John von Neumann Institute for Computing

Monte Carlo Methods: Overview and Basics

Marius Lewerenz

published in

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes,
J. Grotendorst, D. Marx, A. Muramatsu (Eds.),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **10**, ISBN 3-00-009057-6, pp. 1-24, 2002.

http://www.fz-juelich.de/nic-series/volume10

Monte Carlo Methods: Overview and Basics

Marius Lewerenz

LADIR/Spectrochimie Moléculaire, UMR 7075
Bât F74, BP 49, Université Pierre et Marie Curie
4, Place Jussieu, F-75252 Paris Cedex 05
E-mail: lewerenz@spmol.jussieu.fr

The first part of this introductory lecture on Monte Carlo methods gives an overview of the essential ideas of Monte Carlo, discusses the relation between the basic sampling concept and statistical mechanics and the probabilistic interpretation of quantum mechanics, and provides a classification of existing Monte Carlo methods. This part is followed by a summary of essential concepts of probability and statistics, the construction of random walks, and the application of random sampling in the estimation of integrals.

1 Introduction

The original title of this lecture was supposed to be "Classical Monte Carlo", which is somewhat surprising for a winter school about quantum simulations. Since this is the first scientific lecture of this week I decided to interpret the title in a more general sense of (i) providing a sort of classification of the existing Monte Carlo methods, (ii) discussing certain algorithmic elements which were originally invented for classical Monte Carlo simulations but turn out to be useful also for quantum simulations, and last but not least (iii) giving some guidelines of how to handle the stochastic nature of the results. Many of these subjects will be revisited by other speakers in specialized lectures during this week.

The speakers in this workshop come from chemistry, physics, mathematics, and computer science and this is probably true about the audience as well. This fact illustrates the generality of both Monte Carlo and Molecular Dynamics methods and reflects the algorithmic challenges implicit in the simulation of ever larger systems on computers of increasing complexity. The requirements on the stability of trajectories over long simulation times or on the quality of random numbers are continuously increasing.

The section on elementary aspects of random variables and random walks is important because quantum chemists in the broad sense, ranging from electronic structure to reaction dynamics, are mostly not used to error bars in their data. Of course this does not imply the absence of errors in these deterministic calculations, but at least for certain computational techniques we know that we should observe monotonic convergence to a well defined result if a computational parameter (basis set size, grid resolution etc.) is changed. The presence of statistical noise in the result of a computation requires a different spirit in the assessment of the reliability of a calculation and a redefinition of the idea of convergence. The finite error margin on the result from any kind of Monte Carlo calculation can also make it hard to detect subtle errors in a code or deficiencies in the quality of the underlying source of randomness[1,2].

In production mode a proper Monte Carlo code will give a different result each time it is run. For test purposes, however, perfect reproducibility has to be possible as an option. A typical Monte Carlo code is far less complex than an electronic structure code but this

apparent simplicity should not lead to an underestimate of the possible troubles. We want to use random numbers to produce results, but we do not want these results to be random!

1.1 What is Monte Carlo ?

The keyword "Monte Carlo" in the name of a method only indicates that there is some stochastic element in it[3–8]. Stochastic methods are clearly appropriate for the description of stochastic processes but we will see that it can be useful to replace a deterministic problem by an equivalent stochastic problem. We will restrict ourselves here to stochastic methods of physical or chemical relevance.

The central idea of Monte Carlo methods is to represent the solution of a mathematical (or in our case physical) problem by a parameter of a true or hypothetical distribution and to estimate the value of this parameter by sampling from this distribution. This idea provides a direct tie to the ensemble concept introduced into statistical mechanics by Gibbs and the probabilistic interpretation of wave functions in quantum mechanics. Since we know how to translate a thermodynamic problem into its corresponding ensemble we have, at least in principle, a direct prescription for the computation of an observable, namely to construct a sufficiently large number of microstates compatible with the specified macroscopic variables and to average their properties. The distribution function for the microstates in generally known for this type of problem.

Once we know the wave function of a physical system we can imagine to compute expectation values by drawing samples from the associated probability density. This quantum Monte Carlo approach is very interesting because it provides a very practical interpretation of the abstract concept of a wave function. Unfortunately exact wave functions are known for only a very limited number of systems. However, as you will learn during this workshop, there are several Monte Carlo methods which achieve the difficult goal of simultaneously creating and sampling from the a priori unknown quantum distribution.

Monte Carlo methods play a very important role in statistical physics[9–14] and have led to a very high technical standard of algorithms and methods to quantify systematic and statistical errors. While classical Monte Carlo has a well established place in chemistry in particular in the (numerical) theory of liquids[15,16], the group of methods customarily referred to as quantum Monte Carlo is used by only a relatively small, but rapidly increasing number of people. Unfortunately the transfer of methodological knowledge between statistical physics, electronic structure applications of quantum Monte Carlo, and the rapidly growing group of people interested in dynamical applications is still relatively poor. Specifically in my own field of the application of quantum Monte Carlo techniques to vibrational problems a number of far from optimal algorithms are in widespread use. Hopefully this workshop can contribute to change this situation.

1.2 Monte Carlo vs Molecular Dynamics

Monte Carlo methods, which are usually based on random walks and thus sequences of events, contain an apparent dynamics. It is, however, crucial to recognize that the sequence of events does not correspond to a sequence in real time, as opposed to the role of time as an explicit variable in molecular dynamics methods. The sequence of events is generally independent of the physical phenomena which we wish to describe by a Monte

Carlo method and depends on the chosen algorithm. The creation of smart sequences is in fact an important design goal in the construction of efficient Monte Carlo algorithms.

The absence of time in Monte Carlo algorithms implies the fundamental inability to describe non equilibrium processes. We will, however, encounter the concept of imaginary time, which should rather be regarded as inverse energy, in several quantum Monte Carlo methods in the lectures of my colleagues J. Anderson and D. Ceperley.

Why then do we need Monte Carlo methods if the most common ensembles of statistical mechanics (microcanonical and canonical) are equally well accessible by molecular dynamics methods which give access to time dependent and time independent properties? Besides being easily able to describe situations, which are difficult (e.g. grand canonical ensemble) or impossible (e.g. discrete models, see F. Assaad, S. Sorella) to cast into equations of motion, well designed Monte Carlo methods offer more chances of overcoming one of the most severe problems of simulation methods, namely exhaustive sampling of the relevant configuration space. The local configuration updates imposed by the small time steps in the discretized version of the equations of motion (systematic discretisation error!) allow only slow exploration of configuration space by dynamical methods through a sequence of very many small steps. Our only chance for improvement is a faster computer and maybe a better code design.

Monte Carlo methods do not have the equivalent of a time step error (the target distribution is sampled exactly with only statistical errors) and they allow a great degree of freedom in the design of the sequence of steps, potentially including smart global configuration updates which permit rapid exploration of problems which may suffer from significant ergodicity problems in an equivalent dynamical simulation. Monte Carlo methods allow to replace computer power by creative brain power. There are also relatively straightforward applications of Monte Carlo methods to quantum systems, as you will see in the other lectures of this winter school.

The essential conclusion is that molecular dynamics and Monte Carlo methods do not compete but complement each other.

1.3 Classical vs Quantum Monte Carlo

The existing Monte Carlo methods can be divided into methods which assume that classical mechanics is applicable and that consequently energy is a continuous variable and those which are based on the idea of discrete quantum energy levels. While all of the latter methods are sometimes referred to as quantum Monte Carlo methods, I personally prefer to distinguish methods which assume a known set of energy levels (usually from a model hamiltonian) from those in which the Monte Carlo method is actually used to find these energy levels.

This difference becomes evident if we look at the other physically important distinction, namely the role of temperature. In the classical limit the investigation of the system at T=0 corresponds to finding the global minimum on a multidimensional potential surface, which is conceptually simple (but in practice can be formidable problem). At finite temperature, Monte Carlo methods are used to sample points in configuration space according to a known probability distribution for the available energies (e.g. the Boltzmann distribution). Averages over these points are used to estimate the expectation value of any property of interest which is formally defined as an integral in statistical mechanics. Here

	Classical	**Quantum**	
$T = 0$	Locating the minimum of a multidimensional surface, e.g. Simulated annealing	Single occupied quantum state of known energy	Single quantum state(s) with unknown properties: variational Monte Carlo (**VMC**), diffusion Monte Carlo (**DMC**)
$T > 0$	Classical Monte Carlo (**CMC**) in various ensembles (Integration over continuous states)	Summation over discrete states (lattice model Hamiltonians, Ising etc.), technically similar to CMC	Direct averaging over many quantum states: path integral Monte Carlo (**PIMC**)
	known energy levels E_i		unknown energy levels E_i

we have a typical example for the replacement of a deterministic problem (performing a high dimensional integral) by an equivalent stochastic problem.

The same basic techniques can be applied to model quantum systems, where the T=0 situation now corresponds to finding the combination of discrete variables (quantum numbers, spin orientations etc.) which give the lowest energy. This is again a conceptually simple problem, even though in practice the huge size of the available (now discrete) configuration space can still be very troublesome. The computation of thermodynamic averages at T > 0 for this type of system does not require any profound modifications of the algorithms of classical Monte Carlo. Instead of performing a random walk in continuous coordinates one attempts to flip discrete spins or to change quantum numbers (sometimes with very elaborate algorithms[17,18]) before evaluating the quantum energy of the system from a simple formula, which then in turn determines the continuation of the random walk. The Monte Carlo method is used here only to perform the average over available states, but not to actually compute the quantum states or their properties.

The situation becomes much more complicated if we do not even know the states which are available in a given quantum system. Searching for the properties of individual pure quantum states, independent of their occupation, eliminates temperature as a variable, which can formally be set to zero. The simplest of these states, and of course the only one occupied at T=0, is the quantum ground state. The goal of methods like variational Monte Carlo (VMC), Green's function Monte Carlo (GFMC)[19–21], and its most common variant diffusion Monte Carlo (DMC)[22–24] is the calculation of wave functions, energies, and other properties of pure states. Whereas the VMC method relies on exploring a proposed quantum distribution and its subsequent optimisation, the distribution from which

the samples should be taken is a priori unknown in GFMC and DMC and is constructed in the course of the Monte Carlo calculation.

Once a sufficiently complete set of eigenstates has been computed by one of these methods, thermodynamic averages at T>0 could in principle be computed by direct summation or application of sampling techniques as above. Unfortunately even the computation of a few states is still not easy, as we will see in this workshop.

However, if we are willing to sacrifice detailed knowledge of individual quantum states, direct sampling of the density matrix is possible by the numerical implementation of the path integral approach to quantum mechanics[25,26], which again replaces integrals by averages over samples and is known as path integral Monte Carlo (PIMC). This latter method can be mapped onto a classical problem[27] with specific modifications to account for quantum statistics[28].

This last group of methods merits the designation "quantum Monte Carlo" in the more precise sense of attempts to use Monte Carlo techniques to solve the actual quantum problem. You will hear more about the theory underlying these methods and their implementation and application to actual problems in other lectures of this course. Most of the efforts of advancing quantum Monte Carlo technology in molecular physics (VMC, GFMC, DMC) are oriented towards electronic structure[29,24,30–33] but the methods are equally applicable to rovibrational problems. While the fermion statistics of electrons causes particular trouble in the construction of stable algorithms for electronic structure problems, the rovibrational ground state of most molecular systems can be computed exactly with these methods.

1.4 Why do we need Quantum Monte Carlo ?

Quantum Monte Carlo methods are not here to replace other methods in electronic or vibrational structure, but are an interesting complement to these more conventional methods. The GFMC approach, and DMC as its most common implementation, in principle only needs a potential energy surface. While basis set methods have problems to provide enough flexibility to span large configuration spaces and usually require basis sets specially tailored to a given problem in order to remain technically feasible, DMC is an intrisically global method.

Due to its general applicability its value is probably even higher in vibrational problems where the different shapes of interaction potentials require a lot of care in basis function design as opposed to electronic structure, where the potential is always just a sum of Coulomb terms and where there is a well developed general technology based on Gaussian basis functions. The construction of the random walks allows the treatment of large amplitude motion and of wave functions which are spread over a large number of potential minima. Of course we have to pay a price for this: DMC will only give a single wave function and not a whole spectrum, and even this single wave function has an unusual representation and the extraction of unbiased expectation values is rather complicated.

Being an exact method except for statistical errors, one can easily follow the evolution of the quantum properties of systems of increasing complexity (e.g. cluster size effects) without biasing the results through the introduction of more and more approximations or more stringent basis limitations. This attractive scaling of the accuracy is accompanied by a relatively modest growth of the computational effort. The only serious, and often

underestimated, problem is the practical achievement of the full exploration of the relevant configuration space, in other words the ergodicity of a random walk of finite length.

Presenting this topic to an audience of young researchers is probably a good occasion to give a brief list of open problems which might be interesting to look into. On the NCSA web site http://archive.ncsa.uiuc.edu/Apps/CMP/topten/topten.html you can find the following list of reasons why quantum Monte Carlo methods are still not very popular in electronic structure theory.
Top reasons why quantum Monte Carlo is not generally used in chemistry:

- We need forces, dummy!
- Try getting O_2 to bind at the variational level.
- How many graduate students lives have been lost optimizing wavefunctions?
- It is hard to get 0.01 eV accuracy by throwing dice.
- Most chemical problems have more than 50 electrons.
- Who thought LDA or HF pseudopotentials would be any good?
- How many spectra have you seen computed by QMC?
- QMC is only exact for energies.
- Multiple determinants. We can't live with them, we can't live without them.
- After all, electrons are fermions.
- Electrons move.
- QMC isn't included in Gaussian 90. Who programs anyway?

From my own experience with rovibrational problems I might add:

- How to construct trial wave functions for arbitrary potentials?
- We want a lot of excited states.
- We want rigid body constraints.
- How to handle almost degenerate states?

2 Review of Probability and Statistics

This section introduces several basic notions which we will need to describe the data produced by a Monte Carlo calculation and to specify individual components in the construction of random walks. The presentation follows essentially the discussion given in Ref. 4.

2.1 Probabilities and Random Variables

We consider a reservoir of possible outcomes $\{E\}$ for a random event.

$$\{E\} = \{E_1, E_2, E_3 \dots E_n\} \tag{1}$$

We associate a probability p_k with each E_k:

$$P(E_k) = p_k \qquad 1 \geq p_k \geq 0 \tag{2}$$

Properties of p_k:

1. The following relations hold for any pair of E_i, E_j.

$$P(E_i \wedge E_j) \leq p_i + p_j \tag{3}$$

$$P(E_i \vee E_j) \leq p_i + p_j \tag{4}$$

2. If E_i and E_j are mutually exclusive
($E_i \Rightarrow \neg E_j, E_j \Rightarrow \neg E_i$):

$$P(E_i \wedge E_j) = 0 \tag{5}$$

$$P(E_i \vee E_j) = p_i + p_j \tag{6}$$

3. For a class of mutually exclusive events, which contains all possible events we have:

$$P(\mathit{some}\ E) = 1 = \sum_i p_i \tag{7}$$

2.2 Joint and Marginal Probabilities

Suppose that the events E_i and F_j satisfy the conditions defined above with associated probabilities p_{1i} and p_{2j},

$$P(E_i) = p_{1i} \qquad P(F_j) = p_{2j} \quad , \tag{8}$$

and we are interested in the probability of the combined event (E_i, F_j). We define the probability of this event as the *joint probability*

$$P(E_i, F_j) = p_{ij} \tag{9}$$

The events E_i and F_j are called independent if the probability of the combined event can be expressed as

$$p_{ij} = p_{1i} p_{2j} \quad . \tag{10}$$

If the events E_i and F_j are not independent, i. e. $p_{ij} \neq p_{1i} p_{2j}$, it is useful to decompose the *joint probability* as follows:

$$p_{ij} = \left(\sum_k p_{ik} \right) \left[\frac{p_{ij}}{\sum_k p_{ik}} \right] \tag{11}$$

$$p_{ij} = p(i) \left[\frac{p_{ij}}{\sum_k p_{ik}} \right] \tag{12}$$

The quantity $p(i)$ is called the *marginal probability* for the event E_i, the probability of observing E_i combined with any event in the reservoir $\{F\}$. Clearly $p(i) = p_{1i}$ and $\sum_i p(i) = \sum_i \sum_k p_{ik} = 1$.

The second factor $p_{ij}/\sum_k p_{ik} = p(j|i)$ defines the *conditional probability* of observing F_j, provided E_i has occurred. Since we are certain to observe one of the possible F_j in combination with E_i we clearly have

$$\sum_j p(j|i) = \sum_j \frac{p_{ij}}{\sum_k p_{ik}} = \frac{\sum_j p_{ij}}{\sum_k p_{ik}} = 1 \tag{13}$$

The concept of conditional probability is an important element in the construction of random walks, because certain events will be possible only if they are preceeded by particular other events.

2.3 Random Variables and Expectation Values

The random events discussed above E, F can be anything of numerical or non numerical character (e.g. a noise amplitude or a logical decision). If we can associate a numerical value x_i with each random event E_i, we call x a random variable.

We define the *expectation* value $E(x)$ of a random variable x as

$$E(x) = \langle x \rangle = \sum_i p_i x_i \tag{14}$$

Assume that g is a function of x, $g(x_i) = g_i$. Then also g_i will be a random variable and we define

$$E(g(x)) = \langle g(x) \rangle = \sum_i p_i g(x_i) \tag{15}$$

Suppose that $g(x_i) = g(x) = const$:

$$E(g(x)) = \sum_i p_i g(x_i) = g(x_i) \sum_i p_i = g(x) \tag{16}$$

We conclude that the expectation value of a constant is a constant.

In the next step we prove the linearity of the expectation value of two random functions $g_1(x)$ and $g_2(x)$ by substitution of the definition of an expectation value in terms of probabilities:

$$\begin{aligned} E(\lambda_1 g_1(x) + \lambda_2 g_2(x)) &= \langle \lambda_1 g_1(x) + \lambda_2 g_2(x) \rangle \\ &= \sum_i p_i \left(\lambda_1 g_1(x_i) + \lambda_2 g_2(x_i) \right) \\ &= \lambda_1 \sum_i p_i g_1(x_i) + \lambda_2 \sum_i p_i g_2(x_i) \\ &= \lambda_1 \langle g_1(x) \rangle + \lambda_2 \langle g_2(x) \rangle \end{aligned}$$

2.4 Moments of a Distribution

We define the nth moment of a distribution as

$$\mu_n = \langle x^n \rangle = \sum_i p_i x_i^n \tag{17}$$

These powers of x are nothing but special cases of the random functions $g(x)$.

Principal moments:

$$\mu_1 = \sum_i p_i x_i \quad \text{mean of the distribution}$$
$$\mu_2 = \sum_i p_i x_i^2$$

Central moments:

$$\begin{aligned} \langle m_n(x) \rangle &= \langle (x - \mu_1)^n \rangle \\ &= \sum_i p_i (x_i - \langle x \rangle)^n \end{aligned}$$

The special case of $n = 2$ is called the *variance*:

$$Var\{x\} = \langle m_2(x) \rangle = \langle x^2 \rangle - \langle x \rangle^2 \tag{18}$$

The variance is particularly important because $Var\{x\}$ and μ_1 are sufficient to uniquely specify the important Gaussian distribution which usually results from the superposition of a sufficiently large number of random events from arbitrary underlying distributions.

