ'_T defined in Eq.1.26, one needs to solve the SR conditions:

$$\langle\Psi_T|O^k(\Lambda - H)|\Psi_T\rangle = \langle\Psi_T|O^k|\Psi'_T\rangle \text{ for } k = 0,\ldots,p \tag{1.27}$$

that can be rewritten in a linear system:

$$\sum_l \delta\alpha_l \, s_{l,k} = f^k, \tag{1.28}$$

where $s_{l,k} = \langle\Psi_T|O^l O^k|\Psi_T\rangle$ is the covariance matrix and $f^k = \langle\Psi_T|O^k(\Lambda - H)|\Psi_T\rangle$ is the known term; both $s_{l,k}$ and f^k are computed stochastically by a Monte Carlo integration. These linear equations (1.28) are very similar to the ones introduced by Filippi and Fahy [34] for the energy minimization of the Slater part. In our formulation, there is no difficulty to optimize the Jastrow and the Slater part of the wavefunction at the same time. The present scheme is also much simpler because does not require to deal with an effective one body Hamiltonian, but is seems to be less efficient, since it treats all energy scales at the same footing (see Subsection 1.4.1 and Ref. [35]).

After the system (1.28) has been solved, we update the variational parameters

$$\alpha_k = \alpha_k^{(0)} + \frac{\delta\alpha_k}{\delta\alpha_0} \text{ for } k = 1, \ldots, p \tag{1.29}$$

and we obtain a new trial wavefunction $\Psi_T(\alpha)$. By repeating this iteration scheme several times, one approaches the convergence when $\frac{\delta\alpha_k}{\delta\alpha_0} \to 0$ for $k \neq 0$, and in this limit the SR conditions (1.27) implies the Euler equations of the minimum energy. Obviously, the solution of the linear system (1.28) is affected by statistical errors, yielding statistical fluctuations of the final variational parameters α_k even when convergence has been reached, namely when the $\{\alpha_k\}_{k=1,\ldots,p}$ fluctuate without drift around an average value. We perform several iterations in that regime; in this way, the variational parameters can be determined more accurately by averaging them over all these iterations and by evaluating also the corresponding statistical error bars.

It is worth noting that the solution of the linear system (1.28) depends on Λ only through the $\delta\alpha_0$ variable. Therefore the constant Λ indirectly controls the rate of change in the parameters at each step, i.e. the speed of the algorithm for convergence and the stability at equilibrium: a too small value will produce uncontrolled fluctuations for the variational parameters, a too large one will allow convergence in an exceedingly large number of iterations. The choice of Λ can be controlled by evaluating the change of the wavefunction at each step as:

$$\frac{|\Psi_T' - \Psi_T|^2}{|\Psi_T|^2} = \sum_{k,k'>0} \delta\alpha_k \, \delta\alpha_{k'} \, s_{k,k'} \tag{1.30}$$

By keeping this value small enough during the optimization procedure, one can easily obtain a steady and stable convergence. Moreover, we mention that the stochastic procedure is able in principle to perform a global optimization, as discussed in Ref. [20] for the SR and in Ref. [30] for the Stochastic Gradient Approximation (SGA), because the noise in the sampling can avoid the dynamics of the parameters to get stuck into local minima.

Stochastic reconfiguration versus steepest descent method

SR is similar to a standard SD calculation, where the expectation value of the energy $E(\alpha_k) = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}$ is optimized by iteratively changing the parameters α_i

according to the corresponding derivatives of the energy (generalized forces):

$$f_k = -\frac{\partial E}{\partial \alpha_k} = -\frac{\langle \Psi | O_k H + H O_k + (\partial_{\alpha_k} H) | \Psi \rangle}{\langle \Psi | \Psi \rangle} + 2 \frac{\langle \Psi | O_k | \Psi \rangle \langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle^2}, \quad (1.31)$$

namely:

$$\alpha_k \rightarrow \alpha_k + \Delta t f_k. \quad (1.32)$$

Δt is a suitable small time step, which can be taken fixed or determined at each iteration by minimizing the energy expectation value. Indeed the variation of the total energy ΔE at each step is easily shown to be negative for small enough Δt because, in this limit

$$\Delta E = -\Delta t \sum_i f_i^2 + O(\Delta t^2).$$

Thus the method certainly converges at the minimum when all the forces vanish. Notice that in the definition of the generalized forces (1.31) we have generally assumed that the variational parameters may appear also in the Hamiltonian. This is particularly important for the structural optimization since the atomic positions that minimize the energy enter both in the wave function and in the potential.

In the following we will show that similar considerations hold for the SR method, that can be therefore extended to the optimization of the geometry. Indeed, by eliminating the equation with index $k = 0$ from the linear system (1.28), the SR iteration can be written in a form similar to the steepest descent:

$$\alpha_i \rightarrow \alpha_i + \Delta t \sum_k \bar{s}^{-1}_{i,k} f_k \quad (1.33)$$

where the reduced $p \times p$ matrix $\bar{s}$ is:

$$\bar{s}_{j,k} = s_{j,k} - s_{j,0} s_{0,k} \quad (1.34)$$

and the Δt value is given by:

$$\Delta t = \frac{1}{2(\Lambda - \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} - \sum_{k>0} \Delta \alpha_k s_{k,0})}. \quad (1.35)$$

From the latter equation the value of Δt changes during the simulation and remains small for large enough energy shift Λ. However, using the analogy with the steepest descent, convergence to the energy minimum is reached also when

the value of Δt is sufficiently small and is kept constant for each iteration (we have chosen to determine Δt by verifying the stability and the convergence of the algorithm at fixed Δt value). Indeed the energy variation for a small change of the parameters is:

$$\Delta E = -\Delta t \sum_{i,j} \bar{s}_{i,j}^{-1} f_i f_j.$$

It is easily verified that the above term is always negative because the reduced matrix $\bar{s}$, as well as $\bar{s}^{-1}$, is positive definite, being s an overlap matrix with all positive eigenvalues.

For a stable iterative method, such as the SR or the SD one, a basic ingredient is that at each iteration the new parameters α' are close to the previous α according to a prescribed distance. The fundamental difference between the SR minimization and the standard steepest descent is just related to the definition of this distance. For the SD it is the usual one defined by the Cartesian metric $\Delta_\alpha = \sum_k |\alpha'_k - \alpha_k|^2$, instead the SR works correctly in the physical Hilbert space metric of the wave function Ψ, yielding $\Delta_\alpha = \sum_{i,j} \bar{s}_{i,j}(\alpha'_i - \alpha_i)(\alpha'_j - \alpha_j)$, namely the square distance between the two normalized wave functions corresponding to the two different sets of variational parameters $\{\alpha'\}$ and $\{\alpha_k\}$ [1]. Therefore, from the knowledge of the generalized forces f_k, the most convenient change of the variational parameters minimizes the functional $\Delta E + \bar{\Lambda}\Delta_\alpha$, where ΔE is the linear change in the energy $\Delta E = -\sum_i f_i(\alpha'_i - \alpha_i)$ and $\bar{\Lambda}$ is a Lagrange multiplier that allows a stable minimization with small change Δ_α of the wave function Ψ. The final iteration (1.33) is then easily obtained.

The advantage of SR compared with SD is obvious because sometimes a small change of the variational parameters correspond to a large change of the wave function, and the SR takes into account this effect through the Eq. 1.33. In particular the method is useful when a non orthogonal basis set is used as we have done in this work. Indeed by using the reduced matrix $\bar{s}$ it is also possible to remove from the calculation those parameters that imply some redundancy in the variational space. As shown in the Appendix A, a more efficient change in the wave function can be obtained by updating only the variational parameters that remain independent within a prescribed tolerance, and therefore, by removing the

[1] Δ_α is equivalent to the quantity of Eq. 1.30, but the variation of the wave function is expressed in the orthogonal basis $\{(O_k - < O_k >)|\Psi_T\rangle\}_{k=1,\ldots,p}$

parameters that linearly depend on the others. A more stable minimization is obtained without spoiling the accuracy of the calculation. A weak tolerance criterium $\epsilon \simeq 10^{-3}$, provides a very stable algorithm even when the dimension of the variational space is large. For a small atomic basis set, by an appropriate choice of the Jastrow and Slater orbitals, the reduced matrix $\bar{s}$ is always very well conditioned even for the largest system studied, and the above stabilization technique is not required. Instead the described method is particularly important for the extension of QMC to complex systems with large number of atoms and/or higher level of accuracy, because in this case it is very difficult to select - e.g. by trial and error - the relevant variational parameters, that allow a well conditioned matrix $\bar{s}$ for a stable inversion in (1.33).

Once all the parameters are independent, that can be checked by explicit calculation of the spectrum of the reduced matrix $\bar{s}$, the simulation is stable whenever $1/\Delta t > \Lambda_{\rm cut}$, where Λ_{cut} is an energy cutoff that is strongly dependent on the chosen wave function and is generally weakly dependent on the bin length. Whenever the wave function is too much detailed, namely has a lot of variational freedom, especially for the high energy components of the core electrons, the value of Λ_{cut} becomes exceedingly large and too many iterations are required for obtaining a converged variational wave function. In fact a rough estimate of the corresponding number of iterations P is given by $P\Delta t >> 1/G$, where G is the typical energy gap of the system, of the order of few eV in small atoms and molecules. Within the SR method it is therefore extremely important to work with a bin length rather small, so that many iterations can be performed without much effort.

Statistical bias of forces

In a Monte Carlo optimization framework the forces f_k are always determined with some statistical noise η_k, and by iterating the procedure several times with a fixed bin length the variational parameters will fluctuate around their mean values. These statistical fluctuations are similar to the thermal noise of a standard Langevin equation:

$$\partial_t \alpha_k = f_k + \eta_k, \tag{1.36}$$

where

$$\langle \eta_k(t) \eta_{k'}(t') \rangle = 2T_{noise} \delta(t-t') \delta_{k,k'}. \tag{1.37}$$

Within a QMC scheme, one needs to control T_{noise}, by increasing the bin length as clearly $T_{noise} \propto 1/\mathrm{Bin\ length}$, because the statistical fluctuations of the forces, obviously decreasing by increasing the bin length, are related to the thermal noise by Eq. 1.37. On the other hand, the number of iterations necessary to reach the convergence is weakly dependent on the bin length, but it depends mainly on the energy landscape of the system. The optimal value for the bin length is the smallest one that provides T_{noise} within the desired accuracy.

The variational parameters α_k, averaged over the Langevin simulation time will be close to the true energy minimum, but the corresponding forces $f_k = -\partial_{\alpha_k} E$ will be affected by a bias that scales to zero with the thermal noise T_{noise}, due to the presence of non quadratic terms in the energy landscape. The systematic bias on the forces should be controlled through an appropriate choice of the bin length in order to not exceed the statistical error of the averaged parameters.

Structural optimization

In the last few years remarkable progresses have been made to develop Quantum Monte Carlo (QMC) techniques which are able in principle to perform structural optimization of molecules and complex systems [11, 12, 36, 37]. Within the Born-Oppenheimer approximation the nuclear positions $\mathbf{R}_i$ can be considered as further variational parameters included in the set $\{\alpha_i\}$ used for the SR minimization (1.33) of the energy expectation value. For clarity, in order to distinguish the conventional variational parameters from the ionic positions, in this section we indicate with $\{c_i\}$ the former ones, and with $\mathbf{R}_i$ the latter ones. It is understood that $R_i^\nu = \alpha_k$, where a particular index k of the whole set of parameters $\{\alpha_i\}$ corresponds to a given spatial component ($\nu = 1, 2, 3$) of the $i-$th ion. Analogously the forces (1.31) acting on the ionic positions will be indicated by capital letters with the same index notations.

The purpose of the present section is to compute the forces $\mathbf{F}$ acting on each of the T nuclear positions $\{\mathbf{R}_1, \ldots, \mathbf{R}_T\}$, being T the total number of nuclei in the system:

$$\mathbf{F}(\mathbf{R}_a) = -\nabla_{\mathbf{R}_a} E(\{c_i\}, \mathbf{R}_a), \tag{1.38}$$

with a reasonable statistical accuracy, so that the iteration (1.33) can be effective for the structural optimization. In this work we have used a finite difference

operator $\frac{\mathbf{\Delta}}{\mathbf{\Delta R}_a}$ for the evaluation of the force acting on a given nuclear position a:

$$\mathbf{F}(\mathbf{R}_a) = -\frac{\mathbf{\Delta}}{\mathbf{\Delta R}_a} E = -\frac{E(\mathbf{R}_a + \mathbf{\Delta R}_a) - E(\mathbf{R}_a - \mathbf{\Delta R}_a)}{2\Delta R} + O(\Delta R^2) \quad (1.39)$$

where $\mathbf{\Delta R}_a$ is a 3 dimensional vector. Its length ΔR is chosen to be 0.01 atomic units, a value that is small enough for negligible finite difference errors. In order to evaluate the energy differences in Eq. 1.39 we have used the space-warp coordinate transformation [38, 39] briefly summarized in the following paragraphs. According to this transformation also the electronic coordinates $\mathbf{r}$ will be translated in order to mimic the right displacement of the charge around the nucleus a:

$$\overline{\mathbf{r}}_i = \mathbf{r}_i + \mathbf{\Delta R}_a \, \omega_a(\mathbf{r}_i), \quad (1.40)$$

where

$$\omega_a(\mathbf{r}) = \frac{F(|\mathbf{r} - \mathbf{R}_a|)}{\sum_{b=1}^{T} F(|\mathbf{r} - \mathbf{R}_b|)}. \quad (1.41)$$

$F(r)$ is a function which must decay rapidly; here we used $F(r) = \frac{1}{r^4}$ as suggested in Ref. [39].

The expectation value of the energy depends on $\mathbf{\Delta R}$, because both the Hamiltonian and the wave function depend on the nuclear positions. Now let us apply the space-warp transformation to the integral involved in the calculation; the expectation value reads:

$$E(\mathbf{R} + \mathbf{\Delta R}) = \frac{\int d\mathbf{r} J_{\mathbf{\Delta R}}(\mathbf{r}) \Psi^2_{\mathbf{\Delta R}}(\overline{\mathbf{r}}(\mathbf{r})) E_L^{\mathbf{\Delta R}}(\overline{\mathbf{r}}(\mathbf{r}))}{\int d\mathbf{r} J_{\mathbf{\Delta R}}(\mathbf{r}) \Psi^2_{\mathbf{\Delta R}}(\overline{\mathbf{r}}(\mathbf{r}))}, \quad (1.42)$$

where J is the Jacobian of the transformation and here and henceforth we avoid for simplicity to use the atomic subindex a. The importance of the space warp in reducing the variance of the force is easily understood for the case of an isolated atom a. Here the force acting on the atom is obviously zero, but only after the space warp transformation with $\omega_a = 1$ the integrand of expression (1.42) will be independent of $\mathbf{\Delta R}$, providing an estimator of the force with zero variance.

Starting from Eq. 1.42, it is straightforward to explicitly derive a finite difference differential expression for the force estimator, which is related to the gradient of the previous quantity with respect to $\mathbf{\Delta R}$, in the limit of the displacement tending to zero:

$$\mathbf{F}(\mathbf{R}) = -\Big\langle \lim_{|\mathbf{\Delta R}| \to 0} \frac{\mathbf{\Delta}}{\mathbf{\Delta R}} E_L \Big\rangle \quad (1.43)$$

$$+ \quad 2\Big(\big\langle H\big\rangle\big\langle \lim_{|\mathbf{\Delta R}|\to 0}\frac{\mathbf{\Delta}}{\mathbf{\Delta R}}\log(J^{1/2}\Psi)\big\rangle - \big\langle H \lim_{|\mathbf{\Delta R}|\to 0}\frac{\mathbf{\Delta}}{\mathbf{\Delta R}}\log(J^{1/2}\Psi)\big\rangle\Big),$$

where the brackets indicate a Monte Carlo like average over the square modulus of the trial wave function, $\frac{\mathbf{\Delta}}{\mathbf{\Delta R}}$ is the finite difference derivative as defined in (1.39), and $E_L = \frac{\langle\Psi|H|x\rangle}{\langle\Psi|x\rangle}$ is the local energy on a configuration x where all electron positions and spins are given. In analogy with the general expression (1.31) of the forces, we can identify the operators O_k corresponding to the space-warp change of the variational wave function:

$$O_k = \frac{\Delta^\nu}{\Delta R}\log(J^{1/2}_{\mathbf{\Delta R}}\Psi_{\mathbf{\Delta R}}) \tag{1.44}$$

The above operators (1.44) are used also in the definition of the reduced matrix $\bar{s}$ for those elements depending on the variation with respect to a nuclear coordinate. In this way it is possible to optimize both the wave function and the ionic positions at the same time, in close analogy with the Car-Parrinello[40] method applied to the minimization problem. Also Tanaka [41] tried to perform Car-Parrinello like simulations via QMC, within the less efficient steepest descent framework.

An important source of systematic errors is the dependence of the variational parameters c_i on the ionic configuration $\mathbf{R}$, because for the final equilibrium geometry all the forces f_i corresponding to c_i have to be zero, in order to guarantee that the true minimum of the potential energy surface (PES) is reached [42]. As shown clearly in the previous subsection, within a QMC approach it is possible to control this condition by increasing systematically the bin length, when the thermal bias T_{noise} vanishes. In Fig. 1.1 we report the equilibrium distance of the Li molecule as a function of the inverse bin length, for two different basis sets, so that an accurate evaluation of such an important quantity is possible even when the number of variational parameters is rather large, by extrapolating the value to an infinite bin length.

We have not attempted to extend the geometry optimization to the more accurate DMC, since there are technical difficulties [43], and it is computationally much more demanding.

Different energy scales

The SR method performs generally very well, whenever there is only one energy scale in the variational wave function. However if there are several energy scales

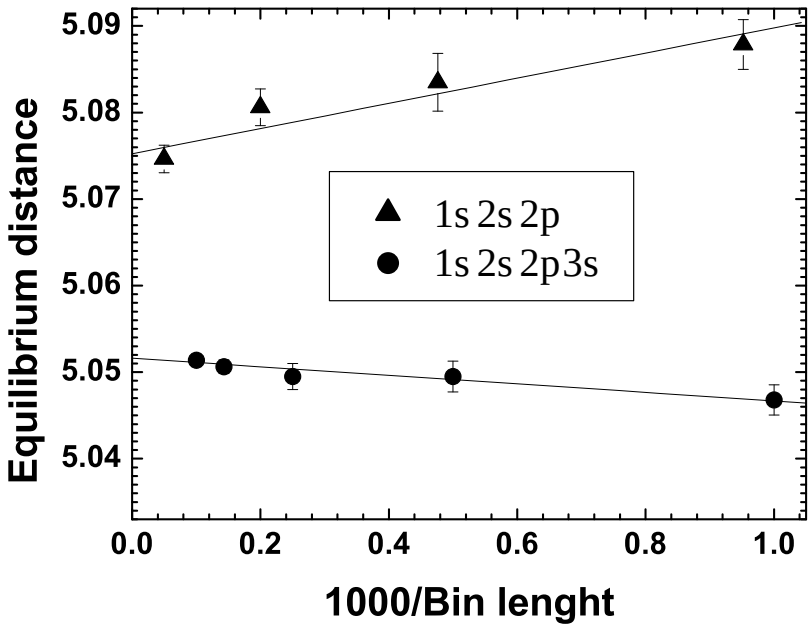


Figure 1.1: Plot of the equilibrium distance of the Li_2 molecule as a function of the inverse bin length. The total energy and the binding energy are reported in Chapter 2 in Tables 2.4 and 2.5 respectively. The triangles (full dots) refer to a simulation performed using 1000 (3000) iterations with $\Delta t = 0.015 H^{-}1$ ($\Delta t = 0.005 H^{-}1$) and averaging over the last 750 (2250) iterations. For all simulations the initial wavefunction is optimized at $Li - Li$ distance 6 a.u.

in the problem, some of the variational parameters, e.g. the ones defining the low energy valence orbitals, converge very slowly with respect to the others, and the number of iterations required for the equilibration becomes exceedingly large, considering also that the time step Δt necessary for a stable convergence depends on the high energy orbitals, whose dynamics cannot be accelerated beyond a certain threshold.

If the interest is limited to a rather small atomic basis, the SR technique is efficient, and general enough to perform the simultaneous optimization of the Jastrow and the determinantal part of the wave function, a very important feature

that allows to capture the most non trivial correlations contained in our variational ansatz. Moreover, SR is able to perform the structural optimization of a chemical system, which is another appealing characteristic of this method. However, to optimize an extended atomic basis, it is necessary to go beyond the SR method, and to use the novel SRH, which will be presented in the next subsection.

1.4.2 Stochastic reconfiguration with Hessian accelerator

With the aim to solve the problem of the "slowing down" of the SR optimization with the increase of the energy spread in the system, another and much more powerful technique has been developed: the stochastic reconfiguration with Hessian accelerator. It is rather clear that the knowledge of the second order derivatives and their inclusion in a quadratic minimization procedure yields an hint on the energy scales of the systems and provides a faster convergence to the minimum of the total energy. Indeed, let us suppose that our variational ansatz defines in the Hilbert space a subspace spanned by the vectors $\{O_k|\Psi_T\rangle\}_{k=0,\dots,p}$ (see subsection 1.4.1); in principle, we can optimize the wave function expanded over that basis by performing an exact numerical diagonalization of the matrix Hamiltonian with elements $H_{k,k'} = \langle O_k|H|O_{k'}\rangle$. In this case, we exploit the entire spectrum of the effective Hamiltonian in the reduced basis, in order to find out the lowest state within this finite subspace. The matrix $H_{k,k'}$ is a term of the total Hessian, and it will be included also in the SRH approach. Since the expansion of the variational wave function over the basis defined by the operators O_k is valid only for a small variation of the parameters α and all the terms are obtained via a QMC sampling and are affected by a statistical noise, the convergence to the minimum will be reached by an iterative scheme, as in the standard SR method. In the quadratic regime without noise the Newton method based on the exact Hessian matrix should converge in one step, but in a region not so close to the stationary point also the contributions beyond the second order will be important and the convergence will be not immediate, although much faster than in the SR case.

To derive the SRH method, we expand the trial wave function by taking into account also second order terms:

$$|\psi_{\alpha+\gamma}\rangle \simeq \left[1 + \sum_k \gamma_k (O_k - < O_k >)\right.$$

$$+ \quad \frac{\beta}{2}\sum_{k,k'}\gamma_k\gamma_{k'}(O_k - < O_k >)(O_{k'} - < O_{k'} >)\Bigg]|\psi_\alpha\rangle. \quad (1.45)$$

It should be noticed that this expansion is exact with $\beta = 1$, *only if* the second order derivatives of the wave function with respect the variational parameters can be factorized. Of course, this is not always the case, in particular for an highly correlated functional form. However, for a Jastrow factor of the form $J = \exp[\sum_k \alpha_k(O_k - < O_k >)]$, where the operators O_k can be either the charge-charge or spin-spin correlations, the expansion in Eq. 1.45 is *exact* for $\beta = 1$. Instead, for the determinant it is equivalent to assume that its expansion for small change of parameters can be written as $|\psi_{\alpha+\gamma}\rangle \propto \exp[\sum_k \gamma_k(O_k - < O_k >)]|\psi_\alpha\rangle$, which is clearly an incomplete second order expansion, since it is nothing but a first order cumulant expansion. Nevertheless, it is not necessary to have a correct second order expression, as close to the convergence the second order terms are irrelevant, and far from the minimum the linear system based on the expansion of Eq. 1.45 is accurate enough to capture the different energy scales of the system. Moreover, the expression in Eq. 1.45 does not need the evaluation of the second order derivatives of the wave function, but only the derivatives of the local energy, with a gain in computational time and in the simplicity of the implementation of the scheme. Finally, we consider a more general form, with β another parameter, because its value can be used to improve the efficiency of the minimization scheme with the VMC sampling.

By substituting the wave function in Eq. 1.45 in the expression for the expectation value of the Hamiltonian $E_\alpha = \langle\psi_\alpha|H|\psi_\alpha\rangle/\langle\psi_\alpha|\psi_\alpha\rangle$, we obtain the variation of the energy as a function of the change of the variational parameters up to the second order:

$$\Delta E = -\sum_k \gamma_k f_k + \frac{1}{2}\sum_{k,k'}\gamma_k\gamma_{k'}\left[S_h + (1+\beta)G\right]^{k,k'} \quad (1.46)$$

where:

$$S_h^{k,k'} = < [O_k, [H, O_{k'}]] > \quad (1.47)$$

$$G^{k,k'} = 2 < (H - E_\alpha)(O_k - < O_k >)(O_{k'} - < O_{k'} >) >$$

$$f_k = -\partial_{\alpha_k} E_\alpha = -2 < (H - E_\alpha)O_k > . \quad (1.48)$$

In the above equations we have used the hermitian character of all the operators involved, implying for instance that $< O_k H >=< HO_k >$; f_k indicate as usual the forces acting on the variational parameters and vanishing at the minimum energy condition, $S_h^{k,k'}$ represent the excitation matrix elements corresponding to the operators O_k, $G^{k,k'}$ take into account the remaining contributions appearing when the WF is not exact ($< H - E_\alpha > \neq 0$), whereas the square brackets indicate the commutator.

By imposing the equilibrium condition to Eq. 1.46, i.e. $\partial \Delta E / \partial \gamma_k = 0$, we end up with the linear system:

$$B\gamma = \mathbf{f}, \tag{1.49}$$

where

$$B = S_h + (1+\beta)G. \tag{1.50}$$

The solution of 1.49 provides the change of the variational parameters which hopefully will lower the energy. In practice, it is not guaranteed that the step will be downhill, and the reason is twofold. First, if the matrix B is not positive definite, the approximate energy shape is not bounded from below and the move turns out to be unreliable. Second, the resulting displacement can be too large and raise the energy, also in the case in which B is positive definite. In order to overcome this problem, we used a solution similar to that adopted in the SR approach. We require that the "distance" between the old and the new parameter sets is below a certain threshold r, which is defined as the maximum wave function change $\Delta \mathrm{WF} = (|\psi_{\alpha+\gamma}\rangle - |\psi_\alpha\rangle)/|\psi_\alpha|$, allowed with a single optimization step. Since $|\Delta \mathrm{WF}|^2$ can be written in terms of the positive definite covariance matrix $\bar{s}$ (see Eq. 1.34), the constraint is

$$\sum_{k,k'} \gamma_k \gamma_{k'} \bar{s}_{k,k'} \leq r^2. \tag{1.51}$$

In practice, if B is not positive definite or if the condition in Eq. 1.51 is not fulfilled, the matrix B is changed, $B \rightarrow B + \mu \bar{s}$, with μ a Lagrange multiplier which yields a well defined "Hessian" matrix B and a change of the variational wave function exactly equal to r. Notice that if μ tends to infinity, the SRH scheme reduces to the standard SR optimization, since the true "Hessian" contribution to B will become negligible.

As already pointed out by Umrigar and Filippi in their derivation of the Hessian based optimization[33], within a QMC framework it is crucial to express all quantities involved in the determination of the linear system (Eq. 1.49) in the form $A(x) - < A(x) >$, i.e. by means of fluctuation of the estimator A, since in this way the noise is much more reduced, and therefore the result much more precise. Sorella[19] found that all matrices useful to compute the "Hessian" B can be rewritten as product of fluctuations:

$$f_k = -2 < \delta e_L(x) \delta O_k(x) > \tag{1.52}$$

$$S_h^{k,k'} = < \delta \partial_{\alpha_k} e_L(x) \delta O_{k'}(x) > + (k \leftrightarrow k') \tag{1.53}$$

$$\bar{s}_{k,k'} = < \delta O_k(x) \delta O_{k'}(x) > \tag{1.54}$$

$$G^{k,k'} = 2 < \delta e_L(x) \delta O_k(x) \delta O_{k'}(x) > \tag{1.55}$$

where $e_L(x) = \frac{\langle \psi_\alpha | H | x \rangle}{\langle \psi_\alpha | x \rangle}$ is the local energy. Now, notice that the operators O_k and the local energy are extensive and therefore scale linearly with the number of electrons N_e, whereas their fluctuations δO_k and δe_L scale with $\sqrt{N_e}$. Thus it is clear from the above expressions that the matrix G has a zero signal to noise ratio, since its magnitude is of order N_e but it is affected by fluctuations of order $^3\sqrt{N_e}$. Therefore its value can not be determined within a given accuracy if the system is large enough, but fortunately its contribution is negligible for a very good trial wave function. All the other quantities are well defined, and one can get rid of G by setting $\beta = -1$.

Another important problem, present in all optimization techniques performed within a QMC framework, is the presence of the nodes, which occurs whenever the parameters to be optimized are present in the antisymmetric part of a fermionic wave function. Indeed the vanishing of wave function on the nodes can yield quantities with an ill defined variance, or spoil directly their average values. In our case, both matrices S and G suffer of this last and more serious problem, but if $\beta = 0$ the divergences coming from the two matrices cancel mutually out, since they have the same behaviour close to the nodes (note that $\partial_{\alpha_k} e_L(x) \simeq -O_k(x) e_L(x)$). Therefore, if both the Jastrow and the antisymmetric part need to be optimized, the choice $\beta = 0$ performs much better. On the other hand, if only the Jastrow parameters are optimized, it is more convenient to set $\beta = -1$. The value $\beta = 1$, corresponding to the true "Newton method" (exact evaluation of the

Hessian matrix) [2] has been found always less efficient than the other cases. In particular, if only the Jastrow parameters are involved, the SRH is 50 times more efficient than SR, and 10 times with respect to the Newton method. In the case of determinantal parameters the gain is lower, ranging from a factor 2 to 10 with respect to the standard SR, since the cut-off r has to be kept smaller, in order to dump the fluctuations coming from the noisy estimators close to the nodes. In all cases, the Hessian contribution is crucial to reduce the "slowing down" for the minimization of a complex energy landscape.

Last but not least, we highlight that also within the SRH scheme it is possible to perform structural optimizations of a molecular system. The trick is to replace the S matrix with the simpler $\bar{s}$ matrix for those elements involving only the ionic derivatives and to exploit the symmetry of S for the mixed ionic - non ionic elements, by using $S_h^{k,k'} = S_h^{k',k} = 2 < \delta\partial_{\alpha_k} e_L(x)\delta O_{k'}(x) >$, which do not involve local energy derivatives of ionic positions. This is the so called "mixed method", which allows to include in the SRH framework either those parameters which are present also in the Hamiltonian, like the ionic positions, or parameters for which the computation of the derivatives of the local energy is particularly cumbersome.

1.5 Diffusion Monte Carlo

The diffusion Monte Carlo method[15, 44] is a stochastic approach which "optimizes" the wave function in an automatic manner, by filtering out the high energy components of the initial trial function and by projecting it onto the lowest possible state non orthogonal to the starting guess and compatible with the given boundary conditions.

The stochastic projection is realized following the dynamics driven by the imaginary time Schroedinger equation:

$$-\partial_t \Phi(\mathbf{R}, t) = (H - E_T)\Phi(\mathbf{R}, t), \tag{1.56}$$

where E_T is a proper shift in the energy scale, H and $\mathbf{R}$ are as usual the Hamiltonian and the N particle configuration in a D dimensional continuous space. The

[2] However, this is valid only for Jastrow parameters appearing linearly in the exponent. In all the other cases the Hessian matrix B defined in our approach is still approximated, even with $\beta = 1$.

derivation of the algorithm can be based either on an integral Green function formalism or on a path-integral approach. We prefer to follow the former route, and write Eq. 1.56 in the following integral form:

$$\Phi(\mathbf{R}, t+\tau) = \int d\mathbf{R}' G(\mathbf{R}', \mathbf{R}, \tau)\Phi(\mathbf{R}', t), \tag{1.57}$$

where G is a Green function, i.e. the formal solution of the equation:

$$\partial_t G(\mathbf{R}', \mathbf{R}, t) = (H - E_T)G(\mathbf{R}', \mathbf{R}, t), \tag{1.58}$$

with the initial condition $G(\mathbf{R}', \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$. It can be easily shown that it is possible to write the Green function in terms of an exponential form involving the Hamiltonian operator:

$$G(\mathbf{R}', \mathbf{R}, t) = \langle \mathbf{R}' | e^{-tH} | \mathbf{R} \rangle, \tag{1.59}$$

and by using the spectral decomposition, G can be expressed as function of the eigenstates and eigenvalues of the Hamiltonian:

$$G(\mathbf{R}', \mathbf{R}, t) = \sum_n \phi_n(\mathbf{R}') \exp\left[-t(E_n - E_T)\right] \phi_n(\mathbf{R}). \tag{1.60}$$

From Eq. 1.60, it is apparent that during the evolution, the higher energy components will decay exponentially faster, and by setting $E_T = E_0$ the ground state will be steadily reached in the asymptotic regime. Indeed, after replacing Eq. 1.60 in Eq. 1.57, one obtains in the limit $\tau \to \infty$:

$$\lim_{\tau\to\infty} \Phi(\mathbf{R}, t+\tau) = \langle \phi_0 | \Psi_T \rangle \exp\left[-\tau(E_0 - E_T)\right] \phi_0(\mathbf{R}) + O\left(e^{-\tau(E_1 - E_0)}\right), \tag{1.61}$$

where we have supposed that the ground state of the system is unique. Notice that the prefactor depends also on the overlap of the trial wave function ($\Phi(\mathbf{R}, 0) = \Psi_T$) with the GS, and the speed of convergence depends also on the energy gap between the GS and the first excited states.

So far, we have analyzed the theoretical ground of the DMC algorithm. For practical purposes, we need to find out an explicit form for the Green function, in order to implement it in a numerical procedure. This is accomplished by using the Trotter-Suzuki approximation, as the kinetic operator K and the potential operator V do not commute each other:

$$e^{-\tau(K+V)} \approx e^{-\tau K} e^{-\tau V} + O(\tau^2). \tag{1.62}$$

In this way it is possible to write a functional form of Green function valid up to the second order in the time step τ:

$$\begin{aligned} G(\mathbf{R}', \mathbf{R}, \tau) &= (2\pi\tau)^{-3N/2} \exp\left[\frac{-(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right] \exp\left[-\tau(V(\mathbf{R}') - E_T)\right] \\ &= G_{\text{diff}}\ G_{\text{rate}}. \end{aligned} \tag{1.63}$$

In Eq. 1.63, the first factor is related to a diffusion process, since its Green function is exactly a Gaussian, the second term is the weight factor related to a rate process describing death and birth events. G_{diff} is the exact Green function for the Fokker Plank equation without drift and it comes from the kinetic term of the imaginary time Hamiltonian, G_{rate} is the exact Green function for the rate equation, involving an exponential solution, and it comes from the potential part of the Hamiltonian. Since the separation of these two processes is a consequence of the Trotter approximation, which is valid only for small time steps, the evolution will be realized iteratively, in such a way that the asymptotic convergence will be reached after many time slices.

Now, let us suppose that the initial trial state is nodeless (we will treat particle statistics later on). We can interpret it as a probability distribution, like in other QMC schemes, and represent it using a single or a set of walkers, distributed according to Ψ_T. Since G_{diff} is normalized, the diffusion Green function will play the role of a transition matrix probability, while G_{rate} will be simply a factor accumulated step by step in the weights. Remember that the general definition of the walker, given in Section 1.2, includes its configuration (a DN dimensional vector) and its weight (a scalar), which in this case is not trivial ($w \neq 1$). After the walkers have been thermalized by sampling the initial distribution Ψ_T, we start the DMC dynamics. A single step evolution is given by:

$$\mathbf{R}_i = \mathbf{R}'_i + \chi \qquad \text{(position updating)} \tag{1.64}$$

$$w_i(t+\tau) = w_i(t) G_{\text{rate}}(\mathbf{R}'_i, \tau) \quad \text{(weight updating)}, \tag{1.65}$$

where χ is a normally distributed DN dimensional vector with variance τ and zero mean and i is the walker index. The weights have an exponentially fast evolution, therefore they need to be renormalized from time to time in order to avoid uncontrolled fluctuations. The walker renormalization is called *branching*. Some walkers are replicated, some others are killed, according to the value their

weights have assumed from the previous branching. The new generation will be distributed according to the discrete distribution $\frac{w_i}{\sum_j w_j}$, and the new starting weights will be set to 1. Different schemes for branching have been proposed. We follow the formulation in which the total number of walkers is kept fixed to M, since it has shown to be efficient, less biased, and easier to parallelize (see Sec. 1.7) than the scheme where the population is free to fluctuate. A crucial problem which can definitely spoil the simulation is the divergence of the potential V, since it is included in the exponent of the rate Green function,

$$G_{\text{rate}}(\mathbf{R}_i, \tau) = \exp\left[-\tau(V(\mathbf{R}) - E_T)\right], \tag{1.66}$$

and therefore in the weights. If the potential is not bounded, as usual in electronic structure calculations, the simulation will be unstable. To overcome this problem, the importance sampling Green function must be introduced. In the following subsection we present the importance sampling, which allows a feasible implementation of the DMC algorithm, in the subsection 1.5.2 we introduce the fixed node approximation in order to deal with fermionic systems, while in the last subsection we show how to compute expectation values within the DMC framework.

1.5.1 Importance sampling

The importance sampling is based on the idea that the evolution of the walkers during the Markov process can be guided by a trial wave function, in order to sample more effectively the entire configuration space[45]. In particular, the most significant regions for the stochastic evaluation will be visited more frequently than the others.

In the DMC approach, the importance sampling is built by changing the probability distribution that will be sampled during the diffusion. Instead of sampling the distribution Φ (see Eq. 1.57), the process will sample the "mixed" distribution $f(\mathbf{R}, t) = \Phi(\mathbf{R}, t)\Psi_T(\mathbf{R})$, where Ψ_T is the usual trial wave function, in this case called also "guidance" wave function. The initial condition is $f(\mathbf{R}, 0) = \Psi_T^2(\mathbf{R})$, and the starting walkers will be distributed according to $\Psi_T^2(\mathbf{R})$, as in a simple VMC sampling. While Φ fulfills Eq. 1.56, f will fulfill the following one:

$$-\partial_t f(\mathbf{R}, t) = -\frac{1}{2}\nabla^2 f(\mathbf{R}, t) + \nabla \cdot [\mathbf{v}_{\text{diff}} f(\mathbf{R}, t)] + [E_L(\mathbf{R}) - E_T]\, f(\mathbf{R}, t), \tag{1.67}$$

where ∇ is the DN dimensional gradient operator, $\mathbf{v}_{\text{diff}}$ is the DN-dimensional *drift velocity* (also called "quantum force") defined by:

$$\mathbf{v}_{\text{diff}}(\mathbf{R}) = \nabla \ln |\Psi_T(\mathbf{R})|, \tag{1.68}$$

and $E_L(\mathbf{R}) = H\Psi_T(\mathbf{R})/\Psi_T(\mathbf{R})$ is the local energy calculated on the trial wave function. The Green function relative to Eq. 1.67 is slightly different from the Green function for Eq. 1.56. Indeed the short time approximation for the importance sampling Green function reads:

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = \Psi_T(\mathbf{R}) G(\mathbf{R}', \mathbf{R}, \tau) \Psi_T(\mathbf{R}')^{-1} \approx \tilde{G}_{\text{diff}} \ \tilde{G}_{\text{rate}}, \tag{1.69}$$

where

$$\tilde{G}_{\text{diff}}(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{-3N/2} \exp\left[\frac{-\left(\mathbf{R}' - \mathbf{R} - \tau \mathbf{v}_D(\mathbf{R}')\right)^2}{2\tau}\right] \tag{1.70}$$

$$\tilde{G}_{\text{rate}}(\mathbf{R}', \tau) = \exp\left[-\tau \left(E_L(\mathbf{R}') - E_T\right)\right]. \tag{1.71}$$

The most important consequence of importance sampling is the change in the rate Green function (Eq. 1.71). Indeed, in the exponent of the weighting factor the local energy appears in the place of the bare interaction potential. If Ψ_T satisfies the *cusp conditions* (see Chapter 2), the local energy does not diverge at the coalescence points (where two particles overlap) as the coulomb potential does, and therefore the weights are much better behaved. Notice that the local energy diverges on the nodes of the trial function (where $\Psi_T = 0$), but the importance sampling guarantees that those regions will be never sampled since there the wave function is vanishing (see Subsection 1.5.2 for a discussion about the difference between the nodal surface and the coalescence regions in a fermionic wave function). From Eq. 1.70 it is apparent that the diffusion Green function with importance sampling contains not only a diffusion move but also a drift component, as a consequence of the modification of the imaginary time equation for the mixed distribution f. The drift velocity close to the nodes is orthogonal to the nodal surface and points outwards. Thus the quantum force drives the particles away from the vanishing regions of the trial wave function, by enforcing the importance sampling itself. Notice that not only the local energy but also the drift velocity is divergent on the nodes, and since the Green function of Eq. 1.69 is approximated,

this results in a worse behaviour of the diffusion process close to the nodal surface, with respect to the ideal exact evolution.

The short time approximation based on the first order Trotter expansion of the Green function yields the so called "time step error", which affects the DMC simulations. In order to obtain a result free of this error one needs to perform simulations at different time steps and then to extrapolate the results to the limit $\tau \to 0$. The behaviour of this extrapolation depends strongly on the Green function used to define the diffusion process and on the trial wave function employed to guide the sampling. An accurate study of the time step error has been done by Umrigar and coworkers [44]. There are mainly two issues which have to be taken into account in order to reduce the time step bias and to improve the importance sampling Green function.

- First, the approximated Green function of Eq. 1.69 does not fulfill the detailed balance condition satisfied by the exact importance sampling Green function:

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau)\Psi_T(\mathbf{R}')^2 = \tilde{G}(\mathbf{R}, \mathbf{R}', \tau)\Psi_T(\mathbf{R})^2. \tag{1.72}$$

  Moreover, the limit of perfect importance sampling is not reached, i.e. even if $\Psi_T = \Psi_0$, the eigenstate and eigenvalue sampled by the approximated evolution are still affected by the time step error. In order to restore the detailed balance condition and the limit of perfect importance sampling, a rejection step has to be incorporated into the propagation governed by the approximate Green function. Following this scheme, once the move has been proposed:

$$\mathbf{R}^{\text{try}} = \mathbf{R}' + \chi + \tau \mathbf{v}_D(\mathbf{R}'), \tag{1.73}$$

  where χ is a D-dimensional vector of normally distributed numbers with variance τ and zero mean, the acceptance probability is computed:

$$p_{\text{accept}}(\mathbf{R}' \to \mathbf{R}^{\text{try}}) = \min\left[1, \frac{\tilde{G}_{\text{diff}}(\mathbf{R}^{\text{try}}, \mathbf{R}', \tau)\Psi_T(\mathbf{R}^{\text{try}})^2}{\tilde{G}_{\text{diff}}(\mathbf{R}', \mathbf{R}^{\text{try}}, \tau)\Psi_T(\mathbf{R}')^2}\right]. \tag{1.74}$$

  A random number χ_{fhit} is drawn ($0 \leq \chi_{\text{fhit}} < 1$ with a constant distribution) and compared with p_{accept}. If $\chi_{\text{fhit}} \leq p_{\text{accept}}$ the move will be accepted, otherwise rejected, as in the generalized Metropolis algorithm (see Section 1.2). A further improvement is obtained by replacing the time step in the rate

Green function of Eq. 1.71 with an effective τ_{eff}, which takes into account the change in the diffusion constant due to the rejection procedure [15, 44].

- Second, in order to cure the singularity of the drift velocity present in the Green function, Umrigar *et al.*[44] proposed a correction, which reduces the drift velocity to its usual definition far from the nodes, but which acts as a cut-off when the walker is close to the nodal surface:

$$\bar{\mathbf{v}}_D = \frac{-1 + \sqrt{1 + 2v_D^2\tau}}{v_D^2\tau}\mathbf{v}_D. \tag{1.75}$$

  Moreover, also the local energy in the rate Green function of Eq. 1.71 is cut-off by the factor $\bar{\mathbf{v}}/\mathbf{v}$, which compensates the $1/R$ divergence on the nodes. Finally, the approximate diffusion Green function close to the nuclei is replaced by an exponential form in order to preserve the atomic cusp conditions.

Using an algorithm with small time-step error allows one to perform simulations with a greater time step without losing accuracy. This also results in a better efficiency of the simulations, since the diffusion constant is proportional to $\sqrt{\tau}$ and the autocorrelation time is reduced for a bigger time step. Therefore the statistical error on the measured expectation values is lowered, compared to the error obtained by a simulation with the same number of time slices but with a smaller time step.

1.5.2 Fixed node approximation

From the Fermi statistics it follows that the ground state of a many-body system will have nodes if the number of particles $N > 2$. Indeed the antisymmetry yields regions of the DN-dimensional configuration space on which the wave function is vanishing and across which it changes sign. The subspace where $\Psi(\mathbf{R}) = 0$ is called *nodal surface*, since its dimension is $DN - 1$. Notice that in general the nodes of the ground state wave function are unknown *a priori*, except for the one dimensional case, where the nodes are completely determined by the coalescence conditions between two particles. All the points defined by the coalescence of two fermions belong to a $DN - D$ dimensional subspace, but if $D = 1$ this dimensionality coincides with that of the nodal surface. Ceperley showed that

these two different subspaces coincide exactly in $1D$, and the nodes are therefore completely fixed by the coalescence conditions[46], which can be easily written and handled. Usually in this thesis the systems under consideration will be three dimensional, for which this argument is not valid.

