I. QUANTUM MONTE CARLO METHODS: INTRODUCTION AND BASICS

Markus Holzmann
LPMMC, UJF, Grenoble, and LPTMC, UPMC, Paris
markus@lptl.jussieu.fr
http://www.lptl.jussieu.fr/users/markus

(Dated: January 24, 2012)

I will provide a rough overview of zero temperature Quantum Monte Carlo calculations, and introduce the basics of Monte Carlo sampling, Markov chains, Metropolis algorithm, and error analysis.

A. Quantum Many-Body Problems

In the following we are interested in the properties of many-body systems at very low (zero) temperature and assume that the system is described by a general non-relativistic Hamiltonian

$$H = \sum_{i=1}^{N} \left[-\frac{\nabla_i^2}{2m} + v_{ext}(\mathbf{r}_i) \right] + \sum_{i<j} v(|\mathbf{r}_i - \mathbf{r}_j|) \tag{1}$$

where v is the inter particle interaction (Coulomb potential for electrons) and v_{ext} is an external potential, e.g. the potential created by the ions in electronic structure or chemical physics. Ideally, one would like to know the eigen values/functions of the Hamiltonian for specific boundary conditions which include also the statistics of the particles (symmetric for bosons, anti-symmetric for fermions)

$$H\Psi_i(\mathbf{R}) = E_i \Psi_i(\mathbf{R}) \tag{2}$$

where from now on $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N\}$ collects all state variables, e.g. spin degrees of freedom, although we will not explicitly consider them. In general, the wavefunctions $\Psi_i(\mathbf{R})$ will be complicated functions of dN variables, $\mathbf{R}$, where d is the spatial dimension, and the exact solution of the Schrödinger equation, Eq. (2), in $d = 2, 3$ dimensions and $N \gg 4$ is infeasible. Therefore, one either simplifies the Hamiltonian or tries to map it to a systems whose solution is accessible (e.g. density functional methods), or one works on approximative wave functions. Here I will discuss zero temperature Quantum Monte Carlo methods to obtain the (approximative) ground state wavefunction, $\Psi_0(\mathbf{R})$.

Variational principle. The ground state energy of any quantum system can be bounded by above using the variational principle: The energy expectation value of any *reasonable* (C_2) wave function, $\Psi_T(\mathbf{R})$, is always higher than the ground state energy, E_0:

$$E_0 \leq E_T \equiv \frac{\int d\mathbf{R} \Psi_T^*(\mathbf{R}) H \Psi_T(\mathbf{R})}{\int \mathbf{R} |\Psi_T(\mathbf{R})|^2} \tag{3}$$

Note, that $\Psi_T(\mathbf{R})$ does not need to be normalized.

The proof is elementary using the eigenstates, Eq. (2),

$$\frac{\int d\mathbf{R} \Psi_T^*(\mathbf{R}) H \Psi_T(\mathbf{R})}{\int \mathbf{R} |\Psi_T(\mathbf{R})|^2} = \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} \geq E_0, \tag{4}$$

where

$$c_i = \int d\mathbf{R} \Psi_i^*(\mathbf{R}) \Psi_T(\mathbf{R}) \tag{5}$$

is the overlap of the trial wave function with the exact eigenstates. The variational principle is of fundamental importance in the following, since it allows us to compare and quantify the "quality" of different wave functions without invoking the comparison with experiment. The variational principle can be extended to excited states considering only trial states which are orthogonal to all states with lower energies. In practice, only different symmetries guarantee the orthogonality, so that the variational principle can only be applied to the lowest energy state within the symmetry of Ψ_T.