2.5 Variance of a Random Function

By extension of the definition of the variance of a random variable we can define the variance of a random function $g(x)$:

$$\begin{aligned} Var\{g(x)\} &= \langle (g(x) - \langle g(x) \rangle)^2 \\ &= \sum_i p_i g^2(x) - 2\langle g(x) \rangle \sum_i p_i g(x_i) \\ &+ \langle g(x) \rangle^2 \sum_i p_i \\ &= \langle g^2(x) \rangle - \langle g(x) \rangle^2 \end{aligned}$$

2.6 Variance of a Linear Combination of Random Functions

By insertion of the definition and linearity of expectation values we can find the result for a linear combination of random functions:

$$
\begin{aligned}
&Var\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} = \\
&= \langle(\lambda_1 g_1(x) + \lambda_2 g_2(x) - \langle\lambda_1 g_1(x) + \lambda_2 g_2(x)\rangle)^2\rangle \\
&= \langle(\lambda_1 g_1(x) + \lambda_2 g_2(x) - \lambda_1\langle g_1(x)\rangle - \lambda_2\langle g_2(x)\rangle)^2\rangle \\
&= \langle(\lambda_1[g_1(x) - \langle g_1(x)\rangle] + \lambda_2[g_2(x) - \langle g_2(x)\rangle])^2\rangle \\
&= \langle(\lambda_1^2[g_1(x) - \langle g_1(x)\rangle]^2 + \lambda_2^2[g_2(x) - \langle g_2(x)\rangle]^2 \\
&\qquad + 2\lambda_1\lambda_2[g_1(x) - \langle g_1\rangle][g_2(x) - \langle g_2\rangle]\rangle \\
&= \lambda_1^2\langle[g_1(x) - \langle g_1(x)\rangle]^2\rangle + \lambda_2^2\langle[g_2(x) - \langle g_2(x)\rangle]^2\rangle \\
&\qquad + 2\lambda_1\lambda_2\langle[g_1(x)g_2(x) - g_1(x)\langle g_2(x)\rangle \\
&\qquad - \langle g_1(x)\rangle g_2(x) + \langle g_1(x)\rangle\langle g_2(x)\rangle\rangle
\end{aligned} \tag{19}
$$

In short this result can be expressed through the variances of the random functions g_1 and g_2 and an extra term:

$$
\begin{aligned}
Var\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} = \lambda_1^2 Var\{g_1(x)\} \\
+ \lambda_2^2 Var\{g_2(x)\} \\
+ 2\lambda_1\lambda_2(\langle g_1(x)g_2(x)\rangle \\
- \langle g_1(x)\rangle\langle g_2(x)\rangle)
\end{aligned}
$$

2.7 The Covariance

The mixed term in the preceeding equation is called the *covariance* of $g_1(x)$ and $g_2(x)$.

$$
Cov\{g_1(x), g_2(x)\} = \langle g_1(x)g_2(x)\rangle - \langle g_1(x)\rangle\langle g_2(x)\rangle \tag{20}
$$

This term can be positive or negative and we will show in the next section that this term is related to the mutual dependence between the two random functions g_1 and g_2.

We have the following special cases for simple random variables x and y:

$$
\begin{aligned}
Cov\{x, y\} &= \langle xy\rangle - \langle x\rangle\langle y\rangle \\
Cov\{x, x\} &= \langle xx\rangle - \langle x\rangle\langle x\rangle = Var\{x\}
\end{aligned}
$$

Depending on the sign of the *covariance*, the variance of a linear combination of random functions or variables can be larger or smaller than the sum of the individual variances.

$$
Var\{g_1 + g_2\} = Var\{g_1\} + Var\{g_2\} + Cov\{g_1, g_2\} \tag{21}
$$

The possibility of negative covariance can be exploited in special sampling techniques (correlated sampling, antithetic variates) to achieve *variance reduction*.

$$
Var\{g_1 + g_2\} < Var\{g_1\} + Var\{g_2\} \tag{22}
$$

2.8 Properties of the Covariance

$$\begin{aligned} Cov\{x,y\} &= \langle xy\rangle - \langle x\rangle\langle y\rangle \\ \langle xy\rangle &= \sum_{ij} p_{ij}x_i y_j \end{aligned} \tag{23}$$

If the random variables x and y are *independent*, the p_{ij} can be decomposed according to

$$p_{ij} = p_{1i}p_{2j} \tag{24}$$

$$\begin{aligned} \langle xy\rangle &= \sum_{ij} p_{1i}x_i p_{2j}y_j \\ &= \left(\sum_i p_{1i}x_i\right)\left(\sum_j p_{2j}y_j\right) \\ &= \langle x\rangle\langle y\rangle \\ \Rightarrow Cov\{x,y\} &= 0 \end{aligned} \tag{25}$$

Independence of two random variables x, y is clearly a sufficient but not a necessary condition for $Cov\{x,y\}$ to be zero!

2.9 Correlation and Autocorrelation

The correlation coefficient $r(x,y)$ is the normalized version of the covariance:

$$r(x,y) = \frac{Cov\{x,y\}}{\sqrt{Var\{x\}Var\{y\}}} \tag{26}$$

$$-1 \le r(x,y) \le 1 \tag{27}$$

If one considers the values of y as copies of x with a constant offset δ (in time or some pseudotime establishing an order)

$$y_j = x_i = x_{j-\delta} \tag{28}$$

one can compute a correlation coefficient for each offset δ.

$$r(x,y;\delta) = A(x;\delta) \tag{29}$$

This function $A(x;\delta)$ is called the autocorrelation function and varies between -1 and +1 as a result of the normalisation by the variances of x and y.

The computation of the autocorrelation function is an important tool to assess the statistical independence of events within a sequence of random events. If $A(x;\delta) \neq 0$ we can be sure that there is some serial correlation between events separated by an offset δ. Conversely $A(x;\delta) = 0$ can occur if either the covariance is systematically zero due to statistical independence of the events or accidentally zero in spite of serial correlation.

All random walk methods require careful autocorrelation analysis.

2.10 Continuous Distributions

In the preceeding section we have assumed discrete random events to simplify the presentation, but generally random variables can also be continuous.

For a one-dimensional case we have

$$-\infty \leq x \leq \infty \tag{30}$$

We can define a *cumulative distribution function* $F(x)$ as

$$F(x) = P\{\text{randomly selected } y < x\} \tag{31}$$

Assume that $x_2 > x_1$. Then the events $x_2 > y \geq x_1$ and $x_1 > y$ are mutually exclusive and we conclude:

$$\begin{aligned} P\{x_2 > y \geq x_1\} + P\{x_1 > y\} &= P\{x_2 > y\} \\ P\{x_2 > y\} &\geq P\{x_1 > y\} \end{aligned}$$

It follows that $F(x)$ is a monotonically increasing function.

$$F(-\infty) = 0 \quad , \quad F(\infty) = 1 \tag{32}$$

The function $F(x)$ is not necessarily smooth. In differentiable regions one can define the *probability density function* $\rho(x)$:

$$\rho(x) = \frac{dF(x)}{dx} \geq 0 \tag{33}$$

2.11 Moments of Continuous Distributions

The concept of moments can be generalised to continuous distributions by the replacement of summations by integrations and of probabilities p_i by $dF(x)$.

$$E(x) = \langle x \rangle = \int_{-\infty}^{\infty} x dF(x) \quad \left(= \int_{-\infty}^{\infty} x\rho(x)dx \right) \tag{34}$$

$$\int_{-\infty}^{\infty} \rho(x)dx = F(\infty) = 1 \tag{35}$$

$$E(g(x)) = \langle g(x) \rangle = \int_{-\infty}^{\infty} g(x)dF(x) \tag{36}$$

The variance is now given as

$$\begin{aligned} Var\{x\} &= E(x^2) - E(x)^2 \\ &= \int_{-\infty}^{\infty} x^2\rho(x)dx - \left[\int_{-\infty}^{\infty} x\rho(x)dx \right]^2 \end{aligned}$$

It is important to note that the variance is not a well defined quantity for all probability densities $\rho(x)$. An well known example is the Cauchy-Lorentz-distribution with an arbitrary width parameter a

$$\rho(x) = \frac{1}{\pi} \frac{a}{x^2 + a^2} \tag{37}$$

for which $E(x) = 0$ and $E(x^2) = \infty$.

2.12 Sums of Random Variables

- Suppose we have random variables $x_1, x_2, \ldots, x_n$ which are distributed according to some probability density function $\rho(x)$. The variable x_i may represent a multidimensional point.
- We evaluate functions $g_i(x_i)$ for each x_i where the functions g_i may or may not be identical. The $g_i(x_i)$ are then random variables.
- We define a weighted sum G over these functions and its expectation value $E(G)$:

$$G = \sum_i^n \lambda_i g_i(x_i) \quad \lambda_i \in \mathbb{R} \tag{38}$$

$$E(G) = \langle G \rangle = \sum_i^n \lambda_i \langle g_i(x_i) \rangle \tag{39}$$

- A special choice is to use $\lambda_i = 1/n$ for all weights and to consider all g_i to be identical

$$E(G) = E\left(\frac{1}{n}\sum_i^n g(x_i)\right) = \frac{1}{n}\sum_i^n E(g) = E(g) \tag{40}$$

We find that the expectation value $E(G)$ for the sum G is identical with the expectation value $E(g)$ for the function. Consequently G can serve as an *estimator* for $E(g)$. This is in fact the basis of all Monte Carlo methods: Expectation values, which generally correspond to integrals, are approximated by a sum (here G) over values of the integrand (here $g(x)$) sampled at a finite set of points $x_i, i = 1 \ldots n$. We now have to establish a measure of the rate of convergence of the estimator G to the true expectation value if we increase n.

2.13 Variance of the Sum of Random Variables

- Assume for simplicity that all x_i are independent. In this case the covariance is zero for all combinations of random variables and the variance of G can be expressed simply as the sum of the variances of its terms:

$$Var\{G\} = \sum_i^n \lambda_i^2 Var\{g_i(x)\} \tag{41}$$

- Again assume identical weights and functions $\lambda_i = 1/n, g_i(x) = g(x)$

$$Var\{G\} = \sum_i^n \frac{1}{n^2} \underbrace{Var\{g(x)\}}_{\text{some number}} = \frac{1}{n} Var\{g(x)\} \tag{42}$$

The variance of the estimator G decreases in proportion to $1/n$ with a generally unknown proportionality factor. The determination of this factor would involve the computation of an integral over g, which is of the same complexity as the direct computation of the

expectation value $E(g)$ by integration. If we could compute this integral we would not resort to sampling and summation in the first place!

Statistical convergence:

The deviation of G from $E(g)$ will not decrease monotonically towards zero with increasing n as a result of the random nature of each new term entering the sum. The concept of convergence has to be accordingly redefined: The deviation δ of the estimator from the true value will exceed a specified limit Δ with a probability which diminishes as $n \to \infty$.

The non monotonic convergence of results is probably the single most irritating feature of Monte Carlo methods for newcomers, in particular for those who are used to other quantum mechanical methods using basis set expansions or grids.

3 Sources of Randomness

Monte Carlo methods require the creation of random events according to specified probability densities. There are three classes of sources:

- Natural sources of (true ?) randomness.
  Historically interesting but inefficient, not reproducible, and of hardly quantifiable quality.
- Deterministic algorithms producing a sequence of numbers with properties which are indistinguishable from a true random sequence as measured by a battery of statistical tests.
- Random walks constructed from primitive random events for all complicated multidimensional distributions.

We will look in detail only at **random walk methods** and assume the availability of a good uniform **random number generator**.

3.1 Random Number Generators

We will not attempt an exhaustive discussion of this subject, which is an active subject of research in number theory and cryptography. There is a very rich literature on the subject, including modern developments in the context of parallel computers and the problem of generating independent subsequences on many processors[34,35]. Many useful methods for the generation of random numbers with a variety of distributions are discussed in Refs. 3, 5, 8, 36. Artefacts and diagnostic tests for the quality of random numbers are discussed in Refs. 1, 37–39. A detailed discussion of random number generators is the subject of other special lectures in this series.

The key features of contemporary random number generators can be summarized as follows:

- No true random sequence due to the underlying deterministic algorithm, therefore more precisely called **pseudo random number generators, PRNG**.
- Construction on the basis of number theory and by extensive statistical testing.
- Usually generation of a sequence of numbers $0 < u \leq 1$ with uniform distribution, $F(u) = u$.
- Several common types: linear congruential generators (LCG), lagged Fibonacci generators (LFG), combined generators.
- Other distributions by transformation algorithms.
- High speed and quality, perfect reproducibility of the sequence for test purposes.
- Deterministic algorithm implies some subtle sequential correlation.
- **There is no publicly known PRNG without at least one reported case of failure.**
- In case of doubt, try a generator of different type and verify statistical consistency.
- New challenges posed by the arrival of parallel computers (Creation of uncorrelated parallel sequences).

3.2 Random Walks

The random walk method has been introduced into statistical physics in the hallmark paper of Metropolis et al.[40] and is enormously flexible for the creation of random events with any conceivable distribution.