The fact that the fermionic wave function changes sign implies that in general $f(\mathbf{R}, t)$ is not necessarily non negative everywhere and thus it can not be interpreted as a probability distribution. Moreover the importance sampling Green function (1.69) changes sign if the Gaussian move crosses a node, and it can not be viewed as a transition probability. Everything can be restored if one lets the weights to carry the sign, but even in that case the DMC algorithm will try to project the wave function onto the *lowest* energy state, which is *bosonic*. Therefore the fermionic component will vanish and the signal to noise ratio of the computed quantities will go to zero, as a consequence of the diffusion from positive to negative nodal pockets and vice versa and the loss of a dominant sign in the weights. This is the well known "sign problem", which affects and spoils all naive DMC simulations for fermions.

A solution for this problem is the *fixed node* approximation (FNA) [13–15], which requires that the lowest state Φ will have the *same* signs as the trial wave function Ψ_T. This is equivalent to change the boundary conditions of the Hamiltonian, that can be realized by rejecting the moves which cross the nodal surface of Ψ_T. In this way the diffusion is constrained within the starting nodal pocket, and the algorithm will give the lowest energy and the corresponding eigenstate of the Hamiltonian with the fixed node boundary conditions. This represents an important limitation, as it is difficult to obtain the exact answer even if the wave function is extremely accurate. On the other hand, the FNA has marked the success of the DMC method, since it has allowed the application of this framework to a large variety of systems, and it is characterized by a list of good properties:

- All the nodal pockets are equivalent for the ground state of the fixed node Hamiltonian, since all of them can be connected through a permutation $\mathcal{P}$ between two particles. Therefore the initial position of the walkers does not matter for the final result (this property is also known as the *tiling theorem* [46]) and the fixed node constraint can be easily implemented in the algorithm by a rejection step.

- The fixed node energy, i.e. the lowest energy with the given fixed node boundary conditions, is a variational upper bound of the true ground state energy of the system. This is easily proved. Suppose that $\Psi_0^m(\mathbf{R})$ is the ground state of the $m-th$ nodal pocket with energy E_0^m. By applying the permutations $\mathcal{P}$ on the particles, one can construct an antisymmetric wave function,

$$\bar{\Psi_0} = \frac{1}{N_\mathcal{P}} \sum_\mathcal{P} (-1)^{\chi_\mathcal{P}} \Psi_0^m(\mathcal{P}\mathbf{R}), \tag{1.76}$$

  where $N_\mathcal{P}$ is the total number of permutations and $\chi_\mathcal{P}$ is the parity of the permutation $\mathcal{P}$. For the tiling theorem, the many-body wave function in Eq. 1.76 is well defined on the *entire* configuration space, and its energy expectation value is exactly E_0^m. For the variational principle it follows that $E_{true} \leq E_0^m$, where E_{true} is the ground state energy of the true Hamiltonian (without the FN boundary conditions).

- If the nodes of the trial wave function are exact, the FNA is exact. Therefore if the exact nodal structure is unknown, like in almost all cases, the optimization of Ψ_T is crucial to get an accurate result. As a consequence of the exact limit of the FNA, the error in the energy are normally second order in the errors in the nodal surface.

Notice that the importance sampling described in subsection 1.5.1 is in accordance with the fixed node requirement, since the quantum force pushes the walkers away form the nodal surface, by producing the same effect of the constraint of rigid walls. Indeed, in principle for small enough time steps, the fixed node approximation does not need to be enforced, since just the quantum force and the acceptance-rejection procedure described in Eq. 1.74 would be enough to avoid that the walkers cross the nodes. In practice, the usual time steps are not so small and the nodal crossing rejection is necessary.

If the Hamiltonian contains an imaginary part, and the system does not have the time-reversal symmetry, the wave function is complex. In this case the FNA, which is valid only for real ground states, cannot be applied. Its extension to deal with complex wave functions is the "fixed phase approximation". In this thesis only real states will be considered, therefore it is not worth getting into details on the fixed phase approach.

1.5.3 Averages

As we have seen, the DMC algorithm with importance sampling will sample the asymptotic mixed distribution $f = \Phi\Psi_T$. Therefore, the expectation value of an operator $\mathcal{O}$ computed within this framework will be a *mixed estimate*:

$$\begin{aligned} \langle \mathcal{O} \rangle_{\text{mix}} &= \frac{\langle \Phi | \mathcal{O} | \Psi_T \rangle}{\langle \Phi | \Psi_T \rangle} = \frac{\int d\mathbf{R} \Psi(\mathbf{R}) \mathcal{O} \Psi_T(\mathbf{R})}{\int d\mathbf{R} \Phi(\mathbf{R}) \Psi_T(\mathbf{R})} \\ &= \frac{\int d\mathbf{R} f(\mathbf{R}) \mathcal{O}_L(\mathbf{R})}{\int d\mathbf{R} f(\mathbf{R})} \simeq \frac{\sum_{i=1}^{M} w(\mathbf{R}_i) \mathcal{O}_L(\mathbf{R}_i)}{\sum_{i=1}^{M} w(\mathbf{R}_i)} \end{aligned} \tag{1.77}$$

Notice the presence of the weights in the DMC evaluation of the expectation value $\langle \mathcal{O} \rangle_{\text{mix}}$, which take into account the different energy contribution for each walker $\mathbf{R}_i$. For the Hamiltonian and for operators that commute with it, the mixed estimate coincides with the *pure estimate*:

$$\langle \mathcal{O} \rangle = \frac{\langle \Phi | \mathcal{O} | \Phi \rangle}{\langle \Phi | \Phi \rangle}, \tag{1.78}$$

where Ψ is present in both sides of the bracket. For the other operators $\langle \mathcal{O} \rangle_{\text{mix}}$ is different from the pure fixed node ground state expectation value. However it is possible to show that a reasonable approximation to the pure average is given by the extrapolated estimate:

$$\langle \mathcal{O} \rangle = 2\langle \mathcal{O} \rangle_{\text{mix}} - \langle \mathcal{O} \rangle_{\text{vmc}} + O(|\Phi - \Psi_T|^2), \tag{1.79}$$

which is equal to the pure average up to the second order in the difference between the fixed node ground state and the trial wave function. Another possibility is to resort to *forward walking*, a technique almost never used in this thesis, and therefore not explained here. The interested reader is referred to see Ref. [47]. However, for a good trial wave function the extrapolated estimate of Eq. 1.79 will be enough to give an accurate result.

1.6 Lattice Green function Monte Carlo

The lattice Green function Monte Carlo (GFMC) method[48] relies on the same principles as the DMC framework. It is a numerical technique based on an iterative application of a projection operator, $G = \Lambda - H$, where Λ is an energy shift

and H is the Hamiltonian. In this context G is called "Green function" and selects the low energy eigenstates of the Hamiltonian which have non zero overlap with the initial starting state. After a sufficient number of iterations and for an appropriate choice of the constant Λ, only the lowest state will survive, and the algorithm will sample a steady distribution. Indeed if we expand the initial wave function $|\Psi_T\rangle$ in the basis of eigenstates of H,

$$|\Psi_T\rangle = \sum_i a_i |\phi_i\rangle, \tag{1.80}$$

where $a_i = \langle \Psi_T | \phi_i \rangle$, then we can obtain the spectral evolution of the state, which after n steps will read:

$$G^n |\Psi_T\rangle = (\Lambda - E_0)^n \left[a_0 |\phi_0\rangle + \sum_{n \neq 0} \left(\frac{\Lambda - E_n}{\Lambda - E_0} \right)^n a_n |\phi_n\rangle \right]. \tag{1.81}$$

For Λ sufficiently large, it is apparent that the prefactor in front of the higher states will vanish exponentially in the number of steps. Therefore, if $a_0 \neq 0$, the wave function obtained by G^n will converge to the ground state of H, otherwise to the lowest state with non zero overlap ($a_n \neq 0$).

This approach is related to the *power method*, but the iterative application of the Green function G is realized on a statistical ground. Let $|x\rangle$ be an element of the basis set chosen. $|x\rangle$ can be a spin configuration of the lattice or a position of the particles in the lattice space. A single iteration step is given by the following equation:

$$\Phi_{n+1}(x') = \sum_x G_{x,x'} \Phi_n(x), \tag{1.82}$$

which is the analogous of Eq. 1.57 on a lattice. In this case, the integration is substituted by a sum, since the number of states is *discrete*. $G_{x,x'} = \Lambda \delta_{x,x'} - H_{x,x'}$ is the matrix elements of G in the chosen basis. If the latter recursive equation is evaluated in an exact way, one obtains, after few iterations, transitions to a large number of states. The computation becomes cumbersome and only small systems could be studied. The solution is to statistically sample the evolution of Eq. 1.82, as it has been done for the DMC framework. In order to have a well defined transition probability $p_{x,x'}$, we normalize the matrix G and we take its modulus:

$$p_{x,x'} = \frac{G_{x,x'}}{b_x \; s_{x,x'}}, \tag{1.83}$$

where $b_x = \sum_{x'} G_{x,x'}/s_{x,x'}$ and $s_{x,x'}$ is the sign of the matrix element $G_{x,x'}$. As usual, the stochastic implementation of Eq. 1.82 is based on a Markov chain, traced by walkers. The dynamics of the diffusion process on the lattice is governed by $p_{x,x'}$ and the prefactor included in the Green function, which arises from its normalization, is taken into account by the weight of walker. Therefore, the analogous of Eqs. 1.64 and 1.65, which are the position and weight updates in a DMC procedure, becomes:

$$x' = x + \delta_m \quad \text{(position updating)} \tag{1.84}$$

$$w_i(n+1) = w_i(n) b_x \, s_{x,x'} \quad \text{(weight updating)}. \tag{1.85}$$

Notice that in the lattice GFMC method, the new position x' is selected through an *heat bath* algorithm, in which all the possible next configurations (x included) are taken into account in the transition probability matrix $p_{x,x'}$, and the selection is based on their relative probability. Therefore the evolution is statistically *exact*, since the Green function is exactly sampled and there is no Trotter-like approximation, which instead affects the dynamics in the DMC algorithm. However in the heat bath approach one needs to compute in advance all the possible states connected through the Hamiltonian to the present position x. A lattice Hamiltonian can have in principle an *infinite* and *discrete* connectivity, but it is not usually the case. Indeed the off diagonal matrix elements come either from kinetic terms or non local potential elements and they are usually *finite*. It is clear that the efficiency of the method relies on the extension of the connectivity. The bigger is the number of off diagonal matrix elements (called also "hopping terms"), the lower is the efficiency.

A problem in common with the DMC method is the fluctuation of the weights, which can grow or decrease exponentially, thus causing a divergence in the variance of the computed averages. As already mentioned, in the case the weight is always positive (no sign problem) the solution is the use of many walkers together with the reconfiguration process (branching) that introduces a small but controlled bias in the simulation, due to the finiteness of the walker population. Nevertheless, the renormalization of the walker distribution limits the fluctuation of the overall weight, and the weighted averages (similar to that in Eq. 1.77) can be computed more effectively. In the presence of the sign problem, i.e. for a fermionic lattice system, it is possible to collect negative sign contribution to w, whenever a walker

hops from two configurations where G (or the wave function) changes sign. In practice, as we have already pointed out, the average sign goes exponentially to zero as the number of iteration n is increased:

$$\langle s_n \rangle = \frac{\sum_i w_i}{\sum_i |w_i|} \approx \left(\frac{\Lambda - E_0}{\Lambda - E_0^{bos}} \right)^n , \tag{1.86}$$

where E_0^{bos} is the bosonic ground state energy, which is obviously below E_0, and the weighted averages are affected by the sign problem instability. Also for a lattice Hamiltonian, this problem is overcome by the analogous of fixed node approximation on the lattice, as it will be explained in Subsection 1.6.2.

In order to have non negative diagonal matrix elements $\Lambda - H_{x,x}$, it is necessary to choose the constant Λ large enough. This requirement often determines a very large constant shift, which increases with larger sizes and is not known *a priori*. Moreover, if Λ is large, the probability to remain in the same configuration,

$$p(x) = \frac{\Lambda - H_{x,x}}{\Lambda - \sum_{x'} H_{x,x'}}, \tag{1.87}$$

becomes very close to one, and slows down the efficiency of the algorithm, since the walker is "stuck" in the same configuration for a long time, while at each iteration one computes the same quantities. Following Ref. [48], the problem of working with large Λ can be easily solved with a great increase of efficiency, by performing the limit $\Lambda \to \infty$ of the Green function in an exact way. By the way, if the potential $V(x)$ is not bounded, it is *necessary* to perform such a limit. Now, let us define the infinitesimal time step $\delta\tau = 1/\Lambda$. If the probability $p(x)$ goes to one, the probability $q(x)$ to leave the configuration x goes to zero. Indeed,

$$q(x) = \frac{\sum_{x'(\neq x)} G_{x,x'}}{\Lambda + \sum_{x'} G_{x,x'}} \approx \delta\tau \sum_{x'(\neq x)} G_{x,x'} + O(\delta\tau^2), \tag{1.88}$$

which has been obtained by picking out $1/\Lambda$ in front of the expression. Given the Green function G and the position x, we are going to determine the total time τ during which the walker stays in the configuration x. In particular $\tau = k_{q(x)}\, \delta\tau$, where $k_{q(x)}$ is the number of iterations (or time slices) before the first acceptance of a new configuration x'. The probability to remain k times in x and to leave at the $k+1$-th iteration, is given by $t(k) = (1-q)^k q$. It is straightforward to verify

that the value $k_{q(x)}$, which is distributed according to $t(k)$, is

$$k_{q(x)} \approx -\frac{\log(1-\chi)}{\delta\tau \sum_{x'(\neq x)} G_{x,x'}} + O(1), \tag{1.89}$$

where χ is a uniform deviate in the interval $[0,1)$. Therefore the time τ is given by the formula:

$$\tau = -\frac{\log(1-\chi)}{\sum_{x'(\neq x)} G_{x,x'}} + O(\delta\tau), \tag{1.90}$$

which is statistically *exact* in the limit $\Lambda \to \infty$. The advantage of using the above Equation is that, at each configuration x, only one random number is necessary to take altogether into account a very large sequence of power iterations, with a clear reduction of the computational time. Notice also that this technique allows to rewrite the Green function in an exponential form:

$$G^k = (\Lambda I - H)^k \propto \exp(-\tau H), \tag{1.91}$$

thus the limit is equivalent to apply the exact imaginary time propagator $\exp(-\tau H)$ to the state between two different configurations x and x', and without time step error. In this way, the concept of power method is replaced by a continuous time formulation, which is much more efficient and elegant [49].

1.6.1 Importance sampling

As we have already seen in the DMC method, also in the GFMC algorithm it is extremely convenient to use an importance sampling technique, with the aim to reduce the variance of the energy and to build a more efficient algorithm, by exploiting some information on the ground state wave function. We use a guiding wave function Ψ_T, as close as possible to the true ground state, to improve the sampling of the configuration space, via the transformed Green function:

$$\tilde{G}_{x,x'} = \Psi_T(x') G_{x,x'} / \Psi_T(x). \tag{1.92}$$

The resulting importance sampling Green function $\tilde{G}$ is nonsymmetric, but has the same spectrum of G as for any eigenvector ϕ_i with energy $\Lambda - E_i$, $\Psi_T(x)\phi_i(x)$ is a right eigenvector of $\tilde{G}$ with the same eigenvalue. The Eq. 1.92 is the analogous of Eq. 1.69 for the DMC framework. Similarly, the steady asymptotic distribution of

the walkers will be the mixed state $\Phi_0\Psi_T$, and the evaluation of the ground state energy will be given by the mixed average of the local energy, which in this case, reads:

$$E_L(x) = \sum_{x'} \Psi_T(x') H_{x,x'} / \Psi_T(x). \tag{1.93}$$

Also in the GFMC scheme, the importance sampling leads to the zero variance property. If the guiding wave function approaches an exact eigenstate of H, the method will be free of statistical fluctuations, and by improving the guiding wave function one is able to considerably increase the efficiency of the algorithm.

1.6.2 Fixed node approximation

The lattice fixed node (FN) approximation, introduced like in the DMC framework to overcome the sign problem of fermionic systems, is based on the definition of an effective Hamiltonian, which avoids crossing of regions in the configuration space that yield a sign flip of the Green function $\tilde{G}$. Therefore the random walk is constrained to stay in pockets with fixed sign for $\tilde{G}$, the sign matrix $s_{x,x'}$ is now constant, and the sign problem does not appear.

The FN Green function is designed as follows:

$$\tilde{G}^{FN}_{x,x'} = \Psi_T(x') \left(\Lambda_{x,x'} - H^{FN}_{x,x'} \right) / \Psi_T(x) \tag{1.94}$$

where the effective FN Hamiltonian $H^{FN}_{x,x'}$ is defined as:

$$H^{FN}_{x,x'} = \begin{cases} H_{x,x} + \mathcal{V}_{sf}(x) & \text{if } x = x' \\ H_{x,x'} & \text{if } x \neq x' \text{ and } \Psi_T(x') H_{x,x'} / \Psi_T(x) \leq 0 \\ 0 & \text{if } x \neq x' \text{ and } \Psi_T(x') H_{x,x'} / \Psi_T(x) > 0 \end{cases} . \tag{1.95}$$

Notice the presence in the diagonal part of the so called *sign flip* term [50], which includes all contributions eliminated from the Green function $\tilde{G}$ to satisfy the fixed node constraint:

$$\mathcal{V}_{sf}(x) = \sum_{\substack{\tilde{H}_{x,x'} > 0 \\ \text{and } x \neq x'}} \tilde{H}_{x,x'} > 0, \tag{1.96}$$

where $\tilde{H}_{x,x'} = \Psi_T(x') H_{x,x'} / \Psi_T(x)$. As it will be shown later on for a more general effective Hamiltonian, the sign flip term is crucial to fulfill the upper bound property of the lattice FN ground state energy.

The main difference between the lattice FN approximation and the FN on the continuum is that in the former case not only the sign of the guidance wave function Ψ_T matters to fix the FN constraint, but also the sign of the off diagonal elements of the $\tilde{G}$ matrix must be taken into account. Moreover, the statement that Ψ_T with the exact nodes gives the exact ground state is not valid any more, since not only the nodes are important to guarantee the exact result, but in general also the ratios $\Psi_T(x')/\Psi_T(x)$ must be equal to those of the exact ground state in order to obtain the lowest state of the true Hamiltonian, as it follows from the proof of the upper bound property. Finally, the resulting mixed distribution and thus the FN ground state will have the same nodes as the trial wave function, namely $\Psi_T(x)\Psi_{FN}(x) \geq 0$ for any configuration x: the same property is valid in the continuous case.

A slight generalization of the FN energy can be obtained defining the effective Hamiltonian in such a way that also the negative hopping terms of $\tilde{G}$ are taken into account. Indeed, by reversing the sign of the positive off-diagonal matrix elements of $\tilde{H}$ and multiplying them by a constant $\gamma > 0$, one obtains [49, 51]:

$$H^{\text{eff}}_{x,x'} = \begin{cases} H_{x,x} + (1+\gamma)\mathcal{V}_{sf}(x) & \text{if } x = x' \\ H_{x,x'} & \text{if } x \neq x' \text{ and } \Psi_T(x')H_{x,x'}/\Psi_T(x) \leq 0 \\ -\gamma H_{x,x'} & \text{if } x \neq x' \text{ and } \Psi_T(x')H_{x,x'}/\Psi_T(x) > 0 \end{cases} . \tag{1.97}$$

The standard FN Hamiltonian is recovered with $\gamma = 0$, while the case with $\gamma = -1$ corresponds to the true Hamiltonian, and the nodes are treated exactly ("nodal release"). The sign problem is present for any $\gamma < 0$. In Eq. 1.97, the prefactor of the sign flip term is $(1+\gamma)$, in such a way that the local energy of the effective Hamiltonian H^{eff}_{γ} is equal to the local energy of the true Hamiltonian for any configuration x and the lattice upper bound theorem is still valid, also for this more general H^{eff}_{γ}.

The upper bound property of the ground state energy of the effective FN Hamiltonian has been proved by ten Haaf and coworkers [50]. Here we briefly follow that proof to show that also the effective Hamiltonian H^{eff}_{γ} in Eq. 1.97 satisfies this property. Let us take a variational state Ψ, and compare its energy with respect to H and to H^{eff}_{γ}:

$$\Delta E = \langle \Psi | (H^{\text{eff}}_{\gamma} - H) | \Psi \rangle. \tag{1.98}$$

After some algebra, it is possible to write ΔE as a sum of positive definite terms:

$$\Delta E = (1+\gamma) \sum_{\left(\begin{smallmatrix} \tilde{H}_{x,x'} > 0 \\ \text{and } x \neq x' \end{smallmatrix} \right)} |H_{x,x'}| \left| \Psi(x) \sqrt{\left| \frac{\Psi_T(x')}{\Psi_T(x)} \right|} - sH_{x,x'} \Psi(x') \sqrt{\left| \frac{\Psi_T(x)}{\Psi_T(x')} \right|} \right|^2 \geq 0, \tag{1.99}$$

where $sH_{x,x'}$ is the sign of the matrix element $H_{x,x'}$, and the brackets below the summation label indicate that each pair of configurations x and x', which occurs twice, is taken once, since the other pair is rearranged in the summation. Since ΔE is positive for *any* wave function, it is also positive for the ground state Ψ_γ^{eff} of the effective Hamiltonian H_γ^{eff}. Therefore

$$E_0^{\text{eff}}(\gamma) \geq E_{FN}(\gamma) \geq E_0, \tag{1.100}$$

where $E_0^{\text{eff}}(\gamma)$ is the ground state energy of H_γ^{eff}, $E_{FN}(\gamma)$ is the expectation value of the true Hamiltonian H over Ψ_γ^{eff}, i.e. $E_{FN}(\gamma) = \langle \Psi_\gamma^{\text{eff}} | H | \Psi_\gamma^{\text{eff}} \rangle$, and E_0 is the ground state energy of H. The former upper bound in Eq. 1.100 comes directly from Eq. 1.99, while the latter comes from the variational principle. Notice that $E_0^{\text{eff}}(\gamma)$ equals the mixed average $E_{MA}(\gamma)$ of H over the distribution $\Psi_T \Psi_\gamma^{\text{eff}}$. Indeed the following identities hold:

$$\begin{aligned} E_{MA}(\gamma) &= \langle \Psi_\gamma^{\text{eff}} | H | \Psi_T \rangle / \langle \Psi_\gamma^{\text{eff}} | \Psi_T \rangle / \langle = \langle \Psi_\gamma^{\text{eff}} | H_\gamma^{\text{eff}} | \Psi_T \rangle / \langle \Psi_\gamma^{\text{eff}} | \Psi_T \rangle \\ &= \langle \Psi_\gamma^{\text{eff}} | H_\gamma^{\text{eff}} | \Psi_\gamma^{\text{eff}} \rangle / \langle \Psi_\gamma^{\text{eff}} | \Psi_\gamma^{\text{eff}} \rangle = E_0^{\text{eff}}(\gamma), \end{aligned} \tag{1.101}$$

where the first equality follows from the identity of the local energy for the effective and the true Hamiltonian applied on Ψ_T. Therefore the lattice FN upper bound theorem of Eq. 1.100 can be rewritten as follows:

$$E_{MA}(\gamma) \geq E_{FN}(\gamma) \geq E_0. \tag{1.102}$$

1.6.3 Averages

It is important to highlight that on a lattice the mixed energy estimate E_{MA} is not necessarily equal to the pure estimate E_{FN}, since the FN state Ψ_γ^{eff} is the ground state of the effective Hamiltonian H_γ^{eff}, which *differs* from H. Instead on the continuum the mixed and pure energy estimates are equal, unless an effective Hamiltonian is employed. Indeed, the fixed node constraint on the continuum

does not rely on an effective Hamiltonian, but only modifies the boundary conditions. However, in the presence of non local potential, also on the continuum it is necessary to handle with an effective Hamiltonian, and the difference between the DMC FN mixed and pure averages will play a crucial role, as we will see in Chapter 3.

In general, GFMC and DMC simulations will sample in the asymptotic regime the mixed distribution $\Psi_T\Psi$, and thus the averages over this sampling are mixed (see Subsection 1.5.3). Here, we want to show that using the properties of the lattice effective Hamiltonian in Eq. 1.97 it is possible to estimate with a good accuracy the pure average [52], which usually is not directly accessible. In particular, we can exploit the parameter γ to write the true Hamiltonian H in terms of the effective Hamiltonian H_γ^{eff}. Indeed the following relation holds:

$$H = H_\gamma^{\text{eff}} - (1+\gamma)\frac{d}{d\gamma}H_\gamma^{\text{eff}}. \tag{1.103}$$

Since the final state of the GFMC simulation based on $\tilde{G}_\gamma^{\text{eff}} = \Lambda - \tilde{H}_\gamma^{\text{eff}}$ will be the *ground state* of H_γ^{eff}, it turns out that:

$$\begin{aligned} E_{FN}(\gamma) &= \langle \Psi_\gamma^{\text{eff}} | H_\gamma^{\text{eff}} - (1+\gamma)\frac{d}{d\gamma}H_\gamma^{\text{eff}} | \Psi_\gamma^{\text{eff}} \rangle \\ &= E_0^{\text{eff}}(\gamma) - (1+\gamma)\frac{dE_0^{\text{eff}}(\gamma)}{d\gamma}, \end{aligned} \tag{1.104}$$

where in the latter equality the Hellmann-Feynman theorem has been used. If γ is view as perturbation of the true Hamiltonian, the well known convexity property of $E_0^{\text{eff}}(\gamma)$ follows from the second order perturbative energy expansion:

$$\frac{d^2E_0^{\text{eff}}(\gamma)}{d\gamma^2} \leq 0. \tag{1.105}$$

Therefore the expectation value $E_{FN}(\gamma)$ of the Hamiltonian H on the fixed node state of H_γ^{eff} is a monotonically increasing function of γ, as clearly:

$$\frac{dE_{FN}(\gamma)}{d\gamma} = -(1+\gamma)\frac{d^2E_0^{\text{eff}}(\gamma)}{d\gamma^2} \geq 0. \tag{1.106}$$

From the above inequality, it is apparent that the best variational estimate of the ground energy of H is obtained for $\gamma = 0$.

The extension to finite γ is however convenient, as from Eq. 1.103 it is possible to estimate the pure expectation value $E_{FN}(0)$ from the known mixed averages $E_0^{\text{eff}}(\gamma)$. Indeed, one can compute the derivative with respect to γ exploiting either correlated sampling techniques or finite differences, and finally obtain:

$$E_{FN}(0) \approx E_0^{\text{eff}}(0) - \frac{E_0^{\text{eff}}(\gamma) - E_0^{\text{eff}}(0)}{\gamma}. \tag{1.107}$$

Notice that the estimate in Eq. 1.107 is *variational*, even if the derivative is evaluated in an approximate way, as a consequence of the convexity property of Eq. 1.106. Moreover, since the energy E_0^{eff} as a function of $\gamma(\geq 0)$ is almost linear in all cases a very good estimate can be obtained by evaluating the above relation even for $\gamma = 1$.

1.7 Parallel calculation

Although Quantum Monte Carlo techniques are in general time consuming, they are easy to parallelize, since they are based on averages of collected data, which can be produced by almost independent calculations. In a QMC algorithm the amount of data exchanged among different processors in a parallel machine is relatively small and the inter processor communication takes a marginal fraction of the total computation time. Therefore it is extremely convenient to parallelize a QMC code, especially for electronic structure calculations, which are very heavy from the computational point of view. We chose to parallelize our code exploiting the concept of "distributed calculation"; the walkers are distributed across the nodes of the parallel machine and each processor carries out the various stages of the QMC algorithm with the given number of walkers. Moreover we used the "master-slaves" paradigm, i.e. one processor (the master) orchestrates the whole simulation and controls the work done by the slaves.

In general, the speed up S of a parallel computation scales with the number of processors P following the Amdahl's[53] formula:

$$S = \frac{1}{s + (1-s)/P}, \tag{1.108}$$

where s is the "serial" fraction of the procedure, i.e. the time spent to execute a non-parallel part of the code, communicate among processors and synchronize

the processes. Therefore also s can depend slightly on P. For a Monte Carlo algorithm s is very small. In particular, for the VMC algorithm $s = 0$, since all the walkers are totally independent among each other, and the speed up reaches the perfect linear scaling with P, as one can see from Fig. 1.2. In the case of SRH simulations, we use a multi-walker sampling of the Hessian matrix, but the inversion and the updating of the variational parameters are purely serial, as done only by the master. Nevertheless, the fit for an SRH simulation with 50 parameters and 8 walkers gives $s = 0.008$, thus only the 0.8% of the total single processor time is a residual "serial" fraction. The LRDMC (analogous to DMC and GFMC) case is more complex. In our implementation, the algorithm keeps the total number of walkers N_w fixed. In the parallel version, the walkers are equally shared among the processors, and each process carries out the *same* number of walkers, along the whole simulation. The fairness of the parallelization is therefore maximal, contrary to the scheme in which the branching can change N_w. In the latest case, each process has a fluctuating number of walkers, and the synchronization among the processors can increase the latency time of the simulation. However, even in the case of fixed N_w, the branching is the most expensive part in the parallelization, since sometimes the replicated walkers need to be transferred from a processor to another, by increasing the amount of the communication. In Fig. 1.2, it is clear that is it more convenient to work with a large N_w, since with 128 walkers it turns out that $s = 0.019$, a value smaller than $s = 0.041$ obtained for the case with 32 walkers. Indeed, if N_w is big with respect to the number of processors, the time spent for the walker replication during the branching is dominated by intra-processor exchanges rather than inter-processor ones, and the communication among processors is less relevant.

The possibility to scale almost linearly with the number of processors makes the QMC capable of computing properties of extended and large systems.

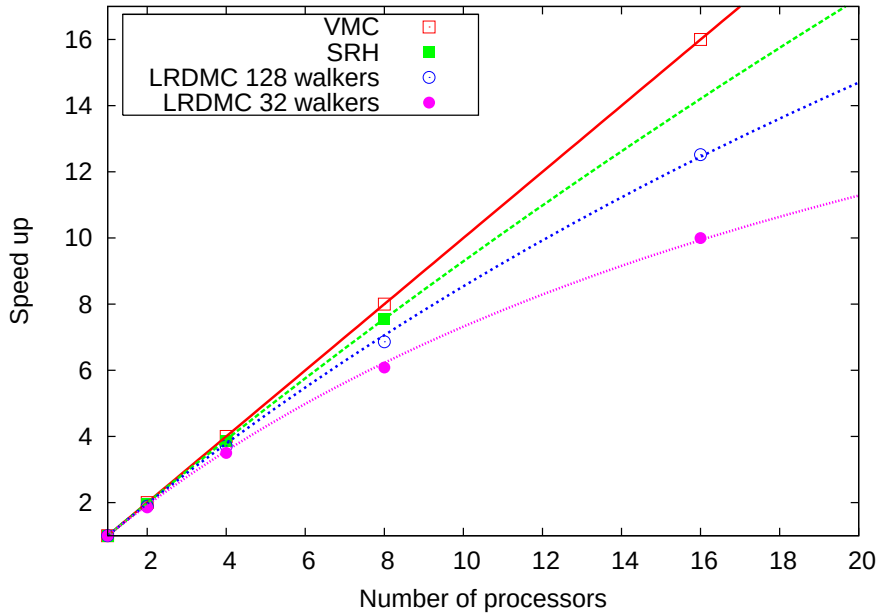


Figure 1.2: Speed up of the parallel simulation versus the number of processors. Typical VMC (16 walkers), SRH (8 walkers) and LRDMC (32 and 128 walkers) simulations are reported. In the SRH calculation we optimized 50 parameters, in both the two LRDMC simulations the time T between two consecutive branchings was set to 0.05 (see Tab. 3.1 and Sec. 3.4). The data have been fitted using the formula in Eq. 1.108. The calculations were carried out on the clx parallel cluster of Xeon $3GHz$ processors at Cineca (Bologna, Italy)

Chapter 2

Jastrow correlated geminal wave function

2.1 Introduction

Since the seminal work by Heitler and London [54], very large steps have been made towards the possibility to predict the quantitative properties of the chemical compounds from a theoretical point of view. Mean field theories, such as HF have been successfully applied to a wide variety of interesting systems, although they fail in describing those in which the correlation is crucial to characterize correctly the chemical bonds. For instance the molecular hydrogen H_2, the simplest and first studied molecule, is poorly described by a single Slater determinant in the large distance regime, which is the paradigm of a strongly correlated bond; indeed, in order to avoid expensive energy contributions - the so called ionic terms - that arise from two electrons of opposite spin surrounding the same hydrogen atom, one needs at least two Slater determinants to deal with a spin singlet wave function containing bonding and antibonding molecular orbitals. Moreover at the bond distance it turns out that the resonance between those two orbitals is important to yield accurate bond length and binding energy, as the correct rate between the ionic and covalent character is recovered. Another route that leads to the same result is to deal with an *antisymmetrized geminal power* AGP wave function, which includes the correlation in the geminal expansion; Barbiellini [55] gave an illuminating example of the beauty of this approach solving merely the simple

problem of the H_2 molecule.

Since 50's, the intensive efforts to explain theoretically the superconductivity have been highlighting the role of pairing in the electronic structure. The BCS wavefunction belongs to an original ansatz in which the correlation is introduced through the product of pairing functions (in this context called Cooper pairs), already exploited in quantum chemistry by the pioneering work of Hurley *et al.* [56] to treat correlation effects in molecular properties. Their wavefunction was called AGP that has been shown to be the particle–conserving version of the BCS ansatz [57]. It includes the single determinantal wavefunction, i.e. the uncorrelated state, as a special case and introduces correlation effects in a straightforward way, through the expansion of the pairing function (geminal): therefore it was studied as a possible alternative to the other multideterminantal approaches. The main advantage of the AGP wave function is that it can be evaluated by computing *only* a single Slater determinant, even beyond the HF approximation. In general, therefore, the computational effort to calculate this correlated wave function is not very demanding, the scaling of the algorithm with the number of particles being comparable with the simplest uncorrelated HF theory. Although this ansatz seemed so appealing, it led to some expensive optimization procedures [58, 59] with numerical problems [60, 61] in particular when applied to large systems, and so it turned out to be non competitive with respect to HF and configuration interaction (CI) methods.

On the other hand the variational approaches based on CI technique, which is able to take into account many Slater determinants, have been shown to be successful for small molecules (e.g. Be_2 [62]). In these cases it is indeed feasible to enlarge the variational basis up to the saturation, the electron correlation properties are well described and consequently all the chemical properties can be predicted with accuracy. However, for interesting systems with a large number of atoms this approach is impossible with a reasonable computational time. Coming back to the H_2 paradigm, it is straightforward to show that a gas with N H_2 molecules, in the dilute limit, can be dealt accurately only with 2^N Slater determinants, otherwise one is missing important correlations due to the antibonding molecular orbital contributions, referred to *each* of the N H_2 molecules. Therefore, if the accuracy in the total energy per atom is kept fixed, a CI-like approach does not scale polynomially with the number of atoms. Although the polynomial

cost of these Quantum Chemistry algorithms - ranging from N^5 to N^7 - is not prohibitive, a loss of accuracy, decreasing exponentially with the number of atoms is always implied, at least in their simplest variational formulations. This is related to the loss of size consistency of a truncated CI expansion. On the other hand, this problem can be overcome by coupled cluster methods, that however in their practical realization are not variational[63].

An alternative approach, not limited to small molecules, is based on DFT. This theory is in principle exact, but its practical implementation requires an approximation for the exchange and correlation functionals based on first principles, like the Local Density approximation (LDA) and its further gradient corrections (GGA), or on semi empirical approaches, like BLYP and B3LYP. For this reason, even though much effort has been made so far to go beyond the standard functionals, DFT is not completely reliable in those cases in which the correlation plays a crucial role. Indeed it fails in describing HTc superconductors and Mott insulators, and in predicting some transition metal compounds properties, whenever the outermost atomic d-shell is near-half-filled, as for instance in the high potential iron proteins (HiPIP)[3]. Also H_2 molecule in the large distance regime must be included in that list, since the large distance Born-Oppenheimer energy surface, depending on Van der Waals forces, is not well reproduced by the standard functionals, although recently some progress has been made to include these important contributions[64–66].

In this chapter we would like to study a different ansatz, which could be both accurate and efficient, so that it would be possible in principle to deal also with complex systems without losing accuracy. We want to exploit the good scaling properties of the AGP approach, with the inclusion of an external correlating factor, the so called Jastrow term, which can overcome the previous difficulties found in the convergence of the optimized AGP basis. Indeed, as already pointed out by Umrigar for the CI expansion[67], the rate of convergence in the basis is increased by the Jastrow factor, just because it allows the wave function to have the correct cusps, otherwise present only asymptotically in the linear combination of determinants or in the geminal expansion. Of course, all the calculations involving the Jastrow-AGP wave function are done within the QMC framework, that can easily deal with explicit correlated wave functions.

The JAGP wave function is the analogous of the resonating valence bond

(RVB) state, first introduced by Pauling[68] in 1949, to describe the chemical structure of molecules such as benzene and nitrogen oxide; the idea behind that concept is the superposition of all possible singlet pairs configurations which link the various nuclear sites of a compound. He gave a numerical estimate of the resonating energy in accordance with thermochemical data, showing the stability of the ansatz with respect to a simple Hartree Fock valence bond approach. Few decades later, Anderson [69] in 1973 developed a mathematical description of the RVB wave function, in discussing the ground state properties of a lattice frustrated model, i.e. the triangular two dimensional Heisenberg antiferromagnet for spin $S = 1/2$. His first representation included an explicit sum over all the singlet pairs, which turned out to be cumbersome in making quantitative calculations, the number of configurations growing exponentially with the system size. Much later, in 1987, motivated to find an explanation to High-Temperature superconductivity by means of the variational approach, he found a much more powerful representation of the RVB state[4], based on the Gutzwiller projection P of a BCS state

$$P|\Psi\rangle = \Pi_{\mathbf{k}} P(u_{\mathbf{k}} + v_{\mathbf{k}} c^{\dagger}_{\mathbf{k},\uparrow} c^{\dagger}_{-\mathbf{k},\downarrow})|0\rangle, \tag{2.1}$$

which in real space and for a fixed number N of electrons takes the form

$$P|\Psi\rangle = P\, \Sigma_{\mathbf{r},\mathbf{r}'} \left[\phi(\mathbf{r}-\mathbf{r}') c^{\dagger}_{\mathbf{r},\uparrow} c^{\dagger}_{\mathbf{r}',\downarrow}\right]^{N/2} |0\rangle, \tag{2.2}$$

where the *pairing function* ϕ is the Fourier transform of $v_{\mathbf{k}}/u_{\mathbf{k}}$. The Cooper pairs described by the BCS wave function are taken apart from each other by the repulsive Gutzwiller projection, which avoids doubly occupied sites; in this way the charge fluctuations present in the superconducting ansatz are frozen and the system can become an insulator even when, according to band theory, the system should be metallic, because there is one electron per site, namely the free electron band is half filled. The wavefunction (2.2) allows a natural and simple description of a superconducting state close to a Mott insulator, opening the possibility for a theoretical explanation of high temperature superconductivity, a phenomenon discovered in 1986[70], but not fully understood until now. Indeed, soon after this important experimental discovery, Anderson[4] suggested that the Copper-Oxygen planes of cuprates could be effectively described by an RVB state, and extensive developements along this lines have subsequently taken place[71].

From the RVB ansatz it is clear that the HTc superconductivity (SC) is essentially driven by the Coulomb and magnetic interactions, with a marginal role played by phonons, in spite of the crucial role they play in the standard BCS theory. As far as the magnetic properties are concerned, the RVB state is quite intriguing because, it represents an insulating phase of an electron model with an odd number of sites per unit cell, with vanishing magnetic moment and without any finite order parameter, namely a completely different picture from the conventional mean field theory, where it is important to break the symmetry in order to avoid the one electron per unit cell condition, incompatible with insulating behavior. This rather unconventional RVB state is therefore called *spin liquid* and it has shown to be a good representative for the GS of some strongly correlated systems on a lattice [5, 72–74].

As the interplay between the Gutzwiller and the BCS state is crucial in determining the accuracy of the RVB ansatz for strongly correlated lattice models, on the same footing the interplay between the Jastrow and the AGP part is expected to be extremely important to set the quality of the JAGP variational description for realistic quantum chemical systems. To highlight the role of the Jastrow, let us coming back to the gedanken experiment of a *gas* of hydrogen dimers: in this case the geminal will contain not only the terms valid for just two sites, but also the contributions from all the other nuclei in the system. It is clear that the AGP wave function will allow strong charge fluctuations around each H pair, and therefore molecular sites with zero and four electrons will be permitted, leading to poor variational energies. For this reason, the AGP alone is not sufficient, and it is necessary to introduce a Gutzwiller-Jastrow factor in order to dump the expensive charge fluctuations.

The structure of this chapter is the following: in Section 2.2 we describe the JAGP ansatz in detail, in Sections 2.3 and 2.4 we present the result obtained for some atoms and molecules respectively, while in Section 2.5 we draw the conclusions.

2.2 Functional form of the wave function

The wavefunction we used in our QMC calculations is defined by the product of two terms, namely a Jastrow J and an antisymmetric part Ψ_{AGP}:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = J(\mathbf{r}_1, \ldots, \mathbf{r}_N)\Psi_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N). \tag{2.3}$$

If the former is an explicit contribution to the dynamic electronic correlation, the latter is able to treat the non dynamic one arising from near degenerate orbitals through the geminal expansion. Therefore our wave function is highly correlated and it is expected to give accurate results especially for molecular systems. The Jastrow term $J(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is further split into a two-body and a three-body factors ($J = J_2 J_3$). The different parts of the wavefunction (2.3) will be described in detail below.

2.2.1 The determinantal part

Let Φ be the pairing function (geminal) which takes into account the correlation between two electrons with opposite spin. If the system is unpolarized ($N^{\uparrow} = N^{\downarrow} = \frac{N}{2}$) and the state is a spin singlet, the AGP wavefunction is

$$\mathbf{\Psi}_{AGP}(\mathbf{r}_1^{\uparrow}, \ldots, \mathbf{r}_N^{\downarrow}) = \hat{A}[\Phi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_1^{\downarrow})\Phi(\mathbf{r}_2^{\uparrow}, \mathbf{r}_2^{\downarrow}) \cdots \Phi(\mathbf{r}_{\frac{N}{2}}^{\uparrow}, \mathbf{r}_{\frac{N}{2}}^{\downarrow})], \tag{2.4}$$

where $\hat{A}$ is an operator that antisymmetrizes the product in the square brackets and the geminal is a singlet:

$$\Phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right), \tag{2.5}$$

implying that $\phi(\mathbf{r}, \mathbf{r}')$ is symmetric under a permutation of its variables. Given this conditions, one can prove [75] that the spatial part of the $\mathbf{\Psi}_{AGP}$ can be written in a very compact form:

$$\Psi_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = det(A_{ij}), \tag{2.6}$$

where A_{ij} is a $\frac{N}{2} \times \frac{N}{2}$ matrix defined as:

$$A_{ij} = \phi(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}). \tag{2.7}$$

We are going to extend the definition of the geminal wavefunction to a polarized system, i.e. a system with a different number of electrons for each spin. This generalization of the geminal model was first proposed by Coleman [76] and called $GAGP$. Without loss of generality, one can assume that the number of up particles ($N^{\uparrow}$) is greater than the down ones ($N^{\downarrow}$). In order to write down the many–body wavefunction with the geminals, one needs to introduce $N^{\uparrow} - N^{\downarrow}$ single particle spin orbitals $\overline{\Phi}_j$ not associated with any geminal and holding unpaired electrons. Once again one recovers the compact notation (2.6) for the spatial part of Ψ_{AGP} (see Appendix B), but this time A_{ij} is a $N^{\uparrow} \times N^{\uparrow}$ matrix defined in the following way:

$$A_{ij} = \begin{cases} \phi(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}) & \text{for } j = 1, N^{\downarrow} \\ \overline{\phi}_j(\mathbf{r}_i^{\uparrow}) & \text{for } j = N^{\downarrow} + 1, N^{\uparrow} \end{cases} \tag{2.8}$$

where the index i ranges from 1 to $N^{\uparrow}$. In the polarized case, the remaining orbitals $\overline{\psi}_j$ may change the total angular momentum and spin quantum numbers with the same rules valid for Slater-type wavefunctions. Within our ansatz it is therefore possible to have definite total spin and angular momenta at least in all cases when the conventional Slater determinant does. On the other hand, whenever a linear combination of Slater determinants is required to have a definite symmetry, e.g. with Clebsch-Gordon coefficients, the same holds for the GAGP wavefunction. Indeed each GAGP can be obtained by antisymmetrizing the product of Slater orbitals (that determine the quantum numbers) and a singlet zero angular momentum term built with the remaining electron pairs. In this way, as far as the quantum numbers of the many-body wavefunction are concerned, the expansion is similar to the conventional one with simple Slater determinants.

The pairing function can be expanded over a basis of single particle orbitals:

$$\Phi_{AGP}(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \sum_{l,m,a,b} \lambda_{a,b}^{l,m} \psi_{a,l}(\mathbf{r}^{\uparrow}) \psi_{b,m}(\mathbf{r}^{\downarrow}), \tag{2.9}$$

where indices l, m span different orbitals centered on atoms a, b, and i,j are coordinates of spin up and down electrons respectively. For spin polarized systems the unpaired orbitals in Eq. 2.8 are expanded as well as the paired ones over the same atomic basis employed in the geminal.

The geminal functions may be viewed as an extension of the simple HF wavefunction, based on molecular orbitals, and in fact the geminal function coincide

with HF only when the number M of non zero eigenvalues of the λ matrix is equal to $N/2$. Indeed the general function 2.9 can be written in diagonal form after an appropriate transformation:

$$\Phi_{AGP}(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \sum_{k}^{M} \lambda^{k} \tilde{\psi}_{k}(\mathbf{r}^{\uparrow}) \tilde{\psi}_{k}(\mathbf{r}^{\downarrow}), \tag{2.10}$$

where $\tilde{\psi}_k(\mathbf{r}) = \sum_{j,a} \mu_{k,j,a} \psi_{j,a}(\mathbf{r})$ are just the molecular orbitals of the HF theory whenever $M = N/2$ (see Appendix C). Notice that with respect to our previous pairing function formulation also off-diagonal elements are now included in the λ matrix, which must be symmetric in order to define a singlet spin orbital state. Moreover it allows one to easily fulfill other system symmetries by setting the appropriate equalities among different $\lambda_{l,m}$. For instance in homo-nuclear diatomic molecules, the invariance under reflection in the plane perpendicular to the molecular axis yields the following relation:

$$\lambda_{m,n}^{a,b} = (-1)^{p_m + p_n} \lambda_{m,n}^{b,a}, \tag{2.11}$$

where p_m is the parity under reflection of the $m-$th orbital.

A further important property of this formalism is the possibility to describe resonating bonds present in many structures, like benzene. A $\lambda_{m,n}^{a,b}$ different from zero represents a chemical bond formed by the linear combination of the *m*-th and *n*-th orbitals belonging to *a*-th and *b*-th nuclei. It turns out that resonating bonds can be well described through the geminal expansion switching on the appropriate $\lambda_{m,n}^{a,b}$ coefficients: the relative weight of each bond is related to the amplitude of its λ.

Another appealing feature of the AGP term is the size consistency. If we smoothly increase the distance between two regions A and B each containing a given number of atoms, the many-electron wave function Ψ factorizes into the product of space-disjoint terms $\Psi = \Psi_A \bigotimes \Psi_B$ as long as the interaction between the electrons coupling the different regions A and B can be neglected. In this limit the total energy of the wave function approaches the sum of the energies corresponding to the two space-disjoint regions. This property, that is obviously valid for the exact many-electron ground state, is not always fulfilled by a generic variational wave function. Strictly speaking, the AGP wave function is certainly size

consistent when both the compound and the separated fragments have the minimum possible total spin, because the geminal expansion contains both bonding and antibonding contributions, that can mutually cancel the ionic term arising in the asymptotically separate regime. Moreover the size consistency of the AGP, as well as the one of the Hartree-Fock state, holds in all cases in which the spin of the compound is the sum of the spin of the fragments. However, similarly to other approaches[63], for spin polarized systems the size consistency does not generally hold, and, in such cases, it may be important go beyond a single AGP wave function. Nevertheless we have experienced that a single reference AGP state is able to describe accurately the electronic structure of the compound around the Born-Oppenheimer minimum even in the mentioned polarized cases, such as in the oxygen dimer.

The last part of this section is devoted to the nuclear cusp condition implementation. The cusp behaviour of the wave function on the coalescence points, i.e. where two electrons overlap (electron cusp) or where an electron and a nucleus overlap (nuclear cusp), is a property fulfilled by the exact wave function[77, 78]. Indeed, the cusps lead to a divergence of the kinetic term, which exactly cancels the divergence arising from the electron-electron or electron-nucleus interactions. Therefore the cusp conditions keep finite the local energy on the coalescence points, with a clear gain in reduction of the variance and in efficiency. Within a Monte Carlo scheme, one is able to satisfy a priori both the nuclear and the electron cusps, with an appropriate choice of the wave function. In this way also the convergence in the basis expansion is expected to be more favorable[67]. A straightforward calculation shows that the AGP wave function fulfills the cusp conditions around the nucleus a if the following linear system is satisfied:

$$\sum_{j}^{(1s,2s)} \lambda_{a,b}^{j,j'} \hat{\psi}'_{a,j}(\mathbf{r} = \mathbf{R}_a) = -Z_a \sum_{c,j} \lambda_{c,b}^{j,j'} \psi_{c,j}(\mathbf{r} = \mathbf{R}_a), \tag{2.12}$$

for all b and j'; in the LHS the caret denotes the spherical average of the orbital gradient. The system can be solved iteratively during the optimization processes, but if we impose that the orbitals satisfy the single atomic cusp conditions, it reduces to:

$$\sum_{c(\neq a),j} \lambda_{c,b}^{j,j'} \psi_{c,j}(\mathbf{R}_a) = 0, \tag{2.13}$$

and because of the exponential orbital damping, if the nuclei are not close together each term in the previous equations is very small, of the order of $\exp(-|\mathbf{R}_a - \mathbf{R}_c|)$. Therefore, with the aim of making the optimization faster, we have chosen to use $1s$ and $2s$ orbitals satisfying the atomic cusp conditions and to disregard the sum (2.13) in Eq. 2.12. In this way, once the energy minimum is reached, also the molecular cusp conditions (2.12) are rather well satisfied.

2.2.2 Two body Jastrow term

As it is well known the Jastrow term plays a crucial role in treating many body correlation effects. One of the most important correlation contributions arises from the electron-electron interaction. Therefore it is worth using at least a two-body Jastrow factor in the trial wave function. Indeed this term reduces the electron coalescence probability, and so decreases the average value of the repulsive interaction. The two-body Jastrow function reads:

$$J_2(\mathbf{r}_1, ..., \mathbf{r}_N) = \exp\left(\sum_{i<j}^{N} u(r_{ij})\right), \tag{2.14}$$

where $u(r_{ij})$ depends only on the relative distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between two electrons and allows to fulfill the cusp conditions for opposite spin electrons as long as $u(r_{ij}) \to \frac{r_{ij}}{2}$ for small electron-electron distance. If we defined u to fulfill also the cusp conditions for parallel spin electrons ($u(r_{ij}) \to \frac{r_{ij}}{4}$ for small r_{ij}), the wave function would be *spin contaminated*, i.e. it would not be an eigenstate of the total spin operator S^2, as pointed out in Ref. [79]. We have chosen to preserve the correct spin symmetry of the total wavefunction, by fulfilling only the condition for antiparallel electrons. Indeed the cusp condition for electrons with parallel spins is much less important because their probability to get close is clearly small, due to the Pauli principle.

The pair correlation function u can be parametrized successfully by few variational parameters. The functional form we used for u, particularly convenient at the chemical bond distance, where we performed most of the calculations, is the one used also by Fahy [80]:

$$u(r_{ij}) = \frac{a_{\sigma_i\sigma_j} r_{ij}}{1 + b_{\sigma_i\sigma_j} r_{ij}}, \tag{2.15}$$

where σ_i is the spin of the i-th electron. The value of $a_{\sigma_i\sigma_j}$ is fixed by the cusp condition at the coalescence points of two electrons and $b_{\sigma_i\sigma_j}$ contains at most three free variational parameters, as $b_{\uparrow\downarrow} = b_{\downarrow\uparrow}$ is implied by the spatial symmetry of the Jastrow factor. We kept $a_{\sigma_i\sigma_j} = 1/2$ and $b_{\sigma_i\sigma_j} = b$, which satisfy the cusp conditions for antiparallel electrons. b is the unique variational parameter of our 2-body Jastrow factor.

For the nitrogen atom, we checked the quality of this wavefunction with respect to a spin contaminated one with two variational parameters, $b_{\uparrow\uparrow} = b_{\downarrow\downarrow} = b_1$ and $b_{\uparrow\downarrow} = b_2$. In both the cases, we kept the geminal expansion to be minimal ($HF + J$ like wavefunction). As reported in Table 2.1, the improvement in energy obtained by contaminating the spin wavefunction is rather negligible, and disappears when the FN DMC simulation is carried out. This implies that it is possible to obtain almost optimal nodes, without spoiling the spin symmetry and by using only one variational parameter for the Jastrow factor.

2.2.3 Three Body Jastrow term

In order to describe well the correlation between electrons the simple Jastrow factor is not sufficient. Indeed it takes into account only the electron-electron separation and not the individual electronic position $\mathbf{r}_i$ and $\mathbf{r}_j$. It is expected that close to nuclei the electron correlation is not accurately described by translationally invariant Jastrow, as shown by different authors, see for instance Ref. [26]. For this reason we introduce a factor, often called three body (electron-electron-nucleus) Jastrow, that explicitly depends on both the electronic positions $\mathbf{r}_i$ and $\mathbf{r}_j$. The three body Jastrow is chosen to satisfy the following requirements:

- The cusp conditions set up by the two-body Jastrow term and by the AGP are preserved.
- It does not distinguish the electronic spins otherwise causing spin contamination.
- Whenever the atomic distances are large it factorizes into a product of independent contributions located near each atom, an important requirement to satisfy the size consistency of the variational wave function.

Analogously to the pairing trial function in Eq. 2.9 we define a three body factor as:

$$\begin{aligned} J_3(\mathbf{r}_1, ..., \mathbf{r}_N) &= \exp\left(\sum_{i<j} \Phi_J(\mathbf{r}_i, \mathbf{r}_j)\right) \\ \Phi_J(\mathbf{r}_i, \mathbf{r}_j) &= \sum_{l,m,a,b} g^{a,b}_{l,m} \psi_{a,l}(\mathbf{r}_i)\psi_{b,m}(\mathbf{r}_j), \end{aligned} \tag{2.16}$$

where indices l and m indicate different orbitals located around the atoms a and b respectively. Each Jastrow orbital $\psi_{a,l}(\mathbf{r})$ is centered on the corresponding atomic position $\mathbf{R}_a$. We have used Gaussian and exponential orbitals multiplied by appropriate polynomials of the electronic coordinates, related to different spherical harmonics with given angular momentum, as in the usual Slater basis. Analogously to the geminal function Φ_{AGP}, whenever the one particle basis set $\{\psi_{a,i}\}$ is complete the expansion (2.16) is also complete for the generic two particle function $\Phi_J(\mathbf{r}, \mathbf{r}')$. In the latter case, however, the one particle orbitals have to behave smoothly close to the corresponding nuclei, namely as:

$$\psi_{a,i}(\mathbf{r}) - \psi_{a,i}(\mathbf{R}_a) \simeq |\mathbf{r} - \mathbf{R}_a|^2, \tag{2.17}$$

or with larger power, in order to preserve the nuclear cusp conditions (2.12).

For the s-wave orbitals we have found energetically convenient to add a finite constant $c_l/(N-1)$. As shown in the Appendix D, a non zero value of the constant c_l for such orbitals $\psi_{a,l}$ is equivalent to include in the wave function a size consistent one body term. As pointed out in Ref. [81], it is easier to optimize a one body term implicitly present in the 3-body Jastrow factor, rather than including more orbitals in the determinantal basis set.

The chosen form for the 3-body Jastrow (2.16) is similar to one used by Prendergast et al. [82] and has very appealing features: it easily allows including the symmetries of the system by imposing them on the matrix $g^{a,b}_{l,m}$ exactly as it is possible for the pairing part (e.g. by replacing $\lambda^{a,b}_{m,n}$ with $g^{a,b}_{m,n}$ in Eq. 2.11). It is size consistent, namely the atomic limit can be smoothly recovered with the same trial function when the matrix terms $g^{a,b}_{l,m}$ for $a \neq b$ approach zero in this limit. Notice that a small non zero value of $g^{a,b}_{l,m}$ for $a \neq b$ acting on p-wave orbitals can correctly describe a weak interaction between electrons such as the the Van der Waals forces.

2.3 Application of the JAGP to atoms

For the application to atoms, in order that the wavefunction Ψ has a definite angular momentum, it is convenient that the geminal is rotationally invariant around the nucleus. This requirement is achieved by taking the generic orbital ψ_i to be an eigenfunction of the single particle angular momentum operators l^2 and l_z; hence, the orbital will be denoted by:

$$\psi_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_{lm}(\theta,\phi), \tag{2.18}$$

where Y_{lm} are spherical harmonics with standard notations and the radial part R_{nl} depends on the principal quantum number n. In this preliminary study of atoms we used a simple geminal expansion in which only the diagonal λ matrix elements are taken into account. Thus the atomic geminal function takes the form:

$$\phi(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) = \sum_{nl}\sum_{m=-l}^{l}\lambda_{nl}(-1)^m\psi_{nlm}(\mathbf{r}^{\uparrow})\psi^*_{nlm}(\mathbf{r}^{\downarrow}). \tag{2.19}$$

In order to optimize the radial part R_{nl} of the the single particle orbitals, we expand these radial functions in a Slater basis, in close analogy with Roothaan–Hartree–Fock calculations [83], namely using functions of the type:

$$r^{n-1}e^{-z_k r} \tag{2.20}$$

with $n \geq 1$, taking in principle as many different z_k's as required for convergence.

In the Roothaan–Hartree–Fock the coefficients of the linear combinations are more involved, since the orthogonality among all single particle states is required. In the Monte Carlo approach we have found that it is much simpler and more efficient to deal with non-orthogonal orbitals, without spoiling the accuracy of the calculation. In fact, for light elements with $Z \leq 15$, studied here, it is possible to obtain almost converged results by using only two exponentials for each radial component (double zeta).

Hence, our single particle orbitals read in general

$$R_{nl}(r) = Cr^{n-1}(e^{-z_1 r} + pe^{-z_2 r}), \tag{2.21}$$

where p is another variational parameter and C is the normalization factor for the radial part R_{nl}:

$$C = \frac{1}{\sqrt{(2n)!}}\left((2z_1)^{-(2n+1)} + 2p(z_1+z_2)^{-(2n+1)} + p^2(2z_2)^{-(2n+1)}\right)^{-\frac{1}{2}}. \tag{2.22}$$

Actually p is not free for all the orbitals: indeed, for a more accurate variational wavefunction, we impose the electron-nucleus cusp condition [77], which implies that each orbital must fulfill the following relation:

$$\frac{\partial \hat{\psi}}{\partial r} = -Z\psi \tag{2.23}$$

at $r = 0$ (the hat denotes the spherical average). That condition is automatically obeyed by all but $1s$ and $2s$ orbitals of the type given in Eq.2.21. Instead, the parameter p of $1s$ orbital must be:

$$p = \frac{z_1 - Z}{Z - z_2}, \tag{2.24}$$

and for the $2s$ state, in order to fulfill Eq. (2.23), we choose a functional form of the type:

$$\psi_{2s}(r) = e^{-z_1 r} + (p + \alpha r)e^{-z_2 r}, \tag{2.25}$$

where α is a further variational parameter and p is given by

$$p = \frac{z_1 - \alpha - Z}{Z - z_2}. \tag{2.26}$$

In our study, we found that the presence of the α term leads to a very marginal improvement of the variational wavefunction, therefore we set $\alpha = 0$ and we kept the $1s$ and $2s$ orbitals to have the same functional form, in order to reduce the total number of parameters.

2.3.1 Results

We have carried out QMC calculations for atoms from Li to P, using the JAGP ansatz to describe the atomic electronic structure. We performed the optimization of both the geminal and the Jastrow part minimizing the energy with the SR method described in Sec. 1.4.1, Chapter 1. For all the atoms, we considered first the minimal geminal expansion, corresponding to the HF single determinant, together with the simplest Jastrow factor with a single parameter reported in Eq. (2.15). To improve the antisymmetric part, we increased the number of orbitals in the geminal expansion, and for Be and Mg atoms we also systematically considered the effect of improving the wave function by using a 3-body Jastrow factor (see Eq. 2.16).

In order to judge the outcome of our calculations, we computed the correlation energies and in particular its fraction with respect to the exact ground state energy for the nonrelativistic infinite nuclear mass Hamiltonian, estimated in Ref. [84]. The quality of the variational wavefunction can be seen by computing the expectation value of the energy by means of the VMC calculations. Furthermore, we carried out DMC simulations within the FN approximation. As we have seen in Sec. 1.5, the DMC energy depends on the quality of the nodal structure of the variational wavefunction and the capability of improving the nodes during the optimization is crucial to obtain almost exact DMC energy values. To that purpose, it is very important to have a variational functional form appropriate to reproduce the correct nodes. We show that the JAGP wavefunction satisfies well this requirement, yielding in all the atoms studied here very good DMC results. The VMC and DMC energies are listed in Table 2.1; in Figures 2.1 and 2.2 we plot the percentage of the correlation energy recovered respectively by VMC and DMC calculations for different atoms and wavefunctions.

The VMC calculations with the minimal geminal expansion and the two body Jastrow factor yield from 60% to 68% of the total correlation energy, with the exception of the Li atom, where 91.4% of the correlation energy is obtained. Therefore, there is a sizable loss of accuracy in going from Li to Be, the worst case being the Boron atom. The corresponding DMC simulations get a large amount of the energy missing in the VMC calculations, recovering from 87.7% to 99.9% of the total correlation energy, but the dependence on the atomic number shows the same behavior found in VMC: the worst results are obtained for Be, B and C atoms, due to the strong multiconfigurational nature of their ground states. As well known, one can improve substantially the variational state of those atoms including not only the $1s^2\ 2s^2$ configuration but also the $2s^2\ 2p^2$, because of the near degeneracy of $2s$ and $2p$ orbitals. In this case the JAGP ansatz is particularly efficient: by adding just one term in the geminal expansion, we are able to remove this loss of accuracy in the correlation energy both in the VMC and the DMC calculations.

In Table 2.2, we summarize some results obtained for Be in previous works and compare them with ours. AGP calculations of atoms have been performed only few times so far, the best one for Be is reported in the last row of the Table 2.2. Kurtz *et al.* [85] were able to recover 84% of correlation energy using a

Table 2.1: Total energies in variational (E_{VMC}) and diffusion (E_{DMC}) Monte Carlo calculations; the percentages of correlation energy recovered in VMC ($E_c^{VMC}(\%)$) and DMC ($E_c^{DMC}(\%)$) have been evaluated using the exact (E_0) and Hartree–Fock (E_{HF}) energies from Ref. [84]. M is the number of terms in the pairing expansion. The energies are in *Hartree*.

	M	E_0	E_{HF}	E_{VMC}	$E_c^{VMC}(\%)$	E_{DMC}	$E_c^{DMC}(\%)$
Li	1	-7. 47806	-7.432727	-7.47415(10)	91.4(2)	-7.47806(5)	100.0(1)
	5			-7.47582(2)	95.06(5)	-7.47801(5)	99.9(1)
Be	2	-14. 66736	-14.573023	-14.63145(5)	61.9(5)	-14.6565(4)	88.5(4)
	5			-14.661695(10)	93.995(11)	-14.66711(3)	99.73(3)
	5[1]			-14.66504(4)	97.54(5)	-14.66726(1)	99.894(11)
B	2	-24. 65391	-24.529061	-24.6042(3)	60.3(2)	-24.63855(5)	87.7(4)
	5			-24.62801(4)	79.25(4)	-24.6493(3)	96.3(3)
C	2	-37. 8450	-37.688619	-37.7848(6)	61.5(4)	-37.8296(8)	90.2(5)
	5			-37.7985(7)	70.3(4)	-37.8359(8)	94.2(5)
N	2	-54. 5892	-54.400934	-54.52180(15)	64.20(8)	-54.57555(5)	92.7(3)
	2[2]			-54.52565(15)	66.20(8)	-54.5753(4)	92.6(2)
	14			-54.5263(2)	66.62(11)	-54.5769(2)	93.47(10)
O	3	-75. 0673	-74.809398	-74.9750(7)	64.2(3)	-75.0477(8)	92.4(3)
	14			-74.9838(6)	67.6(2)	-75.0516(9)	93.9(3)
F	4	-99. 7339	-99.409349	-99.6190(8)	64.6(3)	-99.7145(15)	94.0(5)
	14			-99.6315(7)	68.4(2)	-99.7141(6)	93.91(18)
Ne	5	-128. 9376	-128.547098	-128.8070(10)	66.6(3)	-128.9204(8)	95.6(2)
	14			-128.8159(6)	68.83(17)	-128.9199(7)	95.47(18)
Na	5	-162. 2546	-161.858912	-162.1334(7)	69.37(19)	-162.2325(10)	94.4(2)
	9			-162.1434(7)	71.91(16)	-162.2370(10)	95.5(2)
Mg	6	-200. 053	-199.614636	-199.9113(8)	67.67(19)	-200.0327(9)	95.4(2)
	9			-199.9363(8)	73.38(19)	-200.0375(10)	96.5(2)
	9[1]			-200.0002(5)	87.95(12)	-200.0389(5)	96.79(11)
Al	6	-242. 346	-241.876707	-242.1900(9)	66.77(19)	-242.3215(10)	94.8(2)
	9			-242.2124(9)	71.53(19)	-242.3265(10)	95.8(2)
Si	6	-289. 359	-288.854363	-283.1875(10)	66.0(2)	-289.3275(10)	93.8(2)
	9			-289.1970(10)	67.9(2)	-289.3285(10)	94.0(2)
P	6	-341. 259	-340.718781	-341.0700(10)	65.0(2)	-341.2220(15)	93.2(3)

[1] Wavefunction with a three body Jastrow factor.

[2] Wavefunction with a two body Jastrow factor spin contaminated.

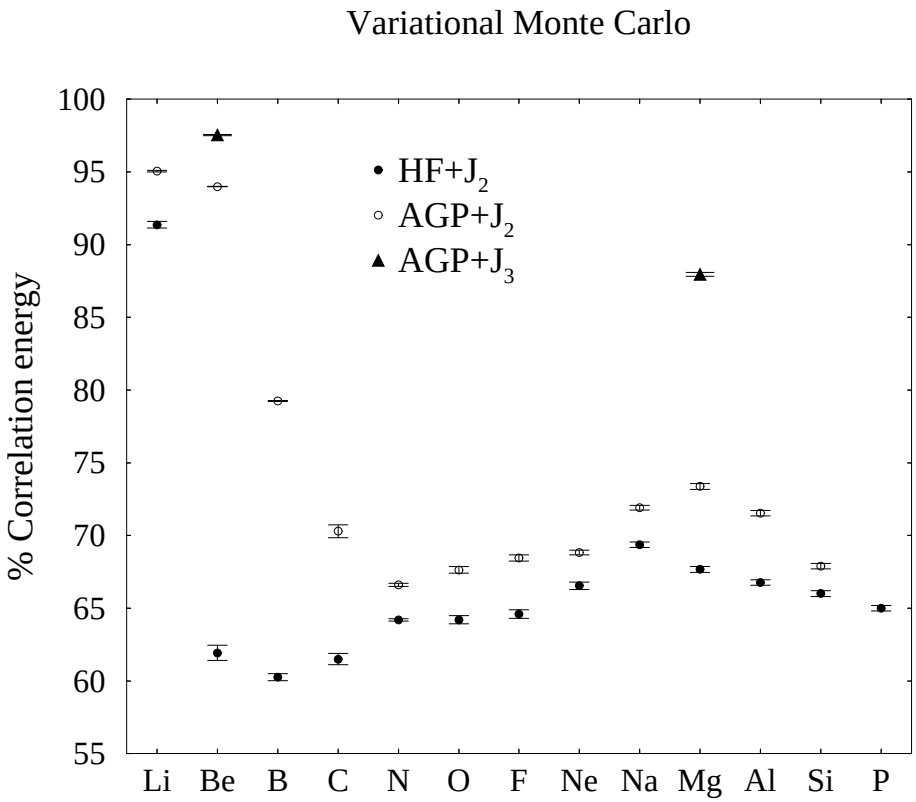


Figure 2.1: VMC correlation energies for $HF + J_2$ (minimal geminal expansion with a two-body Jastrow factor), $AGP + J_2$ (best geminal expansion) and $AGP + J_3$ (best geminal with a three-body Jastrow factor).

geminal expansion with a very large basis; our variational JAGP wavefunction reaches 94% with a much smaller basis ($1s$, $2s$ and $2p$ orbitals). By including a three-body Jastrow factor, 97.5% of the correlation energy is finally obtained, which is comparable with the best multiconfiguration wavefunctions previously studied [86].

This outcome highlights the importance of the Jastrow in reducing the geminal expansion and yielding a better energy. The nodal surface can be substantially improved with the present approach, because the pairing expansion contains implicitly not only the four determinants $1s^2 2s^2$ and $1s^2 2p^2$, but also the other three $2s^2 2p^2$ and three $2p^2 2p^2$, which can improve further the wavefunction. Indeed, the geminal expansion reduces exactly to four determinants in the limit $\lambda_{2s} \to 0$ and $\lambda_{2p} \to 0$ with constant ratio $\frac{\lambda_{2s}}{\lambda_{2p}}$. The fact that the minimum energy is ob-

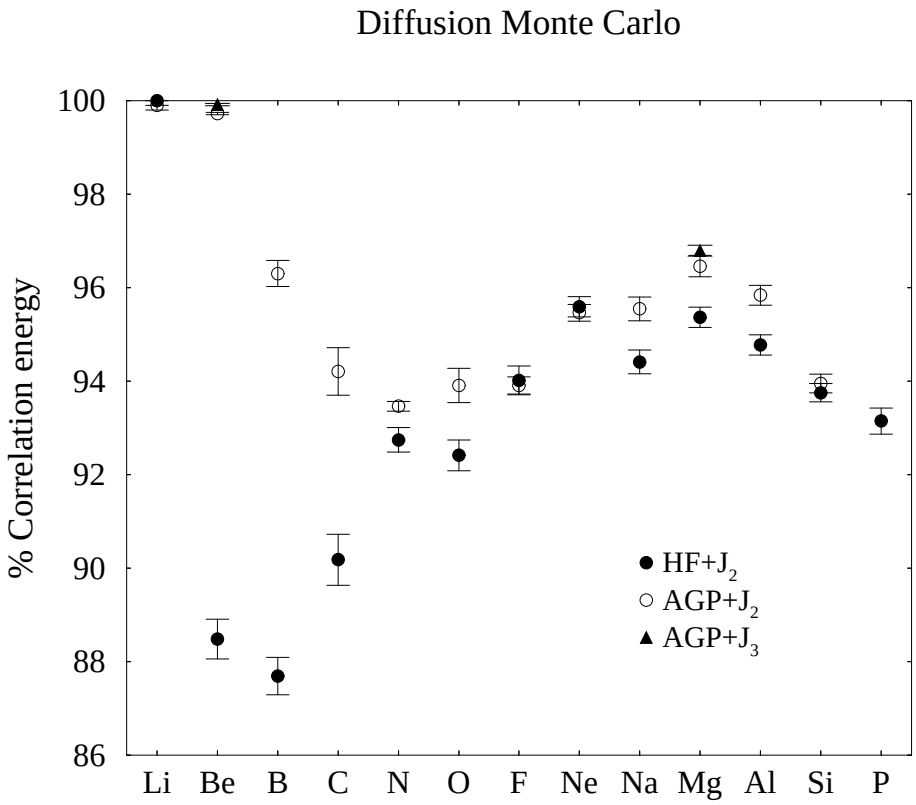


Figure 2.2: DMC correlation energies obtained by $HF + J_2$, $AGP + J_2$ and $AGP + J_3$ wavefunctions.

tained for $\lambda_{2s} \neq 0$ and $\lambda_{2p} \neq 0$ clearly shows that the energy can be lowered by considering the remaining configurations described above. Indeed our DMC

Table 2.2: Comparison of the energies obtained by various authors for Be.

	basis	Jastrow	VMC	DMC
present work	2s1p	two body	-14.661695(10)	-14.66711(3)
Huang *et al.*[86]	2s1p	two body	-14.66088(5)	-14.66689(4)
present work	2s1p	three body	-14.66504(4)	-14.66726(1)
Huang *et al.*[86]	2s1p	three body	-14.66662(1)	-14.66723(1)
Sarsa *et al.*[87]	2s1p	three body	-14.6523(1)	
Kurtz *et al.*[85]	6s3p2d	none	-14.6171	

energies are more than two standard deviations lower than the ones by Huang *et al.*[86], to our knowledge the best available ones obtained with the four determinants: one $1s^2 2s^2$ and three $1s^2 2p^2$. In order to determine accurate nodes for the corresponding DMC calculation they used the two-body Jastrow factor similar to the one of Eq. 2.15 and an highly involved three-body term much more complete than ours (we used a double zeta s-p basis in the an expansion of Eq. 2.16 for the 3-body geminal). For this reason our corresponding VMC energy is slightly worse in this case. We also verified, therefore, that a more accurate description of the Jastrow factor (which do not affect directly the nodes) is crucial to obtain better nodes, provided the variational parameters, belonging to both the Jastrow and the determinantal part, are optimized altogether. For instance, in Ref. [87] the authors optimized only the coefficients in front of the four determinants $1s2s - 1s2p$ and not the orbitals, obtaining clearly a non optimal energy. The JAGP is simple enough to allow a feasible parametrization of the variational state, by capturing the most important correlations.

We found that also Mg, Al and Si, the second row atoms corresponding to Be, B and C in the first row, have a quite strong multiconfigurational character, involving here $3s$ and $3p$ orbitals. Analogously to the Beryllium case, for Mg we have optimized both the two-body and the three-body Jastrow factor, together with the AGP wavefunction containing $3s$-$3p$ resonance. In this case, although at the variational level the three-body wavefunction is much better than the two-body one (see Fig. 2.1), that difference disappears almost completely in the DMC results. This shows that the correction of the nodal surface allowed by the more accurate three-body Jastrow does not seem to be crucial as in the Be atom. On the other hand, the effect of the AGP expansion is significant in improving further the DMC calculation, which already yields good FN energies even with a simple HF+J trial wavefunction for atoms heavier than C (percentage of correlation energy always greater than 92%). By adding the $3p$ contribution to the geminal we were able to recover 96.8% of the correlation energy of Mg (see Fig. 2.2). Also for Al the presence of the $3p$ orbital is significant in lowering the DMC energy, and for Si it seems important only in the VMC value. As clearly shown in Fig. 2.2 the amount of nondynamic correlation is expected to be negligible for large Z: for this reason we have not carried out the geminal optimization for $Z > 14$.

Finally, by using the JAGP wavefunction, we optimized some atoms (from

N to Na) not affected by nondynamic correlation; here, in order to obtain an improvement in the VMC and in the DMC energies, we needed a larger basis ($3s2p1d$) in the geminal expansion.

2.4 Application of the JAGP to molecules

In this section we study correlation, and atomization energies, accompanied with the determination of the ground state optimal structure for a restricted ensemble of molecules. For each of them we performed a full all-electron SR geometry optimization, starting from the experimental molecular structure. After the energy minimization, we carried out all-electron VMC and DMC simulations at the optimal geometry within the fixed node approximation. The basis used here is a double zeta Slater set of atomic orbitals (STO-DZ) for the AGP part (see Eq. 2.18 and Eq. 2.21), while for parameterizing the 3-body Jastrow geminal we used a double zeta Gaussian atomic set (GTO-DZ). In this way both the antisymmetric and the bosonic part are well described, preserving the right exponential behavior for the former and the strong localization properties for the latter. Sometimes, in order to improve the quality of the variational wave function we employed a mixed Gaussian and Slater basis set in the Jastrow part, that allows to avoid a too strong dependency in the variational parameters in a simple way. However, both in the AGP and in the Jastrow sector we never used a large basis set, in order to keep the wave function as simple as possible. The accuracy of our wave function can be obviously improved by an extension of the one particle basis set but this is rather difficult for a stochastic minimization of the energy. Nevertheless, for most of the molecules studied with a simple JAGP wave function, a DMC calculation is able to reach the chemical accuracy in the binding energies, and the SR optimization yields very precise geometries already at the VMC level.

In the first part of this section some results will be presented for a small set of widely studied molecules and belonging to the G1 database. In the second subsection we will treat the benzene and its radical cation $C_6H_6^+$, by taking into account its distortion due to the Jahn-Teller effect, that is well reproduced by our SR geometry optimization.

2.4.1 Small diatomic molecules, methane, and water

Except from Be_2 and C_2, all the molecules presented here belong to the standard G1 reference set; all their properties are well known and well reproduced by standard quantum chemistry methods, therefore they constitute a good case for testing new approaches and new wave functions.

The Li dimer is one of the easiest molecules to be studied after the H_2, which is exact for any Diffusion Monte Carlo (FN DMC) calculation with a trial wave function that preserves the nodeless structure. Li_2 is less trivial due to the presence of core electrons that are only partially involved in the chemical bond and to the $2s - 2p$ near degeneracy for the valence electrons. Therefore many authors have done benchmark calculation on this molecule to check the accuracy of the method or to determine the variance of the inter-nuclear force calculated within a QMC framework. In this work we start from Li_2 to move toward a structural analysis of more complex compounds, thus showing that our QMC approach is able to handle relevant chemical problems. In the case of Li_2, a $3s\ 1p\ STO - DZ$ AGP basis and a $1s\ 1p\ GTO - DZ$ Jastrow basis turns out to be enough for the chemical accuracy. More than 99% of the correlation energy is recovered by a DMC simulation (Table 2.3), and the atomization energy is exact within few thousandth of eV ($0.02\ kcal\ mol^{-1}$) (Table 2.4). Similar accuracy have been previously reached within a DMC approach[79], only by using a multi-reference CI like wave function, that was the usual way to improve the electronic nodal structure. As stressed before, the JAGP wave function includes many resonating configurations through the geminal expansion, beyond the $1s\ 2s$ HF ground state. The bond length has been calculated at the variational level through the fully optimized JAGP wave function: the resulting equilibrium geometry turns out to be highly accurate (Table 2.5), with a discrepancy of only $0.001a_0$ from the exact result. For this molecule it is worth comparing our work with the one by Assaraf and Caffarel [88]. Their zero-variance zero-bias principle has been proved to be effective in reducing the fluctuations related to the inter-nuclear force; however they found that only the inclusion of the space warp transformation drastically lowers the force statistical error, which magnitude becomes equal or even lower than the energy statistical error, thus allowing a feasible molecular geometry optimization. Actually, our way of computing forces (see Eq. 1.43) provides slightly larger variances, without

explicitly invoking the zero-variance zero-bias principle.

The very good bond length, we obtained, is probably due to two main ingredients of our calculations: we have carried out a stable energy optimization that is often more effective than the variance one, as shown by different authors [29, 89, 90], and we have used very accurate trial function as it is apparent from the good variational energy.

Indeed within our scheme we obtain good results without exploiting the computationally much more demanding DMC, thus highlighting the importance of the SR minimization described in Subsection 1.4.1.

Let us now consider larger molecules. Both C_2 and O_2 are poorly described by a single Slater determinant, since the presence of the nondynamic correlation is strong. Instead with a single geminal JAGP wave function, including implicitly many Slater-determinants, it is possible to obtain a quite good description of their molecular properties. For C_2, we used a $2s\ 1p\ STO-DZ$ basis in the geminal, and a $2s\ 1p\ DZ$ Gaussian Slater mixed basis in the Jastrow, for O_2 we employed a $3s\ 1p\ STO-DZ$ in the geminal and the same Jastrow basis as before. In both the cases, the variational energies recover more than 80% of the correlation energy, the DMC ones yield more than 90%, as shown in Tab. 2.3. These results are of the same level of accuracy as those obtained by Filippi *et al*[79] with a multireference wave function by using the same Slater basis for the antisymmetric part and a different Jastrow factor. From the Table 2.4 of the atomization energies, it is apparent that DMC considerably improves the binding energy with respect to the VMC values, although for these two molecules it is quite far from the chemical accuracy ($\simeq$ 0.1 eV): for C_2 the error is 0.60(3) eV, for O_2 it amounts to 0.67(5) eV. Indeed, it is well known that the electronic structure of the atoms is described better than the corresponding molecules if the basis set remains the same, and the nodal error is not compensated by the energy difference between the separated atoms and the molecule. In a benchmark DMC calculation with pseudopotentials [91], Grossman found an error of 0.27 eV in the atomization energy for O_2, by using a single determinant wave function; probably, pseudopotentials allow the error between the pseudoatoms and the pseudomolecule to compensate better, thus yielding more accurate energy differences. As a final remark on the O_2 and C_2 molecules, our bond lengths are in between the LDA and GGA precision, and still worse than the best CCSD calculations, but our results may be considerably

improved by a larger atomic basis set, that we have not attempted so far.

Methane and water are very well described by the JAGP wave function. Also for these molecules we recover more than 80% of correlation energy at the VMC level, while DMC yields more than 90%, with the same level of accuracy reached in previous Monte Carlo studies [92–95]. Here the binding energy is almost exact, since in this case the nodal energy error arises essentially from only one atom (carbon or oxygen) and therefore it is exactly compensated when the atomization energy is calculated. Also the bond lengths are highly accurate, with an error lower then 0.005 a_0.

For Be_2 we applied a 3s 1p STO-DZ basis set for the AGP part and a 2s 2p DZ Gaussian Slater mixed basis for the Jastrow factor. VMC calculations performed with this trial function at the experimental equilibrium geometry yield 90% of the total correlation energy, while DMC gives 97.5% of correlation, i.e. a total energy of -29.33341(25) H. Although this value is better than that obtained by Filippi *et al* [79] (-29.3301(2) H) with a smaller basis ($3s$ atomic orbitals not included), it is not enough to bind the molecule, because the binding energy remains still positive (0.0069(37) H). Instead, once the molecular geometry has been relaxed, the SR optimization finds a bond distance of $13.5(5)\ a_0$ at the VMC level; therefore the employed basis allows the molecule to have a Van der Waals like minimum, quite far from the experimental value. In order to have a reasonable description of the bond length and the atomization energy, one needs to include at least a $3s2p$ basis in the antisymmetric part, as pointed out in Ref. [96], and indeed an atomization energy compatible with the experimental result (0.11(1) eV) has been obtained within the extended geminal model [97] by using a much larger basis set (9s,7p,4d,2f,1g) [98]. This suggests that a complete basis set calculation with JAGP may describe also this molecule, but it is extremely difficult to cope with a very large basis within a QMC framework. Therefore we believe that at present the accuracy needed to describe correctly Be_2 is out of the possibilities of the approach.

2.4.2 Benzene and its radical cation

We studied the ${}^1A_{1g}$ ground state of the benzene molecule by using a very simple one particle basis set: for the AGP, a 2s1p DZ set centered on the carbon atoms

Table 2.3: Total energies in variational (E_{VMC}) and diffusion (E_{DMC}) Monte Carlo calculations; the percentages of correlation energy recovered in VMC ($E_c^{VMC}(\%)$) and DMC ($E_c^{DMC}(\%)$) have been evaluated using the "exact" (E_0) and Hartree–Fock (E_{HF}) energies from the references reported in the table. Here "exact" means the ground state energy of the non relativistic infinite nuclear mass hamiltonian. The energies are in *Hartree*.

	E_0	E_{HF}	E_{VMC}	$E_c^{VMC}(\%)$	E_{DMC}	$E_c^{DMC}(\%)$
Li	-7.47806 [1]	-7.432727 [1]	-7.47721(11)	98.12(24)	-7.47791(12)	99.67(27)
Li_2	-14.9954 [3]	-14.87152 [3]	-14.99002(12)	95.7(1)	-14.99472(17)	99.45(14)
Be	-14.66736 [1]	-14.573023 [1]	-14.66328(19)	95.67(20)	-14.66705(12)	99.67(13)
Be_2	-29.33854(5) [3]	-29.13242 [3]	-29.3179(5)	89.99(24)	-29.33341(25)	97.51(12)
O	-75.0673 [1]	-74.809398 [1]	-75.0237(5)	83.09(19)	-75.0522(3)	94.14(11)
H_2O	-76.438(3) [2]	-76.068(1) [2]	-76.3803(4)	84.40(10)	-76.4175(4)	94.46(10)
O_2	-150.3268 [3]	-149.6659 [3]	-150.1992(5)	80.69(7)	-150.272(2)	91.7(3)
C	-37.8450 [1]	-37.688619 [1]	-37.81303(17)	79.55(11)	-37.8350(6)	93.6(4)
C_2	-75.923(5) [3]	-75.40620 [3]	-75.8293(5)	81.87(10)	-75.8810(5)	91.87(10)
CH_4	-40.515 [4]	-40.219 [4]	-40.4627(3)	82.33(10)	-40.5041(8)	96.3(3)
C_6H_6	-232.247(4) [5]	-230.82(2) [6]	-231.8084(15)	69.25(10)	-232.156(3)	93.60(21)

[1] Exact and HF energies from Ref. [84].

[2] Ref. [99].

[3] Ref. [79].

[4] Ref. [92].

[5] Estimated "exact" energy from Ref. [100].

[6] Ref. [101].

Table 2.4: Binding energies in eV obtained by variational (Δ_{VMC}) and diffusion (Δ_{DMC}) Monte Carlo calculations; Δ_0 is the "exact" result for the non-relativistic infinite nuclear mass hamiltonian. Also the percentages ($\Delta_{VMC}(\%)$ and $\Delta_{DMC}(\%)$) of the total binding energies are reported.

	Δ_0	Δ_{VMC}	$\Delta_{VMC}(\%)$	Δ_{DMC}	$\Delta_{DMC}(\%)$
Li_2	-1.069	-0.967(3)	90.4(3)	-1.058(5)	99.0(5)
O_2	-5.230	-4.13(4)	78.9(8)	-4.56(5)	87.1(9)
H_2O	-10.087	-9.704(24)	96.2(1.0)	-9.940(19)	98.5(9)
C_2	-6.340	-5.530(13)	87.22(20)	-5.74(3)	90.6(5)
CH_4	-18.232	-17.678(9)	96.96(5)	-18.21(4)	99.86(22)
C_6H_6	-59.25	-52.53(4)	88.67(7)	-58.41(8)	98.60(13)

and a 1s SZ on the hydrogen, instead for the 3-body Jastrow, a 1s1p DZ-GTO set centered only on the carbon sites. C_6H_6 is a peculiar molecule, since its highly symmetric ground state, which belongs to the D_{6h} point group, is a resonance among different many-body states, each of them characterized by three double bonds between carbon atoms. This resonance is responsible for the stability of the structure and therefore for its aromatic properties. We started from a non resonating 2-body Jastrow wave function, which dimerizes the ring and breaks the full rotational symmetry, leading to the Kekulé configuration. As we expected, the inclusion of the resonance between the two possible Kekulé states lowers the VMC energy by more than 2 eV. The wave function is further improved by adding another type of resonance, that includes also the Dewar contributions connecting third nearest neighbor carbons. As reported in Tab. 2.6, the gain with respect to the simplest Kekulé wave function amounts to 4.2 eV, but the main improvement arises from the further inclusion of the three body Jastrow factor, which allows to recover the 89% of the total atomization energy at the VMC level. The main effect of the three body term is to keep the total charge around the carbon sites to approximately six electrons, thus penalizing the double occupation of the p_z orbitals. Within this scheme we have systematically included in the 3-body Jastrow part the same type of terms present in the AGP one, namely both $g^{a,b}$ and $\lambda^{a,b}$ are non zero for the same pairs of atoms. As expected, the terms connecting next

Table 2.5: Bond lengths (R) in atomic units; the subscript 0 refers to the "exact" results. For the water molecule R is the distance between O and H and θ is the angle HOH (in deg), for CH_4 R is the distance between C and H and θ is the HCH angle.

	R_0	R	θ_0	θ
Li_2	5.051	5.0516(2)		
O_2	2.282	2.3425(18)		
C_2	2.348	2.366(2)		
H_2O	1.809	1.8071(23)	104.52	104.74(17)
CH_4	2.041	2.049(1)	109.47	109.55(6)
	R_0^{CC}	R^{CC}	R_0^{CH}	R^{CH}
C_6H_6	2.640	2.662(4)	2.028	1.992(2)

nearest neighbor carbon sites are much less important than the remaining ones because the VMC energy does not significantly improve (see the full resonating + 3-body wave function in Tab. 2.6). A more clear behavior is found by carrying out DMC simulations: the interplay between the resonance among different structures and the Gutzwiller-like correlation refines more and more the nodal surface topology, thus lowering the DMC energy by significant amounts. Therefore it is crucial to insert into the variational wave function all these ingredients in order to have an adequate description of the molecule. For instance, in Fig. 2.3 we report the density surface difference between the non-resonating 3-body Jastrow wave function, which breaks the C_6 rotational invariance, and the resonating Kekulé structure, which preserves the correct A_{1g} symmetry: the change in the electronic structure is significant. The best result for the binding energy is obtained with the

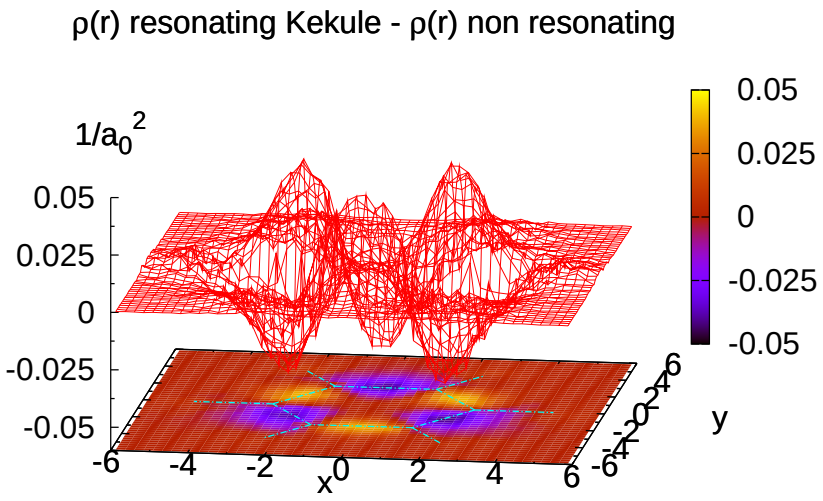


Figure 2.3: Surface plot of the charge density projected onto the molecular plane. The difference between the non-resonating and resonating Kekulé 3-body Jastrow wave function densities is shown. Notice the corresponding change from a dimerized structure to a C_6 rotational invariant density profile.

Kekulé Dewar resonating 3 body wave function, which recovers the $98,6\%$ of the

total atomization energy with an absolute error of 0.84(8) eV. As Pauling [68] first pointed out, benzene is a genuine RVB system, indeed it is well described by the JAGP wave function. Moreover Pauling gave an estimate for the resonance energy of 1.605 eV from thermochemical experiments in qualitative agreement with our results. A final remark about the error in the total atomization energy: the latest frozen core CCSD(T) calculations [100, 102] are able to reach a precision of 0.1 eV, but only after the complete basis set extrapolation and the inclusion of the core valence effects to go beyond the psudopotential approximation. Without the latter corrections, the error is quite large even in the CCSD approach, amounting to 0.65 eV [102]. In our case, such an error arises from the fixed node approximation, whose nodal error is not compensated by the difference between the atomic and the molecular energies, as already noticed in the previous subsection.

Table 2.6: Binding energies in eV obtained by variational (Δ_{VMC}) and diffusion (Δ_{DMC}) Monte Carlo calculations with different trial wave functions for benzene. In order to calculate the binding energies yielded by the 2-body Jastrow we used the atomic energies reported in Tab. 2.1. The percentages ($\Delta_{VMC}(\%)$ and $\Delta_{DMC}(\%)$) of the total binding energies are also reported.

	Δ_{VMC}	$\Delta_{VMC}(\%)$	Δ_{DMC}	$\Delta_{DMC}(\%)$
Kekulé + 2body	-30.57(5)	51.60(8)	-	-
resonating Kekulé + 2body	-32.78(5)	55.33(8)	-	-
resonating Dewar Kekulé + 2body	-34.75(5)	58.66(8)	-56.84(11)	95.95(18)
Kekulé + 3body	-49.20(4)	83.05(7)	-55.54(10)	93.75(17)
resonating Kekulé + 3body	-51.33(4)	86.65(7)	-57.25(9)	96.64(15)
resonating Dewar Kekulé + 3body	-52.53(4)	88.67(7)	-58.41(8)	98.60(13)
full resonating + 3body	-52.65(4)	88.869(7)	-58.30(8)	98.40(13)

The benzene molecule can be idealized by a six site ring Heisenberg model with one electron per site, in order to mimic the out of plane bonds of the real molecule, coming from the p_z electrons and leading to an antiferromagnetic superexchange interaction between nearest neighbor carbon sites. We have studied in this case the spin–spin correlations

$$C(i) = \langle S_0^z S_i^z \rangle, \tag{2.27}$$

where the index i labels consecutively the carbon sites starting from the reference 0, and the dimer–dimer correlations

$$\begin{aligned} D(i) &= D_0(i)/C(1)^2 - 1, \\ D_0(i) &= \langle (S_0^z S_1^z)(S_i^z S_{i+1}^z) \rangle. \end{aligned} \tag{2.28}$$

Both correlation functions have to decay in an infinite ring, when there is neither magnetic ($C(i) \to 0$), nor dimer ($D(i) \to 0$) long range order as in the true spin liquid ground state of the 1D Heisenberg infinite ring. Indeed, as shown in the inset of Fig.(2.4), the dimer–dimer correlations of benzene are remarkably well reproduced by the ones of the six site Heisenberg ring, whereas the spin–spin correlation of the molecule appears to decay faster than the corresponding one of the model. Though it is not possible to make conclusions on long range properties of a finite molecular system, our results suggest that the benzene molecule can be considered closer to a spin liquid, rather than to a dimerized state, because, as well known, the Heisenberg model ground state is a spin liquid and displays spontaneous dimerization only when a sizable next-nearest frustrating superexchange interaction is turned on.[103]

The radical cation $C_6H_6^+$ of the benzene molecule has been the subject of intense theoretical studies[104, 105], aimed to focus on the Jahn-Teller distorted ground state structure. Indeed the ionized $^2E_{1g}$ state, which is degenerate, breaks the symmetry and experiences a relaxation from the D_{6h} point group to two different states, $^2B_{2g}$ and $^2B_{3g}$, that belong to the lower D_{2h} point group. In practice, the former is the elongated acute deformation of the benzene hexagon, the latter is its compressed obtuse distortion. We applied the SR structural optimization, starting from the $^2E_{1g}$ state, and the minimization correctly yielded a deformation toward the acute structure for the $^2B_{2g}$ state and the obtuse for the $^2B_{3g}$ one; the first part of the evolution of the distances and the angles during those simulations is shown in Fig.2.5. After this equilibration, average over 200 further iterations yields bond distances and angles with the same accuracy as the all-electron BLYP/6-31G* calculations reported in Ref. [104] (see Tab. 2.7). As it appears from Tab. 2.8 not only the structure but also the DMC total energy is in perfect agreement with the BLYP/6-31G*, and much better than SVWN/6-31G* that does not contain semi empirical functionals, for which the comparison with our calculation is more appropriate, being fully ab-initio.

Table 2.7: Bond lengths (r) for the two lowest $^2B_{2g}$ and $^2B_{3g}$ states of the benzene radical cation. The angles α are expressed in degrees, the lengths in a_0. The carbon sites are numerated from 1 to 6.

	$^2B_{2g}$	$^2B_{3g}$	Computational method
	acute	obtuse	
$r(C_1 - C_2)$	2.616	2.694	B3LYP/cc-pVTZ [1]
	2.649	2.725	BLYP/6-31G* [2]
	2.659(1)	2.733(4)	SR-VMC [3]
$r(C_2 - C_3)$	2.735	2.579	B3LYP/cc-pVTZ [1]
	2.766	2.615	BLYP/6-31G* [2]
	2.764(2)	2.628(4)	SR-VMC [3]
$\alpha(C_6C_1C_2)$	118.4	121.6	B3LYP/cc-pVTZ [1]
	118.5	121.5	BLYP/6-31G* [2]
	118.95(6)	121.29(17)	SR-VMC [3]

[1] Ref. [105]

[2] Ref. [104]

[3] This work

Table 2.8: Total energies for the $^2B_{2g}$ and $^2B_{3g}$ states of the benzene radical cation after the geometry relaxation. A comparison with a BLYP/6-31G* and SVWN/6-31G* all-electron calculation (Ref. [104]) is reported.

	VMC	DMC	BLYP/6-31G*	SVWN/6-31G*
$^2B_{2g}$	-231.4834(15)	-231.816(3)	-231.815495	-230.547931
$^2B_{3g}$	-231.4826(14)	-231.812(3)	-231.815538	-230.547751

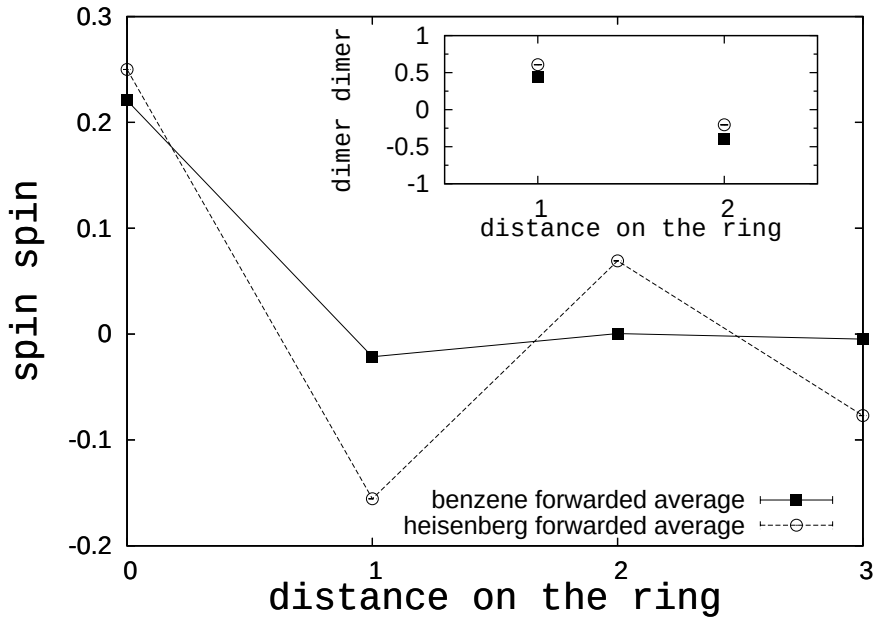


Figure 2.4: Spin–spin correlation function for benzene (full squares) and for the Heisenberg model (empty circles). In the inset, also the dimer–dimer correlation function is reported with the same notation. For the benzene molecule, these correlation are obtained by a coarse grain analysis in which the "site" is defined to be a cylinder of radius $1.3\ a_0$ centered on the carbon nuclei, with a cut off core (i.e. we considered only the points with $|z| > 0.8\ a_0$). All the results are pure expectation values obtained from forward walking calculations.

The difference of the VMC and DMC energies between the two distorted cations are the same within the error bars; indeed, the determination of which structure is the real cation ground state is a challenging problem, since the experimental results give a difference of only few meV in favor of the obtuse state and also the most refined quantum chemistry methods are not in agreement among themselves [104]. A more affordable problem is the determination of the adiabatic ionization potential (AIP), calculated for the $^2B_{3g}$ state, following the experimental hint. Recently, very precise CCSD(T) calculations have been performed in order to establish a benchmark theoretical study for the ionization threshold of benzene [105]; the results are reported in Tab. 2.9. After the correction of the zero point energy due to the different structure of the cation with respect to the neutral

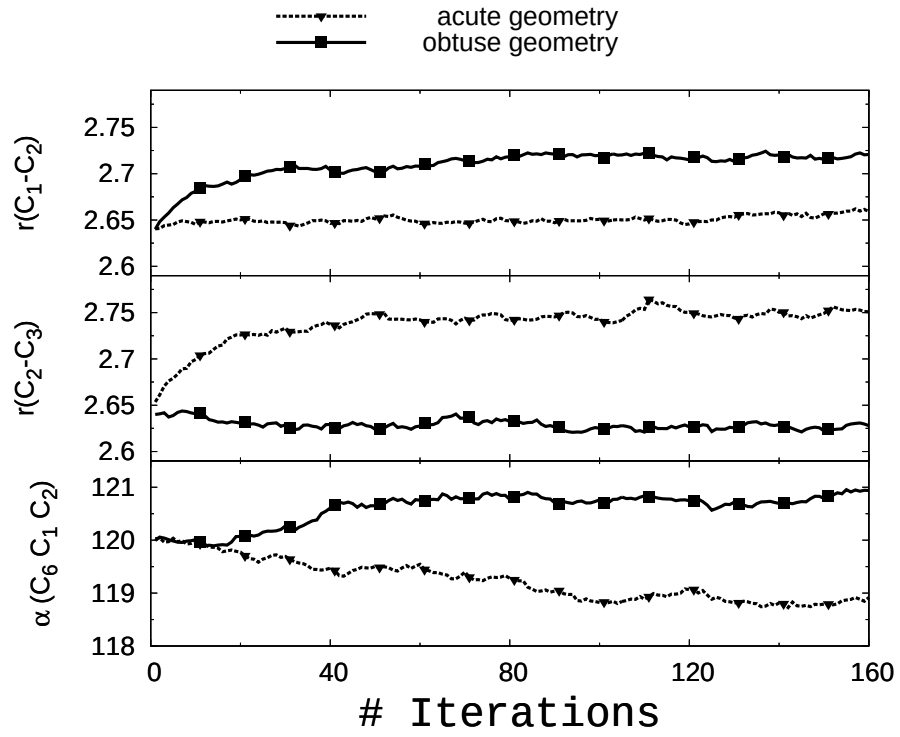


Figure 2.5: Plot of the convergence toward the equilibrium geometry for the ${}^2B_{2g}$ acute and the ${}^2B_{3g}$ obtuse benzene cation. Notice that both the simulations start form the ground state neutral benzene geometry and relax with a change both in the $C-C$ bond lengths and in the angles. The symbols are the same of Tab. 2.7.

molecule and taken from a B3LYP/cc-pVTZ calculation reported in Ref. [105], the agreement among our DMC result, the benchmark calculation and the experimental value is impressive. Notice that in this case there should be a perfect cancellation of nodal errors in order to obtain such an accurate value; however, we believe that it is not a fortuitous result, because in this case the underlying nodal structure does not change much by adding or removing a single electron. Therefore we expect that this property holds for all the affinity and ionization energy calculations with a particularly accurate variational wave function as the one we have proposed here. Nevertheless DMC is needed to reach the chemical accuracy, since the VMC result is slightly off from the experimental one just by few tenths of eV. The AIP and the geometry determination for the $C_6H_6^+$ are encouraging to

pursue this approach, with the aim to describe even much more interesting and challenging chemical systems.

Table 2.9: Adiabatic ionization potential of the benzene molecule; our estimate is done for the $^2B_{3g}$ relaxed geometries of the benzene radical cation, with an inclusion of the zero point motion correction between the $^2B_{3g}$ state and the $^1A_{1g}$ neutral molecule ground state, calculated in Ref. [105] at the B3LYP/6-31G* level.

	VMC [1]	DMC [1]	CCSD(T)/cc-pV∞Z [2]	experiment [3]
AIP	8.86(6)	9.36(8)	9.29(4)	
ΔZPE_{ad}	-0.074	-0.074	-0.074	
best estimate	8.79(6)	9.29(8)	9.22(4)	9.2437(8)

[1] This work

[2] Ref. [105]

[3] Ref. [106]

2.5 Conclusions

In this chapter we have introduced the JAGP variational wavefunction. Both the AGP and the Jastrow play a crucial role in determining the remarkable accuracy of the many-body state: the former permits the correct treatment of the nondynamic correlation effects, the latter allows the wavefunction to fulfill the cusp conditions and makes the convergence of geminal expansion very rapid. Moreover the AGP part is able to include a correlated multiconfiguration state within a numerically feasible scheme, namely by evaluating only a single but appropriately defined determinant, even in the polarized case. In general, therefore, the dependence of the computational time on the number of atoms is comparable with the scaling of the simplest Hartree-Fock calculation.

As is well known the variational energy of the Hartree-Fock wavefunction cannot be improved by extending the variational calculation to a larger basis including all particle-hole excitations applied to the Hartree-Fock state. Analogously, the geminal wavefunction is not only stable with respect to these particle-hole configurations, but also to all possible states obtained by destroying a singlet pair

on some orbital and creating another one on another orbital. Therefore for a two-electrons closed-shell system, like Helium atom, or an ensemble of such non-interacting systems, the AGP wavefunction is formally exact in the asymptotic limit of the geminal expansion. For the other systems, though this wavefunction can take into account a big number of configurations which may allow an energy improvement, obviously it cannot include everything within a single geminal. Indeed there exist multiconfiguration states that are known to be important for atoms like N [107] or for molecules like Be_2, and that involve complicated multi-particle excitations to the Hartree-Fock state. These ones cannot be reduced to creation/destruction of singlet pairs and therefore cannot be handled with a single geminal function. However in our study we have found that the single geminal function with the proper Jastrow factor already provides satisfactory results for all compounds taken into account, yielding more than 91% of the correlation energy in all cases studied by carrying out VMC and DMC simulations. An interesting development of the AGP ansatz has been done by Mitas and coworkers, that used a Jastrow correlated pfaffian wave function, which is able to include not only singlet, but also triplet correlations. It is therefore an extension of the JAGP ansatz, which so far has led to a slight but systematic improvement of the binding energy for some diatomic molecules, whereas in some other cases the energy gain is almost negligible.

The application of the JAGP wave function to atoms is particularly successful for low atomic number, where Hartree-Fock is particularly poor, due to the almost degenerate $2s - 2p$ shells. The case of Beryllium is an interesting benchmark. Indeed, by considering the change of the geminal part altogether with the Jastrow term, we obtained an excellent representation of this correlated atom. Our results are not only comparable but appear even better than the best multideterminantal schemes (using e.g. four Slater determinants), showing that it is possible to represent non trivial correlated states by properly taking into account the interplay of the Jastrow term and the determinantal part of the wavefunction.

The same interplay has proven to be very effective in all molecules studied and particularly in the non trivial case of benzene, where we have shown systematically the various approximations. Only when both the Jastrow and the AGP terms are accurately optimized together, the AGP nodal structure of the wave function is considerably improved. In this way the DMC results can reach the chemical

accuracy and the experimental data for ionization and energy atomization are correctly reproduced. For the above reasons we expect that this wave function should be generally accurate also in complex systems made by many molecules. The local conservation of the charge around each molecule is taken into account by the Jastrow factor, whereas the quality of each molecule is described also by the AGP geminal part.

For an electronic system with many atoms or with an involved molecular structure, the geminal part, when expanded in terms of Slater determinants, yields a very large and non trivial number of configurations, which increases exponentially with the number of atoms considered. The Jastrow factor in this case suppresses the weight of those configurations with two electrons close to the same atomic orbital, correctly describing the effect of the strong Coulomb repulsion. The remarkable advantage of this approach is therefore apparent just for a systems with many atoms, where the JAGP (RVB-like) wavefunction corresponding to an exponentially large number of configurations can be efficiently used for a more accurate description of electron correlation. It is appealing, not only from the computational point of view, that these properties can be obtained by sampling a single determinant wavefunction within the Quantum Monte Carlo techniques.

Chapter 3

Lattice regularized diffusion Monte Carlo

3.1 Introduction

One of the major drawback of the application of QMC techniques to solids and to compounds containing heavy atoms (where for "heavy" we mean an element with atomic number $Z > 10$) is its bad scaling with Z. Ceperley [16] estimated the computational cost for a DMC simulation to grow as $Z^{5.5}$, while Hammond [108] supposed the scaling to be $Z^{6.5}$; very recently, with the increased power of the computational resources, Ma *et al.* [17] verified that the right behaviour is indeed well approximated by $Z^{5.5}$, after performing all electron DMC simulations for noble atoms till Xe. There are two reasons that explain this behaviour. First, close to a nucleus the all electron wave function widely fluctuates, as a consequence of the orthogonality constraint among the core orbitals. Therefore, to sample correctly the core region during a DMC calculation the time step τ must be of the order of Z^{-2}, since the amplitude of the DMC move, proportional to $\sqrt{\tau}$, must have the same magnitude of the characteristic length scale $\xi = 1/Z$ of the core. The smaller is the time step, the greater is the autocorrelation time, and the worse is the efficiency. Second, in the core not only the wave function but also the local energy fluctuates, since the Coulomb attraction between the nucleus and the electrons is strongly divergent and has to be compensated by the kinetic part. Therefore, the variance of the energy increases with the atomic number. The $Z^{5.5}$ behaviour is

obtained by further adding to the correlation-fluctuation contribution also the cubic scaling of the computational cost with the number of electrons (which is taken equal to Z). A way to overcome this drawback is to replace the core electrons by a pseudopotential, also called "effective core potential" (ECP), which has the double effect of reducing the magnitude of the nuclear attraction with a smaller effective charge Z_{eff} and smoothing the single particle orbitals around the nucleus, which now describe only the valence electrons. In this way the scaling of the computational time with the effective charge is much more favorable, being proportional to $Z_{\text{eff}}^{3.4}$ [108], and moreover, unlike Z, Z_{eff} remains a small number for *all* atoms. For example, moving down a column of the Periodic Table, Z_{eff} may generally be held constant. The ECP approximation is usually rather good, since the core electrons are chemically inert and do not significantly contribute to the energetics and the structural properties of a compound. The only role they play is to screen the nuclear charge and to keep the valence electrons in the outer atomic shells. A good pseudopotential has to do exactly the same, and it generally reads:

$$v^P(r) = U_L(r) + \sum_{l=0}^{L-1} \left[U_l(r) - U_L(r)\right] P_l, \tag{3.1}$$

where r is the distance between an electron and the nucleus, $L-1$ is the maximum angular momentum of the core, and P_l is the projection on the momentum l. The ECP usually contains a long range local part (U_L) which screens the nucleus, and a short range non local terms ($U_l - U_L$) which are repulsive and discriminate the electrons on the basis of their angular momentum. The latter sum of Eq. 3.1 takes into account the difference among the angular components of the remaining electrons and is responsible for the non locality of the pseudopotential, since it includes a projection onto the angular momenta.

Also the scalar relativistic effects can be inserted in the ECP form of Eq. 3.1, which instead is too simple to include spin-orbit coupling terms. However spin-orbit effects become apparent on an energy scale usually too small to be seen in a QMC simulation. Another limit of the ECP approximation is the lack of the core-valence correlation. For instance, polarization effects that the core can induce on the valence electrons, are disregarded. Sometimes they are very important, and they can be dealt either by using the core polarization potentials (CPP) beside the ECP, or by reducing the dimension of the core, i.e. by replacing a smaller number

of core electrons.

The ECP of Eq. 3.1 can be easily included in a VMC framework, although the first pseudopotential calculation with the VMC method has been published only in 1990 by Fahy and coworkers [80]. Instead, it is much more difficult to deal with non-local ECP in the standard DMC scheme. The problem comes from the sign of the matrix element $\langle x| \exp(-\tau v^P)|x'\rangle$, where here and in all this chapter x denotes a spatial-spin many-body configuration. Indeed this term is not always positive, and thus it introduces a new kind of "sign problem" in the Green function, which is not related to the antisymmetry of the wave function but to the non-locality character of the potential v^P. In order to overcome this hurdle, the so-called locality approximation (LA) is usually introduced [18, 108–110], which approximates the non-local potential v^P with the local $v^{LA}(x) = \langle x|v^P|\Psi_T\rangle/\langle x|\Psi_T\rangle$. In this way, the off diagonal matrix elements involving v^P are traced in a local potential term, and the sign-flip contributions are now harmless. Therefore, in the case of non-locality the standard DMC approach can be applied only with the LA and the fixed node approximation (FNA) (see Subsection 1.5.2). The price one has to pay is that the mixed-average estimate E_{MA}^{LA} of the total energy, computed during the DMC calculation, is not *variational*. The LA, contrary to the FNA with local potentials, does not provide variational DMC estimates of the true ground state (GS) energy E_{GS} of the Hamiltonian. The only known property of E_{MA}^{LA} is that it equals the exact energy E_{GS} if Ψ_T is exact. Otherwise, it gives no rigorous information about the quality of the approximation. In particular, the ground state Ψ_{FN}^{LA} of the Hamiltonian with the LA can be a worse variational wave function for the true Hamiltonian H, even if it corresponds to a lower E_{MA}^{LA}.

Hence, the use of nonlocal pseudopotentials in DMC is problematic; this has motivated the development of alternative approaches, like the pseudo-Hamiltonian method of Bachelet *et al.* [111], where the non locality is converted in a spatial dependent mass tensor for electrons, and the damped-core method of Hammond [112], in which the core electrons are frozen and treated at the VMC level, while the valence electrons are driven by the DMC process. The former method is not general, since for the most interesting cases, like the transition metals, it is not possible to write a pseudo-Hamiltonian with a positive definite mass tensor, the latter is more computationally demanding than the standard treatment with non local potentials, since the core degrees of freedom are not completely eliminated.

In this Chapter, we present a lattice regularization of the many-electron Hamiltonian which is able to take properly into account the different length scales of the shell atomic structure and removes the above difficulties when using non-local potentials within the FNA. We demonstrate the efficiency of our lattice regularized DMC approach as well as its usefulness in cases where the locality approximation yields inaccurate results.

3.2 Non-local pseudopotentials

ECP are generated within mean field particle theories, such as HF or DFT. Once the all electron orbitals are obtained, the core orbitals are removed and the valence ones are smoothed in the core region, with the norm conserving and continuity constraints [113, 114]. Starting from the new orbitals, the pseudopotentials are generated by requiring that the spectrum of the pseudo hamiltonian matches the single particle energy levels of the all electron problem.

In this thesis only HF pseudopotentials are used since there is a quite general consensus [115] that within a QMC framework they perform better than those generated from DFT atomic calculations. However it is not trivial that effective potentials coming from independent electron theories are well suited for correlated QMC calculations. Alcioli and Ceperley [116] tried to address the problem of generating pseudopotentials within a QMC framework, by taking into account a recipe based on the equivalence of the one body density matrix outside the core region, which is a more general constraint than the usual norm conserving requirement. This scheme turned out to be unfeasible and the problem of the ECP quality for QMC simulations is still an open issue, to be verified a posteriori.

The functions $U_l(r)$ in Eq. 3.1 are generally obtained on a numerical grid, and then fitted to the functional form:

$$v_l(r) = U_l(r) - U_L(r) = \frac{1}{r^2} \sum_k d_{k,l} r^{n_{k,l}} e^{-b_{k,l} r^2}, \tag{3.2}$$

where v_l are the subtracted potentials and $d_{k,l}$, $n_{k,l}$ and $b_{k,l}$ are tabulated parameters. The main difference among the HF pseudopotentials used in this thesis arises from the $n_{k,l}$ and $b_{k,l}$ parameters. The former determine the asymptotic behaviour of the functions v_l at the nucleus, the latter set the core radius r_c, beyond which

the magnitude of the ECP is exponentially negligible. The smaller is r_c, the bigger is the localization of the pseudopotential.

The Stevens, Basch, and Krauss (SBK) [117] pseudopotentials are constructed starting from a single particle radial orbital smoothed in such a away that $\chi_l(r) \propto r^3$ for $r \to 0$. Consequently, the SBK pseudopotentials are "hard"; they diverge as $1/r^2$, i.e. some terms in the non local components have $n_{k,l} = 0$. This behaviour turns out to be cumbersome in a DMC simulation, because the fluctuations of the local energy in the core region are enhanced by the strong divergence of the ECP angular channels, only alleviated by the cusp conditions one has to enforce with a proper selection of the single particle orbital basis, or with a proper choice of the Jastrow factor (see Chap. 4).

The pseudopotentials developed by the Dolg's group are more suited for QMC calculations. The non local components are all finite ($n_{k,l} = 2 \ \forall k \ \forall l$), and only the local part diverges as $-Z_{\text{eff}}/r$ (see Fig. 3.1). The Dolg's ECP have proven to be reliable and accurate enough to describe transition metal atoms [118] and compounds [119], also within a QMC framework. Moreover their overall divergence is milder than SBK's, and the nuclear cusp conditions can be implemented easier into the wave function. In the presence of Dolg's ECP, we chose to work with a Gaussian basis both in the AGP and in the three-body Jastrow geminal expansion (see Section 2.2), and fulfill the nuclear cusp conditions through a proper one body electron-nuclear Jastrow factor (see Subsection 4.2.1).

The Lester's group [115, 120] proposed another kind of pseudopotentials, expressly conceived to fit the QMC necessities. Lester used a functional form for the radial pseudo orbitals χ_l at short range such that $\chi_l(r) \propto r^l$, as $r \to 0$. In this way, he ended up with a "soft" ECP, i.e. not only the non local channels are not singular, but also the local term U_L exactly cancels the Coulomb attractive divergence of the effective nuclear charge (see Fig. 3.1). Thus, the local energy around the nuclear positions will be less fluctuating and a larger time step can be used in DMC simulations. Moreover the nuclear cusp conditions do not need to be applied.

Sometimes we employed also pseudopotentials generated by using the E. Shirley's code, based on the the construction of D. Vanderbilt [121] to build norm-conserving HF ECP.

The general expression of an Hamiltonian containing pseudopotentials reads

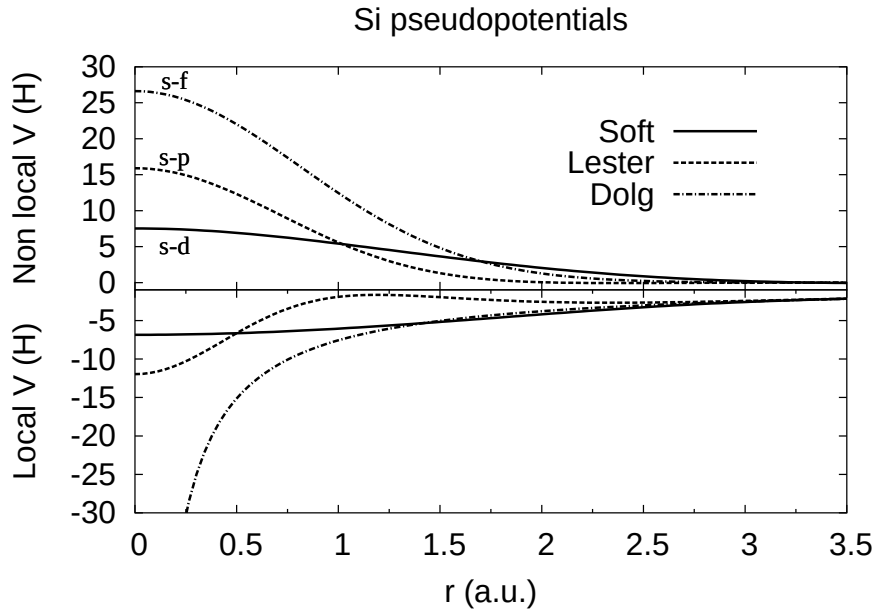


Figure 3.1: Local component (lower panel) and first non local component for the Shirley's (labeled in the figure as "Soft"), Lester's, and Dolg's pseudopotentials for silicon (with a neon core). Since the Shirley's non local components are s and p like, the first non local function is $s-d$ (see Eq. 3.1 and Eq. 3.2), while for Lester's ECP is $s-p$ (only an s like non local component), and for Dolg's ECP is $s-f$ (s, p and d like non local components). Notice that the latter diverges at the origin in the local part. The core radius r_c is 3 a.u. for Shirley's, 2.5 a.u for Dolg's, and 1.8 a.u. for Lester's ECP. The most repulsive is the Dolg's one.

in atomic unit:

$$\begin{aligned} H &= -\frac{1}{2}\sum_i \nabla_i^2 + \sum_{i\neq j}\frac{1}{|x_i - x_j|} + \sum_{n\neq m}\frac{Z^n_{\text{eff}} Z^m_{\text{eff}}}{|R_n - R_m|} - \sum_n\sum_i \frac{Z^n_{\text{eff}}}{|x_i - R_n|} + V^P, \\ V^P &= \sum_n\sum_i v_n^P(|x_i - R_n|), \end{aligned} \tag{3.3}$$

where x_i is the position of the i-th electron, R_n is the position of the n-th ion, Z^n_{eff} is its nuclear charge and v_n^P is the pseudopotential centered on R_n, which has the form reported in Eq. 3.1. The projection onto the l-th angular momentum acting

on the i-th coordinate of a many-body state Ψ can be rewritten as:

$$\begin{aligned} P_l|\Psi\rangle &= \sum_{m=-l}^{l} |lm\rangle\langle lm|\Psi\rangle \\ &= \sum_{m=-l}^{l} Y_{lm}(\Omega_{x_i}) \int d\Omega_{x_i'}\, Y_{lm}^*(\Omega_{x_i'})\Psi(x_1,\ldots,x_i',\ldots,x_N) \\ &= \frac{2l+1}{4\pi} \int d\Omega_{x_i'}\, P_l(x_i\cdot x_i')\Psi(x_1,\ldots,x_i',\ldots,x_N), \end{aligned} \tag{3.4}$$

where for the sake of simplicity the projector is centered on the origin of the reference frame, $\int d\Omega_{x_i'}$ indicates the spherical integration around the origin with radius $|x_i'| = |x_i|$, Y_{lm} and $P_l(x_i \cdot x_i')$ are the spherical harmonics and Legendre polynomials respectively. The last identity of Eq. 3.4 follows from a well known property of the spherical harmonics [122]:$\sum_{m=-l}^{l} Y_{l,m}^*(r_i)\cdot Y_{l,m}(r_i') = P_l(r_i\cdot r_i')$. The general expression for the pseudopotential $v_n^P(|x_i - R_n|)$ in Eq. 3.3 contains both a local and a non-local part, where the angular momentum projectors act on the electron coordinate x_i by means of spherical integrations around the nuclear position R_n.

In the first applications of ECP-QMC calculations [108, 109], the spherical integral in Eq. 3.4 was performed analytically but the formula used was valid only for a pure determinantal trial function. Indeed, both Hurley [109] and Hammond [108] neglected the effect of the Jastrow factor in the pseudopotentials, in order to find a way to integrate out the angular momenta. This further approximation was overcome few years later, in 1990, when Fahy *et al.* [80] proposed to integrate numerically the whole many-body wave function, without discarding the Jastrow factor and with the possibility to include in the spherical integral the most general trial function. They applied a Gaussian quadrature rule [123] suited for integration on a spherical surface, and they proved that an unbiased result can be obtained by randomly selecting the spherical grid each time the integral is performed. Moreover they suggested to compute the same integral more than once at each Monte Carlo step with different random meshes and to take the averaged result, in order to further increase the accuracy of the integration. However the work by Mitas and Ceperley [18] clarified that a larger number of mesh points is more efficient than computing many times the spherical integral with different but smaller meshes, since the accuracy of the numerical integration scales faster with the number of

mesh points than it does with the number of angular samples.

In our thesis, we followed the integration scheme proposed by Mitas and Ceperley [18]. The angular integration to evaluate the projection P_l is performed by a single numerical quadrature on a regular polyhedron defined by N_V vertices, with tetrahedrical ($N_V = 4$), octahedrical ($N_V = 6$, $N_V = 18$), or icosahedrical ($N_V = 12$) symmetry, depending on the required accuracy and on the maximum angular momentum of the core electrons. Meshes with $N_V = 4$, $N_V = 6$, and $N_V = 12$ give exact integration of all functions with angular momentum $l \leq 2, 3,$ and 5, respectively [123]. Notice however that a Jastrow correlated wave function contains angular momenta bigger than the maximum angular momentum in the single particle basis, since the exponent in the Jastrow factor yields a much more complex angular dependence. Therefore, for a general wave function the numerical quadrature is not exact and in order to reduce the bias the mesh is selected randomly, with a uniform deviate for θ ($\in [0, \pi[$) and ϕ ($\in [0, 2\pi[$) angles, where, as usual, $r_z = r\cos\theta$, $r_x = r\sin\theta\ \cos\phi$ and $r_x = r\sin\theta\ \sin\phi$. Once the symmetry of the quadrature is chosen according to the desired numerical precision and to the angular symmetry of the core, the projection in Eq. 3.4 is rewritten as a sum of N_V terms:

$$P_l|\Psi\rangle = (2l+1)\sum_{m=1}^{N_V} a_m P_l(x_i \cdot x_i^m)\Psi(x_1, \ldots, x_i^m, \ldots, x_N), \qquad (3.5)$$

where a_m are the weights of the quadrature, and $x_i^m = x_i + \delta^m$ are the points of the mesh. The Eq. 3.5 is taken as the *definition* of the angular momentum projection.

3.3 Locality approximation

As we have mentioned in the Introduction, if the non local pseudopotentials are included in the Hamiltonian, the standard DMC method cannot be used, since it is not able to deal with non locality. Indeed one can show that the matrix elements $\langle x|\exp(-\tau V^P)|x'\rangle$ of the Green function are not positive definite, even in the simple case of one electron with a non local pseudopotential V^P. In such a case, it is quite easy to prove[124] that:

$$\langle x|\exp(-\tau V^P)|x'\rangle = \sum_l \frac{2l+1}{4\pi}\exp\left[-\tau U_l(x)\right] P_l(x \cdot x'), \qquad (3.6)$$

where $|x| = |x'|$, which is clearly non positive since the Legendre polynomial can take both positive and negative values. Therefore, the Green function cannot be interpreted as a transition probability, and a difficulty similar to the "fermionic sign problem" arises.

In order to overcome the problem, the locality approximation (LA) has been introduced [18, 108, 109], which consists to replace the non local potential V^P in the Hamiltonian 3.3 with a local one, obtained by projecting V^P onto the guidance wave function Ψ_T:

$$V^P_{LA}(x) = \frac{\langle x|V^P|\Psi_T\rangle}{\langle x|\Psi_T\rangle}. \tag{3.7}$$

In Eq. 3.7 the non local operators act on the trial wave function, which is used to integrate out the non diagonal matrix elements $\langle x|V^P|x'\rangle$. The resulting potential is local, since it depends only on the position x, where the wave function is evaluated. Moreover, the Hamiltonian H^{LA} including V^P_{LA} is a local model which *approximates* the true non local Hamiltonian H. Already the earliest ECP-DMC calculations [108, 109] exploited the LA to include non local potentials into the DMC framework, although Mitas and Ceperley[18] were the first to clearly state the nature of the approximation. If we rewrite Eq.1.67 in Subsec. 1.5.1 by taking into account the difference between the local and non local part and inserting the local potential V^P_{LA}, we obtain:

$$\begin{aligned} -\partial_t f(\mathbf{R},t) &= -\frac{1}{2}\nabla^2 f(\mathbf{R},t) + \nabla\cdot[\mathbf{v}_{\text{diff}} f(\mathbf{R},t)] + \left[\frac{H^{LA}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} - E_T\right] f(\mathbf{R},t) \\ &+ \left(\frac{V^P\Phi(\mathbf{R},t)}{\Phi(\mathbf{R},t)} - \frac{V^P\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}\right) f(\mathbf{R},t), \end{aligned} \tag{3.8}$$

where $f(\mathbf{R},t) = \Psi_T(\mathbf{R})\Phi(\mathbf{R},t)$ is the steady distribution sampled in principle by the DMC process. The Eq. 3.8 describes the exact evolution, which includes a non local branching term given by $\frac{V^P\Phi(\mathbf{R},t)}{\Phi(\mathbf{R},t)}$, where the non local potential V^P is applied to the unknown wave function Φ. The LA consists of neglecting the last line of Eq. 3.8: $\left(\frac{V^P\Phi(\mathbf{R},t)}{\Phi(\mathbf{R},t)} - \frac{V^P\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}\right) f(\mathbf{R},t)$. Notice that after disregarding this term, the time evolution for $f(\mathbf{R},t)$ becomes identical to the usual DMC evolution, but with H^{LA} in the place of H. It has proven[18] that if the trial wave function is accurate, the convergence of the approximated energy $E_{LA} = \langle\Psi_T|H^{LA}|\Phi_{LA}\rangle/\langle\Psi_T|\Phi_{LA}\rangle$ to the true GS energy E_0 is quadratic in the difference between the trial and the exact wave function.

An important drawback of the LA is that E_{LA} is not variational, i.e. it can be lower than the GS energy E_0 of H. This is a consequence of the fact that a DMC simulation is able to access only the mixed averages and not the pure ones (see Subsection 1.5.3). However, in the case of local potentials, if Φ is the GS of the fixed node Hamiltonian H_{FN}, i.e. the Hamiltonian with the FN boundary conditions, the mixed average E_{MA} will coincide with the pure expectation value E_{FN}:

$$E_{MA} = \frac{\langle \Psi_T | H_{FN} | \Phi \rangle}{\langle \Psi_T | \Phi \rangle} = \frac{\langle \Phi | H_{FN} | \Phi \rangle}{\langle \Phi | \Phi \rangle} = E_{FN} \geq E_0, \tag{3.9}$$

and the variational property of E_{MA} is verified. On the other hand, if the LA is introduced in a FN DMC calculation, the resulting state Φ_{LA} will be the GS of the LA FN Hamiltonian H_{FN}^{LA} which *differs* from H_{FN}. Now, the following identity holds:

$$E_{LA} = \frac{\langle \Psi_T | H_{FN}^{LA} | \Phi_{LA} \rangle}{\langle \Psi_T | \Phi_{LA} \rangle} = \frac{\langle \Psi_T | H_{FN} | \Phi_{LA} \rangle}{\langle \Psi_T | \Phi_{LA} \rangle} = E_{MA}, \tag{3.10}$$

as the local energy $H^{LA}\Psi_T/\Psi_T$ of the effective LA Hamiltonian is equal to the local energy $H\Psi_T/\Psi_T$ of the true Hamiltonian, but in this case E_{MA} is not equal to the variational E_{FN} since Φ_{LA} is *not* the GS of the FN Hamiltonian H_{FN}. Hence the variational property for E_{MA} in the presence of the LA cannot be proven.

From Eq. 3.7 it is apparent that the local potential diverges on the nodes of the trial wave function Ψ_T. Near a point on the nodal surface one can separate the coordinates into the normal direction (denoted by r_n) and all the other $3N-1$ parallel directions. The local potential close to the node, written in terms of these new coordinates, reads:

$$V_{LA}^P = c_0/r_n + c_1, \tag{3.11}$$

where c_0 and c_1 are constants. Since c_0 can be both positive and negative, depending on the position on the nodal surface, V_{LA}^P includes both attractive and repulsive singularities. Of course, this fictitious behaviour is a consequence of the LA, and will lead to an unstable DMC simulation, if the trial wave function is not vanishing linearly on the same nodal points and the FN approximation is not applied. Indeed, by applying the fixed node boundary conditions, the FN GS wave function Φ vanishes linearly near the nodal surface, i.e. the walkers never reach the node and the simulation turns out to be feasible. However, around the nodal region V_{LA}^P widely fluctuates, and the efficiency and stability of the standard DMC algorithm

can be spoiled, for not so small time steps τ and for poor trial wave functions. This effect is particularly enhanced close to the nucleus of a pseudo atom, where the pseudopotentials are repulsive if not divergent (like for instance the SBK pseudopotentials, see Section 3.2) and the coefficient c_0 can be very large. Therefore an accurate Ψ_T is required, in particular in the core region, in order to avoid large fluctuations of the local energy. In the case of HF pseudopotentials we have found that the HF wave function is a good starting point for the construction and optimization of Ψ_T, since it avoids strong fluctuations in the core. In Fig. 3.2, we plot the local energy during a DMC simulation carried out for the SBK pseudo carbon, using a poor trial wave function, and we compare it with a LRDMC simulation for the same atom. As we will see later, the LRDMC scheme gives always stable simulation, as a consequence of the variational treatment of the non local pseudopotentials.

Not only the stability of the simulation but also the time step error is significantly affected by the LA. Indeed, the DMC energy dependence on the time step, being usually quadratic for a trial wave function which fulfills the cusp conditions, becomes linear in the presence of non local pseudopotentials. Moreover, for large time step the exact Green function is poorly sampled, due to the Trotter approximation, and it can happen that near the nodal surface a walker jumps too close to the node, thus causing a spike in the local energy and the instability of the simulation if the local energy is not properly cut off. From this point of view, the FN approximation, which prevents the walker from crossing the node, is crucial to let the LA works in practice. Moreover the LA is more strong than the FN approximation, since the latter depends only on the nodal structure of Ψ_T, the former depends on the topology of the whole trial wave function, which is exact if the last line of Eq. 3.8 vanishes for *each* $\mathbf{R}$.

With the aim to solve some of the problems related to the LA, Ceperley and Mitas in 1996 [125] suggested to use also on the continuum the same effective Hamiltonian approach used on a lattice to obtain the upper bound property even in the presence of non local operators (see Subsection 1.6.2). They proposed to apply the LA only to those terms which are positive in the Hamiltonian and to sample the hopping terms coming from negative non local elements in the Hamiltonian, i.e. positive non local elements in the Green function. In practice, they defined an effective continuous Hamiltonian with a sign flip term (see Section 1.6.2) to

be treated in a DMC framework with non local operator sampling. They stated that the sign flip term will allow to preserve the upper bound property as in lattice Hamiltonians, but the exact sampling of non local will produce an estimator of the energy with non zero variance, even in the limit of the exact trial wave function.

In the next section we want to show that it is possible to conceive a Monte Carlo projection algorithm, which is able to deal with non local potentials without compromising the validity of both the upper bound and zero variance properties. Moreover, this novel scheme turns out to be efficient and much more stable, if compared to the standard DMC algorithm with the locality approximation.

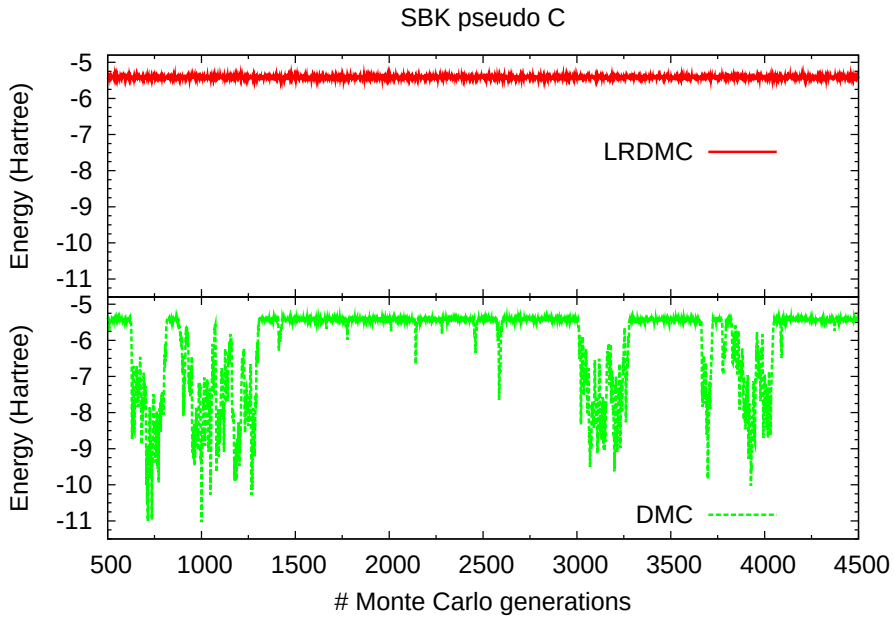


Figure 3.2: Evolution of the local energy during LRDMC (upper panel) and DMC (lower panel) simulations for carbon atom with SBK pseudopotentials (see Section 3.2). Locality approximation is used in the DMC framework, leading to large fluctuations in the local energy. Instead, the LRDMC yields a steady evaluation of the energy.

3.4 Lattice regularized diffusion Monte Carlo

This section is divided into three parts. The first one is devoted to the mapping of a continuous system into a lattice regularized Hamiltonian H_a, which equals the continuous Hamiltonian H as the lattice space a goes to zero. The second part is the description of the lattice regularized diffusion Monte Carlo method, based on the machinery of the GFMC algorithm applied to H_a in order to find out its GS. In the third part we explain how to compute the pure expectation value of H_a (which cannot be evaluated for H using the standard DMC), by exploiting the well known properties of the effective lattice Hamiltonian already described in Sec. 1.6.3.

3.4.1 Regularization of the Hamiltonian

We consider the continuous Hamiltonian in atomic units:

$$H = -\frac{1}{2}\sum_i \Delta_i + V(x)\,, \tag{3.12}$$

where as usual $-\frac{1}{2}\sum_i \Delta_i$ is the kinetic part and $V(x)$ represents the many body potential. We neglect the presence of an external magnetic field, and the wave function and all the terms of the Hamiltonian can be taken real. In the following, first we discretize the kinetic term, and then we regularize the potential part.

Discretization of the Laplacian

We approximate the Laplacian by a finite difference form with two mesh sizes a and $a'(> a)$, where a/a' is a constant:

$$\Delta_i \approx \Delta_i^a = \Delta_i^{a,p} + \Delta_i^{a',1-p} + O(a^2). \tag{3.13}$$

$\Delta_i^{a,p}$ is a Hermitian lattice operator, and p ($0 \leq p \leq 1$) is an arbitrary function included in $\Delta_i^{a,p}$, which depends on the electron position x. In Eq. 3.13 the discretized kinetic operator Δ_i^a has been split into two terms, one with lattice space a and weighted by p, the other with lattice space a' and weighted by $1-p$. In the notation, we have highlighted only the dependence on a for Δ_i^a, since the dependence on a' can be rewritten as function of a, being the ratio a/a' kept constant. The single mesh Laplacian operator $\Delta_i^{a,p}$ is defined as follows:

$$\Delta_i^{a,p} f(x_i, y_i, z_i) \;=\; \frac{1}{a^2}\{p(x_i + a/2)\,[f(x_i + a) - f(x_i)] \tag{3.14}$$

$$+ \quad p(x_i - a/2)\,[f(x_i - a) - f(x_i)]\}$$
$$+ \quad x_i \leftrightarrow y_i \leftrightarrow z_i,$$

where $\mathbf{r}_i \equiv (x_i, y_i, z_i)$ is the position of the i-th electron. For $p = 1$, $\Delta_i^{a,p}$ coincides with the usual discretized form of the Laplacian on a lattice with mesh size a [1], and we end up with only the mesh a, being the second term in Eq. 3.13 vanishing. On the other hand, for $p = 0$ only the terms containing the mesh a' survive, and in this case a' is the unique mesh size. The most interesting situation is the intermediate case, when both $\Delta_i^{a,p}$ and $\Delta_i^{a',1-p}$ play a role. If we rewrite the matrix elements $\langle x| - \frac{1}{2}\sum_i \Delta_i |x'\rangle$ of the continuous kinetic operator by using the definition of the kinetic terms in Eq. 3.13 and the lattice modified Laplacian operator in Eq. 3.14, we obtain an expression valid up to the second order in a or a':

$$\langle x| - \frac{1}{2}\sum_i \Delta_i^a |x'\rangle = \begin{cases} -\frac{1}{2a^2} p(x + \delta_a/2) & \text{if } x' = x + \delta_a \\ -\frac{1}{2a'^2}(1 - p(x + \delta_{a'}/2)) & \text{if } x' = x + \delta_{a'} \\ \sum_i \sum_{\mu\in\{x,y,z\}} \left(\frac{p(\mu_i + a/2) + p(\mu_i - a/2)}{2a^2} \right. & \\ \left. + \frac{2 - p(\mu_i + a/2) - p(\mu_i - a/2)}{2a'^2} \right) & \text{if } x' = x \\ 0 & \text{otherwise} \end{cases}, \tag{3.15}$$

where δ_a is a $3N$ dimensional vector defined as the displacement of *one* particle in *one* direction with length $\pm a$. Therefore there are $6N$ different δ_a, and given the position x, with the double mesh discretization there are $12N$ points x' connected to x through an hopping term in Eq. 3.15. Notice that although p depends on the position x, the discretized kinetic operator $-\frac{1}{2}\sum_i \Delta_i^a$ is Hermitian, since it is easy to show from Eq. 3.15 that $\langle x| - \frac{1}{2}\sum_i \Delta_i^a |x'\rangle = \langle x'| - \frac{1}{2}\sum_i \Delta_i^a |x\rangle$.

Now let us suppose to simulate a Markov process based on the Green function

$$G_{x,x'} = \langle x|\Lambda\delta_{x,x'} + \frac{1}{2}\sum_i \Delta_i^a |x'\rangle, \tag{3.16}$$

constructed by including the discretized kinetic operator and by neglecting for the moment the potential term in Eq. 3.12. The transition probability of such a process is given by:

$$p_{x,x'} = \frac{G_{x,x'}}{b_x}, \tag{3.17}$$

[1] The discretization of the second derivative $\frac{\partial}{\partial x} f(x)$ valid up to the second order in the lattice space a is $\frac{1}{a^2}\left[f(x + a) + f(x - a) - 2f(x)\right] + O(a^2)$

where $b_x = \sum_{x'} G_{x,x'}$ is the normalization, like in the GFMC algorithm (see Eq. 1.83 in Sec. 1.6). Notice that here b_x can be evaluated although we are on the continuum, since after the discretization the connectivity of the Hamiltonian is *finite*, i.e. the number of configurations x' connected to the given position x is finite. If $p = 0$ or $p = 1$, the transition probability matrix of Eq. 3.17 allows only moves with the same amplitude (a' or a), and thus belonging to the same lattice. Instead, if p assumes intermediates values, the transition matrix yields a diffusion sometimes driven sometimes by a, sometimes by a'. Once again, if a' is a multiple of a, the evolution remains on the same lattice, instead if a and a' are incommensurate, the alternate application of $\Delta_i^{a,p}$ and $\Delta_i^{a',1-p}$ yields a randomization of the electron positions all over the continuous space. Therefore we take a/a' irrational. The diffusion process based on these two meshes overcomes the lack of ergodicity present in a strict lattice evolution, and the electron coordinates $\{\vec{r}_i\}$ assume values practically indistinguishable from the continuum. As can be seen in Fig. 3.3, once the double mesh is included in the LRDMC framework, a large reduction of the step-size error occurs, if compared with the use of only one lattice.

Moreover, this scheme is able to account for the different length scales in the system. Indeed, since p can depend on the position x, its functional form can weight the contributions of the two meshes so that the smaller step a is used close to a nucleus and the larger a' far away from the nuclei, with a clear reduction of the autocorrelation time in particular for compounds containing heavy atoms. The optimal function p and ratio a'/a are found to be:

$$p(\vec{r}) = 1/(1 + Z^2|\vec{r} - \vec{R}|^2/4)\,, \tag{3.18}$$

$$a'/a = \sqrt{Z^2/4 + 1}\,, \tag{3.19}$$

where $\vec{R}$ and Z are the position and the atomic number of the nucleus closest to the electron in $\vec{r}$.

Once the discrete Laplacian operator is introduced, the true kinetic term is different from the discretized one by order of a^2. In particular, the virial theorem is no longer satisfied, and this can spoil the results. In order to reduce this bias, the kinetic operator $-\frac{1}{2}\sum_i \Delta_i^a$ is rescaled by a prefactor η, in such a way that:

$$\langle \Psi_T | \sum_i \Delta_i | \Psi_T \rangle = \eta \langle \Psi_T | \sum_i \Delta_i^a | \Psi_T \rangle. \tag{3.20}$$

Thus, η is determined by requiring that the discretized kinetic energy is equal to the continuous one calculated on the state Ψ_T, and it is then simple to show that this prefactor behaves as $1 + O(a^2)$, being irrelevant in the limit $a \to 0$, when the discretized and continuous kinetic operators are equal.

Regularization of the potential

To further improve the accuracy of the approximation and work with larger values of a, we regularize also the potential $V \to V^a$ by requiring that, for the chosen guiding WF Ψ_T, the local energy $\langle x|H^a|\Psi_T\rangle/\langle x|\Psi_T\rangle$ of the Hamiltonian $H^a = -1/2\sum_i \Delta_i^a + V^a$ equals for each value of a the local energy of the continuous Hamiltonian H. This condition leads to the following solution for V^a:

$$V^a(x) = V(x) + \frac{1}{2}\left[\frac{\sum_i(\Delta_i^a - \Delta_i)\Psi_T}{\Psi_T}\right](x). \tag{3.21}$$

Note that the correct limit $H^a \to H$ for $a \to 0$ is preserved and that the regularization of V yields another important property for H^a: if Ψ_T is an eigenstate of H, it is also an eigenstate of H^a for any a, as can be easily derived using that Δ^a is Hermitian. Thus, by improving Ψ_T, a better $a \to 0$ convergence is also expected.

The inclusion of non local pseudopotentials in this novel framework is straightforward. Indeed, from the Eq. 3.5 in Sec. 3.2, it turns out that the non local operator V^P –the total non local pseudopotential in Eq. 3.3– acts on a many-body configuration x by means of a *finite* number of matrix elements equal to $N_V N_{core}$, where N_{core} is the number of electrons in the configuration x within the core radius of a pseudoatom[2]. In particular, the off diagonal matrix element connecting the configurations x and x' reads:

$$\langle x'|V^P|x\rangle = a_{x,x'}\sum_{l=0}^{L-1}(2l+1)\; v^l_{n(x,x')}(|x - \mathbf{R}_{n(x,x')}|)\; P_l(x \cdot x'), \tag{3.22}$$

where $n(x, x')$ is the nucleus around which x is rotated to x', $v_n^l(r)$ is the radial non local component of the pseudopotential, $|x - \mathbf{R}_{n(x,x')}|$ is the distance between the

[2]Here we suppose that we are dealing with only one pseudoatom. For instance, in the case of two pseudoatoms (A and B), given the confi guration x, the number of matrix elements connecting another confi guration x' is $N_V(N^A_{core} + N^B_{core})$, where N^A_{core} (N^B_{core}) is the number of electrons of x within the core radius of atom A (B).

configuration x (or x') and the nucleus $n(x, x')$, and finally $a_{x,x'}$ is the quadrature coefficient relative to the rotation $x \rightarrow x'$.

Therefore, we proceed as if each atomic pseudopotential were by definition discretized with N_V points (see Eq. 3.5) so that V^a becomes:

$$\langle x'|V^a|x\rangle = V^a(x)\delta_{x,x'} + \langle x'|V^P|x\rangle, \tag{3.23}$$

where $V^a(x)$ is the local component of Eq. 3.21 and the configuration x' is defined on a mesh with steps determined by the chosen numerical quadrature. In this way, we have three meshes, the two kinetic meshes with steps a and a', and one pseudopotential mesh.

3.4.2 The algorithm

Although H^a is an Hamiltonian defined on a continuous space, all techniques valid on a lattice can be straightforwardly applied here since H^a acts on a configuration exactly as a lattice Hamiltonian, namely:

$$\langle x|H^a|\Psi_T\rangle = \sum_{x'} H^a_{x,x'}\langle x'|\Psi_T\rangle\,, \tag{3.24}$$

where, for a given x, the number of matrix elements $H^a_{x,x'}$ are *finite* even in the presence of non-local pseudopotentials. In particular, we can resort to the same scheme used in the efficient lattice Green function Monte Carlo algorithm [47, 49, 50]. The resulting algorithm, valid on the continuum, which exploits the lattice GFMC scheme applied to the lattice regularized Hamiltonian H^a, has been called lattice regularized diffusion Monte Carlo (LRDMC) method. The LRDMC evolution is based on the importance sampling Green function:

$$G_{x',x} = \Psi_T(x')(\Lambda\delta_{x',x} - H^a_{x',x})/\Psi_T(x), \tag{3.25}$$

which is a discrete matrix with x and x' defined on the continuum. If all the matrix elements of Eq. 3.25 are non-negative, the positive distribution $\Psi_T(x)\Psi_{GS}(x)$ is statistically sampled by the LRDMC projection, without applying the FN approximation. As in the lattice GFMC algorithm, Λ must be chosen sufficiently large, in order to project the starting trial function Ψ_T to the lowest state Ψ_{GS} of H^a. Note that, in general, the spectrum of a continuum Hamiltonian is not bounded from

above. As a consequence, also the spectrum of the lattice regularized Hamiltonian H^a is not bounded, and hence in the LRDMC framework we need to take the limit $\Lambda \to \infty$, which can be handled with no loss of efficiency as described in Ref. [49] and in Sec. 1.6. The LRDMC algorithm is outlined in Tab. 3.1. Note that nor the GFMC neither the LRDMC algorithm has a time step error, since the Trotter approximation is not employed and the Green function is exactly sampled.

Since the Green function $G_{x',x}$ can be made strictly positive only for bosons, we have to introduce here the analogous of the FNA on a lattice [47, 49, 50] and modify few of the matrix elements of the Hamiltonian H^a. For each configuration x, the matrix elements $H^a_{x',x}$ which yield $G_{x',x} < 0$ are set to zero and included in the so called sign-flip term, $\mathcal{V}_{sf}(x) = \sum_{x' \neq x}, \Psi_T(x')H^a_{x',x}/\Psi_T(x) > 0$, which is then added to the diagonal element $H^a_{x,x}$ [50]. The resulting effective Hamiltonian H^{eff} is:

$$H^{\text{eff}}_{x,x'} = \begin{cases} H^a_{x,x} + \mathcal{V}_{sf}(x) & \text{if } x = x' \\ H^a_{x,x'} & \text{if } x \neq x' \text{ and } \Psi_T(x')H^a_{x,x'}/\Psi_T(x) \leq 0 \\ 0 & \text{if } x \neq x' \text{ and } \Psi_T(x')H^a_{x,x'}/\Psi_T(x) > 0, \end{cases} \tag{3.26}$$

which has the same local energy as H^a and its ground state wave function has the same signs as the trial wave function Ψ_T.

The GS energy of H^{eff} can be efficiently computed with the mixed average estimator E^{eff}_{MA} which also equals the mixed estimator for the Hamiltonian H^a:

$$E^{\text{eff}}_{MA} = \frac{\langle \Psi_T | H^{\text{eff}} | \Psi^{\text{eff}}_{FN} \rangle}{\langle \Psi_T | \Psi^{\text{eff}}_{FN} \rangle} = \frac{\langle \Psi_T | H^a | \Psi^{\text{eff}}_{FN} \rangle}{\langle \Psi_T | \Psi^{\text{eff}}_{FN} \rangle} \tag{3.27}$$

where Ψ^{eff}_{FN} is the GS of H^{eff}. As in Eq. 3.10, the above identity holds since the local energy $H^{\text{eff}}\Psi_T/\Psi_T$ of the effective Hamiltonian H^{eff} equals the local energy $H^a\Psi_T/\Psi_T$ of the regularized Hamiltonian H^a. For a local Hamiltonian H, we recover the standard DMC result $E_{MA} = E_{FN}$ in the limit $a \to 0$ as shown in Fig. 3.3. Indeed if the potential is local, the non local matrix elements of H^{eff} come only from the discretized Laplacian operator, but in the limit $a \to 0$ all the hopping terms connect points within the same nodal region, and the FN constraint is automatically satisfied. Therefore, in the case of local potential, the following identities hold:

$$\lim_{a \to 0} H^{\text{eff}} = \lim_{a \to 0} H^a = H, \tag{3.28}$$

Table 3.1: Schematic chart of the LRDMC method. We use the algorithm with a fixed number of walkers. Between two consecutive branchings, each walker evolves for a time T, chosen a priori in such a way that at least 90% of walkers survive in average after a branching. T_i^{left} is the remaining time for the walker i before the next branching.

initial weights: $w_i = 1\ \forall i$.
initial remaining time: $T_i^{\text{left}} = T\ \forall i$.
loop over LRDMC generations with index g
  loop over the walkers with index i
  ▷ Given the walker i with configuration x, compute the persistence time τ_x, during which the walker stays in that configuration:

$$\tau_x = min[-log(r)/N_x, T_i^{\text{left}}],$$

  where r is a random number with uniform deviate $0 < r \leq 1$, and $N_x = \sum_{x'(\neq x)} G_{x',x}$ is the normalization of the off diagonal Green function $G_{x',x}$.
  ▷ Update the weight of the walker:

$$w_i = w_i\ \exp\left[-\tau_x(E_L(x) - E_T)\right],$$

  where $E_L(x)$ is the local energy, and E_T is the guessed ground state energy, given from input, which avoids overflows or underflows of the weights.
  ▷ Update the remaining time:

$$T_i^{\text{left}} = T_i^{\text{left}} - \tau_x,$$

  If $T_i^{\text{left}} = 0$, stop the evolution for the walker i and start evolving the walker $i+1$ in the loop; otherwise go ahead with the same walker as follows.
  ▷ Move the walker i to a new configuration $x'(\neq x)$, obtained with probability:

$$p_{x',x} = G_{x',x}/N_x,$$

  where $p_{x',x}$ is the transition probability matrix.
  ▷ Update $\Psi_T(x)$, $G_{x',x}$, and $p_{x',x}$.
  ▷ Evaluate the new local operators, like e.g. $E_L(x)$.
▷ Compute the weighted averages $\bar{O}_g = \sum_i O(x_i)w_i/W$, where $W_g = \sum_i w_i$ is the total weight of the generation g.
▷ Use the branching scheme to control the fluctuations of the walker weights.
▷ Set $w_i = 1\ \forall i$ and $T_i^{\text{left}} = T\ \forall i$.

and in this limit E_{MA}^{eff} coincides with the FN energy of the continuous Hamiltonian.

When non-local pseudopotentials are included, the first identity of Eq. 3.28 is not true, since in general

$$\lim_{a\to 0} H^{\text{eff}} \neq \lim_{a\to 0} H^a, \tag{3.29}$$

and the mixed average E_{MA}^{eff} is not equal to the pure expectation value of H^a. However the lattice FN theorem, proved in Ref. [50] and reported in Subsec. 1.6.2 for the lattice GFMC energy, can be applied also to the LRDMC framework, yielding:

$$E_T \geq E_{MA}^{\text{eff}} \geq \langle \Psi_{FN}^{\text{eff}} | H^a | \Psi_{FN}^{\text{eff}} \rangle / \langle \Psi_{FN}^{\text{eff}} | \Psi_{FN}^{\text{eff}} \rangle, \tag{3.30}$$

where E_T is the expectation value of the Hamiltonian H on Ψ_T. Although the Hamiltonian H^{eff} is defined on a continuous space, we can resort to the same scheme used to prove the lattice FN theorem, since for a given x, the number of matrix elements $H_{x,x'}^{\text{eff}}$ and $H_{x,x'}^a$ are finite, as already mentioned. Therefore, it is possible to write the energy difference $\Delta E = \langle \Psi_{FN}^{\text{eff}} | (H_\gamma^{\text{eff}} - H^a) | \Psi_{FN}^{\text{eff}} \rangle$ as a sum of positive contributions:

$$(1+\gamma) \int dx \sum_{\substack{x'(>x) \text{ and} \\ \tilde{H}_{x,x'}^a > 0}} |H_{x,x'}^a| \left| \Psi_{FN}^{\text{eff}}(x) \sqrt{\left| \frac{\Psi_T(x')}{\Psi_T(x)} \right|} - s H_{x,x'}^a \Psi_{FN}^{\text{eff}}(x') \sqrt{\left| \frac{\Psi_T(x)}{\Psi_T(x')} \right|} \right|^2 \geq 0,$$

where, as in Subsec. 1.6.2, $sH_{x,x'}^a$ is the sign of the matrix element $H_{x,x'}^a$, and the condition $x' > x$ below the summation sign indicates that each pair of configurations x and x', which occurs twice, is taken once, since the other pair is rearranged in the summation. The above relation is valid for each a, also in the limit $a \to 0$ when $H^a \to H$. The upper bound therefore follows:

$$E_{MA}^{\text{eff}} \geq E_{FN} \geq E_{GS}, \tag{3.31}$$

where E_{GS} is the ground state energy of the true continuous Hamiltonian H, and E_{FN} is the pure expectation value of H on the ground state of the effective Hamiltonian H^{eff}. This important upper bound property does not generally hold for the mixed-average E_{MA}^{LA} computed in the DMC approach.

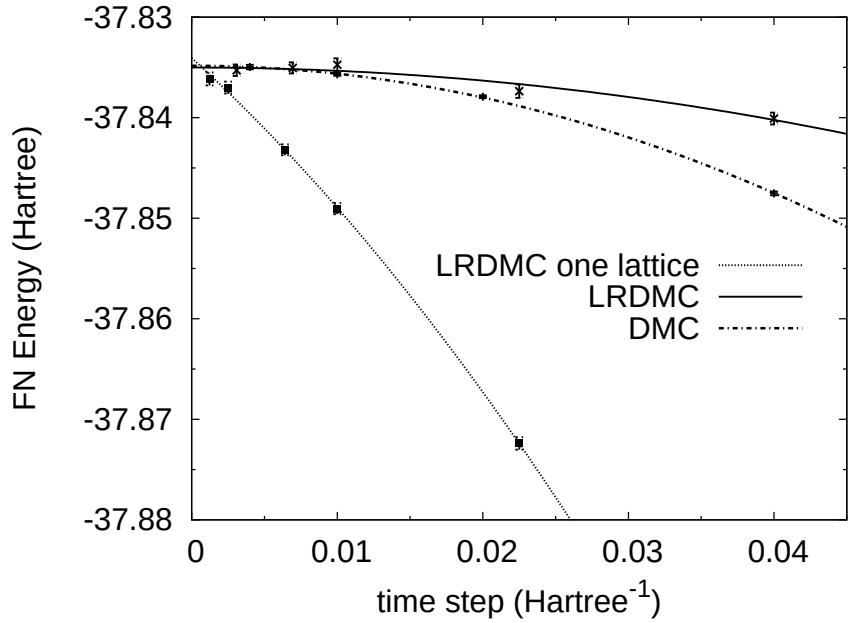


Figure 3.3: FN energies for the all-electron carbon atom computed within DMC, LRDMC and a LRDMC scheme where only one discretization lattice is employed (LRDMC one lattice). The lattice spacing a is here mapped to the time-step τ as $a = \sqrt{\tau}$.

3.4.3 Computation of E_{FN} and pseudopotential localization

In this subsection we extend the effective Hamiltonian H^{eff} to include the parameters γ (≥ 0) and α $(0 \leq \alpha < 1)$ [52], in order to be able to compute the pure expectation value E_{FN} and to recover the locality approximation within the LRDMC framework.

The parameter γ allows us to compute the expectation value E_{FN} of the Hamiltonian H^a on the fixed node solution of H^{eff}. As we have already proven in Eq. 3.29, E^{eff}_{MA}, easily computed during a LRDMC calculation, differs from E_{FN} in the presence of non-local potentials (Eq. 3.30), as in the case studied in Sec. 1.6.

On the other hand, the parameter α sets the degree of localization used in the effective Hamiltonian: it smoothly connects the LA ($\alpha = 1$) to the standard lattice FN approach described above ($\alpha = 0$) where the pseudopotential terms with a non

negative $G_{x',x}$ are not approximated.

The new effective Hamiltonian $H^{\alpha,\gamma}$ is:

$$H^{\alpha,\gamma}_{x',x} = \begin{cases} H^a_{x,x} + (1+\gamma)\mathcal{V}_{sf}(x) + \alpha(1+\gamma)\mathcal{V}^P_{sf}(x) & \text{if } x' = x \\ -\gamma H^a_{x',x} & \text{if } \Psi_T(x')H^a_{x',x}/\Psi_T(x) > 0 \\ (1-\alpha(1+\gamma))H^a_{x',x} & \text{if } \Psi_T(x')V^P_{x',x}/\Psi_T(x) < 0 \\ H^a_{x',x} & \text{otherwise,} \end{cases} \tag{3.32}$$

where $x' \neq x$ and a new sign-flip term is introduced:

$$\mathcal{V}^P_{sf}(x) = \sum_{x' \neq x} \Psi_T(x')V^P_{x',x}/\Psi_T(x) < 0. \tag{3.33}$$

This Hamiltonian satisfies $G_{x',x} > 0$ and reduces to H^{eff} for $\alpha = \gamma = 0$.

Contrary to the usual sign-flip term $\mathcal{V}_{sf}(x)$, the term reported in Eq. 3.33 contains *negative* matrix elements coming exclusively from the non local pseudopotential, which in principle can be treated as hopping terms, since they satisfy the FN constraint. In the effective Hamiltonian $H^{\alpha,\gamma}$, they are placed in the diagonal term with weight α, and subtracted by the same amount from the off diagonal matrix elements with $\Psi_T(x')V^P_{x',x}/\Psi_T(x) < 0$ [3], as we want to include the locality approximation within the LRDMC framework. In this way, while the non local pseudopotential elements which do not fulfill the FN requirement are already traced in the sign-flip diagonal term $\mathcal{V}_{sf}(x)$, the others are traced in the diagonal contribution $\mathcal{V}^P_{sf}(x)$ with weight $\alpha(1+\gamma)$. If we keep $\gamma = 0$ and we set $\alpha = 1$, the non local pseudopotential terms disappear from the off-diagonal matrix elements of the Hamiltonian and are all traced in the diagonal element inside the sign-flip contribution. This is exactly equivalent to the LA, where the local potential is obtained from the non local pseudopotentials by means of the trial wave function Ψ_T (see Eq. 3.7).

Since the lattice regularized effective Hamiltonian $H^{\alpha,\gamma}$ shares the same properties as the effective Hamiltonian valid for lattice systems, within the LRDMC

[3] Notice that by $\Psi_T(x'V^P_{x',x}/\Psi_T(x) < 0$, we mean those negative off-diagonal elements $H^a_{x',x}$ which include non local potentials. Moreover the single element $H^a_{x',x}$ cannot contain both Laplacian and pseudopotential contributions, since the former are converted into *translational* hopping terms, the latter become *rotational* moves. Therefore in practice it is impossible that a couple x and x' is connected both by translational and rotational moves provided by the Hamiltonian.

framework it is possible to evaluate the expectation value of H^a on the GS solution $\Psi_{FN}^{\alpha,\gamma}$ of $H^{\alpha,\gamma}$. Indeed we can follow the same procedure explained in Subsec. 1.6.3, and exploit the identity

$$E_{FN}(\alpha,\gamma) = E_{MA}(\alpha,\gamma) - (\gamma+1)\frac{\partial E_{MA}(\alpha,\gamma)}{\partial\gamma}\,, \tag{3.34}$$

where

$$E_{FN}(\alpha,\gamma) = \langle\Psi_{FN}^{\alpha,\gamma}|H^a|\Psi_{FN}^{\alpha,\gamma}\rangle/\langle\Psi_{FN}^{\alpha,\gamma}|\Psi_{FN}^{\alpha,\gamma}\rangle \tag{3.35}$$

$$E_{MA}(\alpha,\gamma) = \langle\Psi_T|H^a|\Psi_{FN}^{\alpha,\gamma}\rangle/\langle\Psi_T|\Psi_{FN}^{\alpha,\gamma}\rangle. \tag{3.36}$$

This relation is obtained by using that $H^a = H^{\alpha,\gamma} - (\gamma+1)\partial_\gamma H^{\alpha,\gamma}$ and applying the Hellmann-Feynman theorem to the last term. The LRDMC simulation is able to evaluate only $E_{MA}(\alpha,\gamma)$ and not $E_{FN}(\alpha,\gamma)$, as $\langle\Psi_T|H^{\alpha,\gamma}|\Psi_{FN}^{\alpha,\gamma}\rangle/\langle\Psi_T|\Psi_{FN}^{\alpha,\gamma}\rangle$, the quantity actually computed by the LRDMC method, equals $E_{FN}(\alpha,\gamma)$:

$$E_{MA}(\alpha,\gamma) = \langle\Psi_T|H^a|\Psi_{FN}^{\alpha,\gamma}\rangle/\langle\Psi_T|\Psi_{FN}^{\alpha,\gamma}\rangle = \langle\Psi_T|H^{\alpha,\gamma}|\Psi_{FN}^{\alpha,\gamma}\rangle/\langle\Psi_T|\Psi_{FN}^{\alpha,\gamma}\rangle, \tag{3.37}$$

where the last identity follows from the equivalence between the local energy $H^a\Psi_{FN}^{\alpha,\gamma}/\Psi_{FN}^{\alpha,\gamma}$ and $H^{\alpha,\gamma}\Psi_{FN}^{\alpha,\gamma}/\Psi_{FN}^{\alpha,\gamma}$. The best variational E_{FN} energy is for $\gamma = 0$ (see Ref. [52] and Subsec. 1.6.3), which we estimate by computing the derivative with respect to γ in an approximate but variational way:

$$E_{FN}(\alpha,0) \leq E_{MA}(\alpha,0) - [E_{MA}(\alpha,\gamma) - E_{MA}(\alpha,0)]/\gamma, \tag{3.38}$$

where the equality sign holds in the limit of small γ. The proof of the variationality of the estimate given in Eq. 3.38 can be obtained by following the same scheme given in Subsec. 1.6.3 used to prove the upper bound of a similar expression for a lattice Hamiltonian.

The parameter α is used to improve upon the locality approximation and its value is optimized to yield the lowest $E_{FN}(\alpha,0)$. The present scheme can evaluate the variational $E_{FN}(\alpha,0)$ using Eq. 3.38, even though it is not guaranteed that $E_{MA}(\alpha,\gamma)$ is variational for $\alpha > 0$. This can be shown by evaluating the energy difference

$$\Delta E = \langle\Psi|(H^{\alpha,\gamma} - H^a)|\Psi\rangle, \tag{3.39}$$

where Ψ is a variational state. If we follow the proof reported in Ref. [50] and in Subsec. 1.6.2, it is easy to obtain:

$$\begin{aligned}\Delta E &= (1+\gamma)\int dx \sum_{\substack{x'(>x) \text{ and} \\ \tilde{H}^a_{x,x'}>0}} |H^a_{x,x'}| \left| \Psi(x)\sqrt{\left|\frac{\Psi_T(x')}{\Psi_T(x)}\right|} - sH^a_{x,x'}\Psi(x')\sqrt{\left|\frac{\Psi_T(x)}{\Psi_T(x')}\right|}\right|^2 \\ &- \alpha(1+\gamma)\int dx \sum_{\substack{x'(>x) \text{ and} \\ \tilde{V}^P_{x,x'}<0}} |H^a_{x,x'}| \left| \Psi(x)\sqrt{\left|\frac{\Psi_T(x')}{\Psi_T(x)}\right|} + sH^a_{x,x'}\Psi(x')\sqrt{\left|\frac{\Psi_T(x)}{\Psi_T(x')}\right|}\right|^2 ,\end{aligned}$$

where $sH^a_{x,x'}$ is the sign of the matrix element $H^a_{x,x'}$, $\tilde{H}^a_{x,x'}$ and $\tilde{V}^P_{x,x'}$ refer to the importance sampling matrix elements, and the condition $x' > x$ below the summation sign indicates that each pair of configurations x and x', which occurs twice, is taken once, since the other pair is rearranged in the summation. It is apparent that as soon as $\alpha > 0$, there is a negative contribution which can give an overall negative value for ΔE. Therefore for $\alpha > 0$ the variationality of the mixed average E_{MA} is not guaranteed. Instead, for $\alpha = 0$, the upper bound is proved since in that case ΔE contains only a sum of positive terms.

To estimate the variational E_{FN} energy, we use the relation in Eq. 3.38, i.e. we approximate the derivative $\partial_\gamma H^{\alpha,\gamma=0}$ by means of a finite difference expression, with two independent calculations of E_{MA} for $\gamma = 0$ and γ positive. Note that if $\alpha > 0$, we must take $\gamma \leq 1/\alpha - 1$ in order to keep the importance sampling Green function of Eq. 3.25 positive definite. Therefore, for α close to 1, the naive evaluation of the derivative based on finite differences turns out to be unfeasible, since $\Delta\gamma$ must be close to 0 and the error diverges. However, it is possible to overcome this difficulty by using a correlated sampling (see Appendix E) to compute $E_{MA}(\alpha,\gamma)$ for $\gamma = 0$ and $\gamma \in (0, 1/\alpha - 1]$, during the same run and exploiting the same random walk to evaluate the energy differences. In this way, the LRDMC scheme is able to estimate the pure expectation value E_{FN} also in a regime close to the locality approximation ($\alpha \approx 1$), while the standard DMC can only access the mixed average $E_{LA} = \langle\Psi_T|H^{LA}|\Phi_{LA}\rangle/\langle\Psi_T|\Phi_{LA}\rangle$. On the other hand, for $\alpha = 0$, both E_{MA} and E_{FN} computed by the LRDMC method are variational, and in this case E_{FN} can be easily estimated by using two independent evaluations of E_{MA} for $\gamma = 0$ and $\gamma = 1$, without resorting to the correlated sampling (see Appendix E).

3.5 Results

In this section we present some data on the efficiency of the algorithm compared to the standard DMC method (Subsection 3.5.1), the LRDMC results obtained for the silicon and scandium atoms with non local pseudopotentials (Subsection 3.5.2) and finally an application of the LRDMC framework to a quantum wire model, in which standard DMC simulations are affected by a lack of ergodicity during the sampling (Subsection 3.5.3).

3.5.1 Efficiency

In order to check the efficiency of the LRDMC algorithm with respect to the standard DMC method, we have carried out various simulations at different lattice spaces or time steps, taking into account the total CPU time and the statistical error of the energy. To evaluate the efficiency we have computed the following expression:

$$\text{efficiency} = \frac{1}{\text{time} \times \text{error}^2}. \tag{3.40}$$

As test case, we chose the iron atom with Dolg's pseudopotentials to replace the neon core. We believe that it is a good system for testing the efficiency, since it contains a quite big core and a large number of valence electrons, and thus it is possible to check the effect of both the double kinetic move and the variational treatment of non local potentials. The results are reported in Table 3.2 and the energies plotted in Fig. 3.4.

The first outcome is that the LRDMC efficiency is better than the DMC's, mainly due to the double kinetic move, with the smallest step around the core and the greatest one in the valence region. Indeed the pseudopotentials we used for the iron atom replace only the neon like shells, leaving an unscreened nucleus with pseudocharge $Z_{\text{eff}} = 16$, which is relatively large for a pseudoatom. Mitas [118] suggested to use such a small core, in order to reduce the core-valence correlation, very important in this case, and obtain more accurate results. It is clear that an atom with $Z_{\text{eff}} = 16$ shows different characteristic length scales, since its shell structure is complex with a large spread of single particle energy levels. Therefore the double mesh used for the discretization of the Laplacian really matters to set the efficiency of the LRDMC algorithm, by decorrelating the valence dynamics

Table 3.2: Study of the efficiency of the DMC and LRDMC simulations for the 5D atomic state of iron. All the calculations have been carried out using 16 processors of the clx parallel cluster of Xeon $3GHz$ processors located at Cineca (Bologna, Italy). The energy is in Hartree, the cpu time in seconds, the time step τ has been mapped to the lattice space as $a = \sqrt{\tau}$, and the efficiency has been calculated as in Eq. 3.40.

lattice space	method	energy	error	CPU time	efficiency
0.0208	DMC ($\tau = 4.3\ 10^{-4}$)	-123.7886	$1.6\ 10^{-3}$	37230	2.71
	LRDMC ($\alpha = 1, \gamma = 0$)	-123.7973	$1.1\ 10^{-3}$	20104	10.86
	LRDMC ($\alpha = 0, \gamma = 0$)	-123.7885	$1.1\ 10^{-3}$	20331	11.05
0.0245	DMC ($\tau = 6.0\ 10^{-4}$)	-123.7926	$1.9\ 10^{-3}$	18656	3.61
0.0284	DMC ($\tau = 8.1\ 10^{-4}$)	-123.7935	$1.7\ 10^{-3}$	19636	4.18
	LRDMC ($\alpha = 0, \gamma = 0$)	-123.7921	$0.9\ 10^{-3}$	14375	21.96
0.0316	DMC ($\tau = 1.0\ 10^{-3}$)	-123.7893	$1.7\ 10^{-3}$	15224	5.48
0.0391	DMC ($\tau = 1.5\ 10^{-3}$)		unstable		
	LRDMC ($\alpha = 0, \gamma = 0$)	-123.7993	$1.1\ 10^{-3}$	14217	31.30
0.0625	DMC ($\tau = 3.8\ 10^{-3}$)		unstable		
	LRDMC ($\alpha = 0, \gamma = 0$)	-123.8172	$0.8\ 10^{-3}$	13259	55.40

with respect to the highest energy electrons, and thus reducing the autocorrelation time of the sampling. Notice that in order to use a double mesh sampling one has to pay the price of computing in advance twice the number of non diagonal matrix elements computed in the case of only one mesh. Therefore there is a balance between the time spent to evaluate the matrix elements and the efficiency gained from using multiple meshes to decorrelate the electrons. In principle one can use even more than two meshes, by requiring that:

$$\begin{aligned} \Delta^a &= \sum_i \Delta^{a_i, p_i} \\ \sum_i p_i(x) &= 1 \\ a_i/a_1 &= c_i \end{aligned} \tag{3.41}$$

where a_i are the lattice spaces, c_i are irrational numbers, and p_i are the weighting functions appearing in the discretized Laplacian operators Δ^{a_i, p_i} defined in Eq. 3.14. The total number of matrix elements connected to the given configura-

tion x is $2DmN$, where m is the number of meshes, D is as usual the dimension of the space and N is the number of electrons. In the case of non local potentials, also the off diagonal elements in the potential itself become hopping terms of the lattice regularized Hamiltonian, and they contribute to improve the efficiency of the sampling beside those coming from the discretization of the Laplacian. We believe that $m = 2$ is the most effective choice, also in the presence of non local potentials. For very heavy atoms, it could be more continent to add more meshes into the definition of the discretized Laplacian, since the number of characteristic length scales would increase. Anyway, this should be checked, because adding more meshes means computing more matrix element and increasing the total time, as stated above.

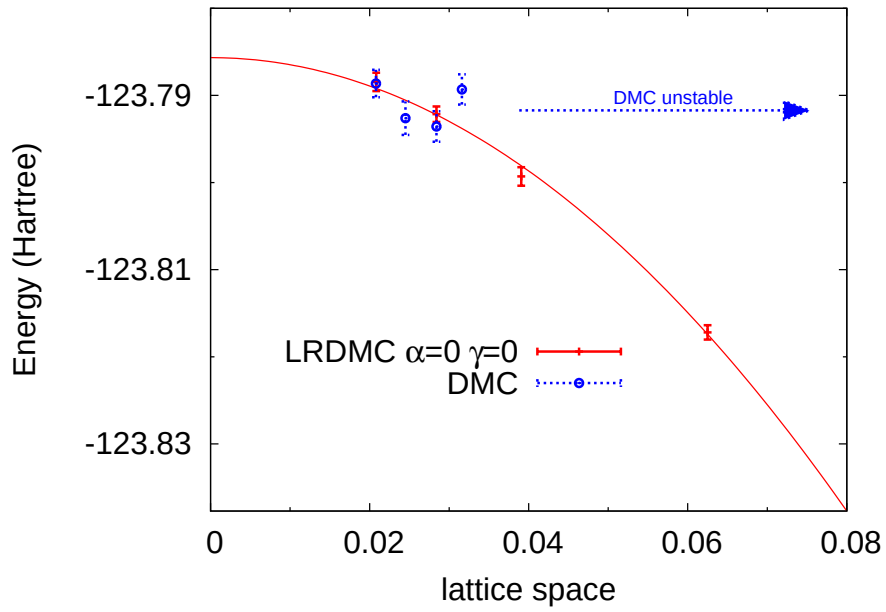


Figure 3.4: Energy dependence on the lattice space of DMC and LRDMC ($\alpha = 0, \gamma = 0$) calculations for the iron atom with Dolg's pseudopotentials. The time step τ has been mapped to the lattice space as $a = \sqrt{\tau}$.

On the other hand, the variational treatment of non local potential does not seem to affect the efficiency, since the LRDMC scheme with the locality approximation ($\alpha = 1, \gamma = 0$) provides almost the same efficiency as the LRDMC algo-

rithm without locality ($\alpha = 0, \gamma = 0$), as it is apparent from Table 3.2. However, the algorithm with $\alpha = 0$ and $\gamma = 0$ yields a stable simulation also for those lattice spaces (time steps) that lead to an unstable DMC evolution. Probably the DMC instability comes from rare but extreme events, in which the walker is close to the node, the localized pseudopotential is extremely attractive and the local energy is too low to be included in a steady branching scheme. Instead, in the LRDMC with $\alpha = 0$ and $\gamma = 0$, which provides variational results, the negative divergent terms of the pseudopotential are converted into hopping terms, which push the walker away from the node, where the localized pseudopotential diverges. In this way the simulation is very stable, even for a hard non local potential or for a poor guiding function.

Finally, the possibility to work with large lattice spaces without compromising the stability of the calculation allows to perform a much more accurate extrapolation to the limit $a \rightarrow 0$. Moreover it has been noticed that in the case of non local pseudopotentials the DMC dependence on the time step is often non monotonic, (see Fig. 3.5) and thus it is cumbersome to extrapolate the DMC result to the limit of $\tau = 0$ without carrying out simulations with very small time step and very low efficiency. Instead, LRDMC results do not show this difficulty, and its energies are always very well fitted by a functional form containing a quadratic dependence on the lattice space.

3.5.2 LRDMC and non locality: application to atoms

We have first tested the performance of the LRDMC approach on the silicon pseudoatom using three Hartree-Fock pseudopotentials which differ in the construction, functional form and core radius. The soft one is a norm-conserving pseudopotential obtained using the code of Shirley and the construction scheme of D. Vanderbilt and it has a quite large core radius. The Dolg's pseudopotential has a smaller core radius, but the magnitude of the non local radial functions $v_l(r)$ is bigger. The Lester's has built to have both a small core radius and flat functions $v_l(r)$ at the origin, and so it is the most local among the three pseudopotentials used. All of them are plotted in Fig. **??** for the local and the s like non local components.

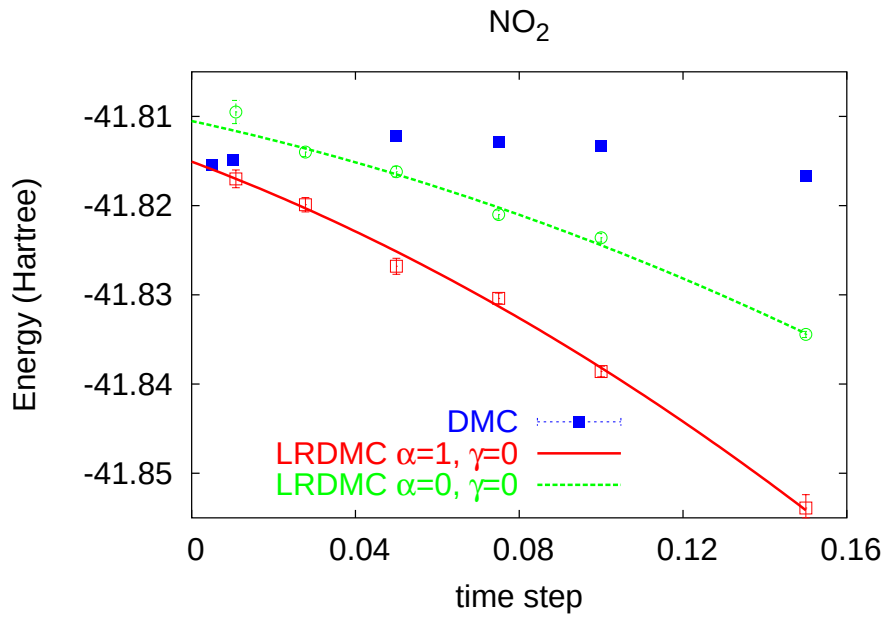


Figure 3.5: Energy dependence on the time step for DMC, LRDMC ($\alpha = 0, \gamma = 0$) and LRDMC ($\alpha = 1, \gamma = 0$) calculations of the nitrogen dioxide NO_2 with norm conserving Hartree-Fock pseudopotentials. The trial wave function has the usual Slater-Jastrow form. The Slater part is a Hartree-Fock determinant, while the Jastrow factor contains three-body correlations, of the form described in Ref. [79]. The lattice space a has been mapped to the time step as $a = \sqrt{\tau}$. The LRDMC framework with the LA has a monotonic dependence on the time step, contrary to the DMC results. The mixed-estimate $E_{MA}(0, 0)$ is higher than the non variational data.

For each pseudopotential, we employ three wave functions with the same determinantal component and, consequently, the same nodes, but with different Jastrow factors. We use functions without Jastrow factor, with a two-body, and a sophisticated three-body Jastrow factor [79]. The determinantal component has been obtained from Hartree-Fock calculations with a Gaussian basis, while the Jastrow factor fulfills both the electron and nuclear cusp conditions. The Jastrow part of the wave function has been optimized using the variance minimization, while the Slater part has been kept fixed.

In this way, we can check the magnitude of the locality error for the pseudopo-

tential and the method used in the calculation. Indeed, if the localization error was absent, only the FN approximation would affect the results, and therefore the energy would depend only on the nodes and not on the *shape* of the guidance wave function. Consequently the wave functions with the same determinantal component would yield the same result. Instead, as shown in Fig. 3.6, the mixed energy estimate E_{LA} computed within DMC changes significantly with the guiding wave function Ψ_T. It means that the locality approximation gives an error, which is relevant at least in the total energies.

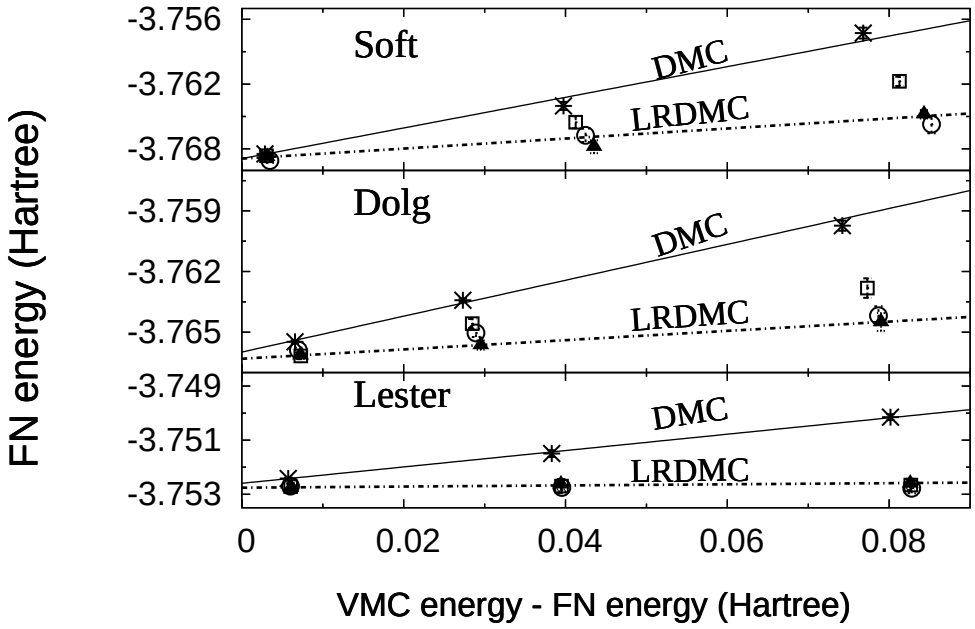


Figure 3.6: FN energies of the silicon pseudoatom computed within DMC (E_{MA}^{LA}) and LRDMC ($E_{FN}(\alpha, 0)$). For different pseudopotentials (Soft, Dolg's [126] and Lester's [120]), we use as guiding wave function's a Hartree-Fock determinant with no Jastrow, a two-body and a three-body Jastrow factor. A more accurate guiding wave function corresponds to a smaller difference between the variational Monte Carlo (VMC) and the FN energies. The LRDMC energies are computed for $\alpha = 0.9$ (filled triangles), $\alpha = 0.5$ (open circles) and $\alpha = 0$ (open squares). The linear fits for the DMC and the LRDMC ($\alpha = 0.9$) data are shown.

The LRDMC framework allows to compute the pure energy expectation val-

ues $E_{FN}(\alpha, 0)$ of the true Hamiltonian, evaluated using different degrees of localization: $\alpha = 0$ (true non local treatment of the pseudopotential matrix elements which do not frustrate the sign of the Green function), 0.5 and 0.9 (close to the locality approximation of the non local potential). All results reported in Fig. 3.6 are obtained using the LRDMC correlated sampling (see Appendix E) to estimate accurately the derivative $\partial E_{MA}(\alpha, \gamma = 0)/\partial\gamma$ in Eq. 3.34. For every α and for every pseudopotential, the $E_{FN}(\alpha, 0)$ energies show a localization error smaller than the mixed averages E_{LA} obtained with the DMC method. For all cases, the statistical uncertainty does not allow us to discriminate between the LRDMC energies obtained for $\alpha = 0.5$ and $\alpha = 0.9$, and a shallow minimum seems to lie between these two values. The localization error is significantly reduced for optimal $E_{FN}(\alpha, 0)$ and the weakest dependence on Ψ_T is obtained for Lester's pseudopotential [120] which has the smallest core radius in the non-local component. Interestingly, since $E_{FN}(\alpha, 0)$ for $\alpha \simeq 1$ is very close to the minimum, the LA seems to provide in this case good wave functions. This is probably due to the fact that the effective Hamiltonian with the LA does not break the rotational invariance of the true Hamiltonian for atoms. Indeed, even if for $\alpha = 0$ some of the non local pseudopotential elements are correctly treated, the presence of the FN approximation, necessary to guarantee the positivity of the Green function, leads to an effective Hamiltonian with part of the pseudopotential contribution in the spin-flip term, the remaining pseudopotential elements in the hopping terms. Therefore, whenever the FN approximation is not necessary, the $\alpha = 0$ lattice regularized Hamiltonian is exact, but since the FN approximation is needed, the effective FN Hamiltonian with $\alpha \neq 1$ turns out to break the rotational invariance. Instead, the case with $\alpha = 1$ approximates completely the pseudopotential, but since it treats on the same footing all the non local elements by localizing the whole pseudopotential, it does not break the rotational invariance, and thus probably yields better wave functions. Notice however that we can say something about the quality of the wave function just because we can compare *variational* expectation values for the *true* Hamiltonian, accessed only by the LRDMC framework and not by the standard DMC method.

A stringent test case for our LRDMC algorithm is the scandium atom: the LA for transition metals yields large errors in the DMC total energies, and performs the worst for the scandium atom [119]. As before, we keep the determinantal part

of the wave function fixed, and employ a 2-body which fulfills both the nuclear and the electron cusp conditions (see Chap. 4), and a 3-body Jastrow factor of the form described in Sec. 2.2.3. The determinantal component is an antisymmetrized geminal function expanded over a ($5s5p5d$) Gaussian-type basis in order to cure near degeneracy effects, and optimized in the presence of the 2-body Jastrow factor by means of the SRH method (see Sec. 1.4.2). We employ Dolg's pseudopotential [126] and compute the $4s^2 3d^n \rightarrow 4s^1 3d^{n+1}$ excitation energy which is reported in Table 3.3. This time we have estimated the pure expectation value $E_{FN}(\alpha, 0)$, by using two independent samples with $\gamma = 0$ and $\gamma = 1$ (see Eq. 3.38).

It is apparent that the LA does not only affect the DMC total energies but also the DMC energy differences: the DMC excitation energy computed with the 2-body Jastrow factor differs from the experimental value by more than three standard deviations. On the other hand, the LRDMC FN results are less sensitive to Ψ_T, and are compatible with the experiment even when a simple 2-body Jastrow factor is employed. Also the LRDMC MA excitations appear to be closer to the experimental value than the DMC ones.

Since estimating $E_{FN}(\alpha, 0)$ is computationally more demanding because require the evaluation of two mixed-average energies (Eq. 3.38), a practical compromise is to use $E_{MA}(0, 0)$ as the energy estimate: $E_{MA}(\alpha, 0)$ at $\alpha = 0$ is variational and its computation is more efficient than the DMC evaluation of E_{MA}^{LA}. For the scandium pseudoatom, which has a large effective charge, we find that, due to the use of a double kinetic mesh within LRDMC, the gain in efficiency over DMC is at least a factor of 2, as already reported in the previous subsection 3.5.1 for the iron atom. Also in this case, we observe that a LRDMC simulation with off-

Table 3.3: Comparison of $4s^2 3d^n \rightarrow 4s^1 3d^{n+1}$ excitation energy (eV) for the scandium atom.

				LRDMC		
	α	VMC	DMC	$E_{MA}(\alpha, 0)$	$E_{FN}(\alpha, 0)$	exp
2-body	0.0	1.099(30)	1.381(15)	1.408(12)	1.417(31)	1.43
2-body	0.5	1.099(30)	1.381(15)	1.394(11)	1.441(25)	
3-body	0.5	1.303(29)	1.436(22)	1.448(9)	1.478(22)	

diagonal pseudopotentials is computationally much more stable than DMC even when a very crude wave function or a very large lattice mesh a are employed. As already discussed, the reason for this better stability is that the negative divergences coming from the pseudopotentials close to the nodes are converted to finite hopping terms in the LRDMC scheme.

3.5.3 LRDMC and ergodicity: application to a quantum wire model

The Hamiltonian of the quantum wire model[127] studied is, in units in which $a_0^* = \hbar^2\epsilon/(m^*e^2) = 1$ and $e^2/(2\epsilon a_0^*) = 0.5$:

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i<j} V_b(|x_i - x_j|), \tag{3.42}$$

$$V_b(x) = (\sqrt{\pi}/2b)\exp\left(x^2/4b^2\right)\mathrm{erfc}\left(|x|/2b\right), \tag{3.43}$$

with b a measure of the wire width. Above, ϵ is the dielectric constant of the semiconductor and m^* the effective mass of the carriers. The pair interaction $V_b(x)$, which is finite at the origin, $V_b(0) = \sqrt{\pi}/(2b)$ and decays as $1/x$ for $x \gg b$, is obtained from an harmonic confinement of the 3DEG, after projection on the lowest subband of transverse motion[127]; it thus provides a good approximation to the 3D system at low density ρ, i.e., $r_s = 1/2\rho \gg \pi b/4$, with the density parameter r_s also providing an estimate of the Coulomb coupling, as ratio between average potential and kinetic energies.

We carry out VMC and DMC simulations for the ground state properties of a thin wire ($b = 0.1$) on a fairly wide coupling range ($1 \leq r_s \leq 10$), thus considering only unpolarized states[128–130]. To this end, we resort to a Slater-Jastrow wavefunction[131] $\Psi_T = JD^{\uparrow}D^{\downarrow}$, with D^{σ} a determinant of N^{σ} plane waves and $J = \exp[-\sum_{i<j} u(|x_i - x_j|)]$, with $u(x)$ a two-body Jastrow function to be optimized. This is the simplest correlated wavefunction for an unpolarized Fermion state with homogeneous density and can be further improved with the inclusion in the Jastrow factor J of higher order terms[86, 132]. To avoid shell degeneracy effects we keep $N^{\uparrow} = N^{\downarrow}$ odd, and in order to further reduce the finite size bias we use periodic boundary conditions and the Ewald's sum[131, 133, 134] of the pair potential $V_b(x)$.

We remind the reader that in 1D the nodes of the Fermionic *ground state* of given symmetry are known exactly[46], as they are fully determined by exchange antisymmetry (see Subsec. 1.5.2), and coincide in particular with those of the above wavefunction for the unpolarized state. Thus, DMC provides in principle exact energies[129, 133, 134] that may serve as benchmark for the VMC simulations when optimizing the Jastrow factor J. As starting point for our simulations we take a two-body Jastrow function of the RPA type,[131, 135, 136] which in Fourier space reads

$$2\rho\tilde{u}_{RPA}(k) = -S_0(k)^{-1} + \sqrt{S_0(k)^{-2} + 4\rho\tilde{V}_b(k)/k^2}, \tag{3.44}$$

with $S_0(k) = (k/2k_F)\theta(2k_F - k) + \theta(k - 2k_F)$ the structure factor of a non interacting 1DEG,

$$\tilde{V}_b(G) = E_1(b^2G^2)\exp(b^2G^2) \tag{3.45}$$

the Fourier transform of the interaction, and $\theta(x)$ and $E_1(x)$ respectively the Heaviside and the exponential integral functions. To get an accurate description of correlation functions we systematically optimize Ψ_T, employing the variance minimization method[26]. As the plane-wave determinants provide the exact nodes for our 1D system, we have chosen to optimize only the two-body Jastrow function $u(x)$, for which we have considered the scaled RPA form $u_\alpha(x) = \alpha\, u_{RPA}(x)$.

The repulsive nature of the pair interaction $V_b(x)$ is directly reflected in u_{RPA}, which is shown in Fig. 3.7. This function is repulsive and the repulsion increases appreciably with decreasing the density (increasing the r_s). Thus, as the density is lowered electrons are kept apart more and more effectively and this results in a quasi long range order which can be described as a quasi Wigner Crystal[137]. Though the two-body Jastrow function remains finite at contact, the Jastrow factor becomes exponentially small yielding what may be seen as pseudonodes of the wavefunction. These pseudonodes have no particular effect on like spin electrons, as the wavefunction is already vanishing at contact for such electrons and most importantly particle exchanges are explicitly summed over in the determinants. On the contrary, the effect on opposite spin electrons when combined with the reduced dimensionality may become dramatic. In a random walk in configuration space with importance sampling given by the Slater-Jastrow wavefunction, the RPA Jastrow function tends to freeze out the exchange between opposite spin

electrons and may cause ergodicity problems (see Sec. 1.2), which show up in the evaluation of spin correlations[133, 134]. Evidently, it is not only the presence of the pseudonodes to cause problems but also the *slope* with which the pseudonodes are approached, on the scale of the interparticle distance. When such a slope becomes sufficiently large, naive algorithms may become inefficient in sampling inequivalent pockets in configuration space, delimited by the pseudonodes.

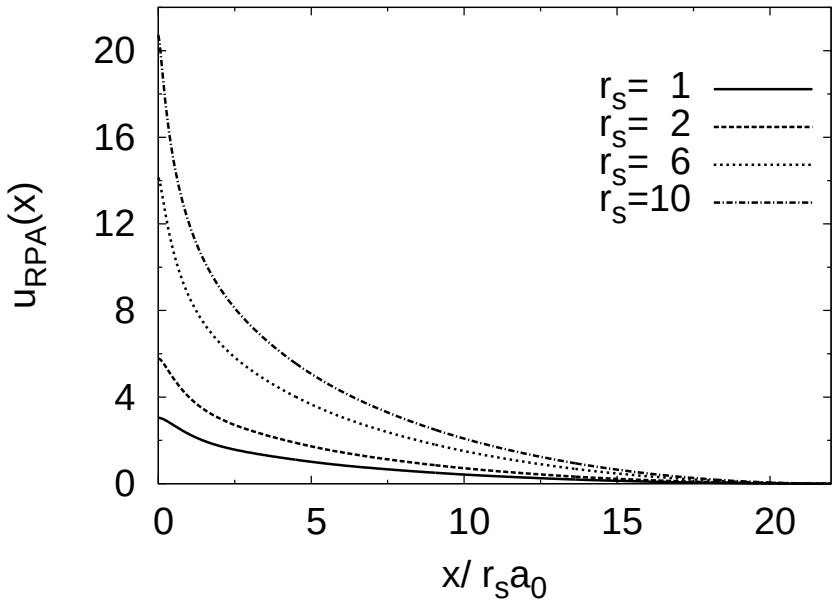


Figure 3.7: RPA two-body Jastrow function for $b = 0.1$ and $r_s = 1$ (solid line), $r_s = 2$ (dashed line), $r_s = 6$ (dotted line), $r_s = 10$ (dot-dashed line).

Ergodicity becomes an issue in the present context because for opposite spin electrons in 1D there is no exchange without crossing. On the other hand, the energy does not seem to depend on the exchange frequency during the sampling, and this is due to the fact that at strong coupling different spin configurations are almost degenerate. The variational Monte Carlo (VMC) algorithm can easily overcome this ergodicity problem, since the proposed move can be forced to flip the spin of an electronic configuration, by explicitly introducing a spin exchange or by allowing the amplitude of the move to be greater than the mean interparticle distance. Instead in the diffusion Monte Carlo (DMC) approach, the random walk has to follow the diffusion process driven by the imaginary time dependent

Schrödinger equation. As we have already seen in Sec. 1.5.1, if the importance sampling is introduced, the resulting Green function, approximated by means of the Trotter expansion up to the second order in the time step τ, includes the drift-diffusion dynamics:

$$x' = x + D\tau\nabla \ln |\Psi_T(x)|^2 + \sqrt{2D\tau}\chi, \tag{3.46}$$

where $\nabla \ln |\Psi_T(x)|^2$ is the quantum force, $D = \hbar^2/2m$ is the diffusion coefficient, and χ is a Gaussian distributed random variable. In order to get rid of the time step error in the final result, one needs to extrapolate the fixed node energies obtained at different time steps for τ going to zero. Thus the mean square displacement, which is related to the variance of the diffusion process, is forced to vanish linearly with τ and in that regime one can not avoid the lack of the ergodicity in the random walk.

Here, we want to apply the LRDMC scheme to the same system, and compare its efficiency of sampling spin flips with respect to the DMC method. Due to the homogeneity of the system, the function p in the discretized Laplacian (see Subsec. 3.4.1) is kept spatial independent, contrary to the general case where the dependence of p on the electronic positions can be exploited to improve the efficiency of the diffusion process. Here $p = 0.5$, and the contributions to the total Laplacian coming from $\Delta^{a,p}$ and $\Delta^{a',1-p}$ are equally weighted. The two terms, with $a/a' = \sqrt{5}$, allow the diffusion to explore all the continuous space, since the two meshes are incommensurate; in this way the lattice space bias due to the discretization of the continuous kinetic operator is greatly reduced, like in the atomic case, and one can work with a reasonably large value of a without a significant lattice step error.

Since the Hamiltonian in Eq. 3.42 does not contain any non local potential, all off diagonal elements of H^a are the kinetic terms, and the number of hopping terms of the discretized kinetic operator defined in Eq. 3.15 is $4N$ in 1D. The fermionic sign problem is still present and needs to be treated by means of the effective fixed node Hamiltonian, although the nodes are exact and the fixed node energy coincides with the true ground state energy of the system.

We did an accurate comparison between the DMC and LRDMC approach, by taking into account the efficiency of the energy estimator, the dependence on the time step and on the lattice space, and the spin flip frequency during the simula-

tions. We applied the two QMC schemes to the quantum wire model with $N = 22$ and for r_s ranging from 1 to 10. For a fair comparison, we chose the DMC time step $\tau = a^2$, so that all two algorithms provide the same amplitude for the diffusion move. We found that the efficiency of the DMC energy estimator is twice better than the efficiency of the LRDMC, since in the latter approach we need to compute in advance all the possible off diagonal moves, by losing a fraction of the computing time. On the other hand, as reported in Table 3.4, the spin exchange frequency is almost the same for the high density model, when the correlation is weak, but the LRDMC becomes more and more effective in sampling the spin flips when the density is lower and the correlation turns to be strong. In particular, for the lowest density ($r_s = 10$) the LRDMC algorithm yields an efficiency in the spin flip sampling which is two orders of magnitude higher than the DMC case.

Table 3.4: Spin exchange frequency (H^{-1}) for the LRDMC and DMC algorithm at different densities for the quantum wire model with $N = 22$ and $b = 0.1$. The frequency is defined as the number of exchanges between opposite spin particles per unit of the imaginary time propagation. Notice that the frequency is reduced when the density lowers, while the efficiency of the LRDMC increases with respect to the DMC. All the simulations have been performed with $a = 0.2 r_s$ and $\tau = a^2$.

	$r_s = 1$	$r_s = 2$	$r_s = 4$	$r_s = 6$	$r_s = 10$
LRDMC	2.36	$1.14\ 10^{-1}$	$3.92\ 10^{-2}$	$6.38\ 10^{-4}$	$1.77\ 10^{-5}$
DMC	2.28	$7.62\ 10^{-2}$	$8.48\ 10^{-3}$	$6.42\ 10^{-5}$	$1.82\ 10^{-7}$
relative frequency	1.03	1.50	4.62	9.94	97.47

Therefore, at given amplitude of the QMC move, the LRDMC is more effective than the DMC scheme; we believe that the reason is related to the Trotter approximation behind the DMC propagator, which spoils the exact dynamics of the diffusion process and apparently reduces the ergodicity of the random walk. Another appealing behaviour of the LRDMC approach is the lattice space dependence of the fixed node energy. As one can see in Fig. 3.8, the LRDMC energies have a quadratic dependence on a with a prefactor much smaller than the slope of the linear fit for the corresponding DMC energies. It means that, in order to obtain an almost converged LRDMC result, one does not need to go to small lat-

tice spaces, with a gain both from the computational point of view and from the efficiency of the spin flips, which of course is reduced as the diffusion move goes to zero.

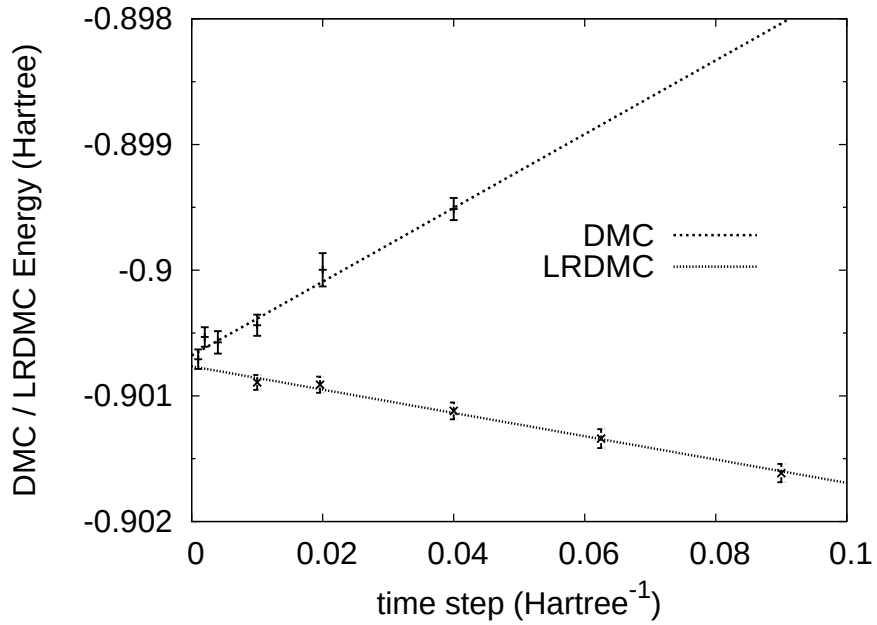


Figure 3.8: DMC and LRDMC energies dependence on the time step. The lattice space a has been mapped into the time step, by means of the relation $a = \sqrt{\tau}$. For both the two cases, the dependence appears to be linear, with a slope of 0.029(2) for the DMC algorithm and -0.0045(5) for the LRDMC approach.

To summarize, it is apparent that both the lattice step bias and the lack of ergodicity is greatly reduced by using the LRDMC algorithm instead of the standard DMC. By performing LRDMC simulations on the quantum wire we were able to evaluate the unbiased spin and charge structure factors, and show the good agreement with the exact predictions for a Luttinger-like Hamiltonian [137] with an interaction similar to the one used here.

3.6 Conclusions

The LRDMC method, presented in this Chapter, has proven to be an efficient and robust projection scheme, which is based on a lattice regularized effective

Hamiltonian, that converges to the continuous one as the lattice space goes to zero. In the LRDMC framework the imaginary time evolution is exact for a given lattice regularized Hamiltonian, and in particular the method is free from the usual time step error of the standard DMC algorithm. The lattice space error takes its place, since the discretization of the Laplacian is exact only in the limit of vanishing mesh size, when the lattice regularized and the continuous Hamiltonian coincide.

We showed that the inclusion of a multiple mesh kinetic operator greatly improves the efficiency of the method. In particular, two meshes with incommensurate sizes are enough to let the sampling span the whole configuration space, with a significant reduction of the lattice space error with respect to the case of a strict one lattice evolution. Moreover, we have seen that the double mesh move can be useful also to alleviate the ergodicity problems of a one dimensional system, which presented a lack of particle exchanges during DMC simulations. Finally, a multiple mesh allows to deal with several length scales, with a reduction of the autocorrelation time of the QMC sampling for heavy atoms.

The LRDMC algorithm is a fully consistent and variational scheme also when non-local potentials are included in the Hamiltonian. Indeed, the non local operator elements which do not frustrate the sign of the Green function can be treated exactly within the LRDMC framework, and in this case it is possible to prove that the mixed energy estimator E_{MA} is variational. Instead the DMC method cannot deal with non-locality, and thus the locality approximation must be applied. However both two methods satisfy the zero-variance property, i.e. if the trial wave function is exact, the energy is exact. In the other cases, the mixed DMC averages do not fulfill the upper bound property, and the locality error affects the DMC calculations, with a dependence on both the shape and the nodes of the trial wave function.

Using the LRDMC method, one can access the pure expectation value E_{FN} of the true Hamiltonian, which is an estimate of the ground state energy more accurate than the usual mixed average, computed in the standard DMC. The LRDMC energy estimate E_{FN} is almost free of the locality error. Only a small dependence on the shape of the trial wave function remains in the LRDMC results due to the fixed node approximation, which limits a totally exact treatment of non-locality.

Finally, we found that LRDMC is much more robust than DMC, since it is stable even for a poor guiding function and a large lattice step. The LRDMC sim-

ulations can be extended over a larger range of lattice space values and therefore the mesh size extrapolation turns out to be more efficient than the DMC time step extrapolation, which usually requires calculations with small time steps.

We believe that this framework can have a wide spread of important applications ranging from nuclear physics [138] to the chemistry of transition metal compounds.

Chapter 4

Iron dimer

4.1 Introduction

Transition elements are present in an astonishing variety of important systems, which range from molecules of biological interest to solids displaying either magnetic or superconducting properties. In particular, iron is of fundamental importance for the life of most species [139], as some of its compounds are the earliest cofactors in enzyme catalysis. On the other hand, its presence in the composition of solids provides peculiar properties, which can usually deserve both theoretical and experimental studies [140, 141], and eventually lead to technological applications.

A lot of work has been done so far in order to understand and predict the properties of iron embedded compounds, and solids, using different computational tools. The large amount of available experimental data allowed to probe the accuracy and reliability of such calculations. It turns out that the density functional theory (DFT), widely used in Quantum Chemistry and Solid State Physics, shows difficulties in treating iron and other near-half-filled d-shell elements, like Cr, Mo, and Ni, because of the tendency of the method to favor d^n over $d^{n-1}s$ configurations. Moreover, the lack of correlation which spoils the proper d-electron description also affects the spin properties of the compound. Indeed, active d-orbitals usually yield an almost degenerate set of levels, which are occupied by maximizing the exchange, and so with a high spin state. In general, DFT is not able to give the correct spin for the eigenstates of the system, when intermediate

transition elements are involved.

Post Hartree Fock methods seem to be much more suitable to describe the physics of transition metal compounds, as they are in principle more accurate than the DFT framework. However, they are computationally heavier, since a large basis is usually needed to get a converged result, in particular when the degeneracy of the system is high, as in the case of iron clusters. In addition to this drawback, their scaling with the system size is unfavorable, and for coupled cluster (CC) methods, it increases significantly with the accuracy of the theory, being of the order of N^7 when also triple correlations are included.

Quantum Monte Carlo (QMC) techniques can be a valid alternative to the above methods for studying iron compounds. In this chapter we focus our attention to the low-lying states of the iron atom and the iron dimer, which has been extensively studied, both from the experimental and theoretical point of view. Therefore, it represents an ideal test case to assess the accuracy of our QMC results with respect to the previous numerical studies. Moreover, a consistent and complete explanation of all experimental data is still missing, despite the simplicity of the molecule, which belongs to a highly symmetric point group ($C_{\infty h}$).

4.2 Computational details

The calculations of the iron atom and the iron dimer are performed using a neon-core pseudopotential, in order to avoid the chemically inert core electrons of the iron atom, and to speed up the QMC simulations. We chose the Dolg's pseudopotential [142], which has previously proven to be reliable at least for atomic Monte Carlo calculations [118]. With the given pseudopotential, we fully optimize the JAGP wave function described in Subsec. 4.2.1 by using the SRH energy minimization scheme, and then we carry out VMC and LRDMC simulations to accurately evaluate the energetics and the structural properties of the systems. Since we have found that the locality effects are negligible, in most cases we use the LRDMC scheme with $\alpha = 0$ and $\gamma = 0$, which is efficient and provides *variational* results. As our goal is to study also the vibrational frequencies of the iron dimer, we perform simulations at different value of the interparticle distance ($R = 3, 3.5, 3.75, 4, 4.5, 5, 6$ a.u.), and we estimate at each point both the force and the energy, at the VMC and LRDMC level. The way to compute forces in

VMC has already been explained in Subsec. 1.4.1. Within a projection Monte Carlo framework, the evaluation of forces is not so trivial due to the presence of the FN constraint. The scheme we use here to compute the interparticle gradient is reported in Subsec. 4.2.2. Finally, our data are fitted using the Morse potential curve, described in Subsec. 4.2.3.

4.2.1 Wave function

The wave function used to describe the electronic structure of the iron atom and dimer has as usual a Jastrow-Slater form. The Slater part is a single reference state (Hartree-Fock) for the atomic case, while it is an AGP correlated wave function for the iron dimer. In both cases, the Slater part requires the evaluation of only a single determinant, as already pointed out in Chap. 2. The single particle atomic states belong to the $[8s5p6d/2s1p1d]$ contracted Gaussian basis set. The AGP geminal function, described in Subsec. 2.2.1, is expanded over this atomic basis, and the chemical bond is the result of the resonance among the various molecular orbitals (MO's) implicitly included in the expansion. In particular, the bonds are related to those coefficients which couple two atomic orbitals belonging to different nuclear sites. Since the geminal is a singlet, the spin and spatial symmetry of the wave function are given by the unpaired MO's, defined as linear combination of atomic orbitals (LCAO).

The Jastrow factor contains up to three-body terms. The two-body term has the form already reported in Eq. 2.14:

$$J_2(\mathbf{r}_1, ..., \mathbf{r}_N) = \exp\left(\sum_{i<j}^{N} u(r_{ij})\right), \tag{4.1}$$

with the usual choice for the Jastrow function u_2:

$$u_2(r_{ij}) = \frac{0.5\, r_{ij}}{1 + b\, r_{ij}}, \tag{4.2}$$

which fulfills the antiparallel electron-electron cusp conditions, where b is a variational parameter, and r_{ij} is the interelectron distance.

In order to fulfill also the nuclear cusp conditions, which are not satisfied by the Slater part if a Gaussian basis set is used, we include explicitly in the Jastrow

factor a one-body term:

$$J_1(\mathbf{r}_1, ..., \mathbf{r}_N) = \exp\left(\sum_{i=1}^{N}\sum_{n=1}^{M} u_1(r_{in})\right), \tag{4.3}$$

where this time $r_{in} = |\mathbf{r}_i - \mathbf{R}_n|$ is the electron nucleus distance, and u_1 is:

$$u_1(r_{in}) = \frac{Z^n_{\text{eff}}\, r_{in}}{1 + b\sqrt[4]{2Z^n_{\text{eff}}}\, r_{in}}. \tag{4.4}$$

The analytic expression of u_1 has already been used by Holzmann *et al.* [143] for QMC simulations of the electron gas and metallic hydrogen, and it provides the correct asymptotic ($r \to \infty$) limit for an homogeneous system. Although a molecule is not homogeneous, we found that this functional form yields good variational energies also for chemical systems. Moreover, b in Eq. 4.4 is the same variational parameter included in Eq. 4.2. Notice in the above Equation that Z^n_{eff} is the charge of the n-th nucleus, which can be either unscreened or effective, if a pseudopotential is used. In particular, the Dolg's pseudopotential for the iron is finite for the non-local components but diverges as $-Z_{\text{eff}}/r$ for the local part. Therefore, it is necessary to include J_1 in the Jastrow factor to satisfy the nuclear cusp condition. Otherwise, if also the local part is finite, as in the Lester's pseudopotentials, Z^n_{eff} must be set to zero, since in that case the nuclear cusp conditions do not need to be applied.

The three-body Jastrow factor J_3 used here has been accurately described in Subsec. 2.2.3. It contains a pairing function analogous to that used in the AGP part, expanded over an atomic Gaussian basis, which includes radial functions of the type: $\exp(-\alpha r^2), r^2\exp(-\alpha r^2), \ldots, r^{2n}\exp(-\alpha r^2)$. For our calculations of iron atom and its compounds, the basis includes s, p, and d angular symmetries.

The number of *independent* variational parameters in the iron dimer wave function are reported below, divided according to the sector where they appear:

- Contracted Gaussian basis set for AGP: 46 parameters;
- Coefficients of AGP geminal expansion and unpaired orbitals: 41 parameters;
- Gaussian basis set for three-body Jastrow geminal: 9 parameters;

- Coefficients of Jastrow geminal expansion: 130 parameters;
- one- and two-body Jastrow factor: 1 parameter.

The total number of independent parameters is 227, fully optimized by using the SRH energy minimization. Notice that we have optimized also the exponents of the Gaussian basis set, in order to reduce the total number of Gaussians in the expansion.

4.2.2 DMC-LRDMC nuclear forces

Computing nuclear forces in the DMC and LRDMC framework is not as trivial as in the VMC case, since the fixed node (FN) approximation, introduced to cope with the fermion sign problem, enforces the fulfillment of boundary conditions which *depend* on the nuclear positions. Therefore, the derivative of the FN energy with respect to a nuclear position must include also the contribution from the variation of the boundary surface, which is not negligible [144]. This contribution is not easy to be taken into account, although a lot of work has been done in order to find a feasible way to compute it exactly.

Here we take into account a more pragmatic approach. Following Ref.[145], we compute the derivative of the mixed estimator of the energy, by explicitly writing the dependence on the local energy appearing in the integral:

$$\begin{aligned}\frac{dE_{MA}}{dR} &= \frac{d}{dR}\frac{\int d\mathbf{x}\,\Phi(\mathbf{x})\Psi_T(\mathbf{x})E_L(\mathbf{x})}{\int d\mathbf{x}\,\Phi(\mathbf{x})\Psi_T(\mathbf{x})} \\ &= \langle \nabla_R E_L\rangle + \left\langle E_L\frac{\nabla_R\Psi_T}{\Psi_T}\right\rangle - E_{MA}\left\langle\frac{\nabla_R\Psi_T}{\Psi_T}\right\rangle \\ &+ \left\langle E_L\frac{\nabla_R\Phi}{\Phi}\right\rangle - E_{MA}\left\langle\frac{\nabla_R\Phi}{\Phi}\right\rangle \end{aligned} \tag{4.5}$$

where R is the interatomic distance. Now, the difficulty in the computation arises from the evaluation of terms like $\left\langle\frac{\nabla_R\Phi}{\Phi}\right\rangle$, which involve the derivative of the projected wave function Φ, whose analytic form is unknown. However, we can rewrite Φ in terms of the Green function projection of Ψ_T, using the formal identity:

$$\Psi(\mathbf{x}') = \int d\mathbf{x} W(\mathbf{x}',\mathbf{x})\Psi_T(\mathbf{x}), \tag{4.6}$$

where $W(\mathbf{x}', \mathbf{x})$ is the imaginary time propagator. The two last terms in Eq. 4.5, after the insertion of the above equality, read:

$$\begin{aligned} \left\langle E_L \frac{\nabla_R \Phi}{\Phi} \right\rangle &= \left\langle E_L \frac{\nabla_R \Psi_T}{\Psi_T} \right\rangle + \left\langle E_L \frac{\nabla_R W}{W} \right\rangle \\ \left\langle \frac{\nabla_R \Phi}{\Phi} \right\rangle &= \left\langle \frac{\nabla_R \Psi_T}{\Psi_T} \right\rangle + \left\langle \frac{\nabla_R W}{W} \right\rangle, \end{aligned} \tag{4.7}$$

where we have isolated the term involving the derivative of the propagator. It includes the dependence on the variation of the boundaries, and gives rise to the difficulties in the exact evaluation of the FN energy derivatives [43]. Here, we approximate the derivatives by *neglecting* this term. Therefore the approximated value of Eq. 4.5 reads:

$$\frac{dE_{MA}}{dR} \approx \langle \nabla_R E_L \rangle + 2 \left\langle E_L \frac{\nabla_R \Psi_T}{\Psi_T} \right\rangle - 2E_{MA} \left\langle \frac{\nabla_R \Psi_T}{\Psi_T} \right\rangle.$$

The above expression is very similar to the Eq. 1.43, written for the evaluation of VMC forces. Indeed the contributions are the same, however the distribution of the QMC sampling is not Ψ_T^2, but the mixed LRDMC (DMC) distribution $\Phi\Psi_T$. Also here we use the space warp transformation to reduce the variance of the Hellmann-Feynman operator, exactly as in the VMC case. Notice that the approximation is exact if the wave function is exact, since in that case the propagator is just the identity. Moreover, Casalegno *et al.* [42] found that this approximation is extremely accurate, at lest for light elements, since the inclusion of the Pulay forces improve significantly its precision, as in the VMC estimate.

4.2.3 Morse potential curve

The analytical function used to represent the interatomic potential curve of many diatomic molecules was proposed by Morse [146] in 1929:

$$V(r) = C + De^{-2a(r-r_0)} - 2De^{-a(r-r_0)}, \tag{4.8}$$

where r_0 is the equilibrium distance or bond length of the molecule, D is its dissociation energy, a is related to its vibrational frequency, and C is an arbitrary energy shift. Usually, this curve fits very well all the potential energy surfaces of diatomic molecules. The only portion where there is a discrepancy is at $r =$

0, where it should be infinity. However this is not a problem, since the lowest vibronic levels are not affected by the behaviour of the potential at the origin, but only around its minimum. The energy E^0 of the zero point motion of the molecule is given by $\frac{1}{2}\hbar\omega_0$, where ω_0 is the vibrational frequency, related to the Morse parameters through the formula:

$$\omega_0 = a\sqrt{\frac{2D}{\mu}}. \tag{4.9}$$

In the above Equation, μ is the reduced mass, which is equal to $M/2$ in the case of a diatomic molecule with nuclear mass M. Once the potential energy curve, yielded by *ab initio* Monte Carlo calculations, is fitted with the Morse potential, the zero point frequency ω_0 is evaluated using Eq. 4.9.

While the energies are fitted with the curve in Eq. 4.8, the forces are fitted with its derivative:

$$F(r) = -\frac{\partial V(r)}{\partial r} = 2aD\left(e^{-2a(r-r_0)} - e^{-a(r-r_0)}\right), \tag{4.10}$$

which provides the parameters a, D and r_0. In order to end up with a more accurate evaluation of the analytic form for the potential energy surface, we perform a multi branch fit of the forces and energies, i.e. we minimize the χ^2 which is the sum of the standard deviations of both the forces and energies.

4.3 Results

4.3.1 Iron atom

The iron atom, being the elementary tile for the construction and calculation of more complex compounds, has been the subject of extensive studies, which tried to assess the accuracy and reliability of various numerical techniques. Indeed, this atom is well known as a system where the correlation effects are extremely important, since there is a strong competition between the $3d$ and $4s$ single particle states, whose relative contribution in the all electron configurations cannot be correctly reproduced at the Hartree-Fock level. For instance, the error in the HF estimate of the $3d^64s^2\ (^5D) \rightarrow 3d^74s\ (^5F)$ excitation energy is larger than 1 eV [83, 118], while the experimental value is only 0.87 eV [147]. Moreover, beside

the strong correlation effects, the relativistic contributions to the low-lying excitation energies of transition metals have proven [148] to be relevant and their inclusion in the calculation is needed in order to reproduce the experimental results. Therefore, already the "simple" iron atom is a challenge and its study requires a very accurate computational tool.

Single and doubles configuration interaction (SDCI) calculations [149] have shown that an extensive basis set $[7s6p4d4f2g]$ and a complete treatment of the excitations are required to account for the differential correlation effects in the $^5D - {}^5F$ separation, whereas the fourth-order Møller-Plesset perturbation theory (MP4) [150] performs poorly for a smaller basis set $(spdf)$, even if the relativistic effects are properly taken into account, with an error which amounts to the 50% of the true experimental value reported above.

Atomic excitation energies of open-shell transition metal elements are difficult to be described also by DFT methods, which favor an s^1d^{n+1} configuration instead of s^2d^n, in order to maximize $d-d$ exchange [151]. Indeed the exchange-correlation functional is approximated and the error can be only partially corrected using an "ad hoc" functional with a self-interaction contribution.

The iron ^5D ground state and its first excited state ^5F, together with the anion and cation ground states, have been studied by Mitas[118], by using QMC techniques. This work clearly shows how the QMC framework can be extremely useful and effective in describing the transition metal compounds, and in particular the iron chemistry. Indeed his DMC results reproduce the experimental data with a mean error of only 0.1 eV. In particular, the discrepancy for the $^5D - {}^5F$ excitation energy (0.84(6) eV) is within the statistical uncertainty of the calculation. For his study, Mitas used a Dolg's pseudopotential, which includes scalar relativistic corrections [142] and eliminates the neon-core electrons. Indeed, the $3s$ and $3p$ shells have a semicore character [118, 149] and need to be included in the calculation in order to reduce the core polarization effects and obtain more accurate results.

In our QMC calculations we use the Dolg's pseudopotential as well. Contrary to the Mitas's variational wave function, which includes the $4s^2 \rightarrow 4p^2$ excitations in the Slater part, our Ψ_T does not contain a multi-configuration state, but it is a single Hartree-Fock determinant multiplied by a three-body Jastrow factor. Both the determinantal and the Jastrow part have been optimized with

the SRH energy minimization. The functional form of the Jastrow factor is reported in Subsec. 4.2.1, while the initial Slater term has been obtained from an Hartree-Fock calculation, performed with the MOLPRO [152] package, and using a $(8s5p6d)/[2s1p1d]$ Gaussian basis set. We compute the ${}^5D - {}^5F$ excitation energy, by carrying out both VMC and LRDMC simulations. The values are reported in Tab. 4.1. The ${}^5D - {}^5F$ transition is correctly reproduced both by the VMC and LRDMC energies, and we obtain the same result as in Ref. [118] Since we have used a single reference wave function, it is apparent that the correlation in the ground state and first excited state of the neutral atom is purely dynamical, and the near degeneracy effects coming from the $4s - 4p$ shells are negligible. Therefore, particularly in this case, a Jastrow correlated wave function is much more suitable and accurate than a multi-determinant expansion, which needs a lot of terms to converge.

Finally we have also estimated the FN LRDMC energies, which are more accurate than the mixed averages, since they reduce a lot the locality error (see Chap. 3). It turns out that the locality error amounts approximately to 0.1 eV in the total energies and it is negligible in the energy differences. Thus, the mixed average estimates are accurate enough to reproduce correctly the experimental transition energies. Indeed, as already pointed out by Mitas [118], the Dolg's pseudopotential is well localized around the nucleus, and hence the locality error is very small.

This first analysis confirms the results obtained by Mitas and shows that the QMC framework should be an effective tool to study the iron compounds.

Table 4.1: Total and excitation energies for the iron atom. The ${}^5F$ and ${}^5D$ energies for the neutral atom are reported. From the total energies we calculate the excitation ${}^5D \rightarrow {}^5F$, expressed in eV and compared with the experiment. The LRDMC results are obtained with $\alpha = 0$.

	VMC	LRDMC MA	LRDMC FN	exp.
$5D$ (Hartree)	-123.7211(15)	-123.7819(11)	-123.7888(25)	
$5F$ (Hartree)	-123.6911(20)	-123.7520(11)	-123.7572(24)	
$5D \rightarrow 5F$ (eV)	0.82(7)	0.81(4)	0.86(9)	0.87 [1]

[1] From Ref. [147]

4.3.2 Iron dimer

In 1979 Harris and Jones carried out the first *ab initio* calculation of the iron dimer [153]. By using the DFT approach with a local spin density approximation (LSDA), they predicted a $^7\Delta_u$ [1] symmetry for the ground state of Fe_2. The single particle configuration of the $^7\Delta_u$ state is $\sigma_g^2(3d)\ \pi_u^4\ \delta_g^3\delta_u^{*,2}\pi_g^{*,2}\sigma_u^{*,1}(3d)\ \sigma_g^2(4s)$, where we specify the occupation and the angular symmetry of the valence MO, which come mainly from the linear combination of $4s$ and $3d$ atomic states. Moreover, we write in brackets the type of atomic orbitals (AO) that predominates in the corresponding σ MO. The antibonding MO are indicated with an asterisk. A subsequent study [154], based on HF-CI calculations, confirmed the overall symmetry of the iron dimer ground state, but the single particle MO occupations were in disagreement, and yielded a state where the chemical bond was mainly due to the $4s$ electrons.

In the 80's the first experimental data were available; a vibrational frequency of 299.6 cm^{-1} was determined for the ground state by Raman spectroscopy[155], while the interatomic distance was found to range from 3.53(24) a.u. [156] to 3.82(4) a.u. [157]. These geometrical properties were obtained from extended x-ray absorption on fine structure (EXAFS) for iron dimers in an inert gas matrix. Argon was used in the former measure, neon in the latter. Also the energetics of Fe_2 was analized. Mass spectrometric investigations provided a first estimate of the dissociation energy, $D^0(Fe_2) = 0.78(18)$eV [154]. Further studies coming from the collision-induced dissociation of Fe_n^+ clusters yielded an higher but more precise value: $D^0(Fe_2) = 1.14(10)$eV [158].

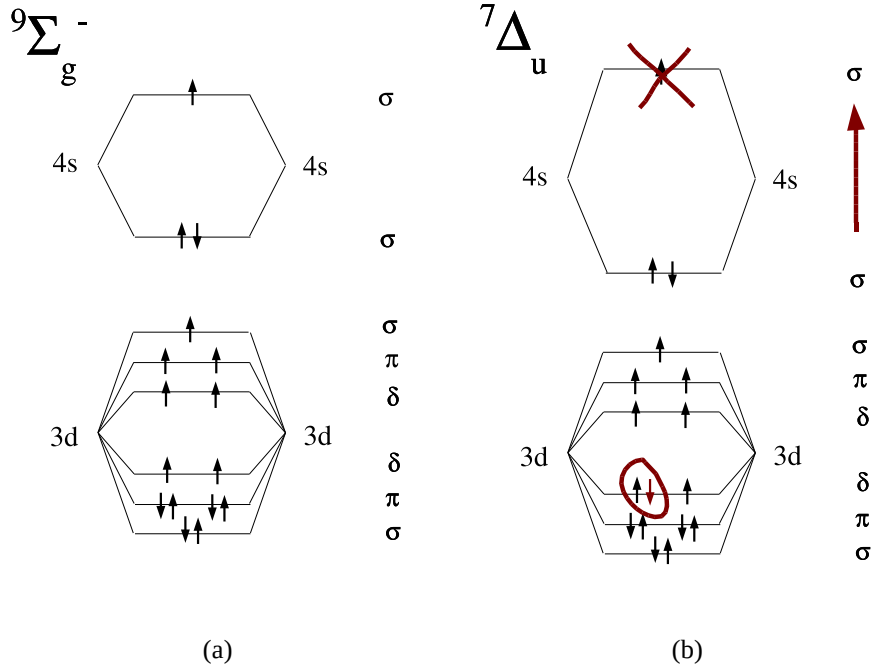
Other theoretical studies were performed, based mainly on DFT methods with improved functionals[151, 159–161] and multi reference CI calculations[162]. All of them found the ground state of Fe_2 to be $^7\Delta_u$, in accordance with the earliest findings. In the most cases, both the energetics and structural properties were taken into account. A common trend of these results is the good agreement with the experimental equilibrium distance, but a poor correspondence with the

[1] The notation $^7\Delta_u$ refers to the total spin $2S+1=7$, the total angular momentum $L=\Delta$, and the total parity under inversion ($\mathbf{r} \to -\mathbf{r}$), where u is the negative parity, and g is the positive one. Moreover, the Σ states can be classifi ed by another quantum number, related to the reflection with respect all planes containing the molecular axis; it can be $+$, if the wave function is invariant, or $-$ otherwise. This sign is written on the top right-hand corner of the label Σ.

vibrational frequencies. In particular, all DFT calculations overestimate the zero point motion, providing a too narrow interatomic potential. Also the ionic states were studied, and for the anion iron dimer all calculations yielded the $^8\Delta_g$ as the ground state of Fe_2^-. Though its computed equilibrium distance is correct, the vibrational frequencies are overestimated as in the case of Fe_2.

In 1986 Leopold *et al.* did an important experiment[163] of negative ion photoelectron spectroscopy, with the aim to study the low-lying electronic states of Fe_2. A sample of Fe_2^- is prepared and excited by an incoming photon. Depending on energy of the photon and on the structure of the excited compound, an electron is detached and its kinetic energy measured. From the difference between the kinetic energy of the electron and the frequency of the photon, it is possible to obtain the spectrum of the compound, and to access the properties of the low-lying electronic states. In this way the EA of Fe_2 has been measured for the first time, and its value is 0.902(8) eV. The spectrum of Fe_2 appears remarkably simple, with only two peaks, one corresponding to the excitation from the Fe_2^- to the Fe_2 ground state, the other related to the transition from the Fe_2 ground state to its first excited state allowed by the selection rules. Both the two states reveal the *same* vibrational frequency and bond length. Few years later, Leopold [164] argued that the simplest explanation of these data is to admit that the $^7\Delta_u$ is not the ground state of Fe_2, and the $^8\Delta_g$ is not the ground state of Fe_2^- as well. This interpretation is based on two hypothesis, verified for the photoelectron spectra of the first-row transition metal ions: The first is that the two band systems observed in the Fe_2^- spectrum are due to detachment from $4s$-like MO's, and the second assumes the $\sigma_g(4s)$ is a strongly bonding MO. Let us suppose that the ground state of Fe_2^- is $^8\Sigma_u^-$. Its configuration turns out to be $\sigma_g^2(4s)\ \sigma_u^{*,2}(4s)3d^{13}$, and the $4s$ electron detachment would produce two possible states, with the same orbital configuration $\sigma_g^2(4s)\ \sigma_u^{*,1}(4s)3d^{13}$ but with the $\sigma_u^*(4s)$ electron high or low coupled to the $3d^{13}$ "bulk". These two states would correspond to the first and second peak of the spectrum, respectively. Indeed, they would display the same structural properties, as the only difference would be the spin coupling among the single particle states. It is easy to see that if the ground state of Fe_2^- was $^8\Delta_g$, whose configuration is $\sigma_g^2(4s)\ \sigma_u^{*,1}(4s)3d^{14}$, it would be impossible to have two states with the same orbital configuration but different spin orientation by detachment of one $4s$ electron. Therefore the ground state of Fe_2 was suggested to be $^9\Sigma_g^-$, its first excited state

to have the ${}^7\Sigma_g^-$ symmetry, while the ground state of Fe_2^- was assigned to ${}^8\Sigma_u^-$. The qualitative scheme of the ${}^9\Sigma_g^-$ and ${}^7\Delta_u$ states is given in Fig. 4.1.



(a) (b)

Figure 4.1: The schematic representations of the energy levels occupation is reported for the ${}^9\Sigma_g^-$ and ${}^7\Delta_u$ states, on the left and right side respectively. Notice that the $4s$ atomic levels are shifted with respect to the $3d$ levels, in order to have a separate representation of $3d$ and $4s$ occupations. Also the hybridization among the orbitals is not reported. If the $\sigma(4s)$ MO's split is enhanced, it is more convenient to occupy a δ_g than a $\sigma_u(4s)$ MO and thus have a ${}^7\Delta_u$ ground state symmetry, as in the case of DFT calculations. Otherwise the ${}^9\Sigma_g^-$ state is the lowest in energy, as found by MRCI and QMC calculations.

Recently a MRCI calculation[165] and another DFT study with coupled cluster corrections[166] supported the Leopold's idea with a numerical verification. In the same period, other DFT works[151, 161] stated the validity of the original belief. Thus, it is worth studying the problem by using QMC methods in order to clarify the nature of the ground state of the anion and the neutral iron dimer. Indeed, the QMC framework is supposed to be very accurate and more suitable

than the other methods to study in particular transition metal compounds. With this purpose, we have computed the energy and the structural properties of $^9\Sigma_g^-$ and $^7\Delta_u$, the two competing configurations for the ground state of Fe_2. Similarly, we have studied the $^8\Sigma_g^-$ and $^8\Delta_u$ states of Fe_2^- to determine which has the lowest energy and the closest properties to the experiment.

Table 4.2: The VMC and LRDMC ($\gamma = 0$,$\alpha = 0$) results are reported for the iron dimer. We calculated total energies at the minimum of the interatomic potential, equilibrium distances R_e, and vibrational frequencies ω_e for two states of the neutral dimer ($^7\Delta_u$ and $^9\Sigma_g^-$) and other two states of Fe_2^- ($^8\Delta_g$ and $^8\Sigma_u^-$). The experimental results for the ground state of Fe_2 are: $R_e = 3.818(11)\ a.u.$[157], and $\omega_e = 300(15)\ cm^{-1}$[163]. Instead for the ground state of the anion Fe_2^- the following valued have been found: $R_e = 3.970(21)\ a.u.$[163], and $\omega_e = 250(20)\ cm^{-1}$[163].

	Energy (Hartree)	R_e (a.u.)	ω_e (cm^{-1})
VMC			
$^7\Delta_u$	-247.358(3)	3.856(12)	388(20)
$^9\Sigma_g^-$	-247.361(2)	4.151(15)	300(16)
$^8\Delta_g$	-247.359(3)	3.936(14)	339(11)
$^8\Sigma_u^-$	-247.377(2)	4.401(29)	192(14)
LRDMC ($\gamma = 0, \alpha = 0$)			
$^7\Delta_u$	-247.532(3)	3.894(18)	373(32)
$^9\Sigma_g^-$	-247.551(2)	4.093(19)	301(15)
$^8\Delta_g$	-247.558(3)	3.908(14)	354(24)
$^8\Sigma_u^-$	-247.580(4)	4.270(33)	217(23)

Interatomic potentials

In order to estimate accurately the equilibrium position, the vibrational frequency and also the dissociation energy of an iron dimer state, we choose to compute the energy and the force at different values of the interatomic distance ($R = 3, 3.5, 3.75, 4, 4.5, 5, 6$), and then fit the data with the Morse potential curve, described in Subsec. 4.2.3. The multi branch fit of both the energy and the force at the same time guarantees a more precise determination of the potential energy

surface, and thus a more accurate evaluation of all structural properties of the compound. The results of the fit are reported in Figures 4.2, 4.3, 4.4, and 4.5, for the VMC and LRDMC data corresponding to the $^9\Sigma_g^-$ and $^8\Sigma_u^-$ states. Analogous fits have been performed for the other states studied here.

Ground states and EA

From both Tab. 4.2 and Fig. 4.8, it is apparent that the ground state of Fe_2 is $^9\Sigma_g^-$, while the lowest energy state of Fe_2^- is $^8\Sigma_g^-$. This result is absolutely non trivial, in particular for the ground state of the neutral dimer, since at the VMC level the two states $^9\Sigma_g^-$ and $^7\Delta_u$ were still very close in energy. Instead the LRDMC energies clearly discriminate between the two configurations, by favoring the $^9\Sigma_g^-$ symmetry. Therefore, our findings support the Leopold's interpretation, beside a previous MRCI calculation[165] and an hybrid DFT-CC study[166]. In the latter work, however, the anion dimer was not even taken into account, instead in the former the authors carried out calculations for both the neutral and the anion ground state. They were able to confirm the Leopold's hypothesis, but they failed to compute the electron affinity, i.e. the energy difference between the Fe_2 and the Fe_2^- ground state. Indeed they stated that the basis set used in their calculations was not large enough to accurately evaluate this quantity[167]. Therefore they published only energy differences between states which belong to the same molecular species. In our approach, the computed electron affinity is 0.44(8) eV at the VMC level, which becomes 0.78(12) eV after the LRDMC simulations, being this value in reasonable agreement with the experiment. For the first time, the LRDMC method allows not only to support the Leopold's idea, but also to reproduce the position of the first peak in the photoelectron spectrum.

Vibrational frequencies

The vibrational frequencies of $^9\Sigma_g^-$ and $^8\Sigma_g^-$ are in good agreement with the experimental data, especially those values obtained from the fit of the LRDMC points. Indeed our best QMC estimate for the vibrational frequency of the Fe_2 ground state is $\omega_e = 301(15) cm^{-1}$, which matches perfectly the value 300(15) cm^{-1}, coming from photoelectron spectroscopy, and the value 299.6 cm^{-1}, provided by Raman spectroscopy. Moreover, though at the VMC level the vibrational fre-

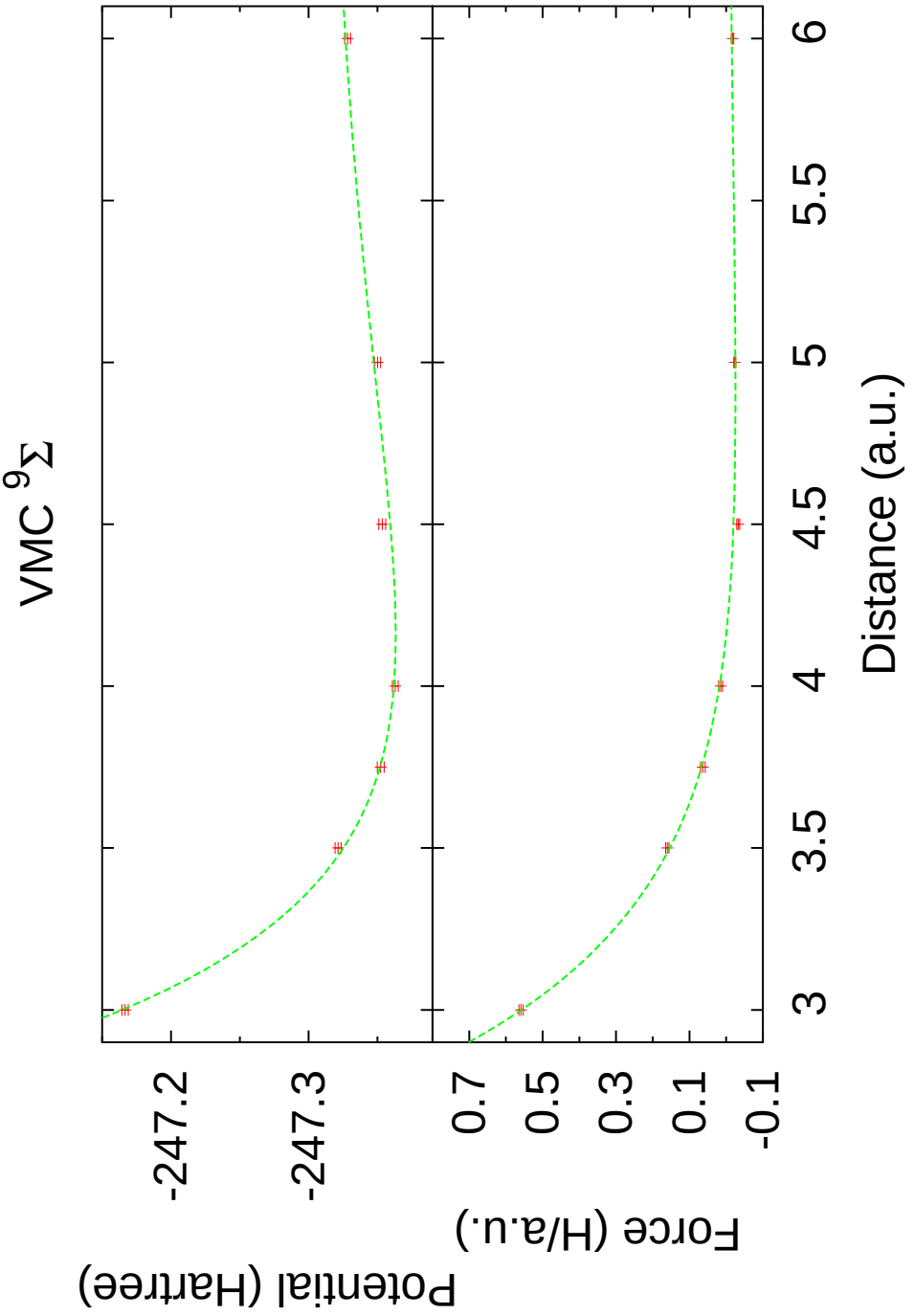


Figure 4.2: Multi branch fit with the Morse function for the interatomic potential of the state $^9\Sigma_g^-$ of Fe_2 calculated at the VMC level. The points represent the results for the energy (upper panel) and the force (lower panel) calculated at different values of the interatomic distance $R = 3, 3.5, 3.75, 4, 4.5, 5, 6$ atomic units.

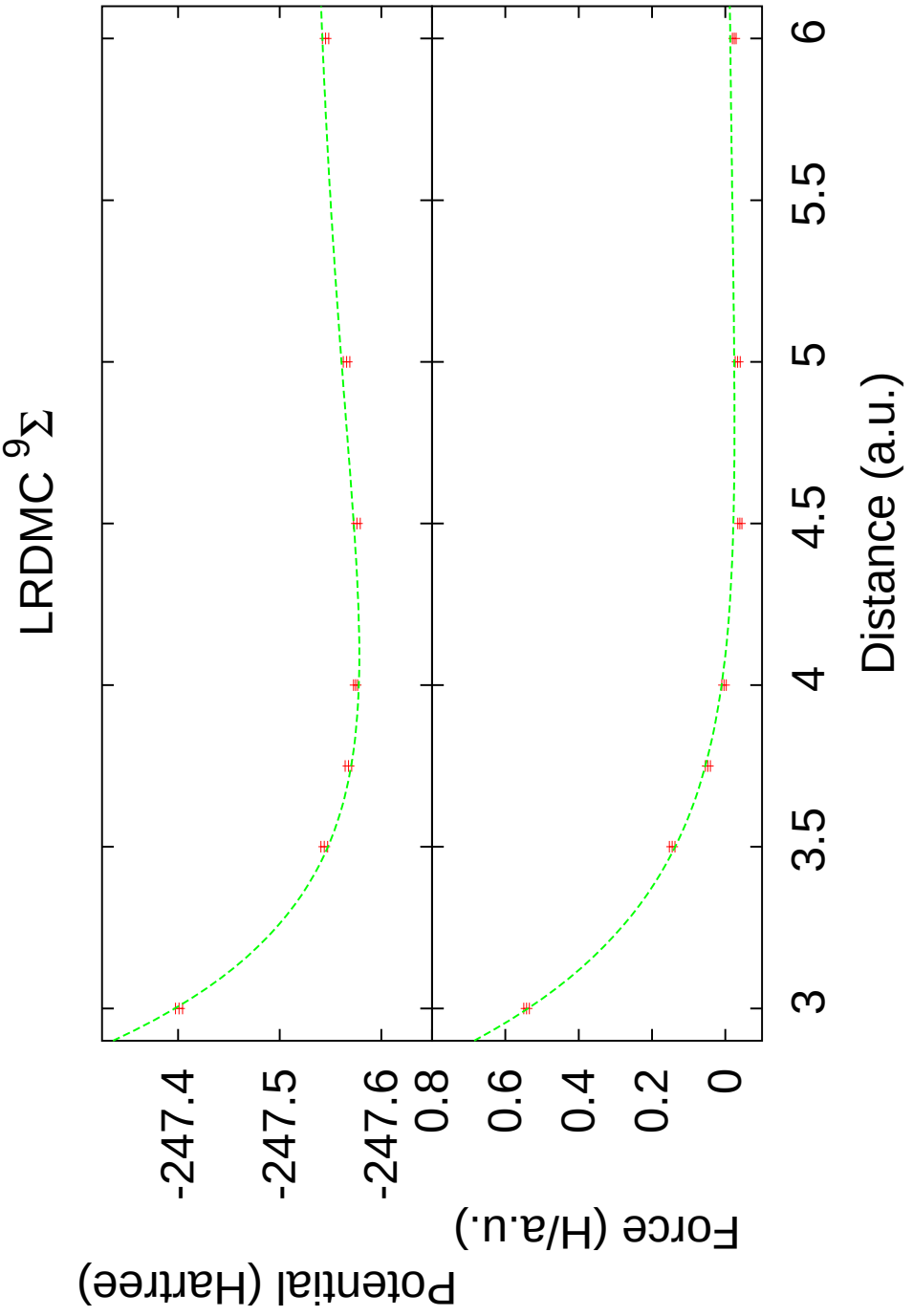


Figure 4.3: Multi branch fit with the Morse function for the interatomic potential of the state $^9\Sigma_g^-$ of Fe_2 calculated at the LRDMC ($\alpha = 0, \gamma = 0$) level. The points represent the results for the energy (upper panel) and the force (lower panel) calculated at different values of the interatomic distance $R = 3, 3.5, 3.75, 4, 4.5, 5, 6$ atomic units.

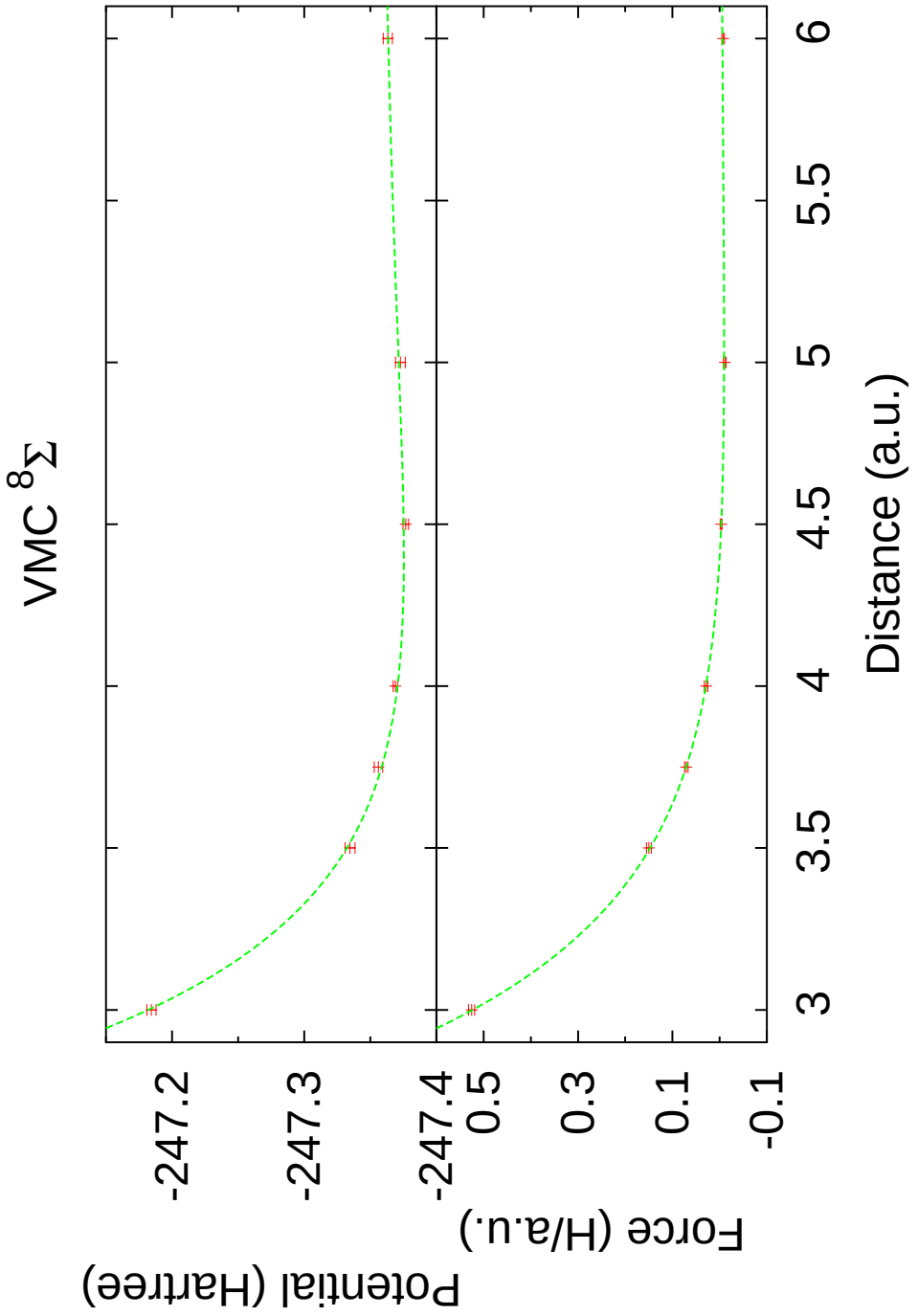


Figure 4.4: Multi branch fit with the Morse function for the interatomic potential of the state $^8\Sigma_u^-$ of Fe_2^- calculated at the VMC level. The points represent the results for the energy (upper panel) and the force (lower panel) calculated at different values of the interatomic distance $R = 3, 3.5, 3.75, 4, 4.5, 5, 6$ atomic units.

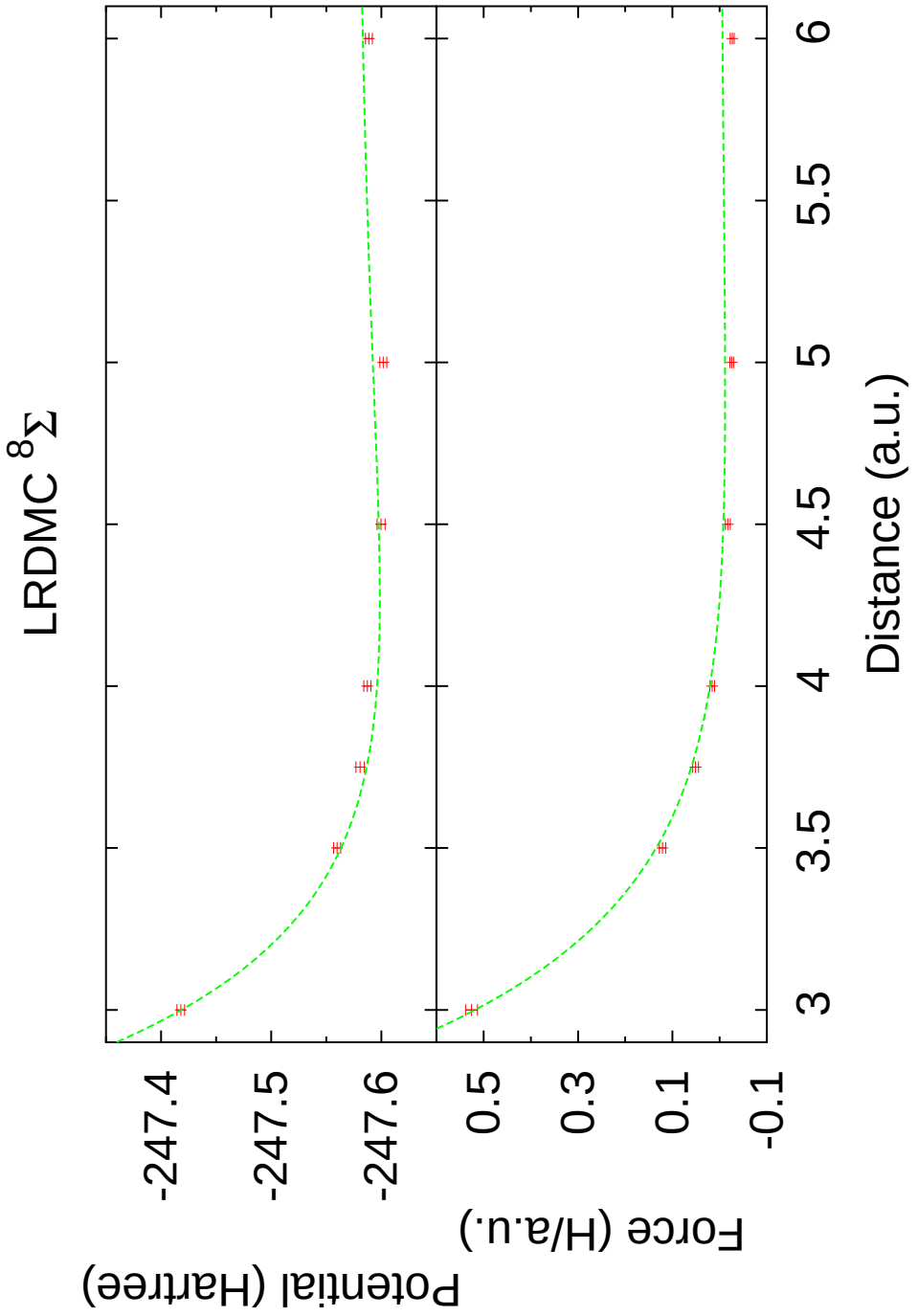


Figure 4.5: Multi branch fit with the Morse function for the interatomic potential of the state $^8\Sigma_u^-$ of Fe_2^- calculated at the LRDMC ($\alpha = 0, \gamma = 0$) level. The points represent the results for the energy (upper panel) and the force (lower panel) calculated at different values of the interatomic distance $R = 3, 3.5, 3.75, 4, 4.5, 5, 6$ atomic units.

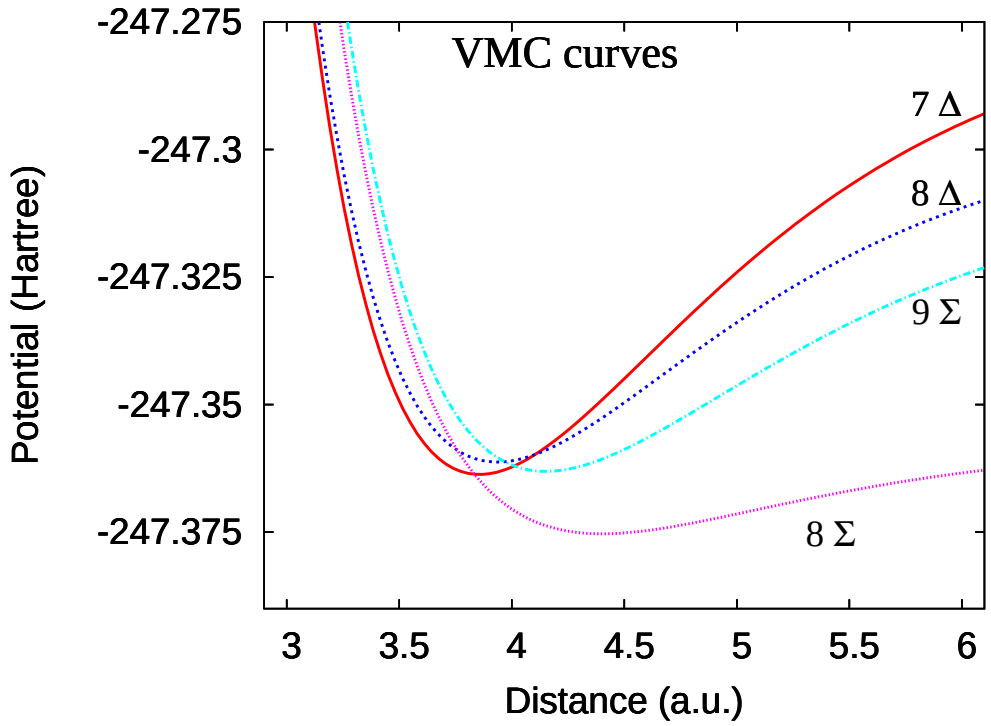


Figure 4.6: Overall view of the energy landscape (interatomic potential) of the following states: $^7\Delta_u$, $^9\Sigma_g^-$ for Fe_2, and $^8\Delta_g$, $^8\Sigma_u^-$ for Fe_2^-. The curves have been obtained from the fit of the VMC data with the Morse function.

quency of the anion ground state, $\omega_e = 192(14)cm^{-1}$, is significantly far from the experimental value 250(25) cm^{-1} yielded by photoelectron spectroscopy, the LRDMC scheme provides a result much closer and compatible with the experimental one. Indeed the best LRDMC value is $\omega_e = 217(23)cm^{-1}$, which is consistent with the experiment within statistical errors. Notice that our calculations correctly reproduce the softening of the vibrational mode going from the neutral to the anion ground state. Finally, the true ground states of the two dimers can be recognized also by this property, since the $^7\Delta_u$ and $^8\Delta_g$ states have much higher vibrational frequency, totally incompatible with the experiment. It is interesting to highlight that the vibrational frequency for $^7\Delta_u$ and $^8\Delta_g$ computed with QMC simulations agree with those calculated by DFT methods for the same states. This probably means that the drawback of DFT simulations for transi-

tion metal compounds is not the computation of the zero point motion, as stated elsewhere [161], which turns out to be *correct* for the given state, but the main problem is the wrong ordering of energy levels, which drives the wrong state to be the lowest in energy. Once again, the wrong energy ordering can be related to a poor exchange-correlation functional included in the DFT scheme.

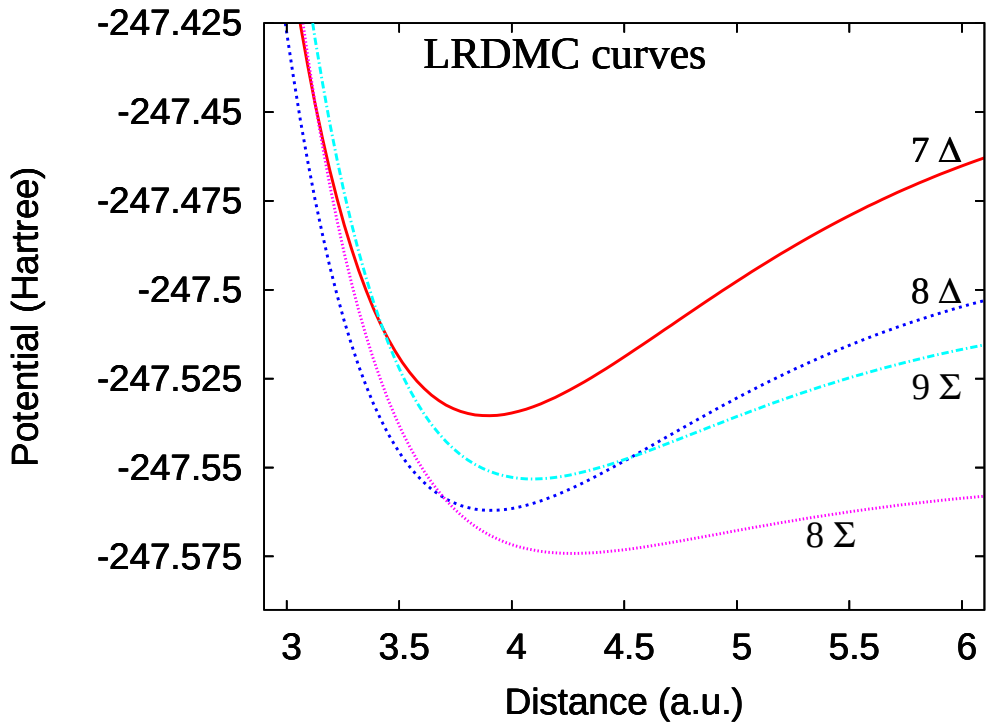


Figure 4.7: Overall view of the energy landscape (interatomic potential) of the following states: ${}^7\Delta_u$, ${}^9\Sigma_g^-$ for Fe_2, and ${}^8\Delta_g$, ${}^8\Sigma_u^-$ for Fe_2^-. The curves have been obtained from the fit of the LRDMC ($\alpha = 0, \gamma = 0$) data with the Morse function.

Equilibrium bond lengths

The equilibrium bond lengths provided by our calculations are generally larger than the experimental data, obtained in 1980 and 1982 for the neutral iron dimer embedded in a matrix of inert gas. The error in the equilibrium distance of ${}^9\Sigma_g^-$ amounts to about 0.3 atomic units. This is probably due to the fact that a percentage of correlation energy is still missing in the LRDMC results. Notice however

that the equilibrium distance of the LRDMC curve is slightly smaller than the distance found in the VMC interatomic potential. Therefore by improving the energy, the bond length moves towards the experimental value. Another source of error can come from the evaluation of forces within the LRDMC framework, which are approximated, as we have seen in Subsec. 4.2.2. A further possible explanation of the discrepancy is that the experiment is not completely reliable, and it must be revised. It is worth mentioning that our findings are supported by Hübner and Sauer [165], that found the same bond lengths for the same ground states, by carrying out accurate MRCI calculations for Fe_2 and Fe_2^-.

The bond length of the anion has been measured only indirectly, since the unique available data are taken from the photoelectron experiment, which revealed a *variation* of the equilibrium distance during the excitation from the anion to the neutral iron dimer. Indeed a harmonic Franck-Condon analysis of the vibronic band intensity profile yielded a bond elongation of 0.15(4) a.u. on electron attachment [163]. Now, the difference between the Fe_2 and Fe_2^- bond length of our LRDMC calculation amounts to 0.18 a.u., which is in perfect agreement with the elongation of the anion dimer measured in the experiment. This is another indication of the quality and reliability of our results.

Dissociation energy

The dissociation energy D^0 has been directly measured only for the neutral dimer. Therefore we will focus our attention only on the dissociation of Fe_2. It it easy to show that the ground state ${}^9\Sigma_g^-$ will dissociate into the ${}^5F$ and ${}^5D$ states of the iron atom. Thus, a possible way to estimate the dissociation energy is to compute the energy difference:

$$D^0(Fe_2) = E({}^5F) + E({}^5D) - E(Fe_2). \tag{4.11}$$

Another way is to extract the value form the fit of the potential surface, but the two results could be different if the wave function is not size consistent as in this case the angular momentum of the compound is not the sum of the momenta of the fragments. The most reliable experimental result provides the value of 1.14(10) eV. At the VMC level our calculations yield $D^0 = -1.39(9)$ eV from energy differences and $D^0 = 1.57(14)$ eV from the fit. These values are totally different

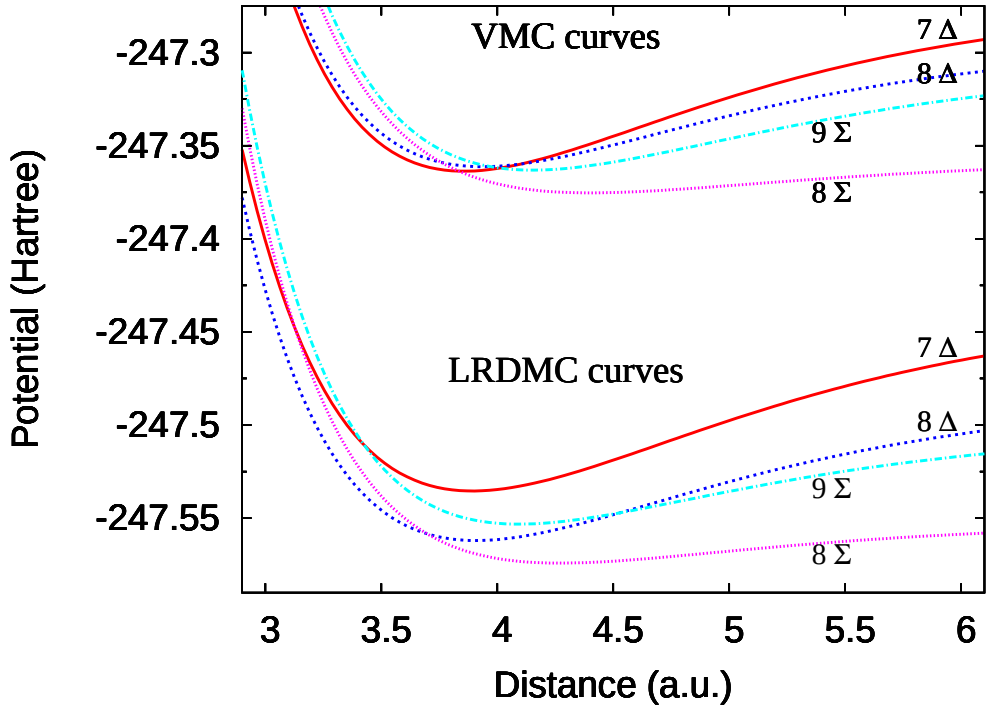


Figure 4.8: Plot of interatomic potentials obtained form *ab initio* VMC and LRDMC ($\alpha = 0, \gamma = 0$) calculations for the following states: $^7\Delta_u$, $^9\Sigma_g^-$ for Fe_2, and $^8\Delta_g$, $^8\Sigma_u^-$ for Fe_2^-.

due to the lack of size consistency, as stated above. At the LRDMC level, we get $D^0 = 0.48(7)$ eV from energy differences and $D^0 = 1.40(12)$ eV from the fit. Here, the values are much closer each other than in the VMC case, since the LRDMC reduces the gap in the accuracy between atomic and molecular calculations, and therefore improves the estimate in Eq. 4.11. Indeed the atoms are very well described by our variational wave function, while it is more difficult to describe the molecules with the same accuracy by using the same variational ansatz, because their electronic structure is more complex and a LRDMC projected wave function is required for an accurate evaluation of their properties. Notice that the values for D^0 obtained from the fit are in reasonable agreement with the experiment both in VMC and LRDMC calculations.

4.4 Conclusions

Our calculations clearly assign the ground state of Fe_2 and Fe_2^- to $^9\Sigma_g^-$ and $^8\Sigma_u^-$ respectively. In particular the LRDMC results favor the $^9\Sigma_g^-$ state with respect to $^7\Delta_u$, which turns out to be 0.5 eV higher in energy. This result is in contrast to a widely accepted assumption, since $^7\Delta_u$ has been thought as the true ground state in many DFT studies[151, 159–161]. Moreover, the failure to observe Fe_2 in electron spin resonance (ESR) experiments[168] supported the idea of an orbitally degenerate ground state, like the $^7\Delta_u$ configuration, but another explanation attributes this failure to the presence of a large zero-field split (larger than 8 cm^{-1}, and thus producing an energy split not detectable by the experimental setup). We believe that the second interpretation of the ESR experiment is valid, and leaves the possibility to have $^9\Sigma_g^-$ as ground state of the iron dimer.

On the other hand, most of the photoelectron spectroscopy data [163] are nicely explained by our findings. The electron affinity, the vibrational frequencies and the bond elongation are correctly reproduced, by assuming that the ground state of Fe_2^- is $^8\Sigma_u^-$, and the first peak in the spectrum corresponds to the transition from $^8\Sigma_u^-$ to $^9\Sigma_g^-$.

The characterization of the states involved in the Fe_2^- photoelectron spectrum was already proposed by Leopold [164], but a complete and consistent theoretical verification of his interpretation was still missing. Indeed recent MRCI calculations[165] found both $^8\Sigma_u^-$ and $^9\Sigma_g^-$ as ground states of Fe_2^- and Fe_2, but the authors did not report any energy difference between these two states. Moreover, one could accept the Leopold interpretation, without inferring that the lowest energy Fe_2 state observed in the photoelectron spectrum is its true ground state. Indeed Bauschlicher[169] suggested that the Fe_2 ground state is $^7\Delta_u$, the Fe_2^- ground state is $^8\Sigma_u^-$, and the $^7\Delta_u$ state is not observed in the photodetachment spectra since it is a two-electron process, and therefore its transition is avoided by the selection rules. Our LRDMC calculations rule out this possibility, by clarifying the nature of the anion and neutral iron dimer ground states, *and* at the same time reproducing the spectroscopic constants of Fe_2 and Fe_2^-.

We are still working to reproduce also the second peak in the photoelectron spectrum, which is related to an excited state of Fe_2. We hope to find a theoretical explanation also of this transition, by performing VMC and LRDMC simulations

for the possible excited states connected to the $^8\Sigma_u^-$ ground state by photodetachment. The most reasonable state is $^7\Sigma_g^-$, but till now we have not found yet a numerical confirmation of this hypothesis.

Conclusions and perspectives

In this thesis, we developed a good variational ansatz and an accurate QMC method to tackle the numerical studies of transition metal compounds.

We introduced the Jastrow correlated geminal (JAGP) wave function [170, 171] to describe accurately a large variety of chemical compounds at the variational level, ranging from the covalent localized bond of the first row diatomic molecules to the delocalized bonds present in the benzene molecule [171, 172], passing through the d-electron correlation of the iron dimer. In all these cases we found a good agreement with the experimental results, both for the energetics and the structural properties. An exception is represented by the academic case of beryllium dimer, where we found only a Van der Walls like minimum, instead of a slightly tighter bond and a smaller equilibrium distance. We believe that the triplet correlations, missing in the AGP part, are crucial to get the correct result, though they should partially accounted for by the Jastrow factor.

The lattice regularized diffusion Monte Carlo (LRDMC) framework has proven to be an efficient and suitable computational tool[173], in particular in the presence of pseudopotentials, where the stability and variationality of the results are guaranteed, even if poor variational guiding functions are employed. Within this framework it is no longer necessary to introduce the locality approximation (LA) when non local potentials are included in the Hamiltonian, contrary to the standard diffusion Monte Carlo method which requires the LA and may provide non variational results. Moreover the possibility to access a pure energy estimate allows to control and reduce the locality error, opening the route to obtain much more reliable and accurate fixed node energies. Since the use of pseudopotentials seems unavoidable at least in the most interesting and non trivial cases, our method is of extreme importance for future Monte Carlo applications to realistic systems.

The applications to the benzene ring, presented in Chap. 2, and the iron dimer, studied in Chap. 4, are encouraging. The former study clearly highlights the importance of the resonance among the chemical bonds, naturally taken into account by the JAGP ansatz. The delocalized p_z orbitals are very well described by the superposition of bonds, while the intra-atomic correlations in the carbon sites are controlled by the Jastrow factor. We have shown the importance of the interplay between the Jastrow and the AGP part, which is essential to have an accurate variational wave function. In this way, the atomization energies, the geometrical structures and the ionization potentials are accurately reproduced by our calculations. It would be interesting to extend this study to the energetics (reduction potential, ionization energies, electron affinity) of DNA bases and base pairs, quantities of great importance to characterize excess electron and hole transfer, which are involved in radiation damage as well as in the development of DNA technologies.

The iron dimer study proves that only a framework which carefully includes the correlation effects in a variational way is able to yield accurate and reliable results. The novel LRDMC method, by projecting the already accurate JAGP wave function to the lowest energy state, allowed to explain the first peak in the photoelectron spectrum of Fe_2^-. The calculation of the spectroscopic constants of Fe_2 and Fe_2^- is an extremely difficult task, and we were able for the first time to reproduce correctly the electron affinity, the bond elongation and the vibrational frequencies provided by the experimental data, after clarifying the nature of the ground states of the anion and neutral dimers.

The iron chemistry beyond the simple case of iron dimer requires a truly multiconfigurational treatment of the correlation, because of the degeneracy among many energy levels. The density functional theory fails in describing the d-correlated electrons and the post Hartree Fock methods are not feasible since an accurate enough multi configuration expansion would include too many terms. Therefore the QMC framework seems to be the only valid alternative to the previous schemes for studying transition metal compounds. Indeed the JAGP wave function allows to include an exponentially increasing number of configurations at the computational cost of only one determinant, and already for the Fe_2S_2 molecule the LRDMC algorithm with a JAGP guiding function should represent a clear advantage with respect to configuration interaction like methods [174]. However a more efficient parametrization of the geminal coefficients would be required to reduce

the total number of parameters and to afford the optimization of more complex systems. In order to verify these assumptions with some important computational success, a lot of work is still needed, since in this thesis we have presented just a first tentative to deal with strongly correlated transition metal compounds, after the development of the accurate JAGP ansatz and the proposal of the promising LRDMC method.

Appendix A

Stabilization of the SR technique

Whenever the number of variational parameters increases, it often happens that the stochastic $(N+1) \times (N+1)$ matrix

$$s_{k,k'} = \frac{\langle \Psi | O_k O_{k'} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{A.1}$$

becomes singular, i.e. the condition number, defined as the ratio $\sigma = \lambda_N / \lambda_1$ between its maximum λ_N and minimum eigenvalue λ_1, is too large. In that case the inversion of this matrix generates clear numerical instabilities which are difficult to control especially within a statistical method. Here $O_k = \frac{d \ln \Psi(x)}{d\alpha_k}$ are the operators corresponding to the variational parameters α_k appearing in the wave function Ψ for $k = 1, \cdots N$, whereas for $k = 0$ the operator O_0 represents the identity one.

The first successful proposal to control this instability was to remove from the inversion problem (Eq.1.28 is Section 1.4.1), required for the minimization, those directions in the variational parameter space corresponding to exceedingly small eigenvalues λ_i.

In this appendix we describe a better method. As a first step, we show that the reason of the large condition number σ is due to the existence of "redundant" variational parameters that do not make changes to the wave function within a prescribed tolerance ϵ. Indeed in practical calculations, we are interested in the minimization of the wave function within a reasonable accuracy. The tolerance ϵ may represent therefore the distance between the exact normalized variational wave function which minimizes the energy expectation value and the approximate

acceptable one, for which we no longer iterate the minimization scheme. For instance $\epsilon = 1/1000$ is by far acceptable for chemical and physical interest. A stable algorithm is then obtained by simply removing the parameters that do not change the wave function by less than ϵ from the minimization. An efficient scheme to remove the "redundant parameters" is also given.

Let us consider the N normalized states orthogonal to Ψ, but not orthogonal among each other:

$$|e_i\rangle = \frac{(O_k - s_{k,0})|\Psi\rangle}{\sqrt{\langle\Psi|(O_k - s_{k,0})^2|\Psi}}. \tag{A.2}$$

where $s_{k,0}$ is defined in Eq. A.1. These normalized vectors define N directions in the $N-$dimensional variational parameter manifold, which are independent as long as the determinant S of the corresponding $N \times N$ overlap matrix

$$\bar{s}_{k,k'} = \langle e_k|e_{k'}\rangle \tag{A.3}$$

is non zero. The number S is clearly positive and it assumes its maximum value 1 whenever all the directions e_i are mutually orthogonal. On the other hand, let us suppose that there exists an eigenvalue $\bar{\lambda}$ of $\bar{s}$ smaller than the square of the desired tolerance ϵ^2, then the corresponding eigenvector $|v>=\sum_i a_i|e_i\rangle$ is such that:

$$\langle v|v\rangle = \sum_{i,j} a_i a_j \bar{s}_{i,j} = \bar{\lambda} \tag{A.4}$$

where the latter equation holds due to the normalization condition $\sum_i a_i^2 = 1$. We arrive therefore to the conclusion that it is possible to define a vector v with almost vanishing norm $|v| = \sqrt{\lambda} \leq \epsilon$ as a linear combination of e_i, with at least some non zero coefficient. This implies that the N directions e_k are linearly dependent within a tolerance ϵ and one can safely remove at least one parameter from the calculation.

In general whenever there are p vectors v_i that are below the tolerance ϵ the optimal choice to stabilize the minimization procedure is to remove p rows and p columns from the matrix (A.3), in such a way that the corresponding determinant of the $(N-p) \times (N-p)$ overlap matrix is maximum.

From practical purposes it is enough to consider an iterative scheme to find a large minor, but not necessarily the maximum one. This method is based on the inverse of $\bar{s}$. At each step we remove the $i-th$ row and column from $\bar{s}$ for which

$\bar{s}_{i,i}^{-1}$ is maximum. We stop to remove rows and columns after p inversions. In this approach we exploit the fact that, by a consequence of the Laplace theorem on determinants, $\bar{s}_{k,k}^{-1}$ is the ratio between the described minor without the $k-th$ row and column and the determinant of the full $\bar{s}$ matrix. Since within a stochastic method it is certainly not possible to work with a machine precision tolerance, setting $\epsilon = 0.001$ guarantees a stable algorithm, without affecting the accuracy of the calculation. The advantage of this scheme, compared with the previous one[20], is that the less relevant parameters can be easily identified after few iterations and do not change further in the process of minimization.

Appendix B

Spin polarized geminal wavefunction

In this appendix, we consider the most general geminal wavefunction with definite spin $S = \frac{N_\uparrow - N_\downarrow}{2}$, where $N_\uparrow$ ($N_\downarrow$) is the number of spin–up (spin–down) electrons and $N_\uparrow > N_\downarrow$ is assumed. To this purpose we introduce second quantized fermionic fields (see e.g. Fetter and Walecka [175]) $\psi^\dagger(\mathbf{r}, \sigma)$ and $\psi(\mathbf{r}, \sigma)$, where $\mathbf{r}$ is the electron position and $\sigma = \pm 1/2$ is its spin projection along the z–axis. These fields satisfy the canonical anticommutation rules:

$$\left\{\psi(\mathbf{r}, \sigma), \psi^\dagger(\mathbf{r}', \sigma')\right\} = \delta_{\sigma\sigma'}\delta(\mathbf{r} - \mathbf{r}'). \tag{B.1}$$

In these notations, the most general wavefunction with definite spin can be formally written in the following way:

$$|\Psi\rangle = P_N \prod_{i=N_\downarrow+1}^{N_\uparrow} \psi^\dagger_{i,\uparrow} \, \exp(\Phi^\dagger)|0\rangle, \tag{B.2}$$

where P_N is the projection on the given number of particles $N = N_\uparrow + N_\downarrow$, $|0\rangle$ denotes the vacuum of electrons and $\psi^\dagger_{i,\uparrow}$ is the most generic (Bogoliubov) orbital with spin $S = 1/2$:

$$\psi^\dagger_{i,\uparrow} = \int d\mathbf{r} \, \left(\phi_i^<(\mathbf{r})\psi(\mathbf{r}, \downarrow) + \phi_i^>(\mathbf{r})\psi^\dagger(\mathbf{r}, \uparrow)\right), \tag{B.3}$$

which is defined by the orbital functions $\phi_i^>$ for the creation of a particle with spin up and $\phi_i^<$ for the annihilation of a particle with spin down. For instance, a

conventional Slater determinant of spin-up particles can be written as $\prod_i \psi^\dagger_{i,\uparrow}|0\rangle$, where $\phi^<_i = 0$. It is clear therefore that this representation is more general and may provide a wavefunction Ψ with more variational freedom than the conventional Slater determinants.

Finally, the pairing creation operator $\Phi^\dagger$ is a singlet, namely $\exp(\Phi^\dagger)|0\rangle$ has spin zero, and is defined by a generic symmetric function $\Phi(\mathbf{r},\mathbf{r}') = \Phi(\mathbf{r}',\mathbf{r})$:

$$\Phi^\dagger = \int d\mathbf{r} \int d\mathbf{r}'\ \Phi(\mathbf{r}',\mathbf{r})\psi^\dagger(\mathbf{r},\downarrow)\psi^\dagger(\mathbf{r}',\uparrow). \tag{B.4}$$

Our purpose is to show here that the value of the wavefunction Ψ can be simply computed, similarly to a conventional Slater determinant, on each configuration $x = \{\mathbf{r}_{1,\uparrow},\ldots,\mathbf{r}_{N_\downarrow,\downarrow}\}$, where $\mathbf{r}_{i,\uparrow}$ are the positions of spin–up particles and $\mathbf{r}_{i,\downarrow}$ are the spin–down ones. These configurations can be generally written as:

$$\langle x| = \langle 0| \prod_{i=1}^{N_\uparrow} \psi(\mathbf{r}_i,\uparrow) \prod_{j=1}^{N_\downarrow} \psi(\mathbf{r}_j,\downarrow). \tag{B.5}$$

Indeed the value F of the wavefunction on $\langle x|$ is:

$$F = \langle x|\Psi\rangle = \left\langle 0 \left| \prod_i \psi(\mathbf{r}_i,\uparrow) \prod_j \psi(\mathbf{r}_j,\downarrow) \prod_{k=N_\downarrow+1}^{N_\uparrow} \psi^\dagger_{k,\uparrow}\ \exp(\Phi^\dagger) \right| 0 \right\rangle. \tag{B.6}$$

Now we insert the identity $\exp(-\Phi^\dagger)\exp(\Phi^\dagger)$ between each fermionic field in the above equation (B.6):

$$\begin{aligned} F &= \Big\langle 0 \Big| \exp(\Phi^\dagger)\exp(-\Phi^\dagger)\psi(\mathbf{r}_1,\uparrow)\exp(\Phi^\dagger)\cdots\exp(-\Phi^\dagger)\psi(\mathbf{r}_{N_\downarrow},\downarrow)\exp(\Phi^\dagger)\cdots \\ &\quad \cdots\exp(-\Phi^\dagger)\psi^\dagger_{N_\uparrow,\uparrow}\exp(\Phi^\dagger) \Big| 0 \Big\rangle. \end{aligned} \tag{B.7}$$

Exploiting the relation valid for generic operators A and B:

$$\exp(-A)\ B\ \exp(A) = B - [A,B] + \frac{1}{2}[A,[A,B]] + \ldots \tag{B.8}$$

one is able to evaluate the following terms:

$$\begin{aligned} \exp(-\Phi^\dagger)\psi(\mathbf{r}_i,\uparrow)\exp(\Phi^\dagger) &= \psi(\mathbf{r}_i,\uparrow) - \int d\mathbf{r}\ \Phi(\mathbf{r}_{i,\uparrow},\mathbf{r})\psi^\dagger(\mathbf{r},\downarrow) \\ \exp(-\Phi^\dagger)\psi(\mathbf{r}_i,\downarrow)\exp(\Phi^\dagger) &= \psi(\mathbf{r}_i,\downarrow) + \int d\mathbf{r}\ \Phi(\mathbf{r},\mathbf{r}_{i,\downarrow})\psi^\dagger(\mathbf{r},\uparrow) \end{aligned}$$

$$\exp(-\Phi^\dagger)\psi^\dagger_{i,\uparrow}\exp(\Phi^\dagger) \;=\; \psi^\dagger_{i,\uparrow} + \int d\mathbf{r}\int d\mathbf{r}'\;\Phi(\mathbf{r},\mathbf{r}')\phi_i^<(\mathbf{r}')\psi^\dagger(\mathbf{r},\uparrow) \quad \text{(B.9)}$$

In order to derive the above relations, notice that all the terms in the RHS of Eq. B.8 are always zero beyond the first two. After substituting the expressions in Eq. B.7 and by using $\langle 0|\exp(\Phi^\dagger) = \langle 0|$, $\psi(\mathbf{r},\sigma)|0\rangle = 0$ and $\langle 0|\psi^\dagger(\mathbf{r},\sigma) = 0$, one can iteratively apply the canonical commutation rules (B.1) and a simplified result for F is obtained:

$$F = \left\langle 0 \left| \prod_{i=1}^{N_\uparrow} \psi(\mathbf{r}_i,\uparrow) \prod_{i=1}^{N_\uparrow} \tilde{\psi}^\dagger_{i,\uparrow} \right| 0 \right\rangle \quad \text{(B.10)}$$

where $\tilde{\psi}^\dagger_{i,\uparrow}$ is the creator of an orbital function of the type (B.3), with transformed orbitals:

$$\begin{aligned} \phi_i(\mathbf{r}) &= \Phi(\mathbf{r},\mathbf{r}_{i,\downarrow}) \quad \text{for } i = 1,\cdots N_\downarrow \\ \bar{\phi}_i(\mathbf{r}) &= \phi_i^>(\mathbf{r}) + \int d\mathbf{r}'\;\Phi(\mathbf{r},\mathbf{r}')\phi_i^<(\mathbf{r}') \quad \text{for } i = N_\downarrow + 1 \cdots N_\uparrow \end{aligned} \quad \text{(B.11)}$$

Then the final value of F can be simply computed by a *single* determinant, as it represents just the value of a $N_\uparrow \times N_\uparrow$ Slater determinant with orbitals given in (B.11) on the spin-up configurations, yielding the final expression (2.8) reported in Chapter 2, Subsection 2.2.1.

Appendix C

Geminal as a multiconfi guration wavefunction

As pointed out in Sec.2.2.1, for the minimal geminal expansion, i.e. $M = N^{\downarrow}$, the AGP wavefunction is reduced to the HF one. Otherwise, if $M > N^{\downarrow}$, it becomes a multiconfiguration wavefunction. We prove this property by starting with the expression for an antisymmetrized geminal wavefunction written in the second quantized form:

$$|\Psi\rangle = P_N \exp(\Phi^\dagger)|0\rangle = \left(\Phi^\dagger\right)^{N^\downarrow}|0\rangle. \tag{C.1}$$

Here, for the sake of simplicity, we assume that the system is unpolarized, but the prove can be easily extended to the polarized case. As reported in Eq.2.10, the expansion for the geminal is:

$$\phi(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) = \sum_{i=1}^{M} \lambda_i \phi_i(\mathbf{r}^\uparrow)\phi_i^*(\mathbf{r}^\downarrow). \tag{C.2}$$

We define the creation operator for a particle in the i-*th* state with spin σ as:

$$\hat{\psi}_{i\sigma}^\dagger = \int d\mathbf{r}\ \phi_i(\mathbf{r})\psi^\dagger(\mathbf{r}, \sigma). \tag{C.3}$$

where $\{\phi_i\}$ are the orbitals defining the states, which we do not assume to be orthogonal among each other. Then we can rewrite the geminal creation operator in Eq.B.4 in terms of the single particle operators:

$$\Phi^\dagger = \sum_{i=1}^{M} \lambda_i \hat{\psi}_{i\uparrow}^\dagger \hat{\psi}_{i\downarrow}^\dagger. \tag{C.4}$$

Therefore, if we put this expression in Eq.C.1 and use the anticommutation rules, the AGP wavefunction reads:

$$|\Psi\rangle = \sum_{\substack{i_1,\ldots,i_{N^\downarrow}\in M \\ i_1\neq i_2\neq\cdots\neq i_{N^\downarrow}}} \lambda_{i_1}\cdots\lambda_{i_{N^\downarrow}}\ \hat{\psi}^\dagger_{i_1\uparrow}\hat{\psi}^\dagger_{i_1\downarrow}\cdots\hat{\psi}^\dagger_{i_{N^\downarrow}\uparrow}\hat{\psi}^\dagger_{i_{N^\downarrow}\downarrow}|0\rangle. \tag{C.5}$$

Each term in the above sum is a Slater determinant with orbitals $\phi_{i_1}\cdots\phi_{i_{N^\downarrow}}$. Hence the antisymmetrized geminal product is equivalent to a linear combination of them, with coefficients given by the product of those λ's selected by the configuration present in each Slater determinant. If $M = N^\downarrow$, the summation has only one term, related to the unique possible configuration of $N^\downarrow$ single particle orbitals: this is the HF case. Otherwise, if $M > N^\downarrow$, one deals with a multiconfiguration wavefunction. Let us consider for instance the geminal for Be. The HF case corresponds to the following geminal expansion:

$$\Phi_{Be}(\mathbf{r},\mathbf{r}') = \phi_{1s}(\mathbf{r})\phi_{1s}(\mathbf{r}') + \phi_{2s}(\mathbf{r})\phi_{2s}(\mathbf{r}'), \tag{C.6}$$

where all the λ have been set to 1, because they affect only the multiplicative constant in front of the total wavefunction. The correlated case just beyond the minimal geminal expansion is:

$$\Phi_{Be}(\mathbf{r},\mathbf{r}') = \lambda_{1s}\phi_{1s}(\mathbf{r})\phi_{1s}(\mathbf{r}') + \lambda_{2s}\phi_{2s}(\mathbf{r})\phi_{2s}(\mathbf{r}') + \lambda_{2p}\sum_{\mu\in\{x,y,z\}}\phi_{2p_\mu}(\mathbf{r})\phi_{2p_\mu}(\mathbf{r}'), \tag{C.7}$$

where we used the real basis representation. The corresponding Configuration Interaction wavefunction is:

$$|\Psi\rangle = \lambda_{1s}\lambda_{2s}|1s^2 2s^2| + \lambda_{1s}\lambda_{2p}\sum_\mu |1s^2 2p_\mu^2| + \lambda_{2s}\lambda_{2p}\sum_\mu |2s^2 2p_\mu^2| + \lambda_{2p}\lambda_{2p}\sum_{\mu\neq\nu}|2p_\mu^2 2p_\nu^2|. \tag{C.8}$$

Here the λ's amplitude sets the relative weight of the corresponding configurations, playing a crucial role in curing nondynamic correlation effects and improving the electronic nodal structure of the wavefunction, with the computational cost of a single determinant.

Appendix D

Size consistency of the 3-body Jastrow factor

In order to prove the size consistency property of the three body Jastrow factor described in Sec. 2.2.3, let us take into account a system composed by two well separated subsystems A and B, which are distinguishable and whose dimensions are much smaller than the distance between themselves; in general they may contain more then one atom. In this case the Jastrow function J_3 (2.16) can be written as $J_3 = e^U$ with:

$$U = \frac{1}{2} \sum_{\substack{i,j \in A \\ i \neq j}} \phi(r_i, r_j) + \frac{1}{2} \sum_{\substack{i,j \in B \\ i \neq j}} \phi(r_i, r_j) + \sum_{i \in A} \sum_{j \in B} \phi(r_i, r_j), \quad \text{(D.1)}$$

where we have explicitly considered the sum over different subsystems. As usual, the two particle function $\phi(r_i, r_j)$ is expanded over a single particle basis ψ, centered on each nucleus of the system:

$$\phi(r_i, r_j) = \sum_{m,n} \lambda^{m,n} \psi^m(r_i) \psi^n(r_j). \quad \text{(D.2)}$$

The indices n and m refer not only to the basis elements but also to the nuclei which the orbitals are centered on.

The self consistency problem arises from the last term in Eq. D.1, i.e. when the electron r_i belongs to A and r_j to B. If the Jastrow is size consistent, whenever A and B are far apart from each other this term must vanish or at most generate a one

body term that is in turn size consistent, as we are going to show in the following. In the limit of large separation all the $\lambda^{m,n}$ off diagonal terms connecting any basis element of A to any basis element of B must vanish. The second requirement is a sufficiently fast decay of the basis set orbitals $\psi(r)$ whenever $r \to \infty$, except at most for a constant term C_n which may be present in the single particle orbitals, and is useful to improve the variational energy.

For the sake of generality, suppose that the system A contains M_A nuclei and N_A electrons. The first requirement implies that:

$$\phi(r_i, r_j) = \sum_{m,n \in A} \lambda^{m,n} \psi^m(r_i)\psi^n(r_j) + \sum_{m,n \in B} \lambda^{m,n} \psi^m(r_i)\psi^n(r_j), \tag{D.3}$$

instead the second allows to write the following expression for the mixed term in Eq. D.1:

$$\sum_{i \in A} \sum_{j \in B} \phi(r_i, r_j) = N_B \sum_{n \in A} C_n P_n + N_A \sum_{m \in B} C_m P_m, \tag{D.4}$$

where the factors P_n are one body terms defined as:

$$P_n = \begin{cases} \sum_{m \in A} \lambda^{n,m} \sum_{i \in A} \psi^m(r_i) & \text{if } n \in A \\ \sum_{m \in B} \lambda^{n,m} \sum_{i \in B} \psi^m(r_i) & \text{if } n \in B \end{cases} \tag{D.5}$$

Notice that if all the orbitals decay to zero, the size consistency is immediately recovered, since the sum in Eq. D.4 vanishes. Analogously to the derivation we have done to extract the one body contribution from the mixed term, the other two terms on the RHS of Eq. D.1 can be rearranged in the following form:

$$\frac{1}{2} \sum_{\substack{i,j \in A \\ i \neq j}} \phi(r_i, r_j) = (N_A - 1) \sum_{n \in A} C_n P_n + \text{ two body terms}, \tag{D.6}$$

and the sum in Eq. D.1 can be rewritten as:

$$U = (N-1) \sum_{n \in A} C_n P_n + (N-1) \sum_{n \in B} C_n P_n + \text{ two body size consistent terms}. \tag{D.7}$$

Therefore the size consistency implies that the scaling of the C_n with the total number of particle N is:

$$C_n = \frac{c_n}{N-1}, \tag{D.8}$$

as mentioned in Subsection 2.2.3 of Chapter 2.

Appendix E

LRDMC correlated sampling

Let us suppose that a LRDMC simulation follows the evolution of the Hamiltonian H, and we want to compute at the same time the ground state energy of the Hamiltonian H^{ϵ}, which differs from H for a perturbation of order ϵ. In order to evaluate the energy of H^{ϵ}, by exploiting the random walk generated by the Green function $G = \Lambda - H$, we resort to a reweighting procedure. There are mainly four processes which characterize the LRDMC framework (see for instance the flow chart reported in Tab. 3.1).

- Timing (evaluation of the persistence time of the random walk in the configuration x):

$$\tau_x = -log(r)/N_x, \tag{E.1}$$

  where r is a random number with uniform deviate $0 < r \leq 1$, and $N_x = \sum_{x'(\neq x)} G_{x',x}$ is the normalization of the off diagonal Green function $G_{x',x}$.

- Weighting (recursive evaluation of the weight w_i the walker i assumes depending on its local energy, its permanence time, and its history):

$$w_i = w_i \ \exp\left[-\tau_x(E_L(x) - E_T)\right], \tag{E.2}$$

  where $E_L(x)$ is the local energy, and E_T is the guessed ground state energy, given from input.

- Moving (the walker is moved to a new configuration x' according to the off-diagonal Green function matrix elements):

$$p_{x',x} = G_{x',x}/N_x, \tag{E.3}$$

where $p_{x',x}$ is the transition probability matrix.

- Branching (reconfiguration process to control the exponential grow of the walkers during the LRDMC evolution). This reconfiguration process amounts to generate a new set of walkers $\{w_j^{\text{new}}, x_j^{\text{new}}\}_{j=1,M}$ in terms of the given M walkers $\{w_j, x_j\}_{j=1,M}$, keeping their total number M fixed. Each new walker $(w_j^{\text{new}}, x_j^{\text{new}})$ will have the same weight $w_j^{\text{new}} = \bar{w} = \sum_j w_j/M$ and an arbitrary configuration x_j^{new} among the possible old ones $\{x_j\}_{j=1,M}$, chosen with a probability $p_k = w_k/\sum_j w_j$. After this reconfiguration the new M walkers have by definition the same weights and most of the irrelevant walkers with small weights are dropped out. In this way, the reconfiguration plays the desired stabilization effect.

All these four processes need to be modified or reweighted in order to correct the dynamics given by $G = \Lambda - H$ on the basis of the evolution driven by $G^\epsilon = \Lambda - H^\epsilon$. Let us analyze step by step all the necessary modifications.

- Timing. Since the perturbed Green function G^ϵ differs from the unperturbed one G, the persistence time needs to be modified:

$$\tau_x^\epsilon = -log(r)/N_x^\epsilon, \tag{E.4}$$

  where r is the same random number in Eq. E.1 generated for the evolution of G, but the normalization N_x' changes, since $N_x^\epsilon = \sum_{x'(\neq x)} G_{x',x}^\epsilon$.

- Weighting. Not only the persistence time has changed, but also the local energy, therefore the weight w_i of the perturbed Green function G^ϵ is updated according to the following relation:

$$w_i^\epsilon = w_i^\epsilon \, \exp\left[-\tau_x^\epsilon (E_L^\epsilon(x) - E_T)\right], \tag{E.5}$$

  where $E_L^\epsilon(x)$ is the local energy of H^ϵ, and E_T is the same as in Eq. E.2. Notice that the whole history of the walker changes, although the random path for the perturbed Hamiltonian H^ϵ is generated according to G.

- Moving. Since this is a correlated sampling scheme, the random walk for the unperturbed and perturbed Hamiltonian *must* be the same, as we want to exploit the same sampling in order to compute small energy differences

related to slightly different Hamiltonians, with an error as tiny as possible. Usually the correlated framework provides errors much smaller than those corresponding to differences of values coming from independent samplings. Hence the path is selected through the transition probability p of Eq. E.3, but we have to correct the dynamics by including in the weights w_i^ϵ a factor which takes into account the difference between p and p^ϵ, the transition probability related to the perturbed Green function G^ϵ:

$$\begin{aligned} w^{\text{correction}} &= \frac{p^\epsilon}{p} = \frac{G^\epsilon_{x',x}}{G_{x',x}} \frac{\sum_{x'(\neq x)} G_{x',x}}{\sum_{x'(\neq x)} G^\epsilon_{x',x}}, \\ w_i^\epsilon &= w_i^\epsilon \, w^{\text{correction}}. \end{aligned} \tag{E.6}$$

This reweighting procedure must be applied for every walker after each move.

- Branching. As in the case of the moving step, the branching is based on the unperturbed evolution, and therefore also this process must be corrected using a reweighting technique. In particular, the new walkers are chosen according to the probability $p_k = w_k / \sum_j w_j$, while for the perturbed evolution they should be chosen according to $p_k^\epsilon = w_k^\epsilon / \sum_j w_j^\epsilon$. Moreover, the new walkers have the same average weight $w_j^{\text{new}} = \bar{w} = \sum_j w_j / M$ as the old ones, but if we followed the perturbed evolution, also this average weight would be replaced by $\bar{w^\epsilon} = \sum_j w_j^\epsilon / M$. Hence, the resulting factor, which takes into account both the reweighted branching probability and the new average weight, is:

$$w_i^{\epsilon \text{ generation}} = \frac{w_i^\epsilon}{w_i} \frac{\sum_j w_j}{M}, \tag{E.7}$$

for each walker and for each generation [1]. In order to extrapolate this reweighted branching procedure to the infinite time evolution, we need to apply the forward walking technique, already described by Calandra and Sorella [47] for GFMC calculations, by propagating the weights in Eq. E.7.

[1] A "generation" is the set of walkers after a branching, *generated* by the LRDMC evolution after an imaginary time propagation T (see Tab. 3.1)

At the end, after calculating and storing the weights and the energies both for the unperturbed and the perturbed evolution, we can compute the following quantity:

$$\frac{\partial E(\epsilon)}{\partial \epsilon} \approx \frac{E(\epsilon) - E(0)}{\epsilon}, \tag{E.8}$$

also in the limit of small ϵ, where $E(\epsilon)$ is the ground state energy of H^{ϵ}, and $E(0)$ is the ground state energy of the unperturbed Hamiltonian H, which has driven the LRDMC projection.

Bibliography

[1] M. York Darensbourg. *Nature*, 433:589, 2005.

[2] C. Tard, X. Liu, S. K. Ibrahim, M. Bruschi, L. De Gioia, S. C. Davies, X. Yang, L.-S. Wang, G. Sawers, and C. J. Pickett. *Nature*, 433:589, 2005.

[3] L. Noodleman, T. Lovell, T. Liu, F. Himo, and R. A. Torres. *Curr. Opin. Chem. Biol.*, 6:259, 2002.

[4] P. W. Anderson. *Science*, 235:1196, 1987.

[5] S. Yunoki and S. Sorella. *Phys. Rev. Lett*, 92:157003, 2004.

[6] D. M. Ceperley and B. J. Alder. *Phys. Rev. Lett.*, 45:566, 1980.

[7] F. Pederiva, C. J. Umrigar, and E. Lipparini. *Phys. Rev. B*, 62:8120, 2000.

[8] J. C. Grossman and L. Mitas. *Phys. Rev. Lett.*, 74:1323, 1995.

[9] C. Pierleoni, B. Bernu, D. M. Ceperley, and W. R. Magro. *Phys. Rev. Lett.*, 73:2145, 1994.

[10] W. L. McMillan. *Phys. Rev.*, 138:A442, 1965.

[11] R. Assaraf and M. Caffarel. *Phys. Rev. Lett.*, 83:4682, 1999.

[12] R. Assaraf and M. Caffarel. *J. Chem. Phys.*, 113:4028, 2000.

[13] J. B. Anderson. *J. Chem. Phys.*, 63:1499, 1975.

[14] J. B. Anderson. *J. Chem. Phys.*, 65:4121, 1976.

[15] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester Jr. *J. Chem. Phys.*, 63:5593, 1982.

[16] D. M. Ceperley. *J. Stat. Phys.*, 43:815, 1986.

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

[18] L. Mitáš, E. L. Shirley, and D. M. Ceperley. *J. Chem. Phys.*, 95:3467, 1991.

[19] S. Sorella. *Phys. Rev. B*, 71:R241103, 2005.

[20] S. Sorella. *Phys. Rev. B*, 64:024512, 2001.

[21] A. D. Sokal. Monte carlo methods in statistical mechanics: Foundations and new algorithms. In *Cours de Troisième Cycle de la Physique en Suisse Romande*. Lausanne, 15, 22 et 29 juin 1989.

[22] J. M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Methuen, London, 1964.

[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. *J. Chem. Phys.*, 21:1087, 1953.

[24] M. H. Kalos and P. A. Withlock. *Monte Carlo methods Volume I: Basics*. Wiley, New York, 1986.

[25] H. Flyvbjerg and H. G. Petersen. *J. Chem. Phys.*, 91:461, 1989.

[26] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins. *Phys. Rev. Lett.*, 60:1719, 1988.

[27] P. R. C. Kent, R. J. Needs, and G. Rajagopal. *Phys. Rev. B*, 59:12344, 1953.

[28] N. D. Drummond and R. J. Needs. *submitted to Phys. Rev. B.*

[29] Martin Snajdr and Stuart M. Rothstein. *J. Chem. Phys.*, 112:4935, 2000.

[30] A. Harju, B. Barbiellini, S. Siljamaki, R.M. Nieminen, and G. Ortiz. *Phys. Rev. Lett.*, 79:1173, 1997.

[31] Xi Lin, Hongkai Zhang, and Andrew Rappe. *J. Chem. Phys.*, 112:2650, 2000.

[32] M. W. Lee, M. Mella, and A. M. Rappe. *J. Chem. Phys.*, 122:244103, 2005.

[33] C. J. Umrigar and Claudia Filippi. *Phys. Rev. Lett.*, 94:150201, 2005.

[34] C. Filippi and S. Fahy. *J. Chem. Phys.*, 112:3523, 2000.

[35] Friedemann Schautz and Claudia Filippi. *J. Chem. Phys.*, 120:10931, 2004.

[36] D. M. Ceperley and M. Dewing. *J. Chem. Phys.*, 110:9812, 1999.

[37] Z. Sun and W. A. Lester. *J. Chem. Phys.*, 97:7585, 1992.

[38] C. J. Umrigar. *Int. J. Quantum Chem. Symp.*, 23:217, 1989.

[39] C. Filippi and C. J. Umrigar. *Phys. Rev. B*, 61:R16291, 2000.

[40] R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985.

[41] S. Tanaka. *J. Chem. Phys.*, 100:7416, 1994.

[42] M. Casalegno, M. Mella, and A. M. Rappe. *J. Chem. Phys.*, 118:7193, 2003.

[43] S. De Palo, S. Moroni, and S. Baroni. *cond-mat/0111486*.

[44] C. J. Umrigar, M. P. Nightingale, and K. J. Runge. *J. Chem. Phys.*, 99:2865, 1993.

[45] M. H. Kalos, D. Levesque, and L. Verlet. *Phys. Rev. A*, 9:257, 1974.

[46] D. M. Ceperley. *J. Stat. Phys.*, 63:1237, 1991.

[47] M. Calandra Buonaura and S. Sorella. *Phys. Rev. B*, 57:11446, 1998.

[48] N. Trivedi and D. M. Ceperley. *Phys. Rev. B*, 41:4552, 1990.

[49] Sandro Sorella and Luca Capriotti. *Phys. Rev. B*, 61:2599, 2000.

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

[51] Sandro Sorella. *Phys. Rev. Lett.*, 80:4558, 1998.

[52] S. Sorella. *arXiv:cond-mat/0201388*.

[53] J. Petersen. *Comp. Phys. Comm.*, 73:72, 1992.

[54] W. Heitler and F. London. *Z. Physik*, 44:455, 1927.

[55] B. Barbiellini. *J. Phys. Chem. Solids*, 61:341, 2000.

[56] A. C. Hurley, J. E. Lennard-Jones, and J. A. Pople. *Proc. R. Soc. London, Ser. A*, 220:446, 1953.

[57] J. R. Schrieffer. *The Theory of Superconductivity*. 5th printing. Addison-Wesley, 1994.

[58] S. Bratoz and P. Durand. *J. Chem. Phys.*, 43:2670, 1965.

[59] B. Weiner and O. Goscinski. *Phys. Rev. A*, 22:2374, 1980.

[60] G. Bessis, C. Murez, and S. Bratoz. *Int. J. Quantum Chem.*, 1:327, 1967.

[61] H. A. Kurtz and N. Elander. *Int. J. Quantum Chem. Quantum Chem. Symp.*, 16:605, 1982.

[62] S. Evangelisti, G. L. Bendazzoli, R. Ansaloni, F. Duri, and E. Rossi. *Chem. Phys. Lett.*, 16:437, 1996.

[63] T. D. Crawford and H. F. Schaefer III. An introduction to coupled cluster theory for computational chemists. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 14, pages 33–136. VCH Publishers, New York, 1991.

[64] W. Kohn, Y. Meir, and D. E. Makarov. *Phys. Rev. Lett.*, 80:4153, 1998.

[65] M. Lein, J. F. Dobson, and E. K. U. Gross. *J. Comp. Chem.*, 20:12, 1999.

[66] H. Rydberg, M. Dion, N. Jacobson, E. Schröder, P. Hyldgaard, S. I. Simak, D. C. Langreth, and B. I. Lundqvist. *Phys. Rev. Lett.*, 91:126402, 2003.

[67] C. J. Umrigar. Variational monte carlo basics and applications to atoms and molecules. In M. P. Nightingale and C. J. Umrigar, editors, *Quantum Monte Carlo Methods in Physics and Chemistry*, volume 525 of *NATO Series C: Mathematical and Physical Sciences*, pages 129–160. Kluwer, Dordrecht, 1998.

[68] L. Pauling. *The nature of the chemical bond*. 3rd edition. Cornell University Press, Ithaca, New York.

[69] P. W. Anderson. *Mater. Res. Bull.*, 8:153, 1973.

[70] J. G. Bednorz and K. A. Muller. *Z. Phys.*, B64:189, 1986.

[71] P. W. Anderson, P. A. Lee, M. Randeria, T. M. Rice, N. Trivedi, and F. C. Zhang. *cond-mat/0311467*.

[72] S. Sorella, G. B. Martins, F. Becca, C. Gazza, L. Capriotti, A. Parola, and E. Dagotto. *Phys. Rev. Lett.*, 88:117002, 2002.

[73] E. Plekhanov, F. Becca, and S. Sorella. *Phys. Rev. B*, 71:064511, 2005.

[74] M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti. *Phys. Rev. Lett*, 94:026406, 2005.

[75] J. P. Bouchaud, AS. Georges, and C. Lhuillier. *J. Phys. (Paris)*, 49:553, 1988.

[76] A. J. Coleman. *J. Math. Phys.*, 13:214, 1972.

[77] T. Kato. *Pure Appl. Math.*, 10:151, 1957.

[78] R. T. Pack and W. B. Brown. *J. Chem. Phys.*, 44:556, 1966.

[79] C. Filippi and C. J. Umrigar. *J. Chem. Phys.*, 105:213, 1996.

[80] S. Fahy, X. W. Wang, and S. G. Louie. *Phys. Rev. B*, 42:3503, 1990.

[81] E. Krotscheck, W. Kohn, and G. X. Qian. *Phys. Rev. B*, 32:5693, 1985.

[82] D. Prendergast, D. Bevan, and S. Fahy. *Phys. Rev. B*, 66:155104, 2002.

[83] E. Clementi and C. Roetti. *Atomic Data and Nuclear Data Tables*. Academic Press, 1974.

[84] S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. Fischer. *Phys. Rev. A*, 47:3649, 1993.

[85] H. A. Kurtz, N. Elander, O. Goscinski, and E. Sangfelt. *Int. J. Quantum Chem. Quantum Chem. Symp.*, 15:143, 1981.

[86] Chien-Jung Huang, C. J. Umrigar, and M. P. Nightingale. *J. Chem. Phys.*, 107:3007, 1997.

[87] A. Sarsa, F. J. Gálvez, and E. Buendía. *J. Chem. Phys.*, 109:7075, 1998.

[88] R. Assaraf and M. Caffarel. *J. Chem. Phys.*, 119:10536, 2003.

[89] D. Bressanini, G. Morosi, and M. Mella. *J. Chem. Phys.*, 116:5345, 2002.

[90] F.J. Galvez, E. Buendia, and A. Sarsa. *J. Chem. Phys.*, 115:1166, 2001.

[91] J. C. Grossman. *J. Chem. Phys.*, 117:1434, 2002.

[92] H. Huang and Z. Cao. *J. Chem. Phys.*, 104:200, 1996.

[93] D.R.Garmer and J.B. Anderson. *J. Chem. Phys.*, 86:4025, 1987.

[94] Shih-I Lu. *J. Chem. Phys.*, 118:9528, 2003.

[95] A. Lüchow, J. B. Anderson, and D. Feller. *J. Chem. Phys.*, 106:7706, 1997.

[96] J. M. L. Martin. *Chem. Phys. Lett.*, 303:399, 1999.

[97] I. Røeggen. *J. Chem. Phys.*, 79:5520, 1983.

[98] I. Røeggen and J. Almlöf. *Int. J. Quantum Chem.*, 60:453, 1996.

[99] D. Felle, C. M. Boyle, and E. R. Davidson. *J. Chem. Phys.*, 86:3424, 1987.

[100] Srinivasan Parthiban and J.M.L. Martin. *J. Chem. Phys.*, 115:2051, 2001.

[101] W. C. Ermler and C. W. Kern. *J. Chem. Phys.*, 58:3458, 1973.

[102] D. Feller and D. A. Dixon. *J. Phys. Chem. A*, 104:3048, 2000.

[103] S. White and I. Affleck. *Phys. Rev. B*, 54:9862, 1996.

[104] K. Muller-Dethlefs and J. B. Peel. *J. Chem. Phys.*, 111:10550, 1999.

[105] M. S. Deleuze, L. Claes, E. S. Kryachko, and J.P. François. *J. Chem. Phys.*, 119:3106, 2003.

[106] http://webbook.nist.gov. Ion energetics data. In P. J. Linstrom and W. G. Mallard, editors, *NIST Chemistry Webbook, NIST Sandard Reference Database*, volume 69. National Institute of Standards and Technology, Gaithersburg, MD, 2001.

[107] Heinz-Jürgen Flad, M. Caffarel, and A. Savin. In W. A. Lester, editor, *Recent Advances in Quantum Monte Carlo Methods*, page 73. World Scientific Publishing, Singapore, 1997.

[108] B. L. Hammond, P. J. Reynolds, and W. A. Lester Jr. *J. Chem. Phys.*, 87:1130, 1987.

[109] M. M. Hurley and P. A. Christiansen. *J. Chem. Phys.*, 86:1069, 1987.

[110] P. A. Christiansen. *J. Chem. Phys.*, 88:4867, 1988.

[111] G. B. Bachelet, D. M. Ceperley, and M. G. B. Chiocchetti. *Phys. Rev. Lett.*, 62:2088, 1989.

[112] B. L. Hammond, P. J. Reynolds, and W. A. Lester. *Phys. Rev. Lett.*, 61:2312, 1988.

[113] D. R. Hamann, M. Schlüter, and C. Chiang. *Phys. Rev. Lett.*, 43:1494, 1979.

[114] P. A. Christiansen, Y. S. Lee, and K. S. Pitzer. *J. Chem. Phys.*, 71:1979, 1979.

[115] C. W. Greeff and W. A. Lester Jr. *J. Chem. Phys.*, 109:1607, 1998.

[116] P. H. Alcioli and D. M. Ceperley. *J. Chem. Phys.*, 100:8169, 1994.

[117] W. J. Stevens, H. Basch, and M. Krauss. *J. Chem. Phys.*, 81:6026, 1984.

[118] L. Mitáš. *Phys. Rev. A*, 49:4411, 1994.

[119] H. J. Flad and Michael Dolg. *J. Chem. Phys.*, 107:7951, 1997.

[120] I. Ovcharenko, A. Aspuru-Guzik, and W. A. Lester Jr. *J. Chem. Phys.*, 114:7790, 2001.

[121] D. Vanderbilt85. *Phys. Rev. B*, 32:8412, 1985.

[122] K. Gottfried and T. M. Yan. *Quantum Mechanics: Fundamentals*. 2nd edition. Springer, 2003.

[123] A. D. McLaren. *Math. Comp.*, 17:361, 1963.

[124] A. Bosin. Monte carlo quantistico per sistemi di sola valenza. Master's thesis, Università degli Studi di Cagliari, 1995.

[125] D. M. Ceperley and L. Mitas. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, volume XCIII, page 1. Wiley, New York, 1996.

[126] A. Bergner, M. Dolg, W. Kuechle, H. Stoll, and H. Preuss. *Mol. Phys.*, 80:1431, 1993.

[127] W. I. Friesen and B. Bergesen. *J. Phys. C*, 13:6627, 1980.

[128] E. Lieb and D. Mattis. *Phys. Rev.*, 125:164, 1962.

[129] A. Malatesta and G. Senatore. *J. Phys. IV*, 10, Pr5:341, 2000.

[130] M. Casula and G. Senatore. *Chem. Phys. Chem.*, 6:1902, 2005.

[131] D. M. Ceperley. *Phys. Rev. B*, 18:3126, 1978.

[132] S. Moroni, S. Fantoni, and G. Senatore. *Phys. Rev. B*, 52:13547, 1995.

[133] A. Malatesta. Quantum monte carlo study of a model one-dimensional electron gas. Master's thesis, Università degli Studi di Trieste, 1999.

[134] M. Casula. Ottimizzazione della funzione d'onda per un gas di elettroni quasi unidimensionale e problemi di ergodicità. Master's thesis, Università degli Studi di Trieste, 2001.

[135] T. Gaskell. *Proc. Phys. Soc.*, 77:1182, 1961.

[136] T. Gaskell. *Proc. Phys. Soc.*, 80:1091, 1962.

[137] H. J. Schulz. *Phys. Rev. Lett.*, 71:1864, 1993.

[138] K. E. Schmidt and S. Fantoni. *Phys. Lett. B*, 446:99, 1999.

[139] H. B. Gray. Biological inorganic chemistry at the beginning of the 21-st century. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 100, page 3563. 2003.

[140] M. Cococcioni, A. Dal Corso, and S. de Gironcoli. *Phys. Rev. B*, 67:094106, 2003.

[141] F. Zhou, C. A. Marianetti, M. Cococcioni, D. Morgan, and G. Ceder. *Phys. Rev. B*, 69:201101, 2004.

[142] M. Dolg, U. Wedig, H. Stoll, and H. Preuss. *J. Chem. Phys.*, 86:866, 1987.

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

[144] K. C. Huang, R. J. Needs, and G. Rajagopal. *J. Chem. Phys.*, 112:4419, 2000.

[145] B. L. Hammond, W. A. Lester Jr., and P. J. Reynolds. *Monte Carlo Methods in ab initio quantum chemistry*. World Scientific, Singapore, 1949.

[146] P. M. Morse. *Ann. d. Physik*, 84:57, 1927.

[147] C. E. Moore. *Atomic Energy Levels*. Natl. Bur. Stand. (U.S.) Circ. No. 467, U.S. GPO, Washington, D.C., 1949.

[148] R. L. Martin and P. J. Hay. *J. Chem. Phys.*, 75:4539, 1981.

[149] C. W. Bauschlicher Jr. *J. Chem. Phys.*, 86:5591, 1987.

[150] K. Raghavachari and G. W. Trucks. *J. Chem. Phys.*, 91:1062, 1989.

[151] S. Chrétien and D. R. Salahub. *Phys. Rev. B*, 66:155425, 2002.

[152] MOLPRO is a package of *ab initio* programs written by H.-J. Werner and P. J. Knowles, with contributions from J. Almlöf, R. D. Amos, M. J. O. Deegan, S. T. Elbert, C. Hampel, W. Meyer, K. Peterson, R. M. Pitzer, A. J. Stone, and P. R. Taylor.

[153] J. Harris and R. O. Jones. *J. Chem. Phys.*, 70:830, 1979.

[154] I. Shim and K. A. Gingerich. *J. Chem. Phys.*, 77:2490, 1982.

[155] M. Moskovits and D. P. DiLella. *J. Chem. Phys.*, 73:4917, 1980.

[156] P. A. Montano and G. K. Shenoy. *Solid State Comm.*, 35:53, 1980.

[157] H. Purdum, P. A. Montano, G. K. Shenoy, and T. Morrison. *Phys. Rev. B*, 25:4412, 1982.

[158] L. Lian, C. X. Su, and P. B. Armentrout. *J. Chem. Phys.*, 97:4072, 1992.

[159] M. Castro and D. R. Salahub. *Phys. Rev. B*, 47:10955, 1993.

[160] M. Castro and D. R. Salahub. *Phys. Rev. B*, 49:11842, 1994.

[161] G. L. Gutsev and C. W. Bauschlicher Jr. *J. Phys. Chem. A*, 107:7013, 2003.

[162] M. Tomonari and H. Tatewaki. *J. Chem. Phys.*, 88:6419, 1988.

[163] D. G. Leopold and W. C. Lineberger. *J. Chem. Phys.*, 85:51, 1986.

[164] D. G. Leopold and J. Almlöf. *J. Chem. Phys.*, 88:3780, 1988.

[165] O. Hübner and J. Sauer. *Chem. Phys. Lett.*, 358:442, 2002.

[166] A. Irigoras, M. del Carmen Michelini, E. Sicili, N. Russo, J. M. Mercero, and J. M. Ugalde. *Chem. Phys. Lett.*, 376:310, 2003.

[167] O. Hübner. *Private communication*, 2005.

[168] C. A. Baumann, R. J. Van Zee, and W. Weltner Jr. *J. Phys. Chem.*, 88:1815, 1984.

[169] C. W. Bauschlicher Jr. and A. Ricca. *Mol. Phys.*, 101:93, 2003.

[170] M. Casula and S. Sorella. *J. Chem. Phys.*, 119:6500, 2003.

[171] M. Casula, C. Attaccalite, and S. Sorella. *J. Chem. Phys.*, 121:7110, 2004.

[172] M. Casula, S. Yunoki, C. Attaccalite, and S. Sorella. *Comp. Phys. Comm.*, 169:386, 2005.

[173] M. Casula, C. Filippi, and S. Sorella. *Phys. Rev. Lett.*, 95:100201, 2005.

[174] O. Hübner and J. Sauer. *J. Chem. Phys.*, 116:617, 2002.

[175] A. L. Fetter and J. D. Walecka. *Quantum theory of many-particle systems*. McGraw-Hill, New York, 1971.