Exact projection. The proof of the variational principle also suggests how any trial wave function can be improved in principle considering

$$\Psi_\beta(\mathbf{R}) = \langle \mathbf{R}|e^{-\beta H}|\Psi_T\rangle = \int d\mathbf{R}' G(\mathbf{R},\mathbf{R}';\beta)\Psi_T(\mathbf{R}') \tag{6}$$

where we have defined the propagator

$$G(\mathbf{R},\mathbf{R}';\beta) = \langle \mathbf{R}|e^{-\beta H}|\mathbf{R}'\rangle \tag{7}$$

We get

$$E_\beta \equiv \frac{\int d\mathbf{R}\Psi^*_\beta(\mathbf{R})H_N\Psi_\beta(\mathbf{R})}{\int \mathbf{R}|\Psi_\beta(\mathbf{R})|^2} = \frac{\sum_i |c_i|^2 e^{-2\beta E_i}E_i}{\sum_i |c_i|^2 e^{-2\beta E_i}} = E_0 + \frac{\sum_{i>0}|c_i/c_0|^2 e^{-2\beta(E_i-E_0)}(E_i-E_0)}{1+\sum_{i>0}|c_i/c_0|^2 e^{-2\beta(E_i-E_0)}} \tag{8}$$

which converges exponentially to the true ground state from above for large projection time β.

Jastrow trial wave function. The simplest wave function which captures the basic features of the ground state of bosonic quantum systems is a Jastrow-type wavefunction

$$\Psi(\mathbf{R}) = e^{-\sum_i u_1(\mathbf{r}_i)-\sum_{ij} u_2(r_{ij})-\sum_{ijk} u_3 \dots} \tag{9}$$

where $u_n(r)$ are one dimensional function which are either given by some approximate theory (Quantum Cluster expansions, Correlated Basis Function) or which can be parametrized with few parameters which are then optimized to lower the energy.

However, already the evaluation of the wavefunction, or the energy, using the Jastrow approximation cannot be done analytically in most cases (and we have not yet optimized it). However, notice that the form (9) is formally just the same as the integrand of the configuration integral of a classical fluid with a corresponding inter-particle potential given by $u(r) = \beta v(r)$. Therefore we have mapped the quantum problem to a purely classical many-body problem.

Slater-Jastrow. The Jastrow ansatz, Eq. (9), is symmetric with respect to particle exchanges and therefore can approximate Bosonic wavefunctions. In order to treat Fermions, we have to use an antisymmetric ansatz for the wavefunction. This is conviniently done using an additional term - a Slater determinant of single particle orbitals

$$D(\mathbf{R}) = \det_{ki} \phi_k(\mathbf{r}_i). \tag{10}$$

For homogeneous systems we are forced to use plane wave orbitals $\phi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ inside the determinant, typically the first N plane-waves with energies $\hbar^2 k^2/2m$ less than the Fermi-energy ϵ_F. The simplest non-trivial ansatz for the fermionic wavefunction, the Slater-Jastrow form, is thus

$$\Psi_F(\mathbf{R}) = \det_{ki} \phi_k(\mathbf{r}_i) e^{-\sum_{ij} u(r_{ij})} \tag{11}$$

and is still one of the most used forms for precise Fermion calculations up to date.

B. Quantum Monte Carlo Integration

Variational Monte Carlo (VMC) uses Monte Carlo methods for the integrations needed to evaluate the energy or other observables of a trial wave function, $\Psi_T(\mathbf{R})$, which is given explicitly. **Projector Monte Carlo** methods, e.g. Diffusion, Reptation, or Variational Pathi Integral Monte Carlo, are based on the additional stochastic sampling of the imaginary-time propagator to sample the true ground state wave function. All methods are based on the representation of the quantum problem similar to classical Monte Carlo simulations. We are interested in calculating expectation values of a given operator, O,

$$\langle O\rangle \ = \ \frac{\int d\mathbf{R} O(R)\pi(\mathbf{R})}{Z}, \quad Z = \int d\mathbf{R}\pi(\mathbf{R}) \tag{12}$$

where the statistical weight $\pi(\mathbf{R})$ is given by

$$\begin{aligned}
\pi(\mathbf{R}) &= \exp[-\beta V(\mathbf{R}), \quad \text{Classical MC} \\
\pi(\mathbf{R}) &= |\Psi_T(\mathbf{R})|^2 = |D(\mathbf{R})|^2 \exp[-2U(\mathbf{R})], \quad \text{Variational MC} \\
\pi(\mathbf{R},\mathbf{R}') &= |\Psi_T|(\mathbf{R})|G(\mathbf{R},\mathbf{R}';\tau)|\Psi_T(\mathbf{R}')|^{-1} \quad \text{Diffusion MC} \\
\pi(\mathbf{R},\mathbf{R}_1,\dots,\mathbf{R}_M,\mathbf{R}') &= |\Psi_T|(\mathbf{R})|G(\mathbf{R},\mathbf{R}_1,\tau)G(\mathbf{R}_1,\mathbf{R}_2;\tau)\cdots G(\mathbf{R}_M,\mathbf{R}';\tau)|\Psi_T(\mathbf{R}')| \quad \text{Reptation MC}
\end{aligned} \tag{13}$$

and Z plays the role of a partition function.

Deterministic integration. Calculating observables involves high dimensional integrations, $D = dN$ dimensions in VMC, and even more in projection methods. For simplicity, we will just consider the general integral

$$I = \int d\mathbf{R}\tilde{f}(\mathbf{R}) \tag{14}$$

Conventional integration is based on discretization of space in finite intervals, let us use P points in each dimension. The integration error, ϵ will scale as a power of $h = 1/P$ ($\epsilon \sim h^2$ for trapezoidal rule) if we can reach the asymptotic region $P \gg 1$. However, the CPU time, T, scales with the total number of points, $T \sim P^D \sim \epsilon^{-D/2}$. Therefore, reducing the error by a factor of two, we need $2^{D/2}$ more computer time.

Stochastic sampling methods. Let us now assume that we can create P configurations $\mathbf{R}(t)$ at different times $t_1, t_2, \ldots, t_P$, and each configuration is distributed by the (normalized) probability $p(\mathbf{R})$. Let us rewrite the integrals under consideration as

$$I = \int d\mathbf{R} f(\mathbf{R}) p(\mathbf{R}), \quad f(\mathbf{R}) = \frac{\tilde{f}(\mathbf{R})}{p(\mathbf{R})} \tag{15}$$

Our best estimate of the integral will be

$$I \approx \bar{f} \equiv \frac{1}{P} \sum_{i=1}^{P} f(\mathbf{R}_i) \tag{16}$$

But do we have any control of the error?

Central limit theorem (Gauss). To estimate the error, we introduce the distribution function of the values f, $c_f(x)$, together with its characteristic function $\tilde{c}_f(k)$, its Fourier transform,

$$c_f(x) = \int d\mathbf{R}\delta(f(\mathbf{R}) - x)p(\mathbf{R}), \quad \tilde{c}_f(k) = \int dx e^{ikx} c_f(x) \tag{17}$$

Since the configurations $\mathbf{R}_i$ are completely uncorrelated obtained, the probability distribution for obtraining the value $\bar{f}$ is given by

$$c_{\bar{f}}(x) = \prod_{i=1}^{P} \int dx_i c_f(x_i) \delta\left(x - \frac{1}{P}\sum_i x_i\right) = \prod_{i=1}^{P} \int dx_i c_f(x_i) \int \frac{dk}{2\pi} e^{-ik\left(x - \frac{1}{P}\sum_i x_i\right)} = \int \frac{dk}{2\pi} e^{-ikx} \left[\tilde{c}_f(-k/P)\right]^P \tag{18}$$

In the limit of large P, we are basically interested in $\tilde{c}_f(k)$ at small $k \to 0$, where we assume that the following expansion holds:

$$\lim_{k\to 0} \tilde{c}_f(k) = e^{ikf_1 - k^2 f_2/2 - ik^3 f_3/6 - \ldots} \tag{19}$$

where f_1 is the mean, f_2 the variance, and f_3 the skewness of the distribution of the observable, e.g.

$$f_1 = \int d\mathbf{R} f(\mathbf{R}) p(\mathbf{R}), \quad f_2 = \int d\mathbf{R} \left[f(\mathbf{R}) - f_1\right]^2 p(\mathbf{R}), \quad \text{etc} \tag{20}$$

Inserting in Eq. (18), we obain the probability distribution of our estimate $\bar{f}$ for large sample sizes $P \gg 1$:

$$c_{\bar{f}}(x) = \left(\frac{P}{2\pi f_2}\right)^{1/2} \exp\left[-\frac{(x - f_1)^2}{2f_2/P}\right] \tag{21}$$

which is a gaussian distribution around the true mean value $f_1 \equiv I$ (our desired integral) with variance $\sigma^2 = f_2/P$. The asymptotic sampling error using $P \gg 1$ points is therefore given by $\sigma = (f_2/P)^{1/2}$, if the variance f_2 exists and is finite. The important consequence is that the error of stochastic sampling methods decreases as $\epsilon \sim P^{-1/2}$ independent of the dimensions of the integral under the fairly large conditions (finite variance independent of D, P are number of *independent* points). The CPU time of Monte Carlo methods thus scales as $T \sim \epsilon^{-2}$, and becomes favourable compared to conventional integration to in $D > 4$ dimensions. Reducing the Monte Carlo error by a factor of two, we only need to wait 4 times longer!

Direct Sampling. Unfortunately, we will not be able to directly create configurations according to our probability distribution $p(\mathbf{R}) = \pi(\mathbf{R})/Z$, Eq. (12) and Eq.(13), for most of our problems. Only a few distributions, uniform, gaussian, exponential..., can be created directly. However, using for example a uniform distribution, one can convince oneself (e.g. looking at classical hard spheres), that the variance of the probability distribution in generally increases exponentially with the dimension D, so that *direct sampling* methods will not work for many-body systems with $D = dN \gg 1$.

Markov chains. At this point we introduce an (artificial) Monte Carlo dynamics in the system, with the aim of creating configurations in time distributed according to $p(\mathbf{R})$, at least asymptotically. Therefore, we consider the conditional probability that the system is in state $\mathbf{R}_{t_n}$ at time t_n

$$P(\mathbf{R}_{t_n}|\mathbf{R}_{t_{n-1}}, \mathbf{R}_{t_{n-2}}, \ldots, \mathbf{R}_{t_1}) \tag{22}$$

given that at the preceding time it was in state $\mathbf{R}_{t_{n-1}}$ at t_{n-1}, $\mathbf{R}_{t_{n-2}}$ at t_{n-2}, etc. In the most simplest case this conditional probability is just given statistically uncorreclated samples, or

$$P_{\text{direct sampling}}(\mathbf{R}_{t_n}|\mathbf{R}_{t_{n-1}}, \mathbf{R}_{t_{n-2}}, \ldots, \mathbf{R}_{t_1}) = p(\mathbf{R}_{t_n})p(\mathbf{R}_{t_{n-1}})\cdots p(\mathbf{R}_{t_1}) \tag{23}$$

However this reduces direct sampling algorithm above which was not efficient enough. A particular dynamics which is still simple to analyse are obtained by Markov chains: we choose certain transition probabilities from one state to another which do not depend on the particular state

$$P_n(\mathbf{R}_{t_n}|\mathbf{R}_{t_{n-1}}, \mathbf{R}_{t_{n-2}}, \ldots, \mathbf{R}_{t_1}) = p(\mathbf{R}_{t_1})T(\mathbf{R}_{t_1} \to \mathbf{R}_{t_2})T(\mathbf{R}_{t_2} \to \mathbf{R}_{t_3})\cdots T(\mathbf{R}_{t_{n-1}} \to \mathbf{R}_{t_n}) \tag{24}$$

The basic object in Markov chains is the transition probability $T(\mathbf{R} \to \mathbf{R}')$ which depends on both states $\mathbf{R}$ and $\mathbf{R}'$ but is independant of time. The transition probability must be normalized

$$\sum_{\mathbf{R}'} T(\mathbf{R} \to \mathbf{R}') = \sum_{\mathbf{R}} T(\mathbf{R} \to \mathbf{R}') = 1 \tag{25}$$

Detailed balance. We can write down a master equation for the probability, $p(\mathbf{R}, t+1)$, to find $\mathbf{R}$ at $t+1$ knowing the distribution at t by considering all possible processes

$$p(\mathbf{R}, t+1) = \sum_{\mathbf{R}' \neq \mathbf{R}} p(\mathbf{R}', t)T(\mathbf{R}' \to \mathbf{R}) + T(\mathbf{R} \to \mathbf{R})p(\mathbf{R}, t) \tag{26}$$

or using the normalization (25)

$$\begin{aligned} p(\mathbf{R}, t+1) - p(\mathbf{R}, t) &= \sum_{\mathbf{R}' \neq \mathbf{R}} p(\mathbf{R}', t)T(\mathbf{R}' \to \mathbf{R}) - \sum_{\mathbf{R}' \neq \mathbf{R}} T(\mathbf{R} \to \mathbf{R}')p(\mathbf{R}, t) \qquad (27) \\ &= \sum_{\mathbf{R}'} p(\mathbf{R}', t)T(\mathbf{R}' \to \mathbf{R}) - \sum_{\mathbf{R}'} p(\mathbf{R}, t)T(\mathbf{R} \to \mathbf{R}') \qquad (28) \end{aligned}$$

The stationary state $\pi(\mathbf{R}) = \lim_{t\to\infty} p(\mathbf{R}, t)$ then satisfies

$$\sum_{\mathbf{R}'} \pi(\mathbf{R}')T(\mathbf{R}' \to \mathbf{R}) = \sum_{\mathbf{R}'} \pi(\mathbf{R})T(\mathbf{R} \to \mathbf{R}') \tag{29}$$

This condition just ensures that once equilibrium is reached it remains unchanged, since losses and gains just compensate. The so called detailed balance condition, is when we impose egality of each term separatly in the sum

$$\pi(\mathbf{R}')T(\mathbf{R}' \to \mathbf{R}) = \pi(\mathbf{R})T(\mathbf{R} \to \mathbf{R}') \tag{30}$$

Exponential convergence. How can be assure that the Markov process is really converging to the stationary distribution $\pi(\mathbf{R})$ and what is the time scale associated to reach stationatity? Let us start from an arbitrary distribution $p(\mathbf{R}_0, t = 0)$, after time $t = n$ we have

$$p(\mathbf{R}_n, t = n) = \sum_{\mathbf{R}_0}\sum_{\mathbf{R}_1}\sum_{\mathbf{R}_2}\cdots\sum_{\mathbf{R}_{n-1}} p(\mathbf{R}_0, t = 0)T(\mathbf{R}_0 \to \mathbf{R}_1)T(\mathbf{R}_1 \to \mathbf{R}_2)\cdots T(\mathbf{R}_{n-1} \to \mathbf{R}_n) \tag{31}$$

which can be simply written as a matrix product introducing a huge vecor to represent the probabilities $p(\mathbf{R}, t)$ and a matrix for $T(\mathbf{R} \to \mathbf{R}')$

$$p(\mathbf{R}_n, t = n) = \langle \mathbf{R}_n, t = n|T^n|\mathbf{R}_0, t = 0\rangle p(\mathbf{R}_0, t = 0) \tag{32}$$

Diagonalizing the matrix T and denoting $|\lambda_i\rangle$ the eigenvector with eigenvalue λ_i we have

$$p(\mathbf{R}_n, t=n) = \sum_i \lambda_i^n \langle \mathbf{R}_n, t=n|\lambda_i\rangle\langle\lambda_i|\mathbf{R}_0, t=0\rangle p(\mathbf{R}_0, t=0) \tag{33}$$

$$\simeq \lambda_0^n a_0[1 + (\lambda_1/\lambda_0)^n a_1 + \ldots] \tag{34}$$

where λ_0 is the largest eigenvector. For large n we will have exponential convergence to this eigenvector since the contributions of the other eigenvectors will decay with e.g.

$$(\lambda_1/\lambda_0)^n = e^{-n\log\lambda_0/\lambda_1} \tag{35}$$

for the first eigenstate.

Using the normalization, Eq. (25), we have

$$\sum_{\mathbf{R}'} \pi(\mathbf{R})T(\mathbf{R}\to\mathbf{R}') = \pi(\mathbf{R}) \tag{36}$$

Substituting the detailed balance inside the rhs we get

$$\sum_{\mathbf{R}'} \pi(\mathbf{R}')T(\mathbf{R}'\to\mathbf{R}) = \pi(\mathbf{R}) \tag{37}$$

we see that the stationary distribution $\pi(\mathbf{R})$ is an eigenstate of the transition matrix with eigenvalue one. If we can show that all other eigenvalues are smaller than one we indeed reach the desired equilibrium. This can indeed be assured if it is possible to go from any of the configurations to all others in a finite number of steps (ergoticity).

Metropolis algorithm. A particular solution for the transition matrix which satisfies detailed balance is the Metropolis algorithm using

$$T(\mathbf{R}\to\mathbf{R}') = \min\left[1, \frac{\pi(\mathbf{R}')}{\pi(\mathbf{R})}\right] \tag{38}$$

To prove detailed balance for general values of $\pi(\mathbf{R})$ and $\pi(\mathbf{R}')$ we write down the tranition matrix $T(\mathbf{R}\to\mathbf{R}')$ and $T(\mathbf{R}'\to\mathbf{R})$. Assume $\pi(\mathbf{R}') > \pi(\mathbf{R})$ first. In that case we have $T(\mathbf{R}\to\mathbf{R}') = 1$, but $T(\mathbf{R}'\to\mathbf{R}) = \pi(\mathbf{R})/\pi(\mathbf{R}')$. We immediately find $\pi(\mathbf{R})T(\mathbf{R}\to\mathbf{R}') = \pi(\mathbf{R})'T(\mathbf{R}'\to\mathbf{R})$ to be fullfilled. The case $\pi(\mathbf{R}) > \pi(\mathbf{R}')$ goes through just the same. Therefore we enforces detailes balance for each term in the summation of Eq. (30) (all elements there are non negative)

$$\pi(\mathbf{R}')T(\mathbf{R}'\to\mathbf{R}) = \pi(\mathbf{R})T(\mathbf{R}\to\mathbf{R}') \tag{39}$$

A priori probabilities. Sometimes it is useful to choose a-priori probabilities to propose moves from $\mathbf{R}$ to $\mathbf{R}'$ which are different from the a-priori probability for the return move $\mathbf{R}'$ to $\mathbf{R}$. In this case we write

$$T(\mathbf{R}\to\mathbf{R}') = \mathcal{A}(\mathbf{R}\to\mathbf{R}')p(\mathbf{R}\to\mathbf{R}') \tag{40}$$

where $\mathcal{A}(\mathbf{R}\to\mathbf{R}')$ is the a-priori probability and $p(\mathbf{R}\to\mathbf{R}')$ the final acceptance probability. Enforcing detailed balance we are led to a generalized Metropolis algorithm for the acceptance probability

$$p(\mathbf{R}\to\mathbf{R}') = \min\left[1, \frac{\pi(\mathbf{R}')}{\mathcal{A}(\mathbf{R}\to\mathbf{R}')}\frac{\mathcal{A}(\mathbf{R}'\to\mathbf{R})}{\pi(\mathbf{R})}\right] \tag{41}$$

The a-priori probability is choosen such that it is fast to evaluate and bigger moves can be done for a reasonable acceptance probability.

C. Simulation

Let us sketch the basic steps of Monte Carlo program

- program **MonteCarlo**
- call **initialize**($\mathbf{R}$)

```
• LOOP
    – call MCstep($\mathbf{R}$)
    – call observables($\mathbf{R}$)
• END LOOP
```

Inside the subroutine **MCstep** we propose a new configuration and accept or reject the new configurations according to any rule which satisfies ergodicity and detailed balance. After each move we can write our observables or average over them, independently if the previous Monte Carlo move was accepted or rejected.

A simple Monte Carlo step based on a uniform displacement of the configuration by a random vector inside a cubic volume of length Δ and subsequent Metropolis algorithm for acceptance writes

```
• subroutine MCstep
• # fist move the old configuration $\mathbf{R}$ to sample a new one $\mathbf{R}_{new}$
    – $\mathbf{R}_{new} = \mathbf{R} + \Delta(\mathrm{rnd}() - 0.5)$
• # MC acceptance/ rejection using Metropolis algorithm
    – $p = \pi(\mathbf{R}_{new})/\pi(\mathbf{R})$
    – IF $p >$rnd():
        ∗ # accept move
        ∗ use $\mathbf{R}_{new}$ as new configuration: $\mathbf{R} = \mathbf{R}_{new}$
• return configuration $\mathbf{R}$
```

Note that rnd() creates random number with $0 \leq \mathrm{rnd}() < 1$.

Acceptance ratio. In the above example, Δ was a parameter which determines the maximal displacement of the configurations and must be adapted to improve the efficiency of the Monte Carlo sampling. It is important to measure always the acceptance ratio of the MC step. In general, one should vary the step size Δ such that one obtains $0.1 \lesssim$ acceptance ratio $\lesssim 0.9$. Optimizing the acceptance probability does not lead to important improvements in the efficiency.

Optimizing a-priori probabilities. In the above example, we have chosen a uniform a-priori probability $\mathcal{A}_u(\mathbf{R} \to \mathbf{R}') = \Delta^{-D}$ if $\mathbf{R}'$ is inside the cube of length Δ centered around $\mathbf{R}$, and zero outside. However, we can use any transition rule as long as you can go anywhere in space within a finite number of steps (ergodicity). Changes in the a-priori probability can lead to important improvements in the efficiency as one may spend more time in the relevant region of phase space (importance sampling). It is clear if one can create an apriori-probability $\mathcal{A}(\mathbf{R} \to \mathbf{R}') \approx \pi(\mathbf{R}')/C$, one samples close to the exact distribution. As a consequence, the acceptance ration will get close to one. More generally, we might chose $\mathcal{A}(\mathbf{R} \to \mathbf{R}') \approx \pi(\mathbf{R}')/\pi(\mathbf{R})$ and expand $\pi(\mathbf{R}')$ for $\mathbf{R}'$ around $\mathbf{R}$. From the first order term, we obtain a drift-force (*Force-Bias MC*), expanding $\log \pi(\mathbf{R}')$ up to second order we obtain the best gaussian a-priori probability. To compare the efficiency of a simulation, we should optimize $1/(\text{CPU time} \times \text{Error}^2)$ for different algorithms.

Error estimation. The central limit theorem provides the basis for error estimations: For uncorrelated data, for large T, we have

$$\langle O \rangle = \overline{O}_T \pm \epsilon_T \tag{42}$$

$$\overline{O}_T = \frac{1}{T} \sum_t O(\mathbf{R}_t) \tag{43}$$

$$\epsilon_T = \langle \left[\overline{O}_T - \langle O \rangle \right]^2 \rangle = \sqrt{\frac{\sigma^2(O)}{T}} \approx \sqrt{\frac{1}{T(T-1)} \sum_t \left[O(\mathbf{R}_t) - \overline{O}_T \right]^2} \tag{44}$$

However, our random walk will in general create data $\mathbf{R}_t$ which are correlated in time, and the true error will be given by

$$\epsilon_T \approx \sqrt{\frac{\tau}{T(T-1)} \sum_t \left[O(\mathbf{R}_t) - \overline{O}_T \right]^2} \tag{45}$$

where τ introduces the exact correlation time (T/τ are the number of true independent configurations created).

To estimate the correlation time, we can average over M subsequent data points, or, equivalently consider the operator

$$O^{(M)} = \frac{1}{M} \sum_{i=1}^{M} O(\mathbf{R}_{(t-1)M+i}) \tag{46}$$

for $t = 1, \dots T/M$. Whereas $\overline{O}^{(M)} = \overline{O}_T$ for all M, we can assume that for large blocking time $M \approx M^* \gg 1$, the averaged values $O_t^{(M)}$ are decorrelated and gaussian distributed

$$p(\overline{O}^{(M^*)}) = \left(\frac{P}{2\pi\sigma_{M^*}^2} \right)^{1/2} \exp\left[-\frac{(\overline{O}^{(M^*)} - \langle O \rangle)^2}{2\sigma_{M^*}^2/P} \right] \tag{47}$$

where $\sigma_{M^*}^2$ is the variance of $O^{(M^*)}$ and $P = T/M^*$ are the number of configurations we have. The error is therefore given by

$$\epsilon_T^{(M^*)} = \sqrt{M^* \sigma_{M*}^2 / T} \tag{48}$$

corresponding to the correlation time $\tau = M^* \sigma_{M*}^2 / \sigma_1^2$. Since the average of two gaussian distributed variables of variance σ^2 is also a gaussian of variance $\sigma^2/2$, we see that further averaging does not change the error any more. Therefore, blocking together M data point, we can determine the true error by increasing M until the apparent error $\sim \sigma_M \sqrt{M/T}$ does not increase anymore and remains stationary.