The method allows to generate samples from specified probability density functions $\rho(x)$, in particular if the space x has a high dimensionality. In fact we need not even be able to compute the absolute value of $\rho(x)$ for a given x, which implies a normalisation which in itself involves a multidimensional integration. It is generally sufficient to compute ratios between values of $\rho(x)$ at points x_i and x_j. Random walks are a sequence of events $x_1, x_2, x_3, \ldots$, constructed such that the probability $P\{x_{new}\}$ of finding x_{new} is some function of $f(x_{new}, x_{last})$ of the previous events. The function $f(x_{new}, x_{last})$ describes a strategy to propagate the walk and corresponds in fact to a *conditional probability* $p(x_{new}|x_{last})$. This is the key difference to direct sampling methods where $P\{x_{new}\}$ is independent of the previous event x_{last}. The process has a memory and implies serial correlation.

Random walks are a special example for a Markov process.

The general conditions for random walks which are supposed to generate samples with distribution $\rho(x)$ can be summarized as follows

1. Every point x where $\rho(x) \neq 0$ must be accessible.
2. It must be possible to revisit the same point x any number of times.
3. The walk must not *periodically* pass through the same points x again.

These conditions are equivalent to requiring *ergodicity* of the random walk.

We should bear in mind that even if the fundamental construction principle of a random walk may ensure ergodicity, there is no guarantee that a random walk of finite length has explored all relevant parts of configuration space for a given physical problem. Insufficient length of random walks is the probably most common source of incorrect Monte Carlo results.

3.3 The Stochastic Matrix

Consider for a moment a system with discrete 'states' (position, orientations, quantum numbers etc.) $x_1, x_2, \ldots x_n$.

$p(x_j|x_i)$ is the conditional probability to observe x_j provided that we had x_i just before and is the transition probability for a Markov process. The ensemble of all $p(x_j|x_i)$ for all combinations of x_i and x_j can be arranged in matrix form, where we use the short notation $p_{ij} = p(x_j|x_i)$ for the transition from x_i to x_j.

$$\underline{\underline{P}} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & & & \vdots \\ \vdots & & & & \vdots \\ p_{n1} & \cdots & \cdots & \cdots & p_{nn} \end{pmatrix} \tag{43}$$

The matrix P is a *stochastic matrix*

All $p_{ij} \geq 0$ because they represent probabilities and $\sum_j p_{ij} = 1$ for all i because each transition from i must lead to one of the available 'states'.

3.4 Properties of the Stochastic Matrix

Consider a row vector

$$\underline{\rho^{(0)}} = \{\rho_1^{(0)}, \rho_2^{(0)}, \ldots, \rho_n^{(0)}\} \tag{44}$$

which describes an initial state in which $\rho_i^{(0)}$ is the probability of initially finding the system in state i.

Each step in the Markov chain can be formulated as a multiplication of this row vector with the stochastic matrix P:

$$\begin{aligned} \underline{\rho^{(1)}} &= \underline{\rho^{(0)}}\underline{\underline{P}} \\ \underline{\rho^{(2)}} &= \underline{\rho^{(1)}}\underline{\underline{P}} \\ \cdots &\quad \cdots \\ \underline{\rho^{(k)}} &= \underline{\rho^{(0)}}\underline{\underline{P}}^k \end{aligned}$$

The asymptotic distribution $\underline{\rho}$ for $k \to \infty$ is

$$\underline{\rho} = \lim_{k\to\infty} \underline{\rho}^{(0)}\underline{\underline{P}}^k \tag{45}$$

Repeated multiplication of $\rho^{(0)}$ with P converges to a stationary situation if

$$\underline{\rho} = \underline{\rho}\underline{\underline{P}} \tag{46}$$

which implies that $\underline{\rho}$ is an eigenvector of P with eigenvalue 1. Here we already see that $\underline{\rho}$ can be multiplied by any scalar without changing the result. The asymptotic distribution is independent of the initial 'state' $\rho^{(0)}$ and depends exclusively on the matrix P. All initial vectors (except for other eigenvectors of $\underline{\underline{P}}$) converge to the same asymptotic distribution.

Matrices generally have a spectrum of eigenvalues. It is relatively easy to show that repeated multiplication of a vector with a matrix will project out the vector belonging to the eigenvalue with the largest modulus. To complete the proof of the validity of the random walk construction we have to prove that 1 is the largest possible eigenvalue of a stochastic matrix.

3.5 Detailed Balance

In the typical Monte Carlo situation we know which probability density $\underline{\rho}$ we want to sample, while the preceeding argument will generate a probability density according to a given stochastic matrix. The key question is how to construct a matrix $\underline{\underline{P}}$ which has an eigenvector corresponding to the desired probability density ρ.

The eigenvector equation can be written out explicitly as

$$\sum_{i}^{n} \rho_i p_{ij} = \rho_j \tag{47}$$

If we now impose detailed balance according to

$$\rho_i p_{ij} = \rho_j p_{ji} \tag{48}$$

we obtain the result

$$\sum_i \rho_i p_{ij} = \sum_i \rho_j p_{ji} = \rho_j \sum_i p_{ji} = \rho_j \tag{49}$$

This is exactly the condition required for an eigenvector of P with eigenvalue 1.

Detailed balance guarantees $\underline{\rho}\underline{\underline{P}} = \underline{\rho}$ and is therefore a sufficient condition to construct a matrix $\underline{\underline{P}}$ with the desired asymptotic distribution but it is not necessarily the only possible way!

3.6 Decomposition of the Transition Process

We can arbitrarily decompose each p_{ij} into a factor describing the *probability of proposing* a particular transition t_{ij} and a factor a_{ij} describing the *probability of accepting* this choice.

$$p_{ij} = t_{ij} a_{ij} \tag{50}$$

This decomposition is valid if the two processes are successive and independent.

Substitution into the detailed balance relationship yields

$$\frac{\rho_j}{\rho_i} = \frac{p_{ij}}{p_{ji}} = \frac{t_{ij}a_{ij}}{t_{ji}a_{ji}} \tag{51}$$

Since we assume that ρ is a known probability density to be generated by the walk, and since we can pick a transition strategy specifying t_{ij} according to our taste, it is useful to convert this expression into a form which defines the required acceptance probability:

$$\frac{a_{ij}}{a_{ji}} = \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}} \tag{52}$$

Note that the construction of the random walk requires only that we are able to compute the *ratio* of probability densities. We can consequently work with densities ρ which are not normalized!

