Monte Carlo Methods in Quantum Many-Body Theories

Rafael Guardiola

Departamento de Física Atómica y Nuclear, Universidad de Valencia
Avda. Dr. Moliner, 50. 46100 - BURJASSOT (Valencia) Spain
e-mail: Rafael.Guardiola@uv.es

Abstract. This is an introduction of Monte Carlo methods for beginners and their application to some quantum many-body problems. Special emphasis is done on the methodology and the general characteristics of Monte Carlo calculations. An introduction to the applications to many-body physics, specifically the Variational Monte Carlo and the Green Function Monte Carlo, is also included.

1 Basic Concepts of Statistics

1.1 Random Variables

We are going to deal extensively with *random variables*, so it is necessary a precise definition. Usually a random variable is related to the result or outcome of a given physical action of a special nature, such that we have not a precise control on all the conditions which make unique the result of the referred action.

The typical example is the result of throwing a die. If we place ourselves in a purely (like Laplace) deterministic point of view, we could imagine being able of *exactly determining* the motion of the die and, in consequence, the result of the throwing once the initial conditions are fixed (such as height, angle, velocity, and so on). It is, however, simpler to imagine our incapacity and, in consequence, it is commonly stated that the result of throwing a die is *unexpected*, or *random* or some other synonymous. This incapacity does not mean that we have absolute ignorance about the outcome, and the partial knowledge of the action is related to the so-called *probability distribution function*, which will be abbreviated from now on as *pdf*. Insisting again with our example, if it is assumed that the die has a perfect regular form, and is perfectly homogeneous, and other additional requirements, the outcome of the action *throwing a die* is a number of the set $\{1, 2, 3, 4, 5, 6\}$, each result having a *probability* of value 1/6.

It will not escape to the reader that the above description is somehow empty and meaningless. The set of conditions of the die and its throwing are equivalent to the word *equiprobable*, and the full statement may be recast in the obviously fallacious form *if the set of results of throwing a die are equiprobable, then each result has a probability 1/6.*

When we toss two dice simultaneously, then the possible result for the sum of values is an integer between 2 and 12. A simple combinatorial calculation shows that there is a different number of ways of obtaining a different result: to get two points, both dice must show one point, but to get four points there are three different ways ([1,3], [2,2] and [3,1]). So, to this new physical action we must assign a random variable with values between 2 and 12, but with different probabilities for each of the results.

There is a double information to be supplied in order to define a random variable: the *domain* and the *probability distribution function*. The domain will represent the set of the accessible values, and the pdf the probability or frequency with which these values appear.

Example 1.1 *The random variable tossing two dice is characterized by a domain* $[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]$ *and corresponding probabilities*

$$[1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36] .$$

There are some conditions to be taken in consideration with respect to the definition of a random variable. The domain must contain all possible values that the random variable may take. Moreover, the probability related to a given value must be positive or null and, finally, the sum of all probabilities must be unity, indicating certitude.

In addition to variables which take only a discrete set of values, one may define random variables in the continuum. However, it is no longer possible to talk of probability of a given value, and instead one must limit oneself to talk of probability that the result is contained in the interval $[x, x + \Delta x]$. The relevant concept when dealing with continuous variables is then the *density of probability distribution function*, which will also be abbreviated by *pdf*, and represents the limit of the quotient of the probability of having a value in the interval $[x, x + \Delta x]$ by Δx, when $\Delta x \to 0$.

The most basic concepts related with random variables are shown in Table 1, both for discrete and for continuous variables.

In Table 1 a new concept has been introduced, namely the *cumulative distribution function*, P_i or $F(X)$, which represents the probability that the random variable takes a value less than or equal to x_i or X, for the cases of discrete and continuous distributions, respectively.

The previous equations regarding pdf's have been explicitly written for a single random variable. These distributions are termed *univariate*. A logical extension is to consider *multivariate* pdf, just by replacing the random variable x by the random vector $\mathbf{x}$, as well as other related quantities. A multivariate pdf is said to define *uncorrelated* random variables if it may be factorized in the form

$$f(\mathbf{x}) \equiv f(x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i) .$$

Table 1. Basic concepts related to random variables

	Discrete	Continuous
Domain	$\{x_1, x_2, \ldots, x_N\}$	$[a, b]$
pdf	$p_i = \text{Prob}(X = x_i)$	$f(x)dx = \text{Prob}(x \leq X \leq x + dx)$
Cumulative	$P_i = \text{Prob}(X \leq x_i)$	$F(X) = \text{Prob}(x \leq X)$
function	$P_i = \sum_{j=1}^{i} p_j$	$F(X) = \int_a^X f(x)dx$
Positivity	$0 \leq p_i \leq 1$	$f(x) \geq 0$
Bounds	$0 \leq P_i \leq 1$	$0 \leq F(X) \leq 1$
Monotonicity	$P_i \geq P_j$ if $i > j$	$F(X) \geq F(Y)$ if $X > Y$
Normalization	$P_N = \sum_{j=1}^{N} p_j = 1$	$F(b) \equiv \int_a^b f(x)dx = 1$

In the following we are going to deal almost exclusively with continuous random variables. In most of the cases, the new concepts to be introduced or properties to be discussed may be easily adapted for discrete variables from the corresponding continuous variables cases.

Just to close this section, it is worth mentioning that many of the concepts of mathematical statistics have invaded the world of physics. The assumed interpretation of quantum mechanics (the Copenhagen interpretation) assigns a statistical character to the results of any quantum experiment. The Boltzmann and Gibbs statistical mechanics is based in a statistical description of systems of many components. Even the interest in the study of chaotic phenomena has its roots in the analysis by Poincaré of the uncertainties related to random events. The reader interested in the historical evolution of the concepts of probability and statistics will find an interesting description in the book of James Newman, *The World of Mathematics* [1].

We are aware of the fact that our presentation on statistics is too short and condensed. Classical texts in mathematical statistics have been written by Feller [2] and Gnedenko [3]. The introductory part of the book by Hammersley and Handscomb [4] is both concise and precise. The book of Malvin Kalos and Paula Whitlock [5] is an excellent reference for most of the content of this paper.

1.2 Examples of Distribution Functions

The Uniform Distribution $U(a, b)$ corresponds to a domain $[a, b]$ and the related pdf is a constant, with value $1/(b - a)$. By using the step function θ it may be described by a single equation

$$f(x) = \frac{1}{b - a} \theta(x - a)\theta(b - x) . \tag{1.1}$$

The cumulative function is

$$F(x) = \frac{x-a}{b-a}\theta(x-a)\theta(b-x) + \theta(x-b) \ . \tag{1.2}$$

The standard form of the uniform distribution is $U(0,1)$, i.e., $f(x) = \theta(x)\theta(1-x)$. This distribution function is of primary interest in Monte Carlo methods, being the basic tool to generate random deviates from the required distributions.

The Normal or Gaussian Distribution $N(\mu, \sigma)$ characterizes a pdf of gaussian shape, defined in the interval $[-\infty, \infty]$, and given by

$$f(x) = \frac{-1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ . \tag{1.3}$$

The special form of writing this distribution is because the parameters μ and σ have a specific meaning, as the mean and standard deviation of the distribution, to be defined later. The standard form of the normal distribution is $N(0,1)$, i.e., with null mean and variance equal to one. The cumulative distribution function has not a simple formula, but is related to the familiar complementary error function.

This distribution is particularly interesting in relation with the central limit theorem, to be described later.

The Exponential Distribution is given by

$$f(x) = \gamma\exp[-\gamma x] \ , \tag{1.4}$$

and defined in the domain $[0, \infty]$. This distribution is of particular relevance in physics, being related with the radioactive decay and with the collisions of particles moving in material media.

1.3 Expectation Values

Let $f(x)$ be a pdf with domain $[a, b]$, and $h(x)$ a function defined in the same domain. The *expectation value* of $h(x)$ *with respect to* $f(x)$ is defined by the integral

$$\langle h \rangle = \int_a^b f(x)h(x)dx \ . \tag{1.5}$$

In quantum mechanics expectation values of operators with respect to a given quantum state are defined in a similar way. In coordinate representation one has

$$\langle \Psi|H|\Psi \rangle = \frac{\int_a^b \Psi^*(x)H(x)\Psi(x)dx}{\int_a^b \Psi^*(x)\Psi(x)dx} \ . \tag{1.6}$$

For normalized states the denominator may be removed (it is the unity), and for *local operators*, i.e., operators not involving derivatives, the quantum mechanical expectation value is converted into

$$\langle \Psi|H|\Psi \rangle = \int_a^b \left\{ \frac{|\Psi(x)|^2}{\int_a^b |\Psi(x)|^2 dx} \right\} H(x)dx \ . \tag{1.7}$$

In the case of non-local operators (e.g. operators involving derivatives) one may define an auxiliary quantity

$$\mathcal{H}(x) = \frac{1}{\Psi(x)}H\Psi(x) \tag{1.8}$$

and express the quantum mechanical expectation value in the same form as in Eq. (1.7), but putting $\mathcal{H}(x)$ instead of $H(x)$. When the operator H is the Hamiltonian, the quantity $\mathcal{H}(x)$ is known as the *local energy*. The local energy has a very important property: when the wave function corresponds to an eigenstate of the operator H, its value is a constant, and its expectation value is the same constant. If the wave function is close to an eigenstate, one may expect the local energy to have smooth fluctuations around a constant value.

Moments of a pdf. There is a special class of expectation values, known with the name of *moments* of the pdf, and defined as follows

Mean is the expectation value of $h(x) = x$, and is normally represented by the letter μ,

$$\mu = \int_a^b xf(x)dx \ . \tag{1.9}$$

Moments of a pdf are defined as the expectation value of x^n, and usually are represented by μ'_n,

$$\mu'_n = \langle x^n \rangle \ .$$

Centered Moments are usually represented by μ_n, $n = 0, 1, 2, \ldots$, and correspond to the expectation value of the function $h(x) = (x-\mu)^n$. The moment μ_0 is the unity and the moment μ_1 is null. For low values of n the centered moments have names, but we will focus on the special case μ_2.

Variance and Standard Deviation. The centered moment μ_2 is known as the *variance* of a distribution function, and its square root as the *standard deviation*, being usually represented by the letter σ,

$$\sigma = \sqrt{\langle (x-\mu)^2 \rangle} \ . \tag{1.10}$$

A practical way of determining the variance of a pdf is given by the equation

$$\sigma^2 = \langle x^2 \rangle - \mu^2 \ . \tag{1.11}$$

Example 1.2 *The moments of the normal distribution $N(\mu, \sigma)$ have the values μ (mean) and σ (standard deviation).*

For a given pdf it may well happen that the moments have infinite values. One of these cases corresponds to the so called Cauchy pdf, given by

$$f(x) = \frac{1}{1+x^2}$$

and defined in $[-\infty, \infty]$. We will always refer to gentle pdf, with a well defined (i.e., not infinite) set of moments. The Fourier transform of these regular distributions is then expressed in terms of the moments by the equation

$$f(k) \equiv \frac{1}{\sqrt{2\pi}} \int_a^b e^{ikx} f(x) dx = \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{i^n}{n!} \mu_n' k^n \ .$$

For a general function, i.e., not a power of the random variable, we have defined above the analogous of the moment μ_1', under the name of expectation value $\langle g(x) \rangle$. It is also convenient to define the analogous to the variance for a general expectation value, as

$$s^2(g) = \langle (g(x) - \langle g \rangle)^2 \rangle \equiv \int_a^b f(x) g^2(x) dx - \left(\int_a^b f(x) g(x) dx \right)^2$$

1.4 Distribution of Mean Values

In many applications of statistics pdf's are characterized by giving just the values of the lowest moments μ and σ of the distribution. The quantum Monte Carlo theory is not an exception to this rule. In practice this means that instead of studying the full distribution, one limits himself to a less complete information. Certainly, this is almost necessary when dealing with multivariate distributions or with expectation values of functions of many variables. For example, instead of giving the distribution of energies of protons emerging from an accelerator, one may just give the values of the mean energy and the variance. There is a connection between these two values and the full pdf, and to get this connection it is convenient to study the distribution of averages of a given distribution function.

From an univariate pdf $f(x)$ one may define another univariate pdf related to the mean of N random values of f,

$$z = \frac{x_1 + x_2 + \cdots + x_N}{N} \ ,$$

and the question is which is the pdf of the new random variable z.

An operative description will clarify the meaning of the variable z and the posed question. Imagine we toss a die five times in succession, and quote as the result the sum of points divided by five. This is repeated many times, and we want to know the frequency distribution of the mentioned averages. It is important to note that each of the five values of this combined event is related to the same pdf (in the example, the probability of each outcome is 1/6), and that successive tossings are *independent*.

The probability of obtaining an average value z will then be the product of probabilities of obtaining arbitrary individual values $x_1, x_2, \ldots, x_N$, but with the constraint that their average is z. For continuous pfd this statement corresponds to

$$g(z) = \int dx_1 f(x_1) \ldots \int dx_N f(x_N) \delta(z - \frac{x_1 + x_2 + \cdots + x_N}{N}) \ , \quad (1.12)$$

where $g(z)$ is the searched distribution.

The Central Limit Theorem. In general it is not possible to get a closed result for Eq. (1.12) when considering an arbitrary value of n and a general starting distribution $f(x)$. It is possible, however, to get an approximate result for $N \to \infty$, and this limit result is the so called *central limit theorem*. A formal proof may be obtained as follows. Write down the integral representation of the delta function,

$$g(z) = \frac{1}{2\pi} \int dq \int dx_1 f(x_1) \ldots \int dx_N f(x_N) \exp[iq(z - \frac{x_1 + x_2 + \cdots + x_N}{N})]$$

and rearrange terms to obtain

$$g(z) = \frac{1}{2\pi} \int dq \left[\int dx f(x) \exp[\frac{iq(z-x)}{N}] \right]^N$$

Afterwards, add and subtract μ in the exponential, where μ is the mean of the pdf $f(x)$. The resulting integral with respect to x is evaluated by means of a series expansion of the exponential,

$$\int dx f(x) e^{\frac{iq(\mu - x)}{N}} \simeq \int dx f(x) \left[1 + \frac{iq(\mu-x)}{N} - \frac{q^2(\mu-x)^2}{2N^2} + \cdots \right] \ .$$

The integral of the first term of the sum is 1, corresponding to the normalization of the pdf. The integral of the second term is zero, because of the presence of the mean value μ. The integral of the third term is related to the variance σ^2 of $f(x)$. Collecting these results the seeked distribution is converted into

$$g(z) = \frac{1}{2\pi} \int dq \exp[iq(z - \mu)] \left[1 - \frac{q^2 \sigma^2}{2N^2} + \cdots \right]^N$$

$$\approx \frac{1}{2\pi} \int dq \exp[iq(z - \mu)] \exp[-\frac{q^2 \sigma^2}{2N}]$$

$$= \frac{1}{\sqrt{2\pi}[\sigma/\sqrt{N}]} e^{-\frac{(z-\mu)^2}{2[\sigma/\sqrt{N}]^2}} \ .$$

The last row of above equation is the **central limit theorem**: *the pdf $g(z)$ of the average of n random values corresponding to a pdf $f(x)$ is a gaussian distribution whose mean is the mean μ of the pdf $f(x)$ and whose variance is the variance σ^2 of the distribution $f(x)$ divided by the number of values n used to compute the average.*

This theorem is satisfied by a large class of pdf, but not all of them fulfill the conditions for its validity (e.g., the Cauchy distribution is one of the exceptions). It is a theorem of asymptotic or limit validity, the larger n the better the fulfillment. In practice, even quite small values of n satisfy with good precision the central limit theorem.

Example 1.3 *The central limit theorem is exact for a gaussian pdf $f(x)$ and for any value of n. To check this statement just follow the previous derivation but computing exactly the integral which has to be raised to the power n.*

Empirical Determination of pdf. In addition to the central limit theorem, there are two other theorems which serve to connect the empirical study of pdf with their mathematical properties.

The **Chebyshev Inequality** establishes a relation between an outcome of the random variable x and the mean of the distribution. It asserts that the probability that a random variable x takes a value *outside* the interval $[\mu - k\sigma, \mu + k\sigma]$ is bounded by $1/k^2$,

$$Pr\{|x - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \ .$$

The larger k, the smaller the probability that a single sample is outside the domain of width $2k\sigma$ around the mean.

A further refinement of this inequality is the **Law of Large Numbers**, very similar to Chebyshev inequality but now in reference to an estimator of the mean $\bar{x} = (x_1 + \ldots + x_N)/N$,

$$Pr\{|\bar{x} - \mu| \geq \frac{k\sigma}{\sqrt{N}}\} \leq \frac{1}{k^2} \ .$$

The last refinement is the above described *Central Limit Theorem*. All these three properties establish the link between the empirical analysis of a given pdf and the mathematical knowledge of it. In practice the central limit theorem is related to a *confidence level*, based on the values of integrals of the normal distribution function in symmetric intervals around its mean. A selection of these values is shown in the table below.

	$k=1$	$k=2$	$k=3$
$\int_{\mu-k\sigma}^{\mu+k\sigma} N[\mu,\sigma](x)dx$	0.6826	0.9545	0.9973

One standard deviation corresponds to a 68% of confidence level, which grows to 95% when talking of two standard deviations.

The central limit theorem is used in two different ways. After an empirical sampling of mean values one may look to the results as a determination of the mean and the variance of the original distribution, or as a measure of the mean of the original distribution and a confidence limit related to the new variance σ^2/N. The square root of this quantity, $\sigma/\sqrt{N}$, is normally appended to the estimator of the mean with a $\pm$ sign, i.e., $\mu \pm \sigma/\sqrt{N}$. The *error* thus written must be interpreted as a confidence limit.

The dependence on n from the central limit theorem may also inform about the amount of work to be done so as to have a predetermined error. The unpleasant fact is that in order to halve the error the empirical sampling must deal with four times more samples. Brute force is not the best way of improving the quality of the results, and the creativity in Monte Carlo work resides in the search of procedures to lower the variance σ, which is the quantity which fixes the error scale.

Generalization of the Central Limit Theorem. The previous discussion has been limited to the study of the distribution of the mean of a pdf, and we have specifically considered univariate distributions.

The theorem may be generalized to deal with other kind of expectation values as well as with multivariate distributions. For example, consider a function $h(\mathbf{x})$ which depends on a vector random variable $\mathbf{x}$ with pdf $f(\mathbf{x})$. Define the mean and variance similarly as above

$$m = \int h(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

and

$$s^2 = \int [h(\mathbf{x}) - m]^2 f(\mathbf{x})d\mathbf{x} \ .$$

By following analogous steps as previously, it may be proved that the average of N samples, which defines the random variable

$$z = \frac{1}{N}\sum_{k=1}^{N} h(\mathbf{x}_k) \ ,$$

follows a normal distribution function with mean m and variance s^2/N in the limit of large N. Apparently the pdf $f(\mathbf{x})$ does not appear in this statement. In fact, it is present in the definition of the values of the mean m and the variance s^2.

278 Histograms. A way of getting empirically the shape of a pdf is to build up an histogram: the domain is divided into bins of a predetermined size, and the sample is placed in the appropriate bin if it is larger than the lower limit and smaller than the upper limit. *Placing* the sample means to increment by one the number of counts at the given bin.

For a bin $[x, x + \Delta]$, the probability of the trial to be there is given by

$$P = \int_x^{x+\Delta} f(z)dz$$

and for a total number of samples N_T one should expect $P \times N_T$ samples in the bin. When Δ is small and the pdf is smooth, the probability may be approximated by

$$P \simeq \Delta f(x + \Delta/2) \ .$$

This probability may we explicitly written as an expectation value of a function $h(z) = \theta(z - x)\theta(x + \Delta - z)$, and the variance may be defined in the usual manner, by computing the expectation value of $h^2(z)$. This special function is such that $h^2(z) = h(z)$ so the variance is given by

$$\sigma^2 = P - P^2$$

or, in terms of the number of counts in the bin N and the total number of counts N_T, there results

$$\sigma^2 = \frac{N}{N_T}\left(1 - \frac{N}{N_T}\right) \ .$$

This formula may generally be approximated by N/N_T. In conclusion, if the number of counts in a given bin is N one should have fluctuations within the error bar $\sigma \simeq \sqrt{N}$. It is important to have in mind the statistical meaning of the standard deviation (i.e. as a confidence limit) and expect to have sometimes results outside one standard deviation.

1.5 How Far Is the Limit?

The central limit theorem is satisfied in the $N \to \infty$ limit. In practice we may be able to deal with very large numbers of N, but of course not infinity. Nevertheless one may try to analyze if a given number of samples is large enough to satisfy quite reasonably the central limit theorem.

Empirical Analysis. The way of analyzing empirically this property is to study empirically the distribution of averages: one generates M times averages of N samples, and the individual values are binned in an histogram. It is convenient to plot also the expected normal distribution with the parameters fixed according the central limit theorem. An eye estimation or, better, a χ^2 analysis, may measure how close we are to the fulfillment of the theorem.

This kind of analysis, which may seem purely academic, is strongly recommended in practical MC calculations: if you plan to do a total of $\mathcal{N}$ samples, then classify those samples in bunches or blocks of N samples, and collect the information as mentioned above. Of course, the quotient $\mathcal{N}/N$ should be a large integer, otherwise the histogram will have too few points. After the first block of N samples one may have already a reasonable estimate of both the mean and the variance of the related normal distribution, and these values could be used to determine the range of the histogram, typically of three standard deviations around the guessed mean so as to see the shape of the resulting distribution.

Table 2. Fit to the histogram of the averages of 100 samples

pdf	μ	σ	μ (fit)	$\sigma/\sqrt{N}$ (fit)	χ^2	μ (sampling)
$U[0,1]$	$1/2$	$1/\sqrt{12} \approx 0.2886$	0.4998	0.0282	12.3	0.5000 ± 0.0009
$1/2\sqrt{x}$	$1/3$	$2/\sqrt{45} \approx 0.2981$	0.3335	0.029	6.2	0.3337 ± 0.0009
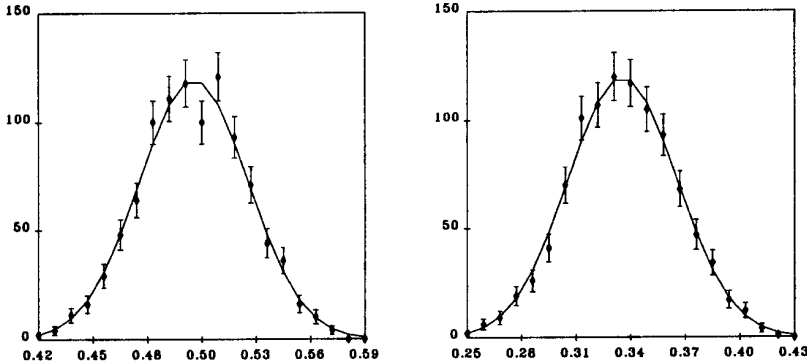
Fig. 1. Distribution of the average of 100 samples drawn from the uniform distribution $U[0,1]$ (left) and from the $1/2\sqrt{x}$ distribution (right). The gaussian in each figure is the central limit theorem prediction.

In Fig. 1 we show this kind of analysis for a total of $M = 1000$ blocks of $N = 100$ samples, for the cases of the uniform distribution $U[0,1]$ and for the inverse square root distribution $f(x) = 1/2\sqrt{x}$ defined in the domain $[0,1]$. From the empirical histogram one may fit a normal distribution and get the parameters. The results corresponding to these two test cases are shown

in Table 2. In the table we include the mean and standard deviation of the original distribution, the mean and $\sigma/\sqrt{N}$ fitted to the gaussian and the χ^2 and, in the last column, the mean and error obtained by using the full set of samples in a single piece. There is an overall agreement between the fitted values and the expected values from the central limit theorem. Also the χ^2 value says that the empirical data are well fitted to a gaussian distribution. In conclusion, the limit of application of the central limit theorem is not very far away, even when the starting distribution is singular, like the inverse square root one.

The Central Limit Theorem for the Exponential Distribution. This is one case in which the distribution of the mean may be exactly computed. The Fourier transform of the exponential e^{-x} distribution is

$$\int_0^\infty e^{-iqx-x}dx = \frac{1}{1+iq}$$

and the distribution of the mean z of N samples is

$$g(z) = \frac{1}{2\pi}\int_{-\infty}^\infty dq\exp(iqz)/[1+iq/N]^N \ .$$

This integral may be computed by going to the complex q − plane, obtaining

$$g(z) = \frac{N^N z^N \exp(-Nz)}{(N-1)!} \ ,$$

known as the Poisson distribution. The relation between Poisson distribution and the gaussian form related to the central limit theorem is very familiar in the statistical analysis of radioactive measurements.

1.6 Probability Distributions of Physical Interest

In section **1.3** we have already mentioned the role of the squared wave function as a probability distribution function. However, the most immediate pdf in physics is provided by Classical Statistical Mechanics, where the probability density of a system of N particles, characterized by a set of positions and momenta $\{\mathbf{r}_k, \mathbf{p}_k\}$ is proportional to $\exp(-H(\mathbf{r}, \mathbf{p})/kT)$, where H is the Hamiltonian of the system. Evaluation of macroscopic properties of the system is done by computing expectation values of the required quantity with respect to this distribution function.

2 Generation of Random Numbers

2.1 Monte Carlo Simulation

Any physical problem tied to a pdf may be dealt with a *simulation* by means of random number generation. The simulation consists in statistically evolving the physical system through the generation of sequences of random numbers which are distributed with the pdf, and it is best done with the help of a computer. So, the unavoidable first step is to design algorithms to generate random numbers related to a given pdf.

Of course, if random numbers are generated by means of a deterministic algorithm, they are no longer random numbers, so the usual name is *pseudo-random* numbers. However, the fact that we can predict the sequence is independent of its randomness.

Algorithms to generate random numbers with a prescribed pdf are based on the generation of random numbers from the uniform distribution $U[0,1]$. So, the scheme of presentation will start with the uniform distribution and afterwards some methods to transform the random numbers to another pdf.

2.2 The Uniform Distribution

The main characteristics of a uniform random generator are the following:

1. First of all, it must be truly random. In other words, the generated sequence must pass all imaginable tests of randomness, including their consideration as pairs, triplets ... nth-tuples of numbers. One of the basic references regarding the analysis of randomness is the book of Knuth [6].
2. Normally, random generators are *periodic*, so that after a certain number of sequential generations the results repeat. Then one should demand a very large period, much larger than any run you need to carry out.
3. Another characteristic of interest is the speed of the generation of a random deviate. This has originated algorithms based on the hardware structure of the processor or the floating point unit tied to it. However, speed may not be so important as one may imagine, and the reason is that after generating a random number or vector one usually does a large number of calculations with it. In other words, the real overhead of the calculations is normally on subsequent processing of the random numbers. For this reason it is probably better to demand *portability*.
4. Portability means that the same sequence of random numbers may be generated by another computer or another language, just by coding appropriately the algorithm. This concept is also related to the *repeatibility* of the sequence, which obviously means the possibility of generating again a previous sequence. The main reason to demand these two properties is that a large part of the computer work is devoted to checking and tuning the code.

Usually you will find a procedure in your preferred computing language to generate uniform random deviates. Using these procedures spoils the portability of the code. One should respectfully admit that the responsibles of the compiler have taken guards with respect to the randomness, but there are known cases in which serious drawbacks have been discovered [7].

Easily coded algorithms are based on congruential methods. They consist in a simple operation on integer numbers modulus a large constant, like

$$N_i = (a \times N_{i-1} + b) \bmod M \to x_i = N_i/M \ .$$

Here a, b and M are predetermined constants, and x_i is the $U[0,1]$ random deviate. The period of the sequence is, at most, M.

Sometimes *shuffling* is added to the algorithm. This consists in maintaining several deviates and updating one of them at random.

There are many public-domain algorithms for the generation of uniform deviates [8, 9, 10].

2.3 The Method of Change of Variables

Let $x \in U[0,1]$ and apply a one-to-one transformation

$$y = h(x)$$

which maps monotonically the domain $[0,1]$ into a new domain $[a,b]$. The question is to determine the pdf corresponding to y, which will be represented by $g(y)$. The link with the starting distribution is the equality of the probabilities

$$f(x)dx = g(y)dy \ .$$

From here it is simple to obtain the form of the new pdf,

$$g(y) = \frac{dx}{dy} = \frac{dh^{-1}(y)}{dy} \ , \tag{2.13}$$

where h^{-1} is the inverse function of $h(y)$.

Unfortunately Eq. (2.13) does not solve our problem, which is to determine the change of variables $h(y)$ appropriate to our pdf $g(y)$. Solving (2.13) for h we obtain

$$h^{-1}(y) = \int_a^y g(z)dz \equiv G(y) \ , \tag{2.14}$$

where $G(y)$ is the cumulative pdf of $g(y)$. The resulting algorithm is simple: the change of variables is the inverse of the cumulative function of the pdf of interest. Of course, the algorithm will only be practical if both the cumulative function and its inverse are simple to obtain.

Some examples of the use of the method of change of variables are the following:

The Uniform Distribution $U[a,b]$, with sampling method $y = a+(b-a)x$, with $x \in [0,1]$.

The Inverse Square Root Distribution given by $1/(2\sqrt{y})$ in $[0,1]$. The cumulative distribution is $\sqrt{y}$ and its inverse is y^2, so the algorithm is to sample $x \in U[0,1]$ and square it.

The Triangular Distribution $2y$, in $[0,1]$. The sampling algorithm is $\sqrt{x}$ with $x \in [0,1]$.

The Exponential Distribution $\gamma \exp(-\gamma y)$ with $y = -\log(x)/\gamma$, for $x \in [0,1]$.

The Gaussian Distribution. The cumulative distribution function is given in terms of the complementary error function, which has not a simple form and its inversion requires the solution of a transcendental equation.

A way to bypass this problem is to consider the bivariate gaussian distribution

$$g(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

in polar coordinates $\{r, \phi\}$, i.e.

$$g(r,\phi) = \frac{1}{2\pi}re^{-r^2/2} \ .$$

The angular part is the uniform distribution $U[0,2\pi]$, and the cumulative distribution of the radial part is $\exp(-r^2/2)$, the inverse being $\sqrt{-2\log y}$. Then one may obtain pairs of gaussian distributed variates by using the pair of equations

$$y_1 = \sqrt{-2\log x_1}\cos(2\pi x_2)$$
$$y_2 = \sqrt{-2\log x_1}\sin(2\pi x_2) \ , \tag{2.15}$$

with both x_1 and x_2 from the uniform distribution. This method is known as the Box-Muller method. Once y_1 and y_2 have been determined, one of the values is taken at random and the other one may be rejected or stored for a subsequent call.

If the gaussian distribution has variance σ^2 the above values must be multiplied by σ. If the mean is μ, it should be added to the generated random numbers (after the scaling).

2.4 Other Algebraic Methods

There is a large collection of methods which may not be classified within a specific scheme. Just for completeness let's mention some of them

- Given $x_1, x_2 \in U[0,1]$, $z = x_1 + x_2$ belongs to a triangular distribution

$$g(z) = z \text{ for } z < 1$$
$$g(z) = 2 - z \text{ for } z > 1$$

defined in $[0,2]$.

- The variable $z = \max\{x_1, x_2, \ldots, x_n\}$, with $x_i \in U[0,1]$, corresponds to the pdf nz^{n-1}.
- If z is the middle value of three uniform deviates, its distribution is the parabola $6z(1-z)$
- **Sum of Random Variables:** when $z = x + y$, with $x \in f(x)$ and $y \in g(y)$, its distribution is the *convolution* product

$$h(z) = \int f(x)g(z-x)dx .$$

- **Product of Random Variables:** In the same conditions as above, the product $z = xy$ corresponds to the pdf

$$h(z) = \int \frac{f(x)}{x}g(z/x)dx .$$

In the last two methods care has to be taken with the domain boundaries of the involved distributions.

2.5 The Acceptance-Rejection Method

This method, due to von Neumann, is a general purpose method valid for *bounded distributions with finite domains*. It may be used also in multivariate distributions, but its efficiency tends to decrease with the number of variables.

Assume the pdf $f(x)$ defined in the domain $[a, b]$, and let M be an upper bound to the distribution, i.e. $f(x) \le M$, $x \in [a,b]$. To obtain a sequence of random numbers of $f(x)$ proceed as follows

1. Draw $z \in U[a, b]$
2. Draw $p \in U[0, M]$
3. If $f(z) \ge p$ then accept z as the random number, else discard both z and p and go to 1

The proof of this method is very simple. One should just write down mathematically the previous statements

$$\overline{f}(x) = \int_a^b \frac{1}{b-a}dz \int_0^M dp \frac{1}{M}\delta(x-z)\theta(f(z)-p) ,$$

where the delta function says that z will be the random number, and the theta function says that this will happen if $f(z) \ge p$.

By carrying out the integral there results the equality

$$\overline{f}(x) = \frac{1}{M(b-a)}f(x)$$

which says that we obtain a function proportional to the wanted pdf. The proportionality constant is always less than one, and simply accounts for the rejected points.

The acceptance-rejection method may be modified so as to have (on the average) efficiency one, i.e., so as to obtain as many random deviates as trials. The new scheme could be as follows:

1. Obtain $x \in U[a, b]$ and compute $f(x)$
2. Count x a total of n times, where n is a random integer with mean $(b-a)f(x)$. The way of determining this number is to compute

$$n = [p + (b-a)f(x)] ,$$

where $p \in U[0, 1]$ and the square brackets mean the *integer part*. The result may be zero (so, x is not accepted), or 1 or larger. In the last case, the random number x must be used as many times as the value of n.
3. Go to 1

The modified von Neumann method may still be generalized to the case of a pdf $h(x)$ which may be written as the product of a pdf $f(x)$ times a *positive* rate term $g(x)$. Note that the three quantities f, g and h must be semipositive definite, and all of them defined in the same domain $[a, b]$. Then, random deviates from h are obtained by counting random deviates of $f(x)$ an integer number of times drawn from a uniform distribution with mean $g(x)$. Mathematically this statement corresponds to evaluating the integral

$$h(z) = \int_a^b f(x)dx \int_0^1 dp\,\delta(z-x)[g(x)+p] \equiv f(z) \int_0^1 dp[g(z)+p] .$$

The integer part appearing in this equation is denoted by the square brackets [ ], and it may be recast as the sum of the integer part of g plus the integer part of the sum of the fractional part of g plus p,

$$[g(z)+p] = [g(z)] + [\text{frac}(g(z))+p]$$

and the second term of this sum is not null only if p is greater than $1 - \text{frac}(g(z))$, its value being 1. So, the integral over p is

$$\int_0^1 dp[g(z)+p] = \int_0^1 dp[g(z)] + \int_{1-\text{frac}(g(z))}^1 dp \equiv [g(z)] + \text{frac}(g(z)) \equiv g(z) ,$$

obtaining the required result.

The part assigned to $g(x)$ has been called above a *rate term*, because it acts like in a creation/annihilation process: the assumed x value may survive or reproduce or just die.

2.6 Other Methods

Just a few lines to mention two capital methods, known as Metropolis method and composition method. These two methods are closely related to the Variational Monte Carlo theory and to the Green Functions Monte Carlo method respectively, and will receive special attention in the rest of this paper.

3 Monte Carlo Quadrature

A large amount of Monte Carlo practical work is devoted to the computation of integrals. There are many numerical methods for the computation of integrals, but in some cases the Monte Carlo method is one of the most efficients. Generally speaking, Monte Carlo method is advantageous in two special circumstances: when the integration domain is complicated and/or when there are many variables involved.

In this section we are going to consider only few techniques related to Monte Carlo, with the goal of arriving to the very useful method called *importance sampling*. The review of methods is very incomplete. We will basically refer to the calculation of the integral of a bounded function of many variables in a finite many-dimensional domain.

3.1 Hit-or-Miss Method

This method is best represented by the computation of the area inside a closed line, or the volume inside a closed surface. These figures are framed with a rectangle or parallelepiped, and uniform random numbers are generated inside the rectangular figure. The ratio of the area of the figure to the area of the enclosing rectangle is the same as the (limit) ratio of the points inside the closed surface to the total number of points drawn.

This method is particularly adequate for integrals where the boundaries are defined in a quite complex form, like the phase space integrals in collision problems. The main disadvantage is that the variance is usually quite large. The method is basically the von Neumann acceptance-rejection algorithm.

3.2 Crude Monte Carlo

Crude MC is the plain application of the central limit theorem to the calculation of the integral

$$\int_a^b h(x)dx = (b-a)\int_a^b \frac{1}{b-a}h(x)dx$$

which may be interpreted as the expectation value of the function $h(x)$ with respect to the uniform distribution function $U[a,b]$.

The algorithm consists in evaluating the average of $\{h(x_i)\}$, where $x_i \in U[a,b]$. Along with the calculation of the average one may also compute the variance by evaluating the expectation value of $h(x)^2$. Practical equations are

$$\langle h \rangle = \int_a^b h(x)dx = \lim_{N\to\infty} \frac{b-a}{N}\sum_{i=1}^{N} h(x_i) \ , \ \ x_i \in U[a,b]$$

and

$$\sigma^2 = \lim_{N\to\infty} \frac{b-a}{N}\left[\frac{1}{N}\sum_i h(x_i)^2 - \left(\frac{1}{N}\sum_i h(x_i)\right)^2\right] \ .$$

3.3 Importance Sampling

There is not an intrinsic difference between the crude MC and the importance sampling MC. Importance sampling is just another application of the central limit theorem but this time to the computation of the expectation value

$$\langle h \rangle = \int f(x)h(x)dx$$

where $f(x)$ is a pdf and $h(x)$ is an arbitrary function. The only difference, thus, with crude MC is the use of a non-uniform pdf.

Any integral may be converted into this form,

$$\int h(x)dx = \int f(x)[h(x)/f(x)]dx.$$

However, the two sides of this equation are very different, from the point of view of MC. On the lhs, one samples the uniform distribution and computes averages of $h(x)$. On the rhs, one samples the *ad hoc* introduced pdf $f(x)$ and samples the quotient of the function $h(x)$ and the pdf.

One should obtain the same value for the expectation value $\langle h \rangle$ by using both procedures. However, the main difference is in the variance of their expectation values. Crude MC will give for the variance the value

$$s^2 = \int h(x)^2 dx - \langle h \rangle^2 \ ,$$

whereas the importance sampling method will give a different value

$$s^2 = \int f(x)[h(x)/f(x)]^2 dx - \langle h \rangle^2 \ .$$

In conclusion, by cleverly choosing the importance sampling function $f(x)$ one may design a better calculation from the MC point of view.

Is there any optimal importance sampling function? The answer to this question is *yes*, the optimal function being $f(x) = C|h(x)|$, C being a proportionallity constant. Unfortunately the answer is, generally, useless. To sample from this importance function one should be able to obtain its cumulative distribution function, in other words, one should be able to compute analytically the integral.

This poses the two main problems related to the importance sampling method: first, to determine the importance sampling function so that the variance is significantly lowered and second, to find a reasonably simple sampling algorithm to draw points from the pdf.

In addition to lowering the variance, there is a side benefit of this method, namely, that it allows to deal with (null-measure) infinite integrands and/or infinite intervals, just by choosing an appropriate pdf.

To give an idea of the power of the importance sampling method the next example describes the calculation of a rather complex function of six variables.

Example 3.1 *Consider the six-dimensional integral of the function*

$$F(\mathbf{x}, \mathbf{y}) = e^{-\mathbf{x}^2 - \mathbf{y}^2 - (\mathbf{x}-\mathbf{y})^2/2}$$

extended to the full space. The exact value is 10.9626.

Table 3 presents the calculation using the crude MC method (each variable integrated between −5 to +5, and the calculation with the importance sampling function $e^{-\mathbf{x}^2 - \mathbf{y}^2}$, adequately normalized. It is absolutely clear the advantage of using the importance sampling algorithm, so that the extra work required to sample from the importance sampling pdf is largely compensated by the smallness of the error. Just by comparing the errors quoted in the last row of the table we may conclude that the importance sampling method is of the order of 250 times more efficient.

Table 3. The integral of the six-dimensional function of example 3.1. The left column shows the crude MC method, and the right column the importance sampling MC method. Each row corresponds to one hundred thousand samples, and the last row is the full average corresponding to one million samples.

Crude MC	Importance Sampling
12.80 ± 4.51	10.46 ± 0.26
8.93 ± 2.96	11.23 ± 0.26
21.90 ± 8.43	11.02 ± 0.26
12.70 ± 3.06	11.15 ± 0.26
13.84 ± 4.48	10.90 ± 0.25
4.58 ± 1.02	10.94 ± 0.26
14.84 ± 4.68	10.72 ± 0.25
8.17 ± 2.38	10.89 ± 0.25
6.79 ± 1.35	10.87 ± 0.26
5.56 ± 1.11	10.75 ± 0.25
11.01 ± 1.27	10.89 ± 0.08

Going back to the question of choosing the importance sampling method, it is already convenient to mention that in the variational MC method the importance sampling function is fixed by the problem itself, see Eq. (1.7). In addition, there is a general purpose algorithm, known as the *Metropolis* or MR²T² method for sampling arbitrary pdf, which will be described below.

3.4 Some Comments

We have limited ourselves to three ways of doing MC integrals. There are many other variance reducing techniques, which however are not of primarily interest for our present discussion. The interested reader may find a variety of methods in Refs. [4, 5].

4 Markov Chains and Metropolis Algorithm

4.1 Markov Chains

A *Markov chain* is defined in terms of a discrete set of states $\{s_1, s_2, \ldots s_N\}$ and a rule governing the transition from a given state s_i to any other state of the set s_j. This rule is expressed in terms of a *transition matrix* $T(j \leftarrow i) \equiv T_{ji}$ so that T_{ji} is the probability of transition from the state s_i to the state s_j. Some transition probabilities may be null, and the matrix may also contain non-null diagonal terms representing the probability of remaining at the initial state.

Given that the matrix elements are transition probabilities and that the transition probability from s_i to *any* final state must be unity, this matrix must fulfill the following two properties

$$0 \leq T_{ji} \leq 1$$
$$\sum_j T_{ji} = 1 \ ,$$

so that the sum of any column is 1. Matrices fulfilling these two conditions are termed *stochastic*. If it happens that a given column has all its elements null, but the diagonal, then the state corresponding to this column is referred to as an *absorbing state*. Once it is reached, it will never move from there.

A way of visualizing a Markov chain is through its interpretation as a *random walk*. One may walk at random many physical systems. An obvious example of random walk is a cards game in which you and the contrary part bet 1 euro. Your *state* is the amount of money remaining after each game. The state with *zero euros* is an absorbing wall.

A more serious example could be the chain of events followed by a neutron traversing an absorbing medium. The neutron may scatter, or may be captured. Here there are two absorbing walls, corresponding to the events *being captured* and *having traversed the medium*.

Both examples clarify the concept of transition probability and the idea of a random walk. There is a variable, which could be called time, which signals a single step of the walk. In the case of the cards game, it is the bet. In the case of neutron absorption, it is the mean free path. Interesting examples of random walks may be found in Ref. [13]

The main difference of the second case with respect to the first case is that in the second case we may have a *continuum* of states: after a collision, the neutron may be deflected to any direction of space. In cases like this, termed *Markov processes*, instead of referring to a transition probability one must deal with a *density of probability*, $T(x', x)$ which must in turn satisfy the two properties

$$T(x', x) \geq 0$$
$$\int T(x', x)dx' = 1 \ .$$

We will be interested in *endless* chains, i.e., with no absorbing states. We will also consider only *irreducible chains*. A chain is called *reducible* if the

set of states may be divided into two disjoint sets so that there cannot be a transition from the first set to the second one, and inversely.

4.2   The Direct Problem

Which is the probability of staying at a given state s_i after a long random walk? The answer is very simple: the probability of arriving to s_i is the product of the probability of arriving to any state s_j times the transition probability from s_j to s_i,

$$P_i = \sum_j T_{ij} P_j \;,$$

and the solution is again simple. This is an homogeneous linear system, which will have a non-trivial solution only if the set of equations is *linearly dependent*. This fact is easily checked by summing up over the free index i, and remembering the normalization condition on the transition matrix.

The solution is not uniquely defined, and it should be supplemented by a *normalization condition*

$$\sum_i P_i = 1 \;.$$

In the case of a Markov process, the equivalent to the matricial equation is the linear integral equation

$$P(x) = \int T(x, x') P(x') dx' \;,$$

with the supplementary normalization condition $\int P(x) dx = 1$.

An alternative way of determining the probabilities P_i related to a Markov chain is to carry out a random walk. The algorithm is very simple

- Initialize P(i)=0, i=1, ... Nstates
- Start at predetermined state I
- For k=1 to Nmoves do
    • J=1
    • Q=RAND()
    • While T(J,I) < Q do
        * Q=Q-T(J,I)
        * J=J+1
        * enddo
    • P(J)=P(J)+1
- For J=1 to Nstates do P(J)=P(J)/Nmoves

Here RAND() is a procedure generating $U[0,1]$ uniform random numbers. The above algorithm is a surprising (and not very efficient) way of solving a class of linear problems. For us the best interpretation is to consider this algorithm

as a method of generating random integers (the labels of the states) corresponding to a pdf P_i, without knowing the actual distribution, the generation being controlled by the Markov chain.

The case of Markov processes is more complicated. Of course, it has no sense to jump from a continuous x to another x'. Instead, the problem function $P(x)$, which is a pdf, is represented by a set of random numbers $\{\xi_i\}$ obtained sequentially, in such a form that the number ξ_{i+1} of the sequence is obtained by drawing a random number from the pfd $f(x) \equiv T(x, \xi_i)$, which is a *conditional probability* of obtaining x from a given ξ_i.

It is convenient to spend some words on the last consideration. Certainly, it is not the same to know the pfd $P(x)$ that to have a sequence of random numbers related to $P(x)$ as its own pdf. However, knowing the set of random numbers permits the determination of expectation values related to $P(x)$: just generate the sequence of random numbers as described above and evaluate the expectation value by using the central limit theorem. Of course, the method will only be practical when it would be possible to sample from the pfd $T(x, \xi_i)$.

This so simple idea is the core of the so called Green's Functions Monte Carlo method, for solving many-body Schrödinger equations.

4.3   The Inverse Problem and Metropolis Algorithm

The inverse problem is stated in very simple words: given a pdf P_i or $P(x)$ may we find a stochastic matrix T_{ij} or $T(x, x')$ defining a random walk which gives rise to the mentioned pdf's?

This question, ant its relevance with physics, was answered by N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. W. Teller in 1953 [12], giving a specific rule to construct it. In addition, the random walk was so designed to be easily carried out. The method is known as *Metropolis* or MR^2T^2 method. It is a general purpose algorithm to generate random numbers from an arbitrary pdf.

For discrete distributions the solution is as follows. Let S_{ji} be an auxiliary *stochastic* and *symmetric* matrix, otherwise arbitrary. The seeked stochastic matrix is given by

$$\begin{aligned}
T_{ji} &= S_{ji} & \text{if } P_j \geq P_i, & \quad i \neq j \\
T_{ji} &= S_{ji} P_j / P_i & \text{if } P_j < P_i, & \quad i \neq j \\
T_{ii} &= S_{ii} + \sum_k' S_{ki}(1 - P_k/P_i) & \text{for } k \in \{P_k < P_i\}
\end{aligned}$$

The prime on the sum indicates that only states k such that $P_k < P_i$ are considered. The values of the diagonal elements have been fixed so as to fulfill the property $\sum_j T_{ji} = 1$.

In order to prove that the matrix so defined satisfies the wanted relation $\sum_i T_{ji} P_i = P_j$, an intermediate step is to prove the *detailed balance* property

$$T_{ji} P_i = T_{ij} P_j \;.$$

In this relation there is no sum over repeated indices. To check this relation it is convenient to distinguish the two cases

$$P_j \geq P_i, \quad T_{ji}P_i = S_{ji}P_i \qquad\qquad T_{ij}P_j = S_{ij}\frac{P_i}{P_j}P_j = S_{ij}P_i$$
$$P_j < P_i, \quad T_{ji}P_i = S_{ji}\frac{P_j}{P_i}P_i = S_{ji}P_j \quad T_{ij}P_j = S_{ij}P_j$$

and to remember the symmetry of the auxiliary transition matrix S_{ij}.

Summing up on i in the detailed balance equation there results the desired condition $P_j = \sum_i T_{ji}P_i$.

In the case of continuous processes one may write down a similar form. Start from a stochastic and symmetric auxiliary transition function $S(y,x)$ and define

$$T(y,x) = S(y,x)\theta(P(y) - P(x))$$
$$+ S(y,x)\frac{P(y)}{P(x)}\theta(P(x) - P(y))$$
$$+ \delta(y - x)\int dz S(z,y)\left(1 - \frac{P(z)}{P(y)}\right)\theta(P(x) - P(z))$$

which again satisfies the detailed balance property

$$T(x,y)P(y) = T(y,x)P(x) \ .$$

Metropolis algorithm is not the only solution to the problem, but it is a procedure which may be easily used. It is interesting to present the algorithm in words. Let's assume that we are at the state s_i. By means of the auxiliary transition matrix S_{ji} we select a candidate s_j for the next state. If the probability of the candidate P_j is greater than the probability of the starting state, the transition is accepted, i.e., the next state of the chain is s_j. If the probability P_j is smaller than P_i, the new state is accepted with probability P_j/P_i. In case of rejection this means that we remain in s_i, i.e., the next state in the chain is the same as the old state, and it should be counted. Here is the algorithm:

– Let I, starting state
– Draw J from S(J,I)
– if P(J)/P(I) > RAND() then I=J
– I is the next state

The algorithm will be practical if we know the way of generating random numbers from the distribution S_{ji}, where i acts as a parameter and j as the random variable. The freedom in choosing the auxiliary transition function S should be used to choose the most convenient one.

In the case of continuous variables the procedure is quite similar. Usually, but not necessarily, the transition operator $S(y,x)$ is taken as a uniform distribution $U[x - \Delta/2, x + \Delta/2]$ centered at x, so the first step is to draw $y \in U[x - \Delta/2, x + \Delta/2]$, and then accept it if $P(y)/P(x) > \xi$, where $\xi \in$

$U[0,1]$. Let me insist again that if the attempted move is rejected this should be interpreted as *take x as the next random number*, so x is counted twice (or more times).

As in any other random generating algorithm, Metropolis algorithm is of asymptotic nature, i.e., it has meaning after many samples have been drawn. Moreover, specially in the continuous case, it is prone to *strong correlations* between successive moves. The main reason for this is that usually the pdf $P(x)$ is peaked around some maximum value, and also the auxiliary distribution $S(y,x)$ is of a rather short range around x. In consequence, successive random numbers are quite close, and to sample the full domain where $P(x)$ is defined it is necessary to carry out many moves.

Here we should note that there are two ways of *not moving appreciably*: when the step Δ is very small, in which case the rate of acceptances is large, but the real displacement is small, or when Δ is large, in which case the rate of rejections is large, and the effective displacement is again small. There are not specific rules to say when a certain rate of acceptances is reasonable. It is normally assumed that a rate between 50% and 80% is acceptable.

4.4 Thermalization in Markov Chains

An interesting property of stochastic matrices is that the product of two stochastic matrices is again a stochastic matrix. The proof is straightforward.

The interesting consequence of this property is that any integer power of a given stochastic matrix, which of course is again stochastic, gives rise to a pdf which is the same as the pdf related to the original transition matrix. Again the proof is very simple, and we show it here for the square of the matrix,

$$\sum_i [T^2]_{ji}P_i = \sum_{ki} T_{jk}[T_{ki}P_i] = \sum_k T_{jk}P_k = P_j \ .$$

This property permits to *thermalize* the random walk, in the following sense. Starting at a given state s_i we proceed carrying out a given number n of steps, $\{s_{i+1}, s_{i+2} \ldots s_{i+n}\}$, but we only count or collect the state s_{i+n} after n steps, and so on. In other words, $n - 1$ intermediate steps are discarded. Nevertheless, the resulting collection of states is a set of random deviates corresponding to the pdf under study.

The objective of this thermalization is to minimize the effect of correlations among successive states when executing the random walk.

5 The Variational Monte Carlo Method

5.1 General Description

Given a Hamiltonian H and a trial wave function Ψ, the variational principle states that the quantity

$$\langle H \rangle = \frac{\int \Psi^* H \Psi}{\int |\Psi|^2}$$

is an upper bound to the energy of the ground state of the Hamiltonian. One may be still more precise, and consider the case in which the Hamiltonian has certain symmetries, so that the eigenstates may be classified as families with precise quantum numbers related to these symmetries. Then, if the trial function has a given symmetry, the expectation value of the Hamiltonian is an upper bound to the energy of the lowest state with that symmetry.

For example, for a harmonic oscillator Hamiltonian $p^2 + x^2$ the eigenstates have a good parity. Then, trial functions like $\Psi(x) = C \exp(-ax^2)$ provide upper bounds to the even parity ground state, and trial functions like $\Psi(x) = Cx \exp(-ax^2)$ will provide upper bounds to the first excited state, which has negative parity.

One of the properties of a many-body Hamiltonian involving identical particles is that it will commute with any permutation of the labels of the particles. Among the operators related to permutations there are two of them of specific importance: the full antisymmetrizer and the full symmetrizer. Note that the symmetry of the trial state is not dictated by the Hamiltonian, and we should impose it externally by selecting the adequate trial function (bosonic or fermionic).

Let $\mathbf{R}$ represent the set of all coordinates of the many-body system. If the system has N particles, $\mathbf{R}$ will have $3N$ components. The expectation value of H is usually written in terms of a multivariate pdf

$$f(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2}$$

and the quantity called *local energy*, already mentioned in section **1.3**, given by

$$E_L(\mathbf{R}) = \frac{1}{\Psi(\mathbf{R})} H\Psi(\mathbf{R}) \ ,$$

i.e., the action of the Hamiltonian on the trial wave function divided by the trial wave function. For the typical Hamiltonians which are the sum of a kinetic energy term, with second derivatives, and a momentum-independent interaction potential, the contribution to the local energy from the potential part is just the numerical value of the potential.

Then the expectation value of the Hamiltonian is given by

$$\langle H \rangle = \int d\mathbf{R} f(\mathbf{R}) E_L(\mathbf{R})$$

and the Monte Carlo way of evaluating this integral is to consider $f(\mathbf{R})$ as the importance sampling distribution function. This is the Variational Monte Carlo (VMC) method.

Here we should mention two *dangerous* shortcuts related to the computation of the local energy, particularly the kinetic energy part. One of them consists in taking advantage of the spherical symmetry of the wave function, so that instead of computing the laplacian $\sum_i \nabla_i^2 \Psi$, one may be tempted to compute just the second derivatives with respect to a coordinate axis (say, X axis) $\sum_i \partial^2\Psi/\partial x_i^2$ and multiply by 3. On the average both procedures will produce the same value, but the variance of the second method will probably be larger.

The other typical shortcut is to transform the expectation value $\langle \Psi | \nabla^2 | \Psi \rangle$ into $-\langle \nabla\Psi | \nabla\Psi \rangle$, by integration by parts, and compute the second form which is simpler and quicker. The transformed contribution to the local kinetic energy will then be proportional to $|\nabla\Psi/\Psi|^2$. Again this second procedure will probably increase the variance of the calculation.

Consider the second case. The basic relation to be used is

$$\int \partial(\Psi\partial\Psi) \equiv 0 = \int \partial\Psi\partial\Psi + \int \Psi\partial^2\Psi$$

and the integral is null because of the boundary conditions corresponding to a localized system. The integrands of the rhs, however, are not the same, but the following relation is still valid

$$-\Psi\partial^2\Psi = -\partial(\Psi\partial\Psi) + \partial\Psi\partial\Psi \ .$$

If the trial state were the exact g.s. wave function, when adding the potential times $|\Psi|^2$ to the above relation, the lhs will be converted into a constant times the squared wave function, $E_0\Psi^2$, and the variance of the integral will be zero. This will not be the case for the rhs, so the variance using the transformation will be larger. One may expect the same to be true for wave functions close to the exact g.s. wave function. For this reason we added above the word *probably*. The argument does not apply to the computation of expectation values of other operators.

The previous discussion may serve to put a general statement about the variance of the calculation of expectation values other than the Hamiltonian. One should expect to have, in general, a variance larger than the variance of the energy, for sufficiently good trial functions. Consequently, if one samples separately the kinetic and the potential part of the Hamiltonian, again one should expect for these observables a variance larger than the variance of the complete Hamiltonian.

In order to run a VMC code one should prepare in advance the required formulae for the calculation. There are two quantities of interest: the calculation of the wave function and the calculation of the local energy. These two procedures are called almost as frequently as the random generator, so one should be very careful in preparing a very efficient code.

The algorithm for the VMC evaluation may go along the following steps

Starting
- Initialize random routine `RAND()`
- Get number of moves `NM`, number of thermalization moves `Ntherm` and Metropolis `STEP`
- Determine an initial point (3N coordinates) `R` and the value of the squared wave function `PSI(R)**2` at this point
- Initialize accumulation variables for energy `SE=0` and energy squared `SE2=0`

Loop of Moves: `for i=1 to NM do`

    Thermalization Loop: `for j=1 to NTHERM do`
- Trial step `RP = R + (RAND()-0.5) * STEP`
- Compute new wave function `PSI(RP)**2`
- Ask Metropolis question. If move is accepted, copy `RP` to `R`, and also the squared wave function
- `enddo` of thermalization loop

    Sampling:
- Compute local energy `EL(R)`
- Update accumulation `SE=SE+EL` and `SE2=SE2+EL**2`
- Eventually, other expectation values

    End Loop of Moves

Final Calculations
- Compute averages and errors
- Print results

5.2 A Simple Example: the Harmonic Oscillator

Consider the one-dimensional harmonic oscillator defined by the Hamiltonian $H = -d^2/dx^2 + x^2$. The exact g.s. wave function is

$$\Psi_0(x) = \frac{1}{\pi^{1/4}} e^{-x^2/2} \ ,$$

the exact energy being 1. To practice with VMC method let use a similar wave function

$$\Psi(x) = \frac{\sqrt{\alpha}}{\pi^{1/4}} e^{-\alpha^2 x^2/2} \ . \tag{5.16}$$

The local energy operator has the value $E_L(x) = \frac{H\Psi}{\Psi} = \alpha^2 + x^2(1 - \alpha^4)$.
The upper-bound requires the computation of the integral

$$E_V = \int_{-\infty}^{\infty} |\Psi(x)|^2 E_L(x) dx \ ,$$

its value being

$$E_V = \frac{\alpha^2}{2} + \frac{1}{2\alpha^2} \ .$$

For the purpose of evaluating the MC method, it is also of interest to determine the variance at the variational upper bound, whose value is

$$\sigma^2 = \frac{(\alpha^4 - 1)^2}{2\alpha^4} \ .$$

If we had used the integration-by-parts way of computing the energy, then the equivalent to the local energy would have been

$$\tilde{E}_L(x) = x^2(1 + \alpha^4)$$

and the new variance

$$\tilde{\sigma}^2 = \frac{(\alpha^4 + 1)^2}{2\alpha^4} \ ,$$

i.e., larger than the previous one.

The MC calculation is very simple: just generate a large number N of random numbers $\{x_i\}$ corresponding to the pdf Ψ^2 (it is a gaussian) and for each value compute the local energy. The MC estimators of interest will be

$$\langle E \rangle = \frac{1}{N} \sum_i E_L(x_i) \ \text{and} \ \langle E^2 \rangle = \frac{1}{N} \sum_i E_L^2(x_i) \ .$$

Here is a quick and dirty `FORTRAN` code for the evaluation of the upper bound:

```
      implicit real*8 (a-h,o-z)
      parameter (seed=1902231.d0, alpha=0.5d0 , N=10000)
      data e,e2/0.d0,0.d0/
      z=randini(seed)                  ! Initialize RANDOM
      do 10 i=1,N                      ! Loop of samples
      x=rangauss()/sqrt(2.d0)/alpha    ! Sample point (scaled!)
      el=alpha**2+x**2*(1-alpha**4)    ! Compute LOCAL energy
      e=e+el                           ! Accumulate for MEAN
10    e2=e2+el**2                      ! and VARIANCE
      write(*,*) e/N, '+/-',sqrt((e2/N-(e/N)**2)/N)
      end
```

This code corresponds to $\alpha = 0.5$, where we know that the upper bound is 2.125 and the variance is $\sigma = 2.6516$. For a total of 10000 samples (as the code above) we should expect a result like 2.13 ± 0.03. However, there is only a 63 % of probability of getting a number between one standard deviation around the mean value. Just running this code 100 times (with different random sequences) I have obtained 56 values within one standard deviation, 98 between two standard deviations and 100 within three standard deviations. These numbers are compatible with the known confidence limits.

Note that this minimal statistical analysis has just shown us the content of the central limit theorem: the set of averages of the local energy is distributed as a normal distribution, with known mean and variance.

It may have some interest to analyze the probability distribution of the local energy. The problem at hand is simple enough to permit the analytic construction of this density, and in other problems one may always study it directly by means of random sampling. The calculation shown above proceeds in two steps. First, a random number x is generated from the pdf $|\Psi(x)|^2$, and then we compute the value of the local energy $E_L(x)$. This quantity is what is ultimately recorded and averaged. In correspondence with the distribution function of x we indirectly generate a distribution function for the energy, and the computed average is just the mean of the later pdf.

The mathematical form of the energy pdf is obtained from the equation

$$g(E) = \int |\Psi(x)|^2 \delta(E - E_L(x)) dx$$

which simply states that the probability of obtaining the value E for the local energy is the integral over all possible x such that $E_L(x)$ equals E. The integral is evaluated by using the known rule of Dirac delta function

$$g(E) = \sum_{i \in \{E_L(x_i) = E\}} \frac{1}{|dE_L(x)/dx|_{x=x_i}} |\Psi(x_i)|^2$$

the sum extending to all roots of the equation $E_L(x) - E = 0$. For $\alpha < 1$ the domain of the new pdf $g(E)$ is $[\alpha^2, \infty]$, and the actual value of the pdf is

$$g(E) = e^{-\alpha^2 \frac{E - \alpha^2}{1 - \alpha^4}} \frac{\alpha}{\sqrt{\pi(E - \alpha^2)(1 - \alpha^4)}} .$$

This is a quite surprising form: it has a (null measure) singularity at $E = \alpha^2$, and decreases exponentially up to infinity.

The way of obtaining the energy pdf and the resulting form is shown in Fig. 2. In the left part we show (arbitrary scale) both the x−space pdf and the local energy. The lines limiting an energy interval establish the correspondence with the coordinate values (the x_i). At the rhs we show the actual pdf for the energy. The VMC method has simply evaluated $g(E)$ in an indirect way, and the quoted upper bound is just the expectation value $\langle E \rangle = \int E\, g(E) dE$. The distribution is far from being a gaussian, but the central limit theorem ensures that the distribution of the mean values is a normal distribution.

5.3 Finding the Minimum

The class of trial functions used in the precedent exercise is very special, because it contains the exact ground state when $\alpha = 1$. In consequence, the variance for this special case is strictly null, and if you run the code for $\alpha = 1$ you will obtain exactly 1 with error 0.
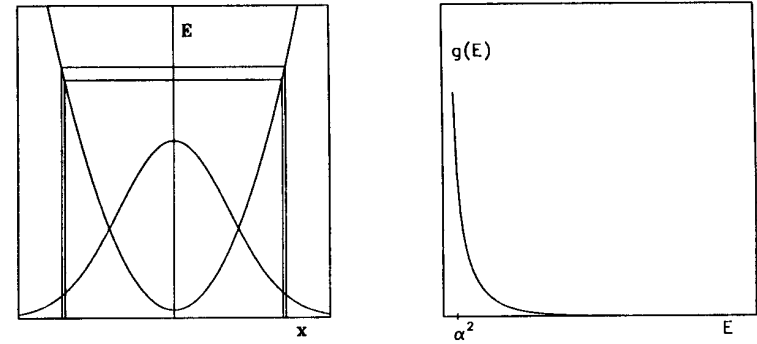

Fig. 2. The pdf related to the squared trial function and the local energy versus the coordinate x (left), and the resulting pdf $g(E)$ of the energy (right).

If we were dealing with other Hamiltonian, but the same family of functions, this will not be the case. In a region around the minimum we should expect something similar to a parabola, within the expected statistical fluctuations. It may very well happen that these fluctuations hide the position of the minimum, giving the aspect of a flat curve or even an inverted parabola.

In order to present the common situation we may consider the problem of determining the optimal upper bound for the Hamiltonian $H = p^2 - 5\exp(-x^2)$ by using gaussian trial functions like the above Eq.(5.16). The minimum may be determined algebraically (solving a transcendental equation), resulting in an energy of -3.123266 at $\alpha = 1.2389$. In the left part of Fig. 3 we show the actual running of a VMC code for several values of the variational parameter α. As mentioned above, the fluctuations around the minimum make it difficult to estimate the lowest point.

A way to overcome this problem is to carry out *correlated samplings* in the region near the minimum. The idea is the following: assume a fixed parameter β close to the minimum, and compute several values of the energy for several values of α by using the following equality

$$E_\alpha \equiv \frac{\int |\Psi(\alpha,x)|^2 E_L(\alpha,x) dx}{\int |\Psi(\alpha,x)|^2 dx} = \frac{\int |\Psi(\beta,x)|^2 [|\Psi(\alpha,x)|^2/|\Psi(\beta,x)|^2 E_L(\alpha,x)] dx}{\int |\Psi(\beta,x)|^2 [|\Psi(\alpha,x)|^2/|\Psi(\beta,x)|^2] dx}$$

so that one evaluates independently numerator and denominator of the rhs equation by taking as importance sampling function the fixed-parameter quantity $|\Psi(\beta,x)|^2$. The crucial point is that random numbers *must be* generated only once, corresponding to the importance sampling function, and used for all values of α. The various computations are not statistically independent, but the hope is that the fluctuations will go always in the same direction. This may be checked by observing the figure at the right of Fig. 3, which is very smooth, compared with the figure at the left which corresponds

to independent samplings. The last consideration is that once the position of the minimum is found, one should carry out a longer VMC calculation, with the variational parameter found, to check the result and to have a smaller variance.
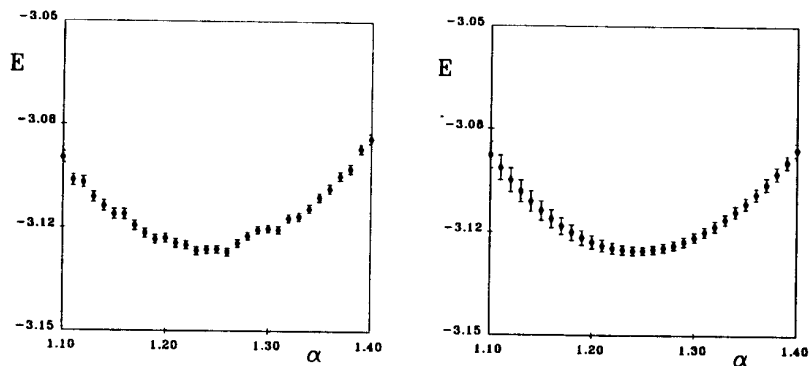

Fig. 3. Search for the minimum by using independent sampling (left) or correlated sampling (right).

5.4 The Structure of Variational Trial Functions

In this section we are going to describe briefly the class of trial functions currently used in the study of many-body systems of strongly interacting particles.

There is a large collection of systems of this kind. Without attempting to be exhaustive, let us mention

1. *Quantum fluids* or *quantum drops*, whose constituents interact through a Lennard-Jones like interaction, with a strongly repulsive core at very short distances and a quite smooth attraction at medium and large distances.
2. *Nuclear systems* and *nuclear matter*, in which there is a soft-core repulsion at very short distances, and attraction at short distances.
3. *Electron systems* like atoms, molecules or the so called *jellium* which describes a cloud of electrons in a substrate of positive charge. Here the main complication comes from the long range of the Coulomb interaction, as well as its singularity at the origin.
4. *Quantum crystals*, *atomic clusters*, and so on.

Some of these systems allow for a zeroth-order description in terms of a mean field or independent particle model. For strongly repulsive interactions this is not possible, because the matrix elements of the potential will be infinite or very large. A simple but very efficient way of describing strongly interacting systems is known with the name of Jastrow [14], which assumes for the trial function the product of an *uncorrelated state*, which takes care of the symmetry or antisymmetry as well as other global properties of the system, and a *correlation factor* which, in its original form, is the product of two-body correlations

$$\Psi(\mathbf{R}) = F_J(\mathbf{R})\Phi(\mathbf{R}) ,$$

with

$$F_J(\mathbf{R}) = \prod_{i<j} f(r_{ij}) .$$

In this form the two-body correlation depends only on the distance of the pair of particles. For interactions having a core one should have correlations which go to zero at short distances and to one at long distances. One may even assume *state-dependent correlations*, i.e., depending on the spin or isospin of the particles involved. For this kind of correlations the Jastrow factor must be properly symmetrized with respect to all orderings of the operators.

5.5 The ⁴He Nucleus

A simple application of the VMC method is the study of the g.s. of the ^{4}He nucleus for a spin/isospin independent nuclear interaction. We will consider the Afnan-Tang [11] S3 interaction, which was adjusted so as to reproduce the low-energy properties of the nucleon-nucleon system and defined in the two $l = 0$ channels, corresponding to spin triplet and singlet. For the description of a saturated nuclear system, it may be simplified to a purely central potential by taking the average of these two channels, having

$$V(r) = 1000e^{-3r^2} - 163.35e^{-1.05r^2} - 21.5e^{-0.6r^2} - 83e^{-0.8r^2} - 11.5e^{-0.4r^2} ,$$

where the potential is given in MeV and the internucleon distances should be measured in fm.

A simple yet efficient trial wave function is a Jastrow factor acting on the $1s^4$ shell-model configuration with radial functions from the harmonic oscillator,

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_4) = \prod_{i<j}^{4} f(r_{ij})e^{-\alpha^2(r_1^2+\ldots+r_4^2)/2}$$

or the explicit translationally invariant form

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_4) = \prod_{i<j}^{4} f(r_{ij})e^{-\gamma\sum_{k<l}(\mathbf{r}_k-\mathbf{r}_l)^2} ,$$

where $\gamma = \alpha^2/8$. In this example we will consider a simple form for the correlation $f(r) = 1 + a\exp[-\beta r^2]$ so that the variational function has three

parameters, α, β and a. Looking to the translationally invariant form, an equivalent description of the variational wave function is just the product of pair correlations $g(r_{ij})$ where the new correlation is

$$g(r) = e^{-\gamma r^2} + ae^{-(\beta+\gamma)r^2} .$$

Here we see the double role of the Jastrow correlation: it may describe both the short range behavior and the long-range confinement property of the finite system. This form is easily generalized to a system of N bosons.

Note that we have not considered the spin-isospin dependent part, which will be a 4×4 Slater determinant built up with the four possible spin-isospin orientations.

Calculations with this kind of correlations and several Wigner nuclear forces may be found in [15]. In that paper the calculations were carried out by a FORTRAN coding of an algebraic algorithm. They may serve as a good check of the MC code.

A simple MC code is presented below. In order to simplify the code the potential and the correlation have been defined in a single sentence. Moreover, the kinetic term of the local energy has been calculated by approximating second derivatives by second differences.

Table 4. Several calculations of ^{4}He with the Wigner part of the S3 interaction and with the Malfliet-Tjon MT-V potential. The g.s. energies are in MeV. ATMS stands for *amalgamation of two-body matrix elements*, CRC is the *coupled rearrangement channel method*, HH is the *hyperspherical harmonics Jastrow correlated method*, CCM is the *coupled cluster method*, FY the *Fadeev-Yakubowsky* method and DMC the *diffusion Monte Carlo method*.

Method	GS energy (S3)[11]	GS energy (MT-V) [17]
HO shell model	−5.89	−6.40
Jastrow: One gaussian [15]	−24.35	−29.01
Jastrow: Two gaussians[15]	−27.15	−30.83
ATMS [18]		−31.36
CRC [19]		−31.357
Jastrow+HH [20]	−27.369	−31.344
CCM [21]		−31.36
FY [22]		−31.36
DMC [15]	−27.35 ± 0.02	−31.32 ± 0.02

Some of the results corresponding to this code are shown in Fig. 4. We have plotted there the first 20 moves (omitting the 10 intermediate thermalization moves of each step), showing how effectively the Metropolis algorithm makes the nucleons to move within the nucleus, and also thousand positions of one particle, with and without thermalization. The latter case shows that the particle is sampling mainly the central part of the wave function, and it is the
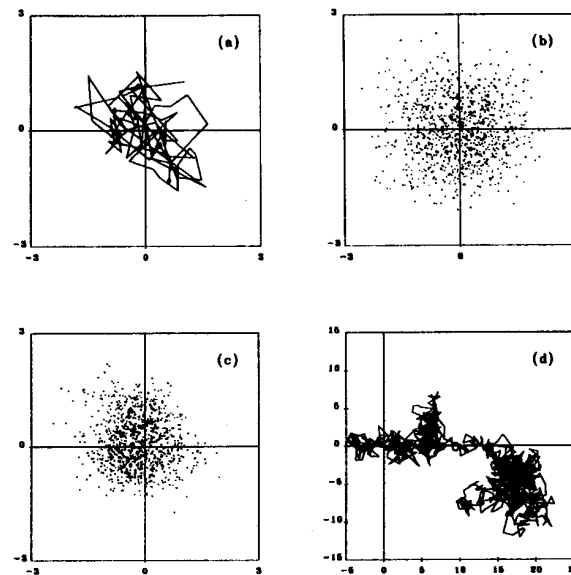
Fig. 4. Some results corresponding to the calculation of ^{4}He. **a:** The first 20 moves of the particles, without showing the intermediate thermalization moves. The diamonds are the starting positions with respect to the cm. **b:** Plot of the position of one particle after 1000 moves, with thermalization, referred to the cm. **c:** Same as b but without thermalization. **d:** Same as b but referred to an external origin. All positions are in fm.

thermalization which extends the sampling also to the tail part. In these three plots, corresponding to parts a, b and c of the figure, the actual positions of the particles have been referred to the center-of-mass of the system. Finally, part d shows the position of the particle, but this time referred with respect to an external system. It shows clearly that the center-of-mass is wandering around, with very large displacements. This is harmless, because the wave function is translationally invariant and there is no binding related to the center-of-mass.

The code may easily be modified to deal with other interactions and correlations. In Ref. [15] you may find parameters adequate for other nucleon-nucleon potentials. Also, the code may be generalized to other finite boson systems, but it will be quite slow. In Ref. [16] one can find results corresponding to invented bose systems of up to 40 particles with the nucleon mass interacting by means of the S3 potential.

How good is this simple Jastrow correlation? Table 4 presents some selected results for the ^{4}He nucleus corresponding to several nucleon-nucleon interactions and to several computational techniques. For a Jastrow wave function of four parameters (sum of two gaussians) the result differs very

little from the exact Diffusion Monte Carlo calculation of the energy. In conclusion, two-body Jastrow correlations contain almost hundred per cent of the structure of the exact wave function. The table also contains the results of other few-body techniques applied to the same problem. It is very rewarding to see the overall agreement between all of them and the high level of precision reached in this (and related) problems.

```fortran
c       =================================================================
c       Variational Monte Carlo for BOSONS
c       Parameters adjusted for He-4 with S3 potential
        include 'varhe.inc'
        step=1                  ! Metropolis step (to be adjusted)
        z=RANDINI(12231.d0)     ! Initialize RAND routine
        do 10 i=1, NA
        do 10  k=1, 3
10      R(i,k)=10*step*(RAND()-0.5d0)   ! Starting  walker
        wfold=WF(R)                     ! Wave function
        Nmoves=1000                     ! Start random walk
        Ntherm=10
        sume=0
        sume2=0
        Nacceptances=0
        do 100 imoves=1, Nmoves
           do 200 itherm=1, Ntherm
              do 300 i=1, NA
              do 300 k=1, 3
300           RP(i,k)=R(i,k)+step*(RAND()-0.5d0)     ! Attempt Move
           wfnew= WF(RP)
c          Metropolis question
           if((wfnew/wfold)**2 .lt. RAND() ) GOTO 200  ! Move rejected
           Nacceptances=Nacceptances+1                 ! Move accepted
              do 400 i=1, NA
              do 400 k=1,3
400           R(i,k)=RP(i,k)                ! Update walker
              wfold=wfnew                   ! and wave function
200     continue
        energy=ELOCAL(WFOLD,R,DRIFT)        ! Compute energy
        sume=sume+energy                    ! Accumulate for
100     sume2=sume2+energy**2               ! averages
c       Print results
        e=sume/Nmoves
        e2=sume2/Nmoves
        write(*,'(A10, F10.3)') 'Energy=   ', e
        write(*,'(A10, F10.3)') 'Error =   ', sqrt((e2-e**2)/Nmoves)
        write(*,'(A10, F10.3)') '% Accept= ',
     1        (100.d0*Nacceptances)/(Nmoves*Ntherm)
        end
c       =================================================================
c       Computes the wave function
```

```fortran
        double precision function WF (R)
        include 'varhe.inc'
c       Correlation function -> R2 is the DISTANCE SQUARED
        f(r2)=exp(-0.08597d0*r2)-0.7191d0*exp(-2.13796d0*r2)
        wf=1
        do 10 i=1, NA-1
        do 10 j=i+1, NA
        r2=(R(i,1)-R(j,1))**2+(R(i,2)-R(j,2))**2+(R(i,3)-R(j,3))**2
10      wf=wf*f(r2)
        return
        end
c       =================================================================
c       Computes the local energy and gradient (to be used in DMC)
        double precision function ELOCAL(wfold , R, DRIFT)
        include 'varhe.inc'
c       Two-nucleon potential-> WIGNER part of S3
        V(r2)=1000.d0*exp(-3.d0*r2)-163.35d0*exp(-1.05d0*r2)-83.d0*
     1  exp(-0.8d0*r2)-21.5d0*exp(-0.6d0*r2)-11.5d0*exp(-0.4d0*r2)
        do 40 i=1, NA             ! To compute derivatives
        do 40 k=1,3               ! define RPlus and Rminus
        RP(i,k)=R(i,k)            ! as R
40      RM(i,k)=R(i,k)
        ekin=0
        do 20 i=1, NA             ! For each particle and each
        do 20 k=1, 3              ! coordinate, compute
        RM(i,k)=R(i,k)-h          ! Second derivatives
        RP(i,k)=R(i,k)+h          ! as centered SECOND differences
        wfplus=wf(RP)             ! and 2*dif(log(WF),R(I,K))
        wfminus=wf(RM)            ! as centered FIRST differences
        ekin=ekin+(WFplus+WFminus-2*WFold)/h**2
        drift(i,k)=(WFplus-WFminus)/(WFold*h)
        RM(i,k)=R(i,k)            ! Restore positions
20      RP(i,k)=R(i,k)
        epot=0                    ! Potential energy
        do 30 i=1, NA-1
        do 30 j=i+1, NA
        r2=(R(i,1)-R(j,1))**2+(R(i,2)-R(j,2))**2+(R(i,3)-R(j,3))**2
30      epot=epot+V(r2)
        ELOCAL=-h2div2m/WFOLD*ekin+epot
        return
c       =================================================================
c       INCLUDE file, to be named VARHE.INC
        implicit real*8 (a-h,o-z)
        parameter (NA=4, h2div2m=20.74d0, h=0.001d0)
        dimension R(NA,3), RP(NA,3), RM(NA,3), DRIFT(NA,3)
```

When dealing with extended systems we cannot attempt the simulation with an infinite number of particles. The usual way of dealing with these systems, inherited from Statistical Mechanics [12, 23], consists in representing the system in a simulation cell with periodic boundary conditions. If the side of the cells is L, the density of the system is determined by putting a given number of particles in the L^3 cell, so that $\rho = N/L^3$.

Metropolis moves are always recast into the simulation cell, but in the actual calculations one should also consider particles in the periodic images. To carry out consistently the Metropolis moves it has to be assumed that the correlation function has a range shorter than $L/2$. Then, to decide if a move of a single particle is accepted or not, only the set of particles contained in a sphere of radius $L/2$, centered in the referred particle, have to be considered, as shown in figure 5. These particles are not necessarily in the simulation cell. If the move is accepted, one imagines that all images have undergone the same move. The evaluation of the potential energy does not require this limitation, but one should carefully check that there is no double counting.
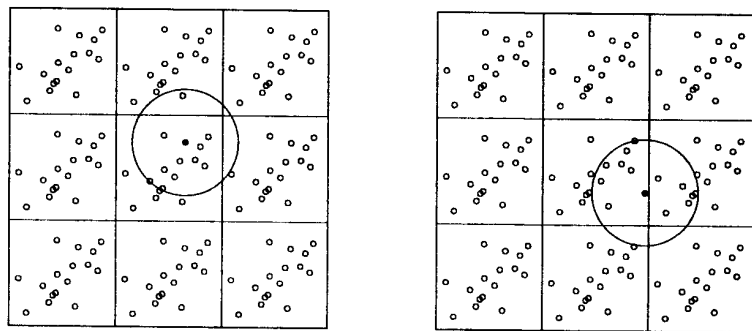
Fig. 5. The simulation cell and its replicas. On the rhs the black dot is the point to be moved, and the circle centered in it determines the selected nearest neighbors which should be counted for the interaction of the particle represented by a black circle. The lhs figure shows the equivalent situation once the selected particle has moved.

There are sets of *magic* numbers in the practical application of the VMC method. In the case of bosonic systems, the idea of choosing a specific number of particles is tied to the possibility of allowing the system *crystallize*, i.e., when particles prefer to place themselves around an ordered and periodic structure. One expects the highest density crystalline order, which corresponds to the face centered cubic structure. The fcc unit cell has 8 particles at the vertices (each particle shared by eight adjacent cells) and six parti-

cles at the faces (each one shared by two cells) which give a total of four *full* particles per unit cell. So, the number of particles is taken as $4n^3$, in such a form that the simulation cell contains an integer number of unit cells. Typical numbers are 108, 256, 500, 864, and so on. For bosonic systems the uncorrelated wave function is simply a constant $(1/L^{3N/2})$, and the integrals are computed, with the prescription above mentioned, in the simulation cell. Once the number of particles has been chosen, to vary the density one simply changes the value of L.

The situation is quite different in the case of fermionic systems. Here the number of particles in the simulation cell is determined by the quantization rules. The single-particle states are eigenvectors of the kinetic energy operator p^2 with periodicity L in each of the X, Y and Z directions. The free solutions are of the form $\psi_k = \exp(i\mathbf{k}\mathbf{r})$, and the possible values of $\mathbf{k}$ must follow the rule

$$\mathbf{k} = \frac{2\pi}{L}(n_x, n_y, n_z)$$

where the n are positive or negative integer numbers. These single-particle levels are filled up in order of increasing energy, up to a maximum value, as indicated in Table 5. The cumulative numbers are 1, 7, 19, 27, 33, 57, ..., to be multiplied by the degeneracy of the system.

Table 5. The values of momenta ordered by increasing energy values. The number in the third column considers the $\pm$ and the reorders of the assignation.

k_{max}^2	(n_x, n_y, n_z)	Number	Cumulative
0	$(0,0,0)$	1	1
1	$(\pm 1, 0, 0)$	2×3	7
2	$(\pm 1, \pm 1, 0)$	4×3	19
3	$(\pm 1, \pm 1, \pm 1)$	8	27
4	$(\pm 2, 0, 0)$	2×3	33
5	$(\pm 2, \pm 1, 0)$	4×6	57
6	$(\pm 2, \pm 1, \pm 1)$	8×3	81

A detailed example of the use of periodic boundary conditions may be found in [24], for the specific case of liquid ^{4}He interacting in terms of a Lennard-Jones potential. There a simple FORTRAN code is also included. This example uses the simple McMillan correlation factor [25],

$$f(r) = \exp(-(b/r)^5/2),$$

where b is the only variational parameter and which gives a quite good upper bound to the g.s. energy. This quantum liquid was extensively studied by Schiff and Verlet [26] and by Kalos and coworkers (see, e.g., [27]).

There are two main complications in the VMC procedures when referring to fermionic systems. The first one is the need of dealing with antisymmetric states (usually Slater determinants) and the second is the role played by the spin and isospin operators which may appear both in the interaction and in the correlation terms.

We may distinguish three levels of complexity, which are briefly described here. For simplicity we will only deal with spin degrees of freedom, the extension to cope with isospin degrees of freedom being straightforward.

Spin Independent Forces and Correlations. This is the simplest case. There are two extreme situations quite easily dealt with: the case of *saturated* systems, in which half of the particles are assumed to have spin up, and the other half spin down, and the case of *fully polarized* systems, in which all particles have the spin in the same direction. This second case is the simplest one, because the wave function is the product of a Slater determinant depending only on the coordinates and the spin part, which is a fully symmetric term.

The first job is to compute the pdf for the Monte Carlo evaluation. Here it is worth mentioning the inadequacy of one of the important simplifications very common in the evaluation of matrix elements with Slater determinants. The usual simplification consists in considering only the antisymmetrization of the, say, ket state, and selecting for the bra state a simple assignment of particles to states. To be more precise, let $\{\phi_1, \phi_2, \dots \phi_N\}$ be the set of occupied orbitals. The antisymmetric wave function is given by

$$\Psi(1, 2, \dots N) = \sum_P \epsilon_P \phi_1(P1)\phi_2(P2)\dots\phi_N(PN) \ ,$$

the sum being extended to the $N!$ permutations. The norm requires the computation of

$$\langle \Psi|\Psi \rangle = \langle \sum_P \epsilon_P \phi_1^*(P1)\phi_2^*(P2)\dots\phi_N^*(PN)| \sum_Q \epsilon_Q \phi_1(Q1)\phi_2(Q2)\dots\phi_N(QN) \rangle$$

and one may just remove the sum over one set of permutations to obtain the equivalent value

$$\langle \Psi|\Psi \rangle = N!\langle \phi_1^*(1)\phi_2^*(2)\dots\phi_N^*(N)| \sum_Q \epsilon_Q \phi_1(Q1)\phi_2(Q2)\dots\phi_N(QN) \rangle \ .$$

The elimination of the sum over P may be done under the integral sign, but the integrands are not the same. In fact, the integrand in the second case is not necessarily positive.

In conclusion, to compute the pdf one must evaluate the Slater determinant and square it. In the case of polarized systems, the Slater determinant is just a function of the space coordinates. In the case of saturated systems one

has two alternatives. The most economical corresponds to consider the spin up and spin down particles as distinguishable, representing the wave function as the product of two Slater determinants of dimension $N/2$. This method has, unfortunately, a quite large variance, because the pdf is not symmetrical under the exchange of any pair of particles. It has, however, partial symmetry under the exchange of particles with the same spin. Full symmetry may be restored by applying a subset of permutations to the pdf (not to the wave functions, and without including the permutation parity) which are the ways of selecting $N/2$ particles from the N particles. The number of these permutations is large even for moderate values of N (for example, for 16 particles there are 12870 permutations) and each of them requires a costly work, namely the evaluation of two determinants. In consequence, unless the number of particles is very small, it may be only possible to deal with the partially symmetrized form of the pdf.

Spin Dependent Forces. The next case corresponds to the situation in which the trial wave function is assumed to be spin *independent*. Here Metropolis moves are carried out as above, but to compute the local energy one should allow for jumps of particles from one spin status to another spin status. For the simple $(\sigma_i \cdot \sigma_j)$ force, its action will be adequately represented with swaps of spins.

Systems with Coupled Channels. The spin dependent problem may be imagined as a *coupled channel problem*. The most familiar coupled channel problem in Nuclear Physics appears in the description of the deuteron with tensor components in the nuclear force [28]. It corresponds to a two-channel differential equation of the form

$$\begin{bmatrix} -\frac{\hbar^2}{2\mu}\nabla^2 & 0 \\ 0 & -\frac{\hbar^2}{2\mu}\nabla^2 \end{bmatrix} \begin{bmatrix} \Psi_s \\ \Psi_d \end{bmatrix} + \begin{bmatrix} V_{ss} & V_{sd} \\ V_{sd} & V_{dd} \end{bmatrix} \begin{bmatrix} \Psi_s \\ \Psi_d \end{bmatrix} = E \begin{bmatrix} \Psi_s \\ \Psi_d \end{bmatrix} \ ,$$

where there is a diagonal kinetic term and a full matrix potential term coupling the s and d channels. The normalization condition of the wave function reads now

$$\int [|\Psi_s|^2 + |\Psi_d|^2] = 1 \ .$$

From the variational point of view one may deal with the coupled-channel problem exactly in the same way as in the ordinary case, considering as pdf the quantity $f(r) = |\Psi_s(r)|^2 + |\Psi_d(r)|^2$ properly normalized. The Metropolis algorithm is simply adapted to this problem: starting at point r one tries a new point r' and is accepted accordingly with the probability quotient $f(r')/f(r)$. The equivalent to the local energy is now the quantity

$$E_L(r) = \frac{1}{|\Psi_s(r)|^2 + |\Psi_d(r)|^2} [\Psi_s^*(r) \ \ \Psi_d^*(r)] \begin{bmatrix} H_{ss} & H_{sd} \\ H_{sd} & H_{dd} \end{bmatrix} \begin{bmatrix} \Psi_s(r) \\ \Psi_d(r) \end{bmatrix} \ ,$$

and the obvious generalization to more channels.

As an illustration of the VMC procedure for coupled channels let's consider the simple problem of two energy-displaced and coupled harmonic oscillators, with a Hamiltonian matrix

$$H = \begin{bmatrix} p^2 + x^2 & x \\ x & p^2 + x^2 + 1 \end{bmatrix}$$

This simple problem is a schematic representation of the nuclear-electronic coupling of diatomic molecules, which is used to explain fluorescence phenomena in diatomic molecules (see e.g. [29, 30]). The ground state has an energy of 0.82867, the upper component being of even parity and the lower component of odd parity. A simple variational function is

$$\Psi_T = \begin{bmatrix} e^{-x^2/2} \\ axe^{-x^2/2} \end{bmatrix} ,$$

where a should be determined variationally. This may be done analytically, with the result $a = 3 - \sqrt{11} \approx -0.31662$ and $E_{var} \approx 0.84168$, not far from the exact result.

Here is the simple FORTRAN code for this evaluation

```
c       Variational coupled oscillators problem
        implicit real*8 (a-h,o-z)
        parameter (Nm=1000, Nt=10, step=2.d0)
        EL(x,a)=4+2/a-(3*a+2)/(a*(1+a**2*x**2))
        f(x,a)=(1+a**2*x**2)*exp(-x**2)
        data a, e, e2/ -0.3166d0, 0.d0, 0.d0/
        x=randini(19923.d0)
        do 10 imoves=1,nm
        do 20 itherm=1, nt
        y=x+step*(rand()-0.5d0)
        if(f(y,a)/f(x,a) .gt. rand()) x=y
20      continue
        e=e+EL(x,a)
        e2=e2+EL(x,a)**2
10      continue
        write(*,*) e/Nm,'+', sqrt((e2/nm-(e/nm)**2)/Nm)
        end
```

Running this code I have obtained the value $E_V = 0.838 \pm 0.006$ for the parameter a corresponding to the minimum.

The main conclusions regarding this discussion is that, even if the trial function has two (in general, more) components, i.e., two coordinate dependent functions, there is not an independent sampling for each one of the components. The other important point is the role of the mixing parameter a of our example. It measures essentially the relative weight of the two components of the wave function, and has a prominent variational role. Even if

the coupled Schrödinger equation is homogeneous, this only allows for an arbitrary global constant and does not permit to fix arbitrarily the value (and even the phase) of the various components of the wave function.

The general spin/isospin variational method may be carried out in the coupled channel formalism described above. However, the enumeration and construction of the spin/isospin channels may be quite involved, and it turns out more convenient to generate the channels by applying a spin/isospin dependent correlation operator to a model state, frequently a shell-model wave function. If Jastrow correlations are used, the correlation operator has the general structure

$$f_{ij} = \sum_k f_k(r_{ij}) \mathcal{O}_{ij}^k ,$$

where the pair operators are of the type spin-spin, or tensor, or any reasonable structure related with the two-body interaction. The N-particle correlation operator is the *symmetrized* product of the pair correlations, the role of the symmetrization being to maintain the fermionic statistics.

I will not enter in more details regarding this question. A quite detailed description of the algebraic treatment of this kind of correlation operators, applied to few-nucleon problems, may be found in [48].

5.8 Additional Comments

The brief presentation of the VMC methods may give a flavor of the power of Monte Carlo methods applied to the variational problem. The required Monte Carlo techniques are conceptually simple, but the practical application may turn out to be both complex and tedious. The main problem is to construct the variational functions and an adequate parameterization, so as to include as much physics of the problem as possible. The tedious part is the search for the variational minimum. A good knowledge of the system is required to carry out reasonable VMC calculations.

6 The Diffusion Monte Carlo Method

6.1 Introduction

The diffusion Monte Carlo (DMC) method is a way of solving exactly (in a statistical sense) the many-body Schrödinger equation by means of a stochastic procedure.

In general, the statement *Quantum Monte Carlo*, abbreviated as QMC, refers to a wide set of methods which attempt to integrate the many-body Schrödinger equation. There was the habitude of referring as *Green Function Monte Carlo* (GFMC) or *Domain Green Function Monte Carlo* the approach in which the time-independent Green Function is approximated by means of a Born series in a certain domain, whereas the name *Diffusion Monte Carlo*

was associated with algorithms which approximate the time-dependent Green Function. However, these two methods do not exhaust the possibilities of stochastic integration.

The basic ideas of QMC were established already around 1950 [31, 32]. The computers at that time were not fast enough to apply the methods to many-body systems. The reborn of the stochastic methods took place in 1962, in a paper by Malvin H. Kalos [33], which actually was involved with the inverse problem, i.e., to determine the depth of a gaussian potential so as to reproduce the binding energy of three and four-body nuclei.

Later, in 1975, Anderson [34] introduced the short-time approximation to the time-dependent Green function, initiating the DMC algorithm. The work of Anderson was concentrated in the application to atomic and molecular systems [35, 36], i.e., to systems with a finite number of particles. The first application to extended systems was carried out by Ceperley and Alder [37] in a famous paper dealing with the electron gas (jellium).

From the didactical point of view we suggest the reader to consult the article of Reynolds et al [38] as well as the book of Hammond, Lester Jr. and Reynolds [39] *Monte Carlo Methods in Ab Initio Quantum Chemistry*.

These lectures will be limited to DMC methods. However, one should cite, for reasons of historical completeness, the work developed in parallel in the GFMC method. Actually, many of the ideas (importance sampling, fixed-node approximation and so on) are shared by the two methods. After the above mentioned article of Kalos [33] in 1962, there were a series of papers also involving Malvin Kalos [40]-[42] which culminated in 1974 with a precise study of the liquid ^{4}He properties [44]. A good review of GFMC method may be found in [27, 45].

The list of references is not complete, and since the 80's there have been many papers dealing with extensions and applications of QMC methods. Some of the applications will be described by J. Boronat and S. Pieper in this volume, so we refer to their lectures for more recent developments. Nevertheless, there have been several methodological advances since that time, some of them being

- Improvement of the short-time approximation in DMC, by S.A. Chin [47]
- Dealing with coupled-channel equations and discrete degrees of freedom, by J. Carlson [48, 49]
- Description of the excitation spectrum, by Ceperley and Bernu [50] and Chin and Krotscheck [51].

Books are not common in QMC methods. I have already referred to the applications to Quantum Chemistry [39] as well as my own didactical paper [24]. Another review-like book is [46].

6.2 Basic Ideas of DMC

The basic ingredients of DMC are the following

1. It considers the Schrödinger equation in imaginary time,

$$-\frac{\partial \Psi(\mathbf{R}, t)}{\partial t} = [H(\mathbf{R}) - E]\Psi(\mathbf{R}, t) , \qquad (6.17)$$

where $\mathbf{R}$ represents the set of all coordinates. The formal solution of (6.17) is

$$\Psi(\mathbf{R}, t) = e^{-[H - E]t}\Psi(\mathbf{R}, 0) \qquad (6.18)$$

where $\exp[-(H - E)t]$ is called the *Green function*, and E is a convenient energy shift.

2. The wave function is positive definite everywhere, as it happens with the ground state of a bosonic system, so it may be considered as a probability distribution function.

3. The wave function is represented by a set of *random vectors* or *walkers* $\{\mathbf{R}_1, \mathbf{R}_2 \ldots \mathbf{R}_{N_c}\}$, in such a form that the time evolution of the wave function is actually represented by the evolution of the set of walkers.

4. The actual computation of the time evolution is done in small time steps τ, and the Green function is accordingly approximated,

$$e^{-[H - E]t} = \prod_{i=1}^{n} e^{-[H - E]\tau} ,$$

where $\tau = t/n$.

5. The imaginary time evolution of an arbitrary starting state $\Psi(\mathbf{R}, 0)$, once expanded in the basis of stationary states of the Hamiltonian

$$\Psi(\mathbf{R}, 0) = \sum_{\nu} C_{\nu}\Phi_{\nu}(\mathbf{R})$$

is given by

$$\Psi(\mathbf{R}, t) = \sum_{\nu} e^{-[E_{\nu} - E]t} C_{\nu}\Phi_{\nu}(\mathbf{R}) ,$$

in such a way that the lowest energy components will have the largest amplitudes after a long elapsed time, and in the $t \to \infty$ limit the most important amplitude will correspond to the ground state (if $C_0 \neq 0$).

6. An important improvement of this scheme is the introduction of the *importance sampling*, which will be described below.

In conclusion, the scheme is quite simple: once one has found an appropriate approximation for the short time Green function, and a starting state has been determined, the job consists in representing the starting state by a collection of walkers and letting them evolve in time, i.e., obtaining a new

collection of walkers from the old collection of walkers, up to a time large enough so that only the g.s. amplitude is of importance.

There is an important question related to the imaginary time evolution. The evolution operator *is not unitary*, so that the norm of the state is not conserved in time. Actually, it may grow without limit or it may go to zero, depending on the value of E. If E equals the g.s. energy, $E = E_0$, then all amplitudes will go to zero at large time, with the exception of the amplitude of Φ_0. On the other hand, if $E > E_0$, all amplitudes will go to infinity and, conversely, for $E < E_0$, all amplitudes will go to zero. In any case, always the C_0 amplitude will dominate. A way of determining the g.s. energy is then to adjust E so that the norm of the state tends to a constant. This way is called, in MC language, the *growth estimator* of the energy.

6.3 The Green Function and the Small-Time Approximation

In coordinate representation the Green function is given by the matrix element

$$G(\mathbf{R}', \mathbf{R}, t) = \langle \mathbf{R}' | e^{-(H-E)t} | \mathbf{R} \rangle ,$$

and the time evolution equation is

$$\Psi(\mathbf{R}', t) = \int d\mathbf{R}\, G(\mathbf{R}', \mathbf{R}, t) \Psi(\mathbf{R}, 0) . \qquad (6.19)$$

At $t = 0$ the value of the Green function is

$$G(\mathbf{R}', \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R}) . \qquad (6.20)$$

From the operatorial representation of the Green function one may easily obtain the formal differential equation

$$-\frac{\partial G}{\partial t} = [H - E]G ,$$

which is transformed in coordinate representation into the differential equation

$$-\frac{\partial G(\mathbf{R}', \mathbf{R}, t)}{\partial t} = \left[-\frac{\hbar^2}{2m} \sum_i \nabla_i^2 + \sum_{i<j} V_{ij} - E \right] G(\mathbf{R}', \mathbf{R}, t)$$

with the boundary condition (6.20). This equation is very similar to (6.17), so we have not progressed too much. The important point, however, is that one may obtain approximations to the value of the Green function for short time intervals.

The main problem in obtaining the Green function is the non commutativity of the kinetic energy operator and the potential energy operator. However, the individual pieces, i.e., the Green functions related exclusively to the kinetic or potential operators, may be readily determined.

For the kinetic operator we have the coordinate representation

$$G_K(\mathbf{R}', \mathbf{R}, t) = \langle \mathbf{R}' | e^{-D\nabla^2 t} | \mathbf{R} \rangle \equiv \frac{1}{(2\pi)^{3N}} \int d\mathbf{k}\, e^{-i\mathbf{k}\mathbf{R}'} e^{-Dtk^2} e^{i\mathbf{k}\mathbf{R}} ,$$

where last equation corresponds to the momentum representation of the position bra and ket states. Here $D = \hbar^2/2m$ is called the *diffusion constant* and all vectors have $3N$ components.

The integral is easily carried out, obtaining

$$G_K(\mathbf{R}', \mathbf{R}, t) = \frac{1}{[4\pi Dt]^{3N/2}} e^{-(\mathbf{R} - \mathbf{R}')^2/4Dt} .$$

It is easily checked that this Green function is the Dirac delta function $\delta(\mathbf{R} - \mathbf{R}')$ in the $t \to 0$ limit.

The case of the Green function related to the potential is even simpler, for momentum-independent interactions. The potential is then a *local operator* and there results

$$G_V(\mathbf{R}', \mathbf{R}, t) = e^{-V(\mathbf{R})t} \delta(\mathbf{R} - \mathbf{R}') .$$

Approximate forms for the full Green function are obtained by using the following approximations

$$\begin{aligned}
\langle \mathbf{R}' | e^{-(K+V)\tau} | \mathbf{R} \rangle &= \langle \mathbf{R}' | e^{-K\tau} | \mathbf{R}'' \rangle \langle \mathbf{R}'' | e^{-V\tau} | \mathbf{R} \rangle + O(\tau^2) \\
&= \langle \mathbf{R}' | e^{-V\tau} | \mathbf{R}'' \rangle \langle \mathbf{R}'' | e^{-K\tau} | \mathbf{R} \rangle + O(\tau^2) \\
&= \langle \mathbf{R}' | e^{-V\tau/2} | \mathbf{R}'' \rangle \langle \mathbf{R}'' | e^{-K\tau} | \mathbf{R}''' \rangle \langle \mathbf{R}''' | e^{-V\tau/2} | \mathbf{R} \rangle + O(\tau^3)
\end{aligned}$$

In all cases an integration over the internal coordinates $\mathbf{R}''$ or $\mathbf{R}'''$ must be understood, being easily carried out by using the delta functions appearing in G_V. The practical equations are

$$\begin{aligned}
G(\mathbf{R}', \mathbf{R}, \tau) &= G_K(\mathbf{R}', \mathbf{R}, \tau) e^{[E - V(\mathbf{R})]\tau} + O(\tau^2) \\
&= e^{[E - V(\mathbf{R}')]\tau} G_K(\mathbf{R}', \mathbf{R}, \tau) + O(\tau^2) \\
&= e^{[E - \{V(\mathbf{R}') + V(\mathbf{R})\}/2]\tau} G_K(\mathbf{R}', \mathbf{R}, \tau) + O(\tau^3)
\end{aligned}$$

In these equations we have also added the energy shift appearing in the original time-dependent equation. It should also be noted that the average of the first two forms also generates an $O(\tau^3)$ approximate Green function,

$$G(\mathbf{R}', \mathbf{R}, \tau) = G_K(\mathbf{R}', \mathbf{R}, \tau) \frac{e^{[E - V(\mathbf{R})]\tau} + e^{[E - V(\mathbf{R}')]\tau}}{2} + O(\tau^3) .$$

The Small-Time Limit. It is quite simple to check formally the validity of the previous forms of the short time Green function in the $\tau \to 0$ limit.

For this purpose we require the expansion of the kinetic part of the Green function in the small-time limit. This expansion, in the simpler one-dimensional problem, is given by

$$\frac{1}{\sqrt{4\pi D\tau}}e^{-x^2/4D\tau} = \delta(x) + D\tau\delta''(x) + \frac{1}{2}D^2\tau^2\delta^{iv}(x) + O(\tau^4) \ .$$

To obtain this strange expansion one should consider the integral of the gaussian times a smooth function $f(x)$, and Taylor expand this function in powers of x. The resulting integrals are of the form $\int_{-\infty}^{\infty} x^n exp(-x^2)$ multiplied by the n-th derivatives of $f(x)$. Only the $n =$ even terms survive.

By using the expansion of the kinetic Green function, and when stability has been reached, i.e., $\Psi(x,\tau) = \Psi(x,0) \equiv \Psi(x)$, the time evolution equation becomes

$$\Psi(x') = e^{[E-V(x')]\tau}\Psi(x') + D\tau\frac{d^2}{dx^2}\left[e^{[E-V(x')]\tau}\Psi(x)\right]_{x=x'} + O(\tau^2)$$

Expanding the term with second derivatives up to $O(\tau^2)$, there results

$$\Psi(x') = \Psi(x') + [E - V(x')]\tau\,\Psi(x') + D\tau\,\Psi''(x') + O(\tau^2) \ .$$

The coefficient in τ^0 of this equation is null, so that effectively the equality becomes an $O(\tau)$ expansion. The leading term is exactly the Schrödinger equation

$$-D\Psi''(x) + V(x)\Psi(x) = E\Psi(x) + O(t) \ .$$

In conclusion, an $O(\tau^2)$ Green function has generated energies which are only $O(\tau)$ accurate.

A similar procedure, but more involved, may be carried out for the $O(\tau^3)$ approximations, showing that the energies obtained are accurate up to $O(\tau^2)$.

6.4   The One-Dimensional Harmonic Oscillator

The purpose of this section is to understand the meaning and use of the short-time Green functions defined above. We will not use here MC methods. Instead, all calculations will be done analytically.

Consider the simple Hamiltonian

$$H = p^2 + x^2 \ ,$$

which corresponds to $D = 1$, and a starting wave function, normalized as a probability distribution function, of gaussian shape

$$\Psi(x,0) = \sqrt{\frac{a}{\pi}}e^{-ax^2} \ .$$

Here a is an arbitrary parameter, but not of variational nature: it just fixes the starting state.

By applying the $O(\tau^3)$ Green function there results a new wave function also of gaussian form, i.e.,

$$\Psi(x,\tau) = \sqrt{\frac{a'}{\pi}}e^{-a'x^2}\left[e^{E\tau}\sqrt{\frac{a}{a + \tau + 2a\tau^2 + \tau^3}}\right] \ , \qquad (6.21)$$

where

$$a' = \frac{a + \tau + 2a\tau^2 + \tau^3}{1 + 2\tau^2 + 4a\tau} \ ,$$

so that, apart from the change in the normalization, a time step is equivalent to the replacement

$$a \leftarrow \frac{a + \tau + 2a\tau^2 + \tau^3}{1 + 2\tau^2 + 4a\tau} \ . \qquad (6.22)$$

Repeated time steps correspond to repeated application of this replacement rule. If the iteration tends to a limiting value, the limiting value is obtained by solving the equation obtained from (6.22) by putting an equal sign instead of the arrow. The solution, i.e., the stabilized gaussian parameter $\tilde{a}$, is

$$\tilde{a} = \frac{\sqrt{1 + \tau^2}}{2}$$

and if the time-step is close to 0, it turns out to be $\tilde{a} = 1/2$, the well known exact result. The first correction has the value $\tau^2/4$, so even if the approximate Green function has an $O(\tau^3)$ error, the resulting wave function, after stability, has an error $O(t^2)$, confirming the results of the calculation of $\tau \to 0$ limit of the last section. For example, when $\tau = 0.01$, the stabilized value is $\tilde{a} = 0.500025$, and for $\tau = 0.1$ its value is $\tilde{a} = 0.5025$.

There are some interesting questions easily answered with the help of this simple example, which are briefly commented.

The Stability Time. The question is how many iteration steps have to be done in order to arrive to stability. Of course, the answer will depend on the starting value of a but *does not depend on the value of the time step τ*.

This is shown in Fig. 6.4 for two starting values of the gaussian parameter a and for two values of the time step. It is worth noting that the case corresponding to $a = 1$ at the beginning (lhs plot) reaches the stability noticeably before the case corresponding to start at $a = 0$ (rhs plot). This simply indicates that the starting wave function with $a = 1$ is closer to the exact g.s. wave function than the case of $a = 0$.
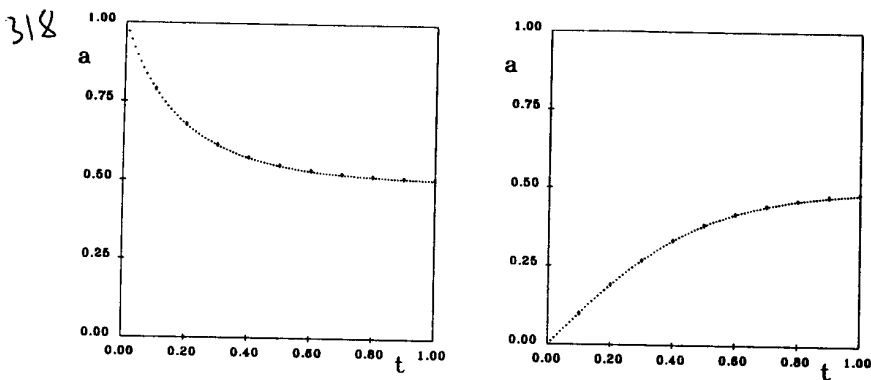
Fig. 6. Time evolution of the parameter a. The figure at the left starts at $a = 1$, and the figure of the right starts at $a = 0$. In Monte Carlo language, the second case corresponds to a starting wave function represented by random points uniformly distributed in $[-\infty, \infty]$. The small dots correspond to the calculation with a time step $\tau = 0.01$, and the big dots to $\tau = 0.1$. The x axis is the total evolved time. The lines end up at different values of a, but this cannot be appreciated in the figure.

The Mixed Estimator for the Energy. The value of the energy is readily obtained from the stabilized solution by computing the integral

$$E = \frac{\int H\Psi \, dx}{\int \Psi \, dx}$$

and applying the Hamiltonian on the bra state, whose value is 1. Note that this equation is exact when Ψ is the *exact ground state* wave function. If it is not, the equation is not even an upper bound, in the variational sense.

Given that the second derivative of 1 is null, one may replace H by $V(x) = x^2$ and compute the integral. The result is

$$E = \frac{2}{2 + \tau^2} \ ,$$

and the Taylor expansion is $1 - \tau^2/2 \cdots$, showing the $O(\tau^2)$ error term. Here we check that the energy estimator has given a value *smaller* than the exact energy (which is 1), showing the non-variational character of the mixed estimator.

Of relevance for the Monte Carlo way of dealing with this method is the computation of the variance. This requires to compute the matrix element of x^4, and the resulting value for the variance is

$$\sigma = \frac{1}{\sqrt{2}a} \ ,$$

and the stabilized limit is $\sigma = \sqrt{2/(1 + \tau^2)}$. Knowing a priori the value of the variance will help in deciding the number of samples to compute, in Monte

Carlo method, so as to have a prescribed value for the error. In our case, the variance is quite large, and for sufficiently small values of the time step τ its value is close to $\sqrt{2}$. If we would like to obtain the energy with an error of 10^{-3} then the required number of samples will be of two millions. This drawback shown in the simple example may be corrected by using an appropriate importance sampling algorithm.

This way of computing the expectation value of the energy has been called a *mixed estimator*, because one could have introduced as bra state any other trial wave function instead of just putting 1. In general, the mixed estimator is given by

$$E = \frac{\int \Psi_T^* H\Psi \, dx}{\int \Psi_T^* \Psi \, dx} \ ,$$

or, in terms of the same local energy introduced in the description of the variational method, the mixed estimator is

$$E = \frac{\int \Psi_T^* E_L(x)\Psi \, dx}{\int \Psi_T^* \Psi \, dx} \ .$$

If the trial function is very close to the exact g.s. wave function, $E_L(x)$ will be a rather flat function and the variance of the integral will be smaller than in the case of putting $\Psi_T = 1$. This will not affect the analytic calculations, however, but will be of primarily importance in the MC evaluation.

We suggest the reader to carry out a simple calculation using for the trial function $\Psi_T = \exp(-x^2)$, and checking that the energy expectation value is the same, but the variance has changed.

There is another alternative for this algebraic calculation which, however, cannot be used in Monte Carlo methods, namely to act with the Hamiltonian on the ket state. In the $\tau = 0$ limit the ket state is an eigenstate of the Hamiltonian, with eigenvalue 1 and the variance is null. This has no interest in DMC calculations because one does not know how to compute the second derivative of a function represented by a set of walkers.

The Normalization of the Solution. After each step the resulting gaussian is not, in general, normalized. There is a correcting factor which appears in Eq. (6.21) in square braces. This factor may be used to determine the energy E by requiring it to be 1. This way of evaluating the eigenvalue is known as the *growth estimator* of the energy. Once the stability has been reached, the equation to be solved is

$$e^{E\tau} = \tau + \sqrt{1 + \tau^2}$$

and the Taylor expansion gives $E = 1 - \frac{\tau^2}{6} + \cdots$, so that again an $O(\tau^2)$ error for the energy with this new method.

Importance sampling has a relevant role in QMC calculations. As we have shown in the previous section, its use may help in lowering the variance of the energy estimators, reducing the required number of samples for a given value of the desired error.

However, this is not the only role of importance sampling. Actually, the potential part of the Green function depends on the potential $V(\mathbf{R})$, and when the potential is large and negative, the exponential may take unbounded values.

One of the ways of dealing with importance sampling in QMC is to consider the (imaginary) time evolution of the new quantity

$$f(\mathbf{R}, t) = \Psi_T(\mathbf{R})\Psi(\mathbf{R}, t) , \tag{6.23}$$

where Ψ is the wave function satisfying the Schrödinger equation and Ψ_T is a trial wave function, time independent, which should preferably be close to the exact ground state wave function. It is not necessarily the starting wave function, even if normally it is taken also as the starting wave function.

The time evolution of $f(\mathbf{R}, t)$ is given by the differential equation

$$-\frac{\partial f(\mathbf{R}, t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 f(\mathbf{R}, t) + \frac{\hbar^2}{2m}\nabla \cdot [\mathbf{F}(\mathbf{R}) f(\mathbf{R}, t)] + [E_L(\mathbf{R}) - E]f(\mathbf{R}, t) , \tag{6.24}$$

where we have used again the *local energy*, which was already defined in the section devoted to variational MC methods.

A new quantity appearing in the evolution equation is the so called *drift force* (it should better be called drift velocity), given by

$$\mathbf{F} = \frac{2}{\Psi_T(\mathbf{R})}\nabla \Psi_T(\mathbf{R}) . \tag{6.25}$$

The transformed Hamiltonian acting on f in the rhs of Eq. (6.24) may be written as the sum of three terms

$$H = K + F + L$$
$$K = -\frac{\hbar^2}{2m}\nabla^2$$
$$F = \frac{\hbar^2}{2m}[(\nabla \cdot \mathbf{F}(\mathbf{R})) + \mathbf{F}(\mathbf{R}) \cdot \nabla]$$
$$L = E_L(\mathbf{R})$$

corresponding respectively to the kinetic part, the drift part and the local energy part. To obtain the previous non importance sampling formula, just put $\Psi_T = 1$, $F = 0$ and $E_L(\mathbf{R}) = V(\mathbf{R})$.

One may define as above a Green function corresponding to this transformed operator, obeying a similar, but more complex, differential equation. Formally, the new Green function is

$$G = e^{-[K+F+L-E]t} ,$$

and an $O(\tau^3)$ approximation may be obtained from

$$G = e^{E\tau}e^{-L\tau/2}e^{-F\tau/2}e^{-K\tau}e^{-F\tau/2}e^{-L\tau/2} + O(\tau^3) .$$

It is important to maintain the order of the operators.

In order to get the coordinate representation one must first obtain the representation of the three pieces, referred as G_L, G_F and G_K. The first has the same form as G_V above, simply putting E_L instead of V. The G_K term is the same as above.

The most complicated term corresponds to the drift force. Its explicit form is (see [52])

$$\langle \mathbf{R}'|G_F|\mathbf{R}\rangle = \delta(\mathbf{R}' - \mathcal{R}(t)) , \tag{6.26}$$

where $\mathcal{R}(t)$ is the *trajectory* obtained by integrating the differential equation

$$\frac{d\mathcal{R}(t)}{dt} = D\mathbf{F}(\mathcal{R}(t)) \tag{6.27}$$

with the boundary condition $\mathcal{R}(0) = \mathbf{R}$. In equation (6.27) the quantity $\mathbf{F}(\mathcal{R}(t))$ of the rhs must be understood as a (vector) function $\mathbf{F}$ of the function to be integrated $\mathcal{R}(t)$. On the other hand, $\mathbf{R}$ is the initial position. Eq. (6.27) shows that the right name for $\mathbf{F}$ is *velocity*, instead of *force*.

Understanding the Drift Green Function. The form just discussed for the Green function G_F related to the drift force is quite involved, and an explicit example will clarify its structure.

Assume a one-dimensional problem with a drift force $F(x) = -x$. This kind of force will appear when the trial wave function has a gaussian shape, specifically for $\Psi_T = \exp(-x^2/4)$, accordingly with Eq. (6.25). Note that the drift always points in the direction in which the trial function grows, and its actual role will be to displace the walkers towards regions where Ψ_T is large.

To obtain the Green function we must solve the differential equation (6.27). Taking $D = 1$ for simplicity and given that the value of the drift force is just its argument with a minus sign, there results the simple differential equation

$$\frac{d\mathcal{R}(t)}{dt} = -\mathcal{R}(t) ,$$

whose solution is

$$\mathcal{R}(t) = Re^{-t} ,$$

satisfying the boundary condition that at $t = 0$ its value is R. So, the drift Green function Eq. (6.26) is

$$\langle R'|G_F|R\rangle = \delta(R' - Re^{-t}) .$$

To check this result we must go back to the integral form of the time evolution equation, $f(R',t) = \int \langle R'|G_F|R \rangle f(R,0)dR$. To carry out the integration on R we should follow the usual steps of solving the argument of the delta function for R and introducing the jacobian of the argument. In this way there results

$$f(R',t) = e^{-t}f(R'e^{-t},0) \ ,$$

where $f(R',0)$ is the boundary condition, i.e., the value of $f(R',t)$ at $t=0$. This is an *arbitrary function*, which may be explicitly defined by saying, e.g., that at $t=0$ our function is some $h(x)$. Then the solution at t is

$$f(R',t) = e^{-t}h(R'e^{-t}) \ .$$

The last step required to be convinced that the obtained result is correct is to go back to the importance-sampling differential equation (6.24), removing the first and the last terms of the rhs, and check that our $f(R,t)$ satisfies it. For our force $F = -R$ the equation reads

$$-\frac{\partial f(R,t)}{\partial t} = \frac{\partial[-Rf(R,t)]}{\partial R}$$

and it is satisfied for any value of the starting function $h(R)$.

6.6 The Explicit Importance-Sampling Green Function

The form of the Green function involving the computation of the classical trajectory is not very much practical. One may obtain a simpler form by solving the classical equation up to some given order in τ, resulting in the explicit Green function [52]

$$\langle R'|G|R \rangle = \frac{1}{(4\pi D\tau)^{3N/2}} e^{-\{R' - R - D\tau[F(R') + F(R)]/2\}^2/4D\tau} \quad (6.28)$$

$$e^{\tau E - \tau[E_L(R') + E_L(R)]/2} \det[I - D\tau H(R')/2] + O(\tau^3) \ ,$$

where I is the $3N \times 3N$ unit matrix, H contains the second derivatives of the logarithm of the importance sampling function

$$H_{ij} = \frac{\partial[2\log \Psi_T(R')]}{\partial R_i \partial R_j}$$

and the determinant of the matrix with I and R' is the $O(\tau^3)$ normalization condition of the gaussian part of the Green function (when integrated with respect to R).

A simpler $O(\tau^2)$ form of the importance-sampling Green function is [38]

$$\langle R'|G|R \rangle = \frac{1}{(4\pi D\tau)^{3N/2}} e^{-\{R' - R - D\tau F(R)\}^2/4D\tau} \quad (6.29)$$

$$e^{\tau E - \tau[E_L(R') + E_L(R)]/2} + O(\tau^2) \ .$$

The use of any of these forms of the Green function was already discussed in section **2.5** as an improvement of von Neumann acceptance-rejection method. In both cases there is a gaussian-like part, properly normalized, and a rate term represented by the exponential involving the energy and the local energies. The Monte Carlo interpretation of the integral equation with these Green functions is an algorithm to obtain new points R' (none, one or more) from each member in the old set $\{R\}$ by applying the exponential rate term to R' obtained after a gaussian isotropic diffusion centered at $R + D\tau F(R)$ (case of eq (6.29)). This last step requires first to displace the starting point R by $D\tau F(R)$ and afterwards to select another displacement governed by the normal distribution characterized from the parameter $\sigma = \sqrt{2D\tau}$.

The first displacement is known with the name of *drift*, and is related to our trial function Ψ_T: it points always in the growing direction of the trial function, so as to bias the distribution of final points by increasing their number in the regions where the trial function is larger. This drift is modulated by the rate term, which will favor the areas where the local energy is smaller, i.e., favoring the g.s. components.

In the case of the more complex $O(\tau^3)$ Green function (6.28) there is a double drift, before and after the gaussian diffusion, with a half step.

These forms of the Green function are the most frequently used. However, they have not been generalized for the use in the coupled-channel problems. An alternative way of introducing importance sampling is by multiplying the time evolution integral equation (6.19) by the time-independent trial function,

$$\Psi_T(R')\Psi(R',\tau) = \int [\Psi_T(R')\langle R'|G|R \rangle \Psi_T^{-1}(R)] \Psi_T(R)\Psi(R,0)dR \ ,$$

where $\langle R'|G|R \rangle$ is one of the forms shown in section **6.3**. As in the previous formalism, the quantity with evolves in time is again $f(R,\tau) = \Psi_T(R)\Psi(R,\tau)$, but the new Green function is $\Psi_T(R')\langle R'|G|R \rangle \Psi_T^{-1}(R)$.

We have already mentioned the role of the importance sampling trial function in regard with the bias of the random walk which selects the areas where the trial function is larger and the trial energy is smaller. Another important result is that the stabilized population of walkers refers to the quantity $\Psi_T(R)\Psi(R,t)$, which is required to compute the mixed estimator of the energy introduced in section **6.4**. To compute the energy one simply averages the values of the local energy $E_L(R)$ over the set of walkers. At the same time, one may also obtain an estimate of the variance.

A final role of the importance sampling method is the ability of dealing with states whose wave function is *not* positive definite. The way of proceeding is to assume that the new probability distribution function $f(R,t)$ will be positive or null, i.e., tying the nodes of the seeked wave function to the nodes of the trial function. This means to put impenetrable walls along the nodal surfaces defined by the zeros of the trial function, and this means that any time a walker tries to cross one of these walls, it is killed. This approach,

known as the *fixed node* approximation, is of fundamental importance in the case or fermionic systems.

6.7 QMC and Numerical Methods

The MC evaluation of the time evolution equation by means of a combination of drift and diffusion is a very clever way of computing a quite complicated integral, and here we would like to mention the relation of the MC method with the Gauss-Hermite quadrature rule.

The integral to be carried out has a piece consisting in a very narrow gaussian, characterized by a width $\sigma = \sqrt{2D\tau}$. The validity of the step-by-step time evolution requires σ to be a small quantity, much smaller than the typical dimension characterizing the system, so that it is a very peaked function. If one attempts to compute numerically the integral in some one-dimensional problem by means of the usual trapezoidal or Simpson rules, the required number of points must be very large, making the calculation quite time consuming.

The proper way of computing numerically the evolution integral is by means of an algorithm like the one shown in the section **6.5**, i.e., using the delta-function expansion of the gaussian. This method is very similar to the Gauss-Hermite quadrature rule, where a given integral involving a gaussian is approximated by

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx = \sum_k w_k f(x_k) \ ,$$

the weights w_k and the points x_k being chosen such as to have an exact rule for the highest possible polynomial. This method is conceptually equivalent to the MC sampling algorithm, where the new random points are selected accordingly with the (conditional) gaussian distribution.

The small-time approximation to the Green function is not very much adequate for ordinary quadrature rules, but is well tailored for MC methods.

6.8 The DMC Algorithm

Here we collect all results presented in the previous section and present the basic algorithm of DMC method.

Previous to the preparation of a computer code an importance sampling trial function must be determined. Adequate procedures must be devised to compute the local energy and the logarithmic gradient of the trial function.

The remaining tasks are written orderly:

Initialize:
- Get a value for the time step. This time step should be varied and from the set of pairs (time step, energy) carry out a least-squares extrapolation to zero.

- Guess a value for the energy shift E
- Create a starting set of walkers, representing the function $f(\mathbf{R}, t = 0)$. Normally, the importance trial function is used as a variational trial function, and both E and the starting set of walkers are determined by a previous VMC Metropolis–like algorithm.
- For each walker $\mathbf{R}$ compute and store the local energy and the drift function
- Initialize accumulators for energy and its square, as well as for other observables

Time Evolution Loop: Repeat up to large enough time the next steps

Thermalization Loop: Successive positions are strongly correlated, because the time step has to be very small. So, go on doing moves but only count a subset of them to get the averages

Loop over Old Walkers:
- Generate a trial move

$$\mathbf{R}' = \mathbf{R} + Dt\mathbf{F}(\mathbf{R}) + \boldsymbol{\xi}$$

where $\boldsymbol{\xi}$ is drawn from the multivariate gaussian distribution with null mean and $\sigma = \sqrt{2Dt}$
- Compute the *replication factor*

$$n = [\exp\{t(E_L(\mathbf{R})/2 + E_L(\mathbf{R}')/2 - E)\}] \ ,$$

where the square brackets mean *integer part*.
- Unless $n = 0$ make n copies of $\mathbf{R}'$, $E_L(\mathbf{R}')$ and $\mathbf{F}(\mathbf{R}')$ to the new set of walkers

End of Loop over Walkers

New Generation:
- Copy the new set to the old set
- Test the growth of the population, and adjust the energy E

End of Thermalization Loop

Update Accumulators: Once the stabilization time has elapsed, update the accumulators for averages and errors

End of Time Iterations Loop

Final Work: Compute averages, errors ... and print results

6.9 Example: the ⁴He Nucleus

Below there is a simplified code for a practical calculation of the g.s. energy of the ⁴He nucleus described by a Wigner nucleon interaction. It is the same problem described in section **5.5** and uses the same routines described there for the calculation of the local energy as well as the gradient required for the determination of the drift term.

The code has been simplified by defining many parameters in a FORTRAN PARAMETER statement, just to reduce the number of lines. Among the

326 required improvements, there is one of particular concern, namely the control on the population growth. There are two variables, NWstart and NWend, representing the number of walkers before and after a time step respectively, which may grow or decline in correspondence with the value of the energy E0, and it should be convenient to add a procedure to vary E0 so as to maintain a stabilized population.

Two sets of time steps have been distinguished, labeled by a positive or negative value of the counter itime, so as to discard the energy averages obtained during the stabilization time. There is also a call to a RANGAUSS routine which generates random numbers from the $N[0,1]$ normal distribution.

The code has two phases, the initialization and the time evolution. The starting set of walkers has been obtained by generating nucleon coordinates at random. In a serious application the Metropolis algorithm based on the trial function should have been used to get both a good starting set of walkers and an estimate of the parameter E0.

The main MC operation is carried out by the statement labeled with the number 40, which selects a possible destination point obtained from the previous point by adding the drift (Grad) and the isotropic diffusion (rangauss). Branching is decided by computing the variable Nsons which is an integer number from a pdf taking two values, $n_1 = [\exp(tE - t[E_L(\mathbf{R}) + E_L(\mathbf{R}')]/2)]$ and $n_2 = n_1 + 1$ with mean value $\exp(tE - t[E_L(\mathbf{R}) + E_L(\mathbf{R}')]/2)$ (see section 2.5).

```
c      ================================================
c      Diffusion Monte Carlo for He-4 with S3 force
c      Must be linked with ELOCAL and WF subroutines
       include 'varhe.inc'
       parameter(NWmax=3000, Nstab=20, Ntime=100, Nth=10)
       dimension Walker (NA,3,NWmax,2), EL(NWmax,2),
     1           Grad(Na,3,NWmax,2)
       parameter(D=h2div2m, t=0.001d0)
c      ============Starting configuration ==================
       sig=sqrt(2*D*t)              ! Parameter for gauss distrib.
       z=randini(1234.5d0)          ! Initialize gaussian generator
       NWstart=1000                 ! Starting population
       E0=-27                       ! and guess energy
       IN=1                         ! To swap OLD and
       IO=2                         ! new generations
       EAV=0
       do 10 kw=1, Nwstart          ! Raw initial set of
       do 20 i=1, NA                ! walkers
       do 20 j=1, 3
20     Walker(i,j,kw,IN)=rangauss()
       wfold=WF (Walker(1,1,kw,IN))
10     EL(kw,IN)=ELOCAL(wfold, Walker(1,1,kw,IN), Grad(1,1,kw,IN))
c      ============= START ITERATIONS ===========================
       EF=0                         ! To compute averages and
```

```
       EF2=0                        ! error
       do itime=-Nstab, Ntime       ! Nstab to reach stability
       do ith=1, Nth                ! Thermalization loop
       eav=0
       NWend=0                      ! New set of walkers
       do 30 kw=1, NWstart
       do 40 ip=1, Na               ! Diffusion and drift
       do 40 j= 1, 3                ! O(t**2) Green function
40     R(ip,j)=Walker(ip,j,kw,IN)+D*t*Grad(ip,j,kw,IN)+sig*rangauss()
       wfold=WF(R)
       ELnew=ELOCAL(wfold, R , DRIFT)
       Nsons=exp((EO-(ELnew+EL(kw,IN))/2)*t)+RAND()     ! Branching
           do 60 isons=1, nsons
           NWend=NWend+1
           do 50 ip=1, Na
           do 50 j=1, 3
           walker(ip,j,NWend,IO)=R(ip,j)
50         GRAD  (ip,j,NWend,IO)=DRIFT(ip,j)
           EL(NWend,IO)=ELnew
60         eav=eav+ELnew
30     continue
       IN=3-IN                      ! Swap OLD/NEW generations
       IO=3-IO
       NWstart=NWend
       write(*,*) NWend, eav/NWend  ! Monitor evolution
       enddo                        ! End of thermalization loop
       if(itime.gt.0) then
       EF=EF+eav/NWend              ! Compute averages only
       EF2=EF2+eav**2/NWend**2      ! after stability
       endif
       enddo                        ! End of TIME iterations
c      ==============   Compute and print results ============
       EF=EF/Ntime
       EF2=EF2/Ntime
       write(*,*) ' Energy=', EF        ! Write results
       write(*,*) ' Error =', SQRT((ef2-ef**2)/Ntime)
       end
```

Results obtained with the above code are shown in figure 7. The wavy-like aspect of the time evolution is typical of this specific Markov process with a small time step, and shows how strong are the correlations between two successive moves. Thermalization is required in order to have a good estimate of the variance. The left part of the figure shows the $t = 0$ limit extrapolation, which results in a value of $E = -27.18 \pm 0.12$ MeV. To obtain the value mentioned in Table 4 of section 5.5 and in Ref. [15] a very long run will be required with this simplified code. It would be better to use a more precise importance sampling wave function [15].
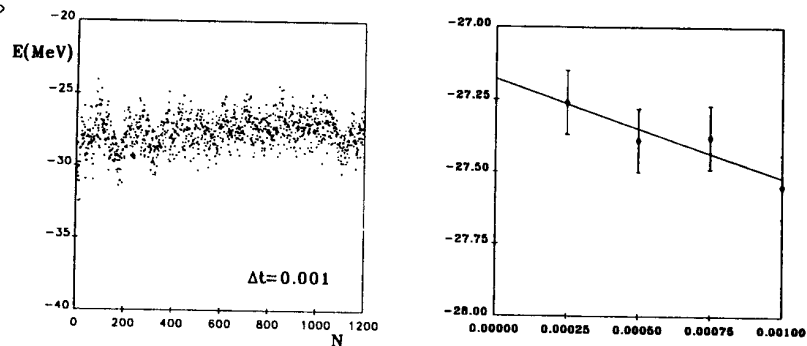
Fig. 7. The time evolution of the average energy (left) as a function of the number of time steps N and the evaluation of the average for different time steps (right). The continuous line represents a linear fit for $t = 0$ extrapolation.

6.10 Some Physical Systems

Here I briefly present some calculations which have mainly an historical interest, because they show the effort of many research groups to attain the maximum precision by using various many-body techniques. The possibility of obtaining essentially exact results by QMC methods was a strong stimulus to improve other theories, as we shall see.

More complete descriptions of nowadays studies in the QMC framework are discussed in this volume by S. Pieper, referred to the Argonne–Los Alamos–Urbana collaboration on the nuclear few-body problem, and by J. Boronat, applied to the study of zero temperature quantum liquids.

Jellium. One of the most impressive applications of the DMC method was carried out by Ceperley and Alder [37] who studied the system known as *jellium*, consisting in an electron gas on a uniform background of positive charge. This is a simple model of conductivity of metals, and is characterized by a single length parameter r_s, conveniently measured in units of Bohr radius a_0, which defines the electron density as

$$\rho = \frac{3}{4\pi a_0^3 r_s^3} \ .$$

Many years ago Wigner [53] proved that this system will have a crystal structure, body centered cubic, at very low densities (large r_s).

The table below presents a set of results for this system, computed with several many-body theories. This is a nice example of several many-body theories properly working in the domain in which they are adequate: RPA and theories, like CCM, containing RPA, are well suited in the high density regime. FHNC is appropriate at densities not very large.

Table 6. Correlation energy per particle in the jellium model, computed with several theories. RPA is the *random phase approximation*, CCM is the *coupled cluster method*, FHNC is the *fermionic hypernetted chain* approximation and HNC/Lado is a bosonization of the FHNC equation. Energies are quoted in millirydbergs.

Method	$r_s \to 0$	$r_s = 1$	$r_s = 3$	$r_s = 5$	$r_s = 20$
QMC [37]		-120		-56.3	-23.0
RPA [54]	$62.2 \log(r_s)$	-157	-105	-84.9	-42.8
CCM(2) [55]	$62.2 \log(r_s)$	-123	-75.1	-56.8	
CCM(4) [56]	$62.2 \log(r_s)$	-122	-73.8	-56.0	-23.6
FHNC [57]		-138	-79	-58	
HNC/Lado [58]	$57 \log(r_s)$	-114	-71.0	-54.1	-21.8
VMC [59]		-122	-72.2	-55.0	-22.5

It should be mentioned that the QMC calculations are not a single calculation: one of the unpleasant properties of the importance sampling trial function is that forces the walkers to adapt to its shape. So, the row labeled QMC is a subset of a series of calculations which assume for the trial function a paramagnetic (unpolarized) Fermi fluid, a ferromagnetic (polarized) Fermi fluid, and a bcc Wigner crystal, mentioned in order of decreasing density. It would be nice to have a theory able to select the appropriate shape for the wave function, but this goal has not yet been reached.

The Diatomic Molecule LiH. Atomic and molecular physicists are used to do calculation with a very high degree of precision. The energy scale in atomic physics is the Rydberg (13.6 eV) or the Hartree (2 Rydberg), and the theoretical studies attain precisions of the order of one thousandth of Rydbergs.

The LiH is a rather simple molecule, with two centers and four electrons, one from H and the other three from Li. In the Born–Oppenheimer approximation the centers are assumed to be fixed and the calculations are carