Quantum Monte Carlo methods for light nuclei.

Stefan Rombouts*
Ghent University, Dept. of Subatomic and Radiation Physics,
Proeftuinlaan 86, B-9000 Gent (Belgium)
(Dated: October 31, 2002)

Over the past decades, quantum Monte Carlo methods have proven to be very valuable for the study of the nuclear many-body problem. They have allowed fully microscopical calculations of the ground state properties of light nuclei. The basic ingredient of these methods is a decomposition of the imaginary-time evolution operator into path integrals. These paths are then sampled using Monte Carlo techniques. Particularly for light nuclei, up to $A=10$, this approach has resulted in an excellent agreement with experimental data, provided that three-body interactions are taken into account.

I. INTRODUCTION

From the theoretical point of view, the structure of atomic nuclei is a complicated many-body problem: even for a limited number of particles there exists a huge number of configurations that can contribute to the ground state structure. Furthermore the interactions between the particles are far from simple. A quantum-mechanical description of this many-body system quickly leads to equations which can not be solved straightforwardly. Apart from that, these systems might exhibit a collective behavior one would not expect from the equations of motion of the individual particles. A way to study such properties is offered by quantum Monte Carlo methods [1-4]: instead of solving the Schrödinger equation, these methods aim at a statistical evaluation of the partition function. This allows an exact numerical calculation of thermodynamical and ground state properties, up to a controllable statistical error. Quantum Monte Carlo methods have been applied with great success for the microscopical calculation of ground-state properties of light nuclei [5-9]. The aim of this text is to give a basic understanding of the methods that were used to achieve these results. Readers who want to learn more can find a good starting point in reference [10].

The basic idea of quantum Monte Carlo methods is to write the imaginary-time evolution operator as a path integral. These paths are then sampled using Monte Carlo methods. Using paths in real-space works fine for bosons. For fermions problems arise because of anti-symmetrization: weights become negative, and therefore the Monte Carlo sampling becomes cumbersome. This is the infamous sign problem. However, using good variational wave functions as a start, and summing explicitly over spin and isospin degrees of freedom, one has been able to perform converged calculations for nuclei of up to 10 particles.

In section II we show how one can obtain accurate information about a highly correlated quantum-mechanical many-body state from knowledge on the imaginary-time evolution operator of the system. For practical calculations the evolution operator has to be broken up into manageable parts, leading to a path-integral representation. Section III gives an overview of several possible decomposition schemes that are widely used in the field of quantum Monte Carlo. The next section discusses the Monte Carlo techniques that are used to sample the paths. For fermions, the method often is hampered by the sign problem, which is the subject of section V. Section VI shows how these methods are put into practice for light nuclei.

II. THE IMAGINARY-TIME EVOLUTION OPERATOR

The central operator in most quantum Monte Carlo methods, is the Boltzmann operator or imaginary-time evolution operator,

$$e^{-eta \hat{H}}$$

with $\hat{H}$ the Hamiltonian of the system and $\beta$ the inverse temperature or imaginary time. This operator has many faces:

*Fund for Scientific Research - Flanders (Belgium)
• **Thermodynamics:** The operator of Eq.(1) is the many-body density matrix of the grand canonical ensemble at a temperature $T = 1/(k\beta)$. This can be easily understood by expanding the operator in eigenstates of $\hat{H}$:

$$e^{-\beta \hat{H}} = \sum_\lambda |\lambda\rangle e^{-\beta E_\lambda} \langle \lambda|.$$  

(2)

It describes a mixed system, that has a probability proportional to the Boltzmann factor $e^{-\beta E_\lambda}$ to occupy the eigenstate $|\lambda\rangle$. Taking the trace one obtains the grand canonical partition function,

$$Z(\beta) = \text{Tr} \left[ e^{-\beta \hat{H}} \right] = \sum_\lambda e^{-\beta E_\lambda}.$$  

(3)

From the function $Z(\beta)$ one can derive all thermodynamic observables, e.g. the energy is given by

$$E(\beta) = -\frac{d \ln[Z(\beta)]}{d\beta} = \text{Tr} \left[ \hat{H} e^{-\beta \hat{H}} \right] / Z(\beta).$$  

(4)

• **Ground-state projector:** Taking the limit of $\beta$ to infinity in the above expressions corresponds to cooling the system to absolute zero temperature. This brings the system to its ground state. Therefore one can efficiently use the operator of Eq.(1) as a ground-state projection operator: applying it to any trial state that has a non-zero overlap with the true ground state will result in a state that approaches the ground state as $\beta$ grows, while excited-state components are suppressed exponentially. Again this is demonstrated by an expansion into the eigenstates $|\lambda\rangle$ of $\hat{H}$: given a trial state

$$|\Psi_T\rangle = \sum_\lambda c_\lambda |\lambda\rangle,$$  

(5)

with $c_{GS} \neq 0$, one obtains

$$e^{-\beta \hat{H}} |\Psi_T\rangle = e^{-\beta E_{GS}} c_{GS} |\Psi_{GS}\rangle + e^{-\beta E_{GS}} \sum_{\lambda \neq GS} e^{-\beta (E_\lambda - E_{GS})} c_\lambda |\lambda\rangle \simeq e^{-\beta E_{GS}} c_{GS} |\Psi_{GS}\rangle.$$  

(6)

If the trial state $|\Psi_T\rangle$ is already a good approximation to the ground state, one does not have to go to too large values of $\beta$ in order to obtain accurate ground-state results. The ground-state energy can be obtained as

$$E_{GS} \simeq \langle \Psi_T | \hat{H} e^{-\beta \hat{H}} | \Psi_T \rangle / \langle \Psi_T | e^{-\beta \hat{H}} | \Psi_T \rangle.$$  

(7)

This is the so called **mixed estimator**, because it takes the overlap between the estimate for the ground state and the trial state. Observables that commute with the Hamiltonian can be evaluated in this way. For other observables one has to apply the ground-state estimate on both sides, which is not always easy to implement.

**Exercise 1** Derive an estimate for the expectation value $\langle \hat{A} \rangle$ of an operator $\hat{A}$ by taking a linear combination of the trial state expectation value and the mixed estimator,

$$\langle \hat{A} \rangle \simeq \alpha_1 \frac{\langle \Psi_T | \hat{A} e^{-\beta \hat{H}} | \Psi_T \rangle}{\langle \Psi_T | e^{-\beta \hat{H}} | \Psi_T \rangle} + \alpha_2 \frac{\langle \Psi_T | \hat{A} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle},$$  

(8)

in such a way that the estimate is exact up to first order in the difference $(|\Psi_{GS}\rangle - |\Psi_T\rangle)$.

• **Imaginary time propagator:** The operator of Eq.(1) has a striking resemblance to the Heisenberg evolution operator $\exp(-itH)$ that propagates a state over an interval $t$ in real time. Interpreting $\beta$ as an imaginary time, $\beta = it$, one can describe the action of the operator $\exp(-\beta \hat{H})$ on a trial state as a **propagation in imaginary time**. As shown above, this evolution in imaginary time is equivalent to a diffusion to the ground state.
III. DECOMPOSITIONS

In the previous section it was shown that the imaginary-time propagator contains a lot of information about the quantum many-body system. Calculating this propagator exactly would be equivalent to solving the many-body Schrödinger equation. For most systems, this is practically impossible. The quantum Monte Carlo approach tries to tackle this problem by breaking up the imaginary-time propagator into parts that can be treated exactly individually, and to generate a limited but relevant sample of these using Monte Carlo techniques. The efficiency and the feasibility of the Monte Carlo algorithms depends crucially on the decomposition that is used. One can divide the decompositions that are used in practice generally into two categories: path-integral decompositions and auxiliary-field decompositions.

A. path-integral decompositions

As a first step one divides the imaginary-time propagator into small time steps, such that one can safely make approximations that are correct only up to first order in the time step \( \epsilon \):

\[
e^{-\beta \hat{H}} = \left( e^{-\epsilon \hat{H}} \right)^M,
\]

with \( \epsilon = \beta / M \). Approximating \( \exp(-\epsilon \hat{H}) \) with an expression that has a leading error of order \( \epsilon^2 \) will result in an overall error on Eq(9) of order \( M(\epsilon)^2 = \beta^2 / M \). By taking the value of \( M \) large enough (making the time steps smaller), one can reduce the error to any desired level of accuracy. At every time step in Eq(9), one can now introduce a complete set of states in coordinate space,

\[
\sum_{\mathbf{R}} \langle \mathbf{R} | \mathbf{R} \rangle = 1,
\]

where \( \mathbf{R} \) is a multi-particle coordinate vector. The imaginary-time propagator can now be written as

\[
\langle \mathbf{R}_0 | e^{-\beta \hat{H}} | \mathbf{R}_M \rangle = \int \int \int d\mathbf{R}_1 d\mathbf{R}_2 \cdots d\mathbf{R}_{M-1} \langle \mathbf{R}_0 | e^{-\epsilon \hat{H}} | \mathbf{R}_1 \rangle \langle \mathbf{R}_1 | e^{-\epsilon \hat{H}} | \mathbf{R}_2 \rangle \cdots \langle \mathbf{R}_{M-1} | e^{-\epsilon \hat{H}} | \mathbf{R}_M \rangle.
\]

The expressions that have to be evaluated take the form

\[
\langle \mathbf{R} | e^{-\epsilon \hat{H}} | \mathbf{R}' \rangle.
\]

For a Hamiltonian that contains a kinetic energy term and a local interaction, \( \hat{H} = \hat{T} + \hat{V} \), one can make the following approximations, because of the small time step:

\[
\langle \mathbf{R} | e^{-\epsilon \hat{H}} | \mathbf{R}' \rangle = \langle \mathbf{R} | e^{-\epsilon \hat{T} + \hat{V}} | \mathbf{R}' \rangle \\
\approx \langle \mathbf{R} | e^{-\epsilon \hat{T}} | \mathbf{R}' \rangle \langle \mathbf{R}' | e^{-\epsilon \hat{V}} | \mathbf{R}' \rangle + O(\epsilon^2) \\
\propto e^{-\frac{\| \mathbf{R} - \mathbf{R}' \|^2}{2 \hbar^2 / m}} e^{-\epsilon \hat{V}(\mathbf{R}') + O(\epsilon^2)}.
\]

Exercise 2 Evaluate the kinetic propagator by inserting complete sets of momentum states |\( \mathbf{K} \)|,

\[
\int \int \langle \mathbf{R} | \mathbf{K} \rangle \langle \mathbf{K}' | e^{-\epsilon \hat{T}} | \mathbf{K}' \rangle \langle \mathbf{K}' | \mathbf{R}' \rangle d\mathbf{K} d\mathbf{K}' = ?
\]

This shows that the imaginary time propagation can be simulated through a Gaussian diffusion in coordinate space, reweighted by an interaction factor. In order to obtain an efficient algorithm, one normally has to apply importance sampling, which amounts to including a drift term

\[
\mathbf{R}' - \mathbf{R} \rightarrow \mathbf{R}' - \mathbf{R} - 2(\hbar^2 / m) \nabla \ln \Psi_T(\mathbf{R})
\]

in the diffusion step, that is corrected for in the interaction factor, such that the diffusion mainly proceeds to regions in space where \( |\Psi_T(\mathbf{R})| \) is important (see section IV B). Instead of the asymmetric decomposition of Eq(13), one often uses a symmetric decomposition

\[
e^{-\epsilon (\hat{T} + \hat{V})} = e^{-\frac{1}{2} \epsilon \hat{T}} e^{-\epsilon \hat{V}} e^{-\frac{1}{2} \epsilon \hat{T}} + O(\epsilon^3),
\]

because its leading error is one order higher in \( \epsilon \). Together with Eq(9) this results in the so-called Suzuki-Trotter decomposition [11].
Exercise 3 (not a simple one!) Find a decomposition of the form Eq.(17) with five factors such that the error is of order $e^5$. Hint: one exponent will contain a commutator. The following relation, known as the Baker-Campbell-Hausdorff formula, can be useful:

$$e^A e^B = e^{A+B + \frac{1}{2}[A, B] + \frac{1}{12}([A-B],[A,B]) + ...}.$$  

(18)

Another way to improve the short-time propagator can be obtained by making use of the two-body propagator, if this is known exactly (which amounts to solving the two-body problem). Then one can apply the following improved propagator:

$$\langle R | e^{-\mathcal{H}t} | R' \rangle = \prod_{i=1,A} G_0(r_i, r'_i) \mathcal{S} \left[ \prod_{i<j} \frac{1}{\sqrt{2 \pi}} | \langle r_i | r_j \rangle \rangle | \right],$$  

(19)

with $G_0(r_i, r'_i)$ the free (Gaussian) one-particle propagator, $g_{0;i,j}(r_{ij}, r'_{ij})$ the free (Gaussian) pair propagator, and $g_{ij}(r_{ij}, r'_{ij})$ the pair propagator that takes the exact interaction into account and $\mathcal{S}$ the symmetrization operator over all indices. This type of propagator is exact at the two-body level. At the many-body level, its error is strongly reduced compared to the expression of Eq.(14). This means that one can reach the same accuracy with much larger time steps (typically $\times 5$), thereby reducing the fluctuations and improving the efficiency.

The decomposition of Eq.(11) together with Eq.(14) or Eq.(19) can be used to project a ground state onto a trial state as in Eq.(6). Used in this way, the method is called diffusion Monte Carlo or Green's function Monte Carlo [1]. Another possibility is to take the trace of Eq.(11), by equating $R_0 = R_M$, and summing over all possible paths $\{R_0, R_1, ..., R_{M-1}\}$. From Eq.(3) it can be seen that this results in a finite temperature method, that goes by the name of path-integral Monte Carlo [2]. In practice one often uses improved versions of the short-time propagator of Eq.(14), in order to reduce the time step error[6].

A similar way of thinking can be applied to lattice or spin systems, resulting in a method know as world-line Monte Carlo[12, 13]. Because the paths or world lines in this case are strongly restricted by the structure of the lattice, one can use highly optimized updating schemes for the Monte Carlo sampling, such as the loop algorithm[14, 15] or the worm algorithm[16].

All these methods have the advantage that they require few memory to store the individual configurations and that the basic steps are very simple. Furthermore they solve the many-body problem directly in the full coordinate space. This makes them very useful for bosonic systems. For fermions problems appear because of the fermionic antisymmetry: a path that brings two fermions near to each other and lets them move apart again, will have nearly the same weight as the path that lets the two fermions follow nearly the same road, except that they interchange positions when they are near. However, the weight of the latter path will have the opposite sign of the weight of the former. As a result the sum over all paths will be a sum of terms of equal magnitude but with opposite signs, resulting in a bad signal-to-noise ratio for the Monte Carlo sampling. This is the reason for the so called sign problem.

B. auxiliary-field decompositions

Another approach is based on decompositions that try to take the antisymmetry exactly into account. The idea is to decompose the imaginary-time evolution operator into a sum of evolution operators for non-interacting systems [3, 4, 17-19].

This builds on the observation that for non-interacting systems the imaginary-time evolution operator can be evaluated exactly: instead of propagating one $A$-particle wave function, one simply has to propagate $A$ single-particle wave functions, even if one takes the (anti-)symmetry between the particles fully into account. This is related to the Thouless theorem[20]: the exponential of a one-body operator will transform a Slater determinant into another Slater determinant.

The operators that build up the Hamiltonian can normally be grouped into a one-body part $\hat{H}_1$ and a two-body part $\hat{H}_2$ The difficulties come with the two-body part. In order to separate out the two-body part, one splits the imaginary-time interval into a large number of time steps using the Suzuki-Trotter decomposition of Eq.(17). The one-body propagators of the form exp($-\epsilon \hat{H}_1$) are easily expressed in the single-particle representation. For the two-body propagator exp($-\epsilon \hat{H}_2$) this is not possible. Therefore one expands the two-body propagator into a sum over one-body propagators,

$$e^{-\epsilon \hat{H}_2} = \sum_\sigma e^{-\epsilon h_1(\sigma)} w(\sigma),$$  

(20)
where each of the operators $\hat{h}_1(\sigma)$ is a one-body operator and $w(\sigma)$ is a weight factor. The index $\sigma$ is called the auxiliary field. What remains to be shown is how one can decompose the two-body propagator into one-body propagators as in Eq. (20). For this purpose one uses the Hubbard-Stratonovich transformation [21, 22], which linearizes the exponential of the square of any operator $\hat{P}$,

$$e^{\frac{\hat{P}^2}{\epsilon}} = \int e^{-\frac{M}{2\epsilon} \hat{P}^2} e^{\epsilon \hat{P}} d\sigma. \quad (21)$$

This relation follows straightforwardly from the fact that the Fourier transform of a Gaussian is again a Gaussian. One can always write the two-body exponent $-\epsilon \hat{H}_2$ as a sum over squares of one-body operators $\hat{P}_k$

$$-\epsilon \hat{H}_2 = \sum_k \hat{P}^2_k. \quad (22)$$

This is not more difficult than solving a two-body problem, and often the symmetries of the interaction guide this decomposition [3]. Using the Hubbard-Stratonovich transform Eq. (21), the small-time-step approximation, and discrete approximations to the integrals, one obtains a decomposition of the form

$$e^{-\beta \hat{H}} = \sum_{\sigma} W(\sigma) e^{-\hat{h}(\beta, \sigma)}, \quad (23)$$

where $\hat{h}(\beta, \sigma)$ is a (non-Hermitian) one-body operator such that each individual right hand side can be treated exactly. Monte Carlo techniques have to be used to evaluate the multi-(multi-)dimensional summation over all the auxiliary fields $\sigma$. There is still a lot of freedom in the decompositions of the from Eq. (20). This can be put to use to eliminate time-step errors [4] or to reduce sign problems [23, 24]. Another recent development is the combination of path-integral and auxiliary-field decompositions, in order to combine continuous configurations in coordinate space with discrete configurations in the spin-isospin space in the Monte Carlo sampling [25].

### IV. MONTE CARLO TECHNIQUES

The decompositions derived in the previous section lead to too many terms to perform a complete summation. The best one can do is to calculate a limited number of terms and to make a statistical estimate of the true result. For this it is necessary to generate a representative sample of all the terms. The decompositions described above require the evaluation of expressions of the form

$$\langle E \rangle = \sum_x E(x)w(x) \sum_x w(x), \quad (24)$$

with $x$ a multidimensional coordinate, $w(x)$ a weight function and $E(x)$ the contribution of configuration $x$ to the observable $E$.

- For the path-integral decompositions on has configurations of the type $x = (R_0, R_1, \ldots, R_M)$, with a weight function of the form

$$w(x) = \langle \Psi_T | R_0 | e^{-\epsilon \hat{H}} | R_1 | e^{-\epsilon \hat{H}} | R_2 | \ldots | e^{-\epsilon \hat{H}} | R_M | \Psi_T \rangle, \quad (25)$$

and e.g. the contribution to the mixed estimator of the energy, Eq. (7), given by

$$E(x) = \langle R_M | \hat{H} | \Psi_T \rangle / \langle R_M | \Psi_T \rangle. \quad (26)$$

- For the auxiliary-field decompositions $x$ corresponds to a vector of auxiliary fields, $x = (\sigma_1, \sigma_2, \ldots, \sigma_M)$. To set the case, let us consider the grand canonical energy given by Eq. (4). For the Monte Carlo evaluation using Eq. (24) we have to take

$$w(x) = W(x) \text{Tr} e^{-\hat{h}(\beta, x)} \quad (27)$$

$$E(x) = \text{Tr} \left[ \hat{H} e^{-\hat{h}(\beta, x)} \right] / \text{Tr} \left[ e^{-\hat{h}(\beta, x)} \right]. \quad (28)$$
The Monte Carlo strategy for the evaluation of Eq. (24) consists of generating a sample of \( N \) configurations \( x_1, x_2, \ldots, x_N \), distributed according to the weight function \( w(x) \), i.e. the probability for a configuration \( x \) to be present in the sample is proportional to \( w(x) \). Then we can estimate \( \langle E \rangle \) by the sample average,

\[
\langle E \rangle \approx \langle E \rangle_{MC} = \frac{1}{N} \sum_{i=1}^{N} E(x_i). \tag{29}
\]

The Central Limit Theorem guarantees that the sample average will converge to the true expectation value for \( N \) going to infinity, i.e. for big enough sample sizes. If the sample configurations are drawn independently of each other, one can estimate the statistical error \( S_E \) on the Monte Carlo result \( \langle E \rangle_{MC} \) in the standard way,

\[
S_E^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} (E(x_i) - \langle E \rangle_{MC})^2. \tag{30}
\]

The main difficulty is to generate the configurations \( x_1, x_2, \ldots, x_N \) in a random way but such that they are distributed according to \( w(x) \). Several Monte Carlo techniques have been developed to facilitate this, most of them based on the Markov chains or on a scheme with walkers and branching.

### A. The Metropolis algorithm and Markov chain Monte Carlo methods

A Markov chain is a series of configurations \( x_0, x_1, x_2, x_3, \ldots \), where each configuration \( x_n \) was generated from the previous one using a fixed but stochastic transition rule \( P(x_{n-1} \rightarrow x_n) \). If the same set of rules is used over and over again,

\[
x_0 \xrightarrow{P} x_1 \xrightarrow{P} x_2 \xrightarrow{P} x_3 \xrightarrow{P} \ldots, \tag{31}
\]

the configurations will tend to be distributed following a series of probability distribution functions (p.d.f.’s),

\[
\pi_0(x_0) \rightarrow \pi_1(x_1) \rightarrow \pi_2(x_2) \rightarrow \pi_3(x_3) \rightarrow \ldots, \tag{32}
\]

related by

\[
\pi_{n+1}(x) = \int \pi_n(y)P(y \rightarrow x)dy. \tag{33}
\]

Under very general conditions one finds that the series of p.d.f.’s \( \pi_n(x) \) converges to one particular function \( \pi(x) \), independent of the initial p.d.f.,

\[
\lim_{n \to \infty} \pi_n(x) = \pi(x). \tag{34}
\]

This distribution is called the stationary distribution of the Markov chain. A sufficient condition for a given distribution \( \pi(x) \) to be the stationary distribution for a Markov transition rule \( P(x \rightarrow y) \), is given by the condition of detailed balance:

\[
\pi(x)P(x \rightarrow y) = \pi(y)P(y \rightarrow x), \tag{35}
\]

for all configurations \( x \) and \( y \). This is checked easily by taking \( \pi_n = \pi \) in Eq.(33), applying Eq.(33) and the normalization condition

\[
\int P(x \rightarrow y)dy = 1. \tag{36}
\]

To demonstrate that the Markov chain will converge to \( \pi(x) \), even if it is started from a different initial p.d.f. \( \pi_0 \neq \pi \), is a bit more complicated. It can be understood by considering the ratio \( r_n(x) = \pi_n(x)/\pi(x) \). Then one can write, in line with Eq.(33),

\[
r_{n+1}(x) = \int \pi_n(y)P(y \rightarrow x)/\pi(x)dy \\
= \int r_n(y)\pi(y)P(y \rightarrow x)/\pi(x)dy \\
= \int r_n(y)\pi(x)P(x \rightarrow y)/\pi(x)dy \\
= \int r_n(y)P(x \rightarrow y)dy. \tag{37}
\]

\[
\int r_n(y)P(x \rightarrow y)dy = 1.
\]
Because $P$ is normalized, Eq.(36), this means that $r_{n+1}(x)$ is equal to the average of $r_n(y)$ over all configurations $y$ that can be reached from $x$ in one Markov step. Under the condition of ergodicity, i.e. that any configuration can be reached from any other configuration with a finite probability in a finite number of Markov steps, one can understand that after a large number of Markov steps, this procedure will tend to make $r_n(x)$ a uniform function of $x$. To assure convergence one should also check that the Markov chain is aperiodic, i.e. that the number of steps needed to go from one configuration to another does not have to be a multiple of some fixed integer number. Otherwise the function $r_n$ might oscillate with that periodicity between several functions. If $r_n(x)$ becomes uniform, it has to converge to 1, because $\int r_n(x)\pi(x)dx = 1$ for all $n$. Therefore one can conclude that $\pi_n$ will converge to the stationary distribution $\pi$. In practice one always has to start with a number of thermalization steps in order to let the distribution converge to the desired target distribution.

The decompositions of the previous section provide a huge number of configurations and a weight for these configurations that is given by their contribution to the partition function or to the ground-state overlap. In order to sample these configurations proportionally to their weight, we would like to construct a Markov chain in the configuration space whose stationary distribution $\pi(x)$ is proportional to the weight of the configuration $x$. Therefore we will try to find a transition rule $P(x \rightarrow y)$ that satisfies the detailed balance condition Eq.(35). This is achieved elegantly with the Metropolis algorithm[26]. It starts from a simple symmetric rule, $P_0(x \rightarrow y) = P_0(y \rightarrow x)$, to go from one configuration to another.

**Exercise 4 (this one is easy!) What is the stationary distribution of $P_0$?**

The Metropolis algorithm goes as follows:

- Given a configuration $x_n$, one draws a configuration $y$ according to $P_0(x_n \rightarrow y)$.
- If $\pi(y) > \pi(x)$, then one takes $x_{n+1} = y$.
- Otherwise one draws a uniform random number $u$ between 0 and 1.
  - If $\pi(y)/\pi(x_n) \geq u$ on takes $x_{n+1} = y$, else on takes $x_{n+1} = x_n$.

It is a simple set of rules that guarantees condition Eq.(35). Note that one only needs to consider the ratio between $\pi(y)$ and $\pi(x_n)$. This means that the target distribution does not have to be normalized to unity, which is often a big advantage. The art of the method exists in choosing the transition rule $P_0(x \rightarrow y)$ such that a reasonable fraction of all the moves is accepted (typically 10% to 50%), such that not too much time is wasted in rejected trial moves and such that the moves are not too small either.

**Exercise 5 The method has been extended to asymmetric trial moves by W.K. Hastings [27]. Given that $P_0(x \rightarrow y) \neq P_0(y \rightarrow x)$, derive the factor $q(x \rightarrow y)$ such that the following acceptance/rejectance rules result in a transition rule $P$ that satisfies the detailed balance condition Eq.(35) for the target distribution $\pi(x)$:

- If $q(x_n \rightarrow y) > 1$, then take $x_{n+1} = y$.
- Otherwise draw a uniform random number $u$ between 0 and 1.
  - If $q(x_n \rightarrow y) \geq u$ on takes $x_{n+1} = y$, else on takes $x_{n+1} = x_n$.

Note that for the symmetric Metropolis algorithm one finds

$$q(x \rightarrow y) = \frac{\pi(y)}{\pi(x)}.$$ 

(38)

Often this allows a more efficient sampling, because one can try to work with trial move $P_0$ whose stationary distribution is similar to $\pi(x)$, such that most of the moves will be accepted. Other acceptance rules exist that result in the correct stationary distribution too, e.g. the ‘heat-bath rule’:

$$q(x \rightarrow y) = \frac{\pi(y)}{\pi(x) + \pi(y)}.$$ 

(39)

in case of summetric trial moves. However, the Metropolis rule is known to be the most efficient one [28].

Note that the configurations generated by the Markov chain are not statistically independent: strong correlations exists between successive configurations. It can be shown that the correlations decrease exponentially with the number of Markov steps between two configurations. The Central Limit Theorem is still valid, such that one can safely use the Markov chain configurations to estimate the true results, be it that one has to take the correlations into account when estimating the statistical errors [29]. Therefore some authors suggest to use only a subset of the generated configurations, e.g. one configuration every 200 Markov steps, in order to obtain quasi-independent samples such that the error estimate of Eq.(30) is valid. Generally it is more efficient to run several (typically 10 or 20) independent Markov chains, and then to take for each chain the average over all generated configurations. These averages can then be used as truly independent estimates in Eq.(30) to estimate reliable errorbars.
B. walkers and branching

Another way to sample the terms of the decompositions of section III is based on a scheme of walkers with branching [1]. It is most easily explained for the diffusion Monte Carlo method, though it can be applied to sample any given distribution. Consider first the single-walker version: it starts from a configuration \( R_0 \) that is drawn according to the trial density, \( R_0 \propto |\Psi_T(R_0)|^2 \), and given a weight \( w_0 = 1 \). Then one draws the next configuration \( R_1 \) according to a Gaussian distribution around \( R_0 \),

\[
R_1 \propto e^{-\frac{\|R_1 - R_0\|^2}{2\sigma^2},}
\]

and one multiplies the weight factor,

\[
w_1 = e^{-V(R_1)/\Psi_T(R_1)/\Psi_T(R_0)}w_0.
\]

From Eq.(14) it follows that the walker \( (R_1, w_1) \) is distributed according to a probability \( \pi_1(R_1, w_1) \) such that

\[
\int w_1 \pi_1(R_1, w_1) dR_0 = \langle \Psi_T | R_1 \rangle e^{-V} e^{-\frac{c}{2}|\Psi_T|}.
\]

One can iterating this procedure, sampling \( R_n \) from a Gaussian around \( R_{n-1} \) and updating the weight factor as in Eq.(41). Weighting the resulting configuration \( R_n \) with the weight factor \( w_n \), one obtains a representation of the ground-state overlap \( \langle \Psi_T | R_n \rangle \langle R_n | \Psi_{GS} \rangle \). The problem is that the weight factors \( w_n \) tend to oscillate with an exponentially growing variance, because of their multiplicative nature. This makes the single-walker scheme useless in practice. The trick of importance sampling, Eq.(16) reduces the variance by incorporating the factor \( \Psi_T(R_1)/\Psi_T(R_0) \) approximately in the diffusion step, but the multiplicative nature of the weights remains. In order to avoid this problem, one uses a technique called branching:

- start with a set of many walkers (typically several hundreds).
- propagate each of the walkers as explained before. Keep track of the weight \( w_i \) of each walker.
- at regular intervals, rearrange the set of walkers:
  - if a walker has a large weight, then split it into two walkers with only half the weight, and let them from then on propagate independently (this is the branching).
  - if a walker has a small weight, then try to combine it with another small walker, selecting one or the other proportional to their weights, and propagate the remaining one from then on with the sum of the two weights. (the other is eliminated).

This procedure keeps the weights more or less around the same value, and the rules are constructed such that the expectation values remain correct:

\[
\left\langle \frac{\sum_i w_i O(R_i)}{\sum_i w_i} \right\rangle = \left\langle \frac{\sum_{i'} w_{i'} O(R_{i'})}{\sum_{i'} w_{i'}} \right\rangle.
\]

The number of walkers does not remain constant under this procedure. One should add or subtract a constant to the potential in order to keep the number of walkers more or less constant during the run. To decide what weights are 'large' and which ones are 'small', is somehow arbitrary. One can think of several possible schemes, comparing to the average over all present walkers or comparing to an a-priori given value. Many different schemes have been proposed to decide on the branching, but all of them are more or less equivalent: any scheme of walkers with branching has a systematic error that is inversely proportional to the number of walkers[30]. Furthermore the samples are highly correlated because of the branching. Care has to be taken to estimate the statistical errors[29]. The main advantages of the method are that one avoids the thermalization steps of the Markov chain and that one obtains results for successive imaginary times in one run.

V. SIGN PROBLEMS

Above it was tacitly assumed that all terms in the decomposition of the imaginary-time evolution operator come with a positive weight. However, often this is not the case. As mentioned before, the path-integral decompositions
suffer from a severe sign problem due to the fermionic antisymmetry. Also the auxiliary-field methods are generally confronted with the problem, except when there is a symmetry at work that keeps the terms positive.

Ideally one would generate the configurations $x$ proportional to the weight $w(x)$, but if $w(x)$ can be negative this is not possible, because a probability necessarily has to be positive. The best one can do, is to generate the configurations proportional to the absolute value of the weight, $|w(x)|$. Then one can evaluate an observable $E$ through

$$
\langle E \rangle = \frac{\sum_x E(x)s(x)|w(x)|}{\sum_x s(x)w(x)},
$$

$$
= \frac{\sum_x E(x)s(x)|w(x)|}{\sum_x s(x)|w(x)|} \langle s \rangle,
$$

$$
= \frac{\langle E \rangle}{\langle s \rangle},
$$

(44)

with $s(x)$ the sign of $w(x)$. The quantity $\langle s \rangle$ is called the average sign. Taking a finite sample of configurations will result in approximate values for $\langle E \rangle$ and $\langle s \rangle$. It is interesting to check the consequences of small deviations from the true values. One can expect the deviations to be proportional to $1/\sqrt{N}$, with $N$ the number of sampled configurations,

$$
\langle E \rangle_{MC} = \langle E \rangle (1 + \epsilon_N)
$$

$$
\langle s \rangle_{MC} = \langle s \rangle (1 + \delta_N)
$$

$$
\epsilon_N, \delta_N \propto 1/\sqrt{N}.
$$

(45)

Then one obtains up to first order in $1/N$ that

$$
\langle \langle E \rangle_{MC} \rangle = \left( \frac{\langle E \rangle_{MC}}{\langle s \rangle_{MC}} \right) = \frac{\langle E \rangle}{\langle s \rangle} (1 + \epsilon_N - \delta_N + \delta_N^2)
$$

(46)

$$
= \langle E \rangle \left( 1 + \frac{\delta_N^2}{N} \right).
$$

(47)

Because $s$ only takes on the values $+1$ or $-1$, we find that

$$
\langle \delta_N^2 \rangle = \frac{1 - \langle s \rangle^2}{N \langle s \rangle^2}.
$$

(48)

This shows that for small values of the average sign, the results will have a bias proportional to $1/(N \langle s \rangle^2)$. Even more dramatic is the effect on the variance of $\langle E \rangle_{MC}$. One finds that

$$
\text{var} \left( \langle E \rangle_{MC} \right) = \frac{\langle E \rangle^2}{\langle s \rangle^2} \langle (\epsilon_N - \delta_N)^2 \rangle
$$

$$
= \frac{\langle (\langle E \rangle_{MC} - \langle E \rangle_{MC})^2 \rangle}{\langle s \rangle^2}.
$$

(49)

The statistical errors will be inversely proportional to the average sign. One can try to fight the sign problem by taking larger samples, such that $N \langle s \rangle^2 \gg 1$. In practice one observes that the average sign decreases exponentially with growing system size, interaction strength and inverse temperature[31], forcing this strategy to fail at some point.

**sign problem remedies**

Many sign problem remedies have been proposed, but few of them have proven to be useful.

- *symmetries*: as mentioned before, there is a number of systems where symmetries guarantee that all the weights remain positive. In casu boson systems[1], one-dimensional fermion systems[13], the half-filled Hubbard model[32], even-even and $N = Z$ nuclei with schematic shell-model interactions[3] and Harmonic interactions[33].

- *fixed node approximation*: forbidding the walkers to cross the nodes of the trial wave function keeps all the weights positive[34]. The result is a variational upper bound for the energy, that becomes exact if the nodes of the trial function coincide with the nodes of the exact wave function. For the auxiliary-field methods a similar procedure has been worked out, called the constrained-path approximation[35]. It is no longer variational but seems to give reliable results in a number of cases.
• **explicit summation:** For few-particle systems one can try to sum explicitly over some components for each configuration in order to cancel out most of the sign fluctuations. This approach was very successful for the description of light nuclei, with explicit summation over all spin and isospin degrees of freedom [5, 6]. For more than 10 particles the summation becomes too heavy to be practical.

• **perturbation theory:** In the nuclear shell model, the dominant part of the interaction has a symmetry that guarantees a good sign. The remaining part can be treated perturbatively using a clever extrapolation scheme [36].

• **better decompositions:** The auxiliary-field method allows a lot of freedom in choosing the decomposition of Eq. (23). It has been suggested to shift the auxiliary fields onto *shifted contours* in the complex plain in order to suppress sign fluctuations [23]. Recent work suggests that pair operators can be used to construct decompositions that suffer less sign problems [24].

• Other tricks have been shown to reduce the sign problem in specific situations [37–40]

### VI. QUANTUM MONTE CARLO METHODS FOR LIGHT NUCLEI

The quantum Monte Carlo calculations for light nuclei have three important ingredients: first of all one needs a realistic interaction to describe the physics. Then a lot of effort goes into the construction of a highly optimized trial wave function. Monte Carlo methods are used to evaluate the trial wave functions in the optimization process called *Variational Monte Carlo* (VMC). Finally the Green's function Monte Carlo method (GFMC) is used to obtain an estimate for the exact ground-state energy of the system.

#### A. The nuclear interaction

The parametrization of the nuclear force is constrained by the fact that it should be applicable for GFMC calculations. Therefore it has to be local in coordinate space, otherwise the evaluation of propagators of the form Eq. (14) would be cumbersome.

A useful parametrization has been compiled by the Argonne group [41]. It is the so-called *Argonne v_{18} potential*,

\[
v_{ij} = v_{ij}^E + v_{ij}^R, \tag{50}
\]

with

• \( v_{ij}^E \) : the electromagnetic terms, including one- and two-photon-exchange Coulomb interactions, vacuum polarization, Darwin-Foldy terms, magnetic moment terms, with appropriate proton and neutron form factors.

• \( v_{ij}^R \) : the one-pion exchange potential,

\[
v_{ij}^R = f_{\pi}^2 \frac{1}{3} m_{\pi} c^2 X_{ij}^\pi r_i r_j,
\]

\[
X_{ij}^\pi = Y(r) \sigma_i \sigma_j + T(r) S_i S_j,
\]

\( Y(r) \) and \( T(r) \) are the Yukawa and tensor functions with a cutoff specified by the parameter \( c \):

\[
Y(r) = \frac{e^{-m_{\pi} r}}{m_{\pi} r} C(r)
\]

\[
T(r) = \left[ 1 + \frac{3}{m_{\pi} r} + \frac{3}{(m_{\pi} r)^2} \right] \frac{e^{-m_{\pi} r}}{m_{\pi} r} C^2(r)
\]

\[
C(r) = 1 - e^{-c^2}.
\]

• \( v_{ij}^R \) : a short-range phenomenological term

The one-pion exchange is the dominating contribution to the potential at long distances. At short distances a lot of other effects come into play: other meson exchanges, double pion exchange, Pauli-exclusion effects of the quarks inside the nucleons. Therefore one has introduced the cutoff parameter \( c \), and additional phenomenological terms to
account for the short-range physics. The one-pion exchange and phenomenological parts can be written as a sum over eighteen operators, from where the name $v_{18}$:

$$ v_{ij}^\pi + v_{ij}^R = \sum_{k=1,18} v_k(r_{ij}) \mathcal{O}^k_{ij}. $$

(51)

The first fourteen are charge independent:

$$ \mathcal{O}^k_{ij} = [1, (\sigma_i \cdot \sigma_j), S_{ij}, (\mathbf{L} \cdot \mathbf{S}), \mathbf{L}^2, \mathbf{L}^2(\sigma_i \cdot \sigma_j), (\mathbf{L} \cdot \mathbf{S})^2] \otimes [1, (\tilde{r}_i \cdot \tilde{r}_j)]. $$

The Tensor operator

$$ S_{ij} = 3(\sigma_i \cdot \hat{r}_{ij})(\sigma_2 \cdot \hat{r}_{ij}) - \sigma_i \cdot \sigma_j, $$

(52)

is a crucial operator: it produces a $D$-wave component in the deuteron wave function and it provides most of the binding for the light nuclei. All the free parameters in the interaction are fixed by fitting the results of exact calculations to experimental data for free nucleon-nucleon scattering (the Nijmegen database, [42, 43]) in the energy range 0 - 350 MeV and to the properties of the deuteron. No fit is made to heavier nuclei.

Simpler forms of the interaction, including only some of the above operators, have been used to. They go by the name of Argonne $v_6$, Argonne $v_8$ or other variants. The angular-momentum dependent terms require special attention, because they are not local in the strict sense. In the Monte-Carlo calculations they are evaluated by using finite differences between the coordinates at consecutive time steps. The Argonne $v_{18}$ interaction is non-relativistic. The lowest order relativistic corrections would be proportional to $P/c$, with $P$ the linear momentum operator. This operator is present in the interaction through the angular-momentum operators. Therefore, by fitting the interaction coefficients of the angular-momentum dependent terms to experimental data, one effectively takes into account the lowest order relativistic corrections.

The Argonne $v_{18}$ interaction is often supplemented with a three-body force. The parametrization of the three-body force is much harder to fix, because the phenomenology is much less understood than for the two-body interactions, and there is no free three-nucleon scattering data available. One typically uses phenomenological forces that are based on some meson-exchange model, with parameters that are then fitted to experimental data for $^3H$, $^3He$ and the $\alpha$ particle (Tucson-Melbourne, Urbana IX and Illinois-II are the names of some of these parametrizations[41]).

B. Variational Monte Carlo

The convergence of the GFMC depends crucially on the quality of the initial trial state $|\Psi_T\rangle$. Therefore considerable effort is spent on optimizing this trial wave function.

The simplest form for the trial wave function would be a Slater determinant. One could try to optimize the single-particle wave functions in order to minimize the energy,

$$ E_T = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}. $$

(53)

This is equivalent to a Hartree-Fock-calculation. The Slater determinants used for the GFMC calculations for light nuclei are based on the eigenfunctions of Woods-Saxon potentials. One optimizes the wave functions by varying the potential parameters. Different parameters can be used for each orbital. In principle one could vary the complete wave function in coordinate space. However, the important corrections will not come from better single-particle wave functions but from many-particle correlations that are missing in a Slater determinant wave function. E.g. the hard-core repulsion between nucleons should make it very improbable for one nucleon to be near to another one. Therefore one tries to improve the wave function by using correlating functions. At the two-body level these correlations are taken into account in a Jastrow wave function:

$$ \Psi_T(r_1, r_2, \cdots r_A) = \hat{A} \left[ \prod_{ij} f(r_i, r_j) \right] \Psi_0(r_1, r_2, \cdots r_A), $$

(54)

with $\Psi_0$ e.g. a Slater determinant made up of single particle states $\psi_1, \cdots \psi_A$. The functions are determined from a two-body problem, such that they assure the correct asymptotic behaviour for the $A$-body problem too. Often they
also depend on some parameters, which one tries to optimize in order to minimize the energy. The structure of atomic nuclei is so complicated, that one furthermore needs to include three-body correlation functions of the form

$$f_{ijk} = 1 - t_1 \left( \frac{||r_i - r_j||}{R_3} \right)^{t_2} e^{-t_3 R_3},$$  \hspace{1cm} (55)

with

$$R_3 = ||r_i - r_j|| + ||r_i - r_k|| + ||r_k - r_j||.$$  \hspace{1cm} (56)

The wave function described above is easy to evaluate for a given vector \( \mathbf{R} \), but it is very hard to evaluate its integral or operator expectation values because of the high dimensionality of the function. This complicates the optimization procedure for the wave-function parameters. Therefore one resorts to Monte-Carlo methods: one generates a sample of random (multi-particle) coordinates \( \mathbf{R} \) distributed according to

$$\pi(\mathbf{R}) = \langle \Psi_T | \mathbf{R} | \Psi_T \rangle.$$  \hspace{1cm} (57)

The Metropolis algorithm is well suited for this purpose.

**Exercise 6** One could use a Gaussian transition rule for generating the trial moves for the Metropolis algorithm. A more efficient method is to use an asymmetric transition rule (see exercise 5),

$$P_0(\mathbf{R} \rightarrow \mathbf{R}') = e^{-\frac{(\mathbf{R}' - \mathbf{R} - D(\mathbf{R}))^2}{2s^2}},$$  \hspace{1cm} (58)

with an average step width \( s \) and a drift term \( D(\mathbf{R}) \). Try to find an optimized form for \( D(\mathbf{R}) \) in the limit of a small step width \( s \). Hint: Eq (16) might give some inspiration.

The nuclear many-body wave function not only depends on the particle coordinates \( \mathbf{R} \), but also on the discrete coordinates in spin and isospin space. To guarantee the right symmetry for the trial wave function and to allow a smooth sampling in the Monte-Carlo procedure, these discrete coordinates are summed over explicitly for each configuration of space coordinates \( \mathbf{R} \). For light nuclei this is not to much of a problem, because there are not too many spin-isospin configurations. However, for a nucleus such as \(^{12}\text{C}\), one has to take into account already 43008 configurations. This is at the limit of present computational possibilities. With increasing computer power one can expect \(^{12}\text{C}\) to come into reach in a couple of years, but heavier nuclei probably never will be calculated with this technique because the number of configurations grows exponentially with each particle added.

The sample of real space configurations \( \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M \) can be used to calculate a variational estimate for the energy,

$$E_{\text{VMC}} = \frac{1}{M} \sum_i E(\mathbf{R}_i),$$  \hspace{1cm} (59)

with \( E(\mathbf{R}) \) the local energy, given by

$$E(\mathbf{R}) = V(\mathbf{R}) + \frac{\hbar^2}{2m} \frac{\Delta \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}.$$  \hspace{1cm} (60)

To optimize the variational parameters, one could try to minimize the expectation value of the energy \( E_{\text{VMC}} \). Instead, one often prefers to minimize the variation of the energy: if the variational wave function would be the exact wave function, then the local energy would be a constant,

$$E(\mathbf{R}) = E_{\text{GS}}, \quad \text{forall} \mathbf{R},$$  \hspace{1cm} (61)

and hence the variance on \( E(\mathbf{R}) \) would become zero. Typically one generates a sample of configurations, for which one varies and optimizes the interaction parameters. With the new parameters one generates a new sample, for which one repeats the same procedure, until convergence is reached.

**C. Green’s function Monte Carlo**

Starting from the variational wave functions described in the previous paragraphs, one can start a GFMC calculation using the path-integral decomposition of section III A and the Monte-Carlo scheme with walkers and branching of section IV B. This approach has been applied to light nuclei with great success [5–9].
The application of the GFMC to fermions is far from obvious, because of the sign problems related to the fermion antisymmetry. Luckily, for the wave functions of light nuclei it turns out that the antisymmetry manifests itself mainly in the spin-isospin sector. By explicit summation over all spin-isospin configurations, one obtains a weight function \( w(R) \) for the coordinate-space configuration that is dominated by a symmetric component. This is why the sign problem can be avoided at small imaginary times. When the wave function is propagated further in imaginary time, the sign problem sets in anyway. Therefore one has to start from a very good initial wave function, such that the imaginary-time propagation converges to the ground state quickly, before the sign problems occur. If it is necessary to propagate further in imaginary time, then one can use the constrained-path condition to keep the signs positive, and then release the constraint and observe how the results evolve in the next time steps to get an idea of the error on the constrained-path results.

The GFMC uses the Monte Carlo method with walkers and branching to sample the configurations. Here the trick with the drift term of Eq. (16) and exercise 6 is used to suppress the leading order in the fluctuations of the weights of the walkers. This is another reason why the initial wave function should be highly optimized. Furthermore the GFMC code can reuse a lot of the VMC code, because the single-walker updates are very similar to the VMC Metropolis steps.

The GFMC method was used to calculate the properties of light nuclei using the force described in section VI A. It turned out that it was not possible to get an overall agreement with the experimentally observed binding energies and spectra using a force that correctly describes the free particle scattering data and the two-particle states. This problem was solved by including a three-body force. Even though the three-body force parameters were fitted only to the \( A \leq 3 \) nuclei, it turned out that now one could reach an excellent agreement with the experimental data for all nuclear states that could be calculated (at present up to \( A \leq 10 \)).

![Diagram of nuclear energy levels for realistic potential models](attachment:diagram.png)

**FIG. 1:** Nuclear energy levels for realistic potential models; shading denotes Monte Carlo statistical errors. (taken from reference [44])

**Conclusions**

Quantum Monte Carlo methods offer a wide range of tools to study correlated many-body systems. They are capable to take the many-body correlations exactly into account, up to controllable statistical and discretization errors. For fermions, most applications are hampered by the sign problem. However, by using excellent variational wave functions to start from, and by explicit summation over spin- and isospin-degrees of freedom, one has been able to perform converged calculations for light nuclei, up to \( A \leq 10 \). These results have shown that the structure of light nuclei can be understood from a non-relativistic local interaction, provided that three-body interactions are taken into account. One is still working on the extension to heavier nuclei, but due to the complications in the variational wave-functions and because of the explicit summation over spin and isospin degrees of freedom, it remains to be seen if the method can be extended beyond \( A = 12 \).
Acknowledgments

The author wishes to thank D. Dean and B. Barrett for the interesting discussions and R. Wiringa for providing figure 1. Particularly the author would like to express his gratitude to Joseph Carlson for the fruitful exchanges of ideas over the past years and for providing some of the results shown in the lectures. The author is grateful to the Fund for Scientific Research - Flanders (Belgium), to the NATO Science Programme and to the Research Board of the University of Gent for financial support.

APPENDIX A: SOLUTIONS OF THE EXERCISES

• Solution of exercise 1:
The following estimator is one order better in the difference \(|\Psi_{GS} - |\Psi_T\rangle\) than the mixed estimator of Eq.(7) for operators \(\hat{A}\) that do not commute with the Hamiltonian:

\[
\langle \hat{A}\rangle \simeq \frac{2}{\langle \Psi_T|e^{-\beta H}|\Psi_T\rangle} \left( \frac{\langle \Psi_T|\hat{A}|\Psi_T\rangle}{\langle \Psi_T|\Psi_T\rangle} - \frac{\langle \Psi_T|\hat{A}|\Psi_T\rangle}{\langle \Psi_T|\Psi_T\rangle} \right). \tag{A1}
\]

• Solution of exercise 2:
For a system of \(A\) identical but distinguishable particles of mass \(m\) one can write

\[
\langle K|\hat{\mathbf{R}}\rangle = \frac{1}{(2\pi\hbar)^{3A/2}} e^{-i\mathbf{K}\cdot\mathbf{R}/\hbar}, \tag{A2}
\]

and

\[
\langle K|e^{-i\hat{\mathbf{T}}}|K'\rangle = \exp \left(-\frac{i}{2m} \sum_{j=1}^{A} k_j^2 \right) \delta(K - K'). \tag{A3}
\]

It follows that

\[
\int \int \langle \mathbf{R}|K\rangle \langle K|e^{-i\hat{\mathbf{T}}}|K'\rangle\langle K'|\mathbf{R}'\rangle dKdK' = \left( \frac{m}{2\pi\hbar^2} \right)^{3A/2} e^{-\frac{(n' - n)^2}{2m^2\hbar^2}}. \tag{A4}
\]

For indistinguishable fermions (bosons) the delta function in Eq.(A3) has to be replaced by a determinant (permanent) of single-particle delta functions. However, in practice one includes only the term for distinguishable particles (this corresponds to the diagonal contribution to the determinant/permanent), because the off-diagonal contributions are one order smaller in the time step and because the trial state will impose the right symmetry on the total imaginary-time path anyway.

• Solution of exercise 3:

\[
e^{-i(\hat{\mathbf{T}} + \hat{\mathbf{V}})} = e^{-i\hat{\mathbf{V}}} e^{-\frac{\Delta}{2} \hat{\mathbf{T}}} e^{-\frac{\Delta}{2} \hat{\mathbf{V}}} e^{-i\hat{\mathbf{T}}} e^{-i\hat{\mathbf{V}}} + O(\epsilon^5), \tag{A5}
\]

with

\[
\hat{\mathbf{V}} = \hat{\mathbf{V}} + \frac{\epsilon^4}{48} [[[\hat{\mathbf{V}}, \hat{\mathbf{T}}, \hat{\mathbf{V}}]]. \tag{A6}
\]

For a purely local interaction \(\hat{\mathbf{V}}\) and a purely kinetic one-body operator \(\hat{\mathbf{T}}\), the triple commutator can be incorporated exactly in the diffusion Monte-Carlo decomposition. See reference [46] for a detailed derivation.

• Solution of exercise 4:
Because \(P_0\) is symmetric, \(P_0(x \rightarrow y) = P_0(y \rightarrow x)\), it is obvious that the condition of detailed balance, Eq.(35), is fulfilled for a uniform distribution,

\[
\pi(x) = \text{constant}. \tag{A7}
\]

• Solution of exercise 5:
Using the expression

\[
q(x \rightarrow y) = \frac{P_0(y \rightarrow x)\pi(y)}{P_0(x \rightarrow y)\pi(x)}, \tag{A8}
\]

one can write for \(x \neq y\) that

\[
P(x \rightarrow y) = P_0(x \rightarrow y) \min(1, q(x \rightarrow y)). \tag{A9}
\]
It is straightforward to check that \( P \) fulfills the detailed balance condition of Eq. \((35)\) with the target distribution \( \pi(x) \). If \( P_0 \) has a known stationary distribution \( \pi_0(x) \), one can simplify Eq. \((A8)\) to
\[
q(x \to y) = \frac{\pi(y)/\pi_0(y)}{\pi(x)/\pi_0(x)}.
\] (A10)

\textbf{Solution of exercise 6:}
The function \( q \) of Eq. \((A8)\) takes on the form
\[
q(R \to R') = \exp \left[ -\frac{(R - R' - D(R'))^2 - (R' - R - D(R))^2}{2s^2} \right] \frac{\Psi_T(R')^2}{\Psi_T(R)}.
\]

\[
= \exp \left[ -\frac{(R' - R) \cdot (D(R) + D(R'))}{2s^2} - \frac{D(R')^2 - D(R)^2}{2s^2} + 2 \ln |\Psi_T(R')/\Psi_T(R)| \right].
\] (A11)

By expanding the exponent of Eq. \((A11)\) in \( R' - R \), we find that the lowest order cancels if
\[
D(R) = 2s^2 \nabla \ln |\Psi_T(R)|.
\] (A12)

Because \( D(R) \) is of order \( s^2 \), the displacement \( R' - R \) will be of order \( s \) and the acceptance rate will be close to unity,
\[
q(R \to R') = 1 + \mathcal{O}(s^2).
\] (A13)