3.7 Accepting Proposed Transitions

There are two conventional choices for the acceptance probabilities a_{ij}, a_{ji} which satisfy this relation:

1. Metropolis (1953):

$$a_{ij} = \min\left[1, \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}}\right] \tag{53}$$

Proof by verification of the two possibilities:

(a) $\rho_j t_{ji} \geq \rho_i t_{ij} \Rightarrow$ $a_{ij} = 1$, $a_{ji} = \frac{\rho_i}{\rho_j}\frac{t_{ij}}{t_{ji}}$, $\frac{a_{ij}}{a_{ji}} = \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}}$ $q.e.d.$

(b) $\rho_j t_{ji} < \rho_i t_{ij} \Rightarrow$ $a_{ij} = \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}}$, $a_{ji} = 1$, $\frac{a_{ij}}{a_{ji}} = \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}}$ $q.e.d.$

2. Glauber

$$a_{ij} = \frac{1}{1 + \frac{\rho_i}{\rho_j}\frac{t_{ij}}{t_{ji}}} \tag{54}$$

Verification by substitution:

$$\frac{a_{ij}}{a_{ji}} = \frac{1 + \frac{\rho_j}{\rho_i}\frac{t_{ji}}{t_{ij}}}{1 + \frac{\rho_i}{\rho_j}\frac{t_{ij}}{t_{ji}}} \tag{55}$$

$$\frac{a_{ij}}{a_{ji}} = \frac{(\rho_i t_{ij} + \rho_j t_{ji})\rho_j t_{ji}}{(\rho_j t_{ji} + \rho_i t_{ij})\rho_i t_{ij}} \tag{56}$$

$$\frac{a_{ij}}{a_{ji}} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \quad q.e.d. \tag{57}$$

3.8 Random Walks in Continuous Spaces

The fundamental ideas of the construction of random walks via the stochastic matrix and a detailed balance ansatz can be generalized to an infinite number of states.

$$a(x \to x') = \min\left[1, \frac{\rho(x')t(x' \to x)}{\rho(x)t(x \to x')}\right] \tag{58}$$

3.9 Coordinates of a Random Walk

The coordinates or 'states' of random walks are defined very broadly and can be

- All discrete (e.g. spins on a lattice)
- All continuous (e.g. atomic coordinates in simulations of liquids)
- Any mixture of the two (e.g. particles with spin and continuous space coordinates)

Specifically also the particle number in physical simulations can be a (discrete) coordinate (Grand canonical Monte Carlo).

3.10 The Transition Function t

A common simplification consists in the assumption

$$t(x' \to x) = t(x \to x') \tag{59}$$

and specifically the choice $t(x \to x') = t(|x - x'|)$.

The formula for the Metropolis acceptance probability is then simply

$$a(x \to x') = \min\left[1, \frac{\rho(x')}{\rho(x)}\right] \tag{60}$$

Examples:

- Proposing a new position x' with uniform probability from a volume surrounding x (e.g. a hypercube of predefined size in the majority of simple random walk methods).
- Picking a new position x' according to a multidimensional Gaussian centered on x:

$$t(x \to x') \propto \exp\left(-\frac{|x - x'|^2}{2\sigma^2}\right) \tag{61}$$

Explicit inclusion of $t(x \to x')$ in the formulation allows *guided random walks*. Guided random walks play a major role many smart sampling techniques like Force bias Monte Carlo, J-walking, improved variational quantum Monte Carlo, and diffusion Monte Carlo with importance sampling. The construction of the transition strategy is very often closely tied to the nature of the problem and no general guiding rules can be given here.

A good strategy tends to propose moves which strike a good compromise between large displacement in the underlying coordinate space to minimize the serial correlations and a good acceptance ratio. In diffusion quantum Monte Carlo the transition strategy is given once a trial wave function has been chosen, but in many other cases there is considerable space for human imagination, making this one of the key parts of Monte Carlo algorithms where brain power can beat supercomputer power.

3.11 How to Accept a Proposed Change of State?

The translation of the algebraic relationship for the probability a of accepting a move into a practical algorithm is in fact very simple. The cumulative distribution function $F(x)$ for The probability $P\{u \leq x\}$ that a random number u distributed in the interval $[0, 1]$ with a uniform probability density $\rho(u) = 1$ is less then or equal x ($x \leq 1$) is given by $P\{u \leq x\} = \int_0^x \rho(u)du = \int_0^x du = x$. Consequently $u \leq a$ will be true with probability a and we can accept the proposed move whenever a uniform random number u satisfies $u \leq a$. It is worthwhile to note here that the quality of the uniform random number generator for u has to be very high.

3.12 Summary of Important Random Walk Features

Good features:

- Random sampling from distributions in spaces of high dimension.
- No need to be able to normalize the probability density function (which would involve a multidimensional integration).
- Very general coordinate definition and very broad applicability.

Troublesome Features:

- The desired distribution is reached only asymptotically.
  When is a random walk in its asymptotic regime?
- Serial correlation between sampling positions.
  Requires careful autocorrelation analysis.

4 Monte Carlo Integration

4.1 The Principle of Monte Carlo Integration

We have seen that the expectation value of the sum of random variables $g(x_i)$ with x_i drawn according to the probability density $\rho(x)$ is identical with the expectation value of $g(x)$ over the underlying distribution:

$$E(G) = E(g(x)) \tag{62}$$

$$G = \frac{1}{n}\sum_{i}^{n} g(x_i) \quad , \quad x_i \propto \rho(x) \tag{63}$$

Now recall the original definition of the expectation value as an integral over the distribution:

$$E(g(x)) = \int \rho(x)g(x)dx = \lim_{n\to\infty} \frac{1}{n}\sum_{i}^{n} g(x_i) \tag{64}$$

The following conclusions can be drawn from this formula:

- The statistical error for the integral is independent of the dimensionality of the integral.
- The statistical error diminishes proportional to $1/\sqrt{n}$.
- The proportionality constant controlling the absolute size of the error bar depends on the variance of g.
- The integrand need not even be smooth.

The most important observation is the first one, which indicates that Monte Carlo integration can be interesting for multidimensional integrals. While all true random sampling techniques will lead to an error bar on the result which diminishes in proportion to $n^{-\frac{1}{2}}$ for n samples, there are in fact special methods based on socalled quasirandom numbers which achieve a somewhat faster rate of statistical convergence for problems with a moderate number of dimensions.

What we can influence by smart sampling techniques is the prefactor of $n^{-\frac{1}{2}}$ since it depends on the variance of the function being sampled. If parts of the function can be integrated analytically or factored out as a density accessible for direct sampling we may end up with a residual function to be sampled by Monte Carlo which has a smaller variance than the original one. These *variance reduction* techniques allow significant improvements of the statistical errors but do not affect the $n^{-\frac{1}{2}}$ convergence rate.

The last point is also of certain interest in particular in comparison to highly accurate grid based quadrature rules. The latter (e.g. Gaussian quadratures) are very powerful, but require the existence of high order derivatives of the integrand. Functions with discontinuities can be very troublesome to integrate unless one can subdivide the integration domain according to the location of the discontinuities. Monte Carlo is very robust and works for 'spiky' integrands. So even for low dimensional cases where Monte Carlo is not the most efficient method it may be an interesting way to produce a crude estimate due to its simplicity.

We should point out, however, that Monte Carlo integration of oscillatory function is notoriously difficult, as the computation of any integral which is small compared to the variance of the integrand, unless we can devise a trick to absorb the oscillations.

4.2 Efficiency of Monte Carlo Integration

In order to get an idea about the efficiency of Monte Carlo integration and to derive a criterion for the number of dimensions where Monte Carlo becomes interesting we make the following assumptions

- $g(x)$ is a d-dimensional function.
- We have fixed resources which allow the evaluation of $g(x)$ at a fixed number of points n.
- We dispose of a grid based product integration rule with leading error h^q in the grid step h (e.g. Simpson $q = 5$).
- In a balanced treatment of all dimensions the grid step size will then be $h \propto n^{-1/d}$.

Efficiency analysis:

- The product rule quadrature will have overall accuracy

$$\varepsilon_{Grid} \propto n^{-q/d} \tag{65}$$

- The Monte Carlo quadrature will have overall accuracy

$$\varepsilon_{MC} \propto n^{-1/2} \tag{66}$$

- Monte Carlo will be more efficient if $d > 2q$.

References

1. A. M. Ferrenberg, D. P. Landau, and Y. J. Wong. Monte Carlo simulations: Hidden errors from "good" random number generators. *Phys. Rev. Lett.*, 69:3382, 1992.
2. I. Vattulainen and T. Ala-Nissila. Mission impossible: Find a random pseudorandom number generator. *Comput. in Phys.*, 9:500, 1995.
3. R. V. Rubinstein. *Simulation and the Monte Carlo Method*. J. Wiley & Sons, New York, 1981.
4. M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods, Vol. 1*. J. Wiley & Sons, New York, 1986.
5. P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation, 2nd ed.* Springer, Berlin, New York, 1987.
6. C. W. Gardiner. *Handbook of Stochastic Methods, 2nd ed.* Springer, Berlin, New York, 1990.
7. I. M. Sobol. *Die Monte Carlo Methode, 4. Aufl.* Deutscher Verlag der Wissenschaften, Berlin, 1991.
8. G. S. Fishman. *Monte Carlo, Concepts, Algorithms, and Applications*. Springer, New York, 1996.
9. K. Binder, editor. *Monte Carlo Methods in Statistical Physics, 2nd ed. (Topics in Current Physics Vol. 7)*. Springer, Berlin, New York, 1986.
10. K. Binder, editor. *Application of the Monte Carlo Method in Statistical Physics, 2nd ed. (Topics in Current Physics Vol. 36)*. Springer, Berlin, New York, 1987.
11. D. W. Heermann. *Computer Simulation Methods in Theoretical Physics, 2nd ed.* Springer, Berlin, 1990.
12. K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics, 2nd ed. (Springer Series in Solid-State Sciences Vol. 80)*. Springer, Berlin, New York, 1992.

13. K. Binder, editor. *The Monte Carlo Method in Condensed Matter Physics, (Topics in Applied Physics Vol. 71)*. Springer, Berlin, New York, 1992.
14. M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, Oxford, 1999.
15. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1987.
16. D. Frenkel and B. Smit. *Understanding Molecular Simulation; From Algorithms to Applications*. Academic Press, New York, 1996.
17. R. H. Swendsen and J.-S. Wang. *Phys. Rev. Lett.*, 58:81, 1987.
18. U. Wolff. *Phys. Rev. Lett.*, 62:361, 1989.
19. M. H. Kalos. Monte Carlo calculations of the ground state of three- and four-body nuclei. *Phys. Rev.*, 128:1891, 1962.
20. M. H. Kalos. Energy of a boson fluid with Lennard-Jones potentials. *Phys. Rev. A*, 2:250, 1970.
21. M. A. Lee and K. E. Schmidt. Green's function Monte Carlo. *Comput. in Phys.*, 6:192, 1992.
22. J. B. Anderson. A random walk simulation of the Schrödinger equation: H_3^+. *J. Chem. Phys.*, 63:1499, 1975.
23. P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr. Fixed-node quantum Monte Carlo for molecules. *J. Chem. Phys.*, 77:5593, 1982.
24. B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds. *Monte Carlo Methods in ab initio Quantum Chemistry*. World Scientific, Singapore, 1994.
25. R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367, 1948.
26. R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw Hill, New York, 1965.
27. D. Chandler and P. G. Wolynes. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.*, 74:4078, 1981.
28. D. M. Ceperley. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279, 1995.
29. W. A. Lester, Jr. and B. L. Hammond. *Annu. Rev. Phys. Chem.*, 41:283, 1990.
30. J. B. Anderson. Exact quantum chemistry by Monte Carlo methods. In S. R. Langhoff, editor, *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, page 1. Kluwer, Dordrecht, 1994.
31. J. B. Anderson. Fixed-node quantum Monte Carlo. *Int. Rev. Phys. Chem.*, 14:85, 1995.
32. D. M. Ceperley and L. Mitas. Quantum Monte Carlo methods in chemistry. *Adv. Chem. Phys.*, 93:1, 1996.
33. J. B. Anderson. Quantum Monte Carlo: Atoms, molecules, clusters, liquids, and solids. *Rev. Comp. Chem.*, 13:133, 1999.
34. M. Mascagni, S. A. Cuccaro, D. V. Pryor, and M. L. Robinson. A fast, high quality, and reproducible parallel lagged-Fibonacci pseudorandom number generator. *J. Comp. Phys.*, 119:211, 1995.
35. M. Mascagni. Parallel linear congruential generators with prime moduli. *Parallel Comput.*, 24:923–936, 1998.

36. R. M. Ziff. Four-tap shift-register-sequence random-number generators. *Comput. in Phys.*, 12:385, 1998.
37. P. Grassberger. On correlations in 'good' random number generators. *Phys. Lett. A*, 181:43, 1993.
38. I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical tests for random numbers in simulations. *Phys. Rev. Lett.*, 73:2513, 1994.
39. I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical models as tests of randomness. *Phys. Rev. E*, 52:3205, 1995.
40. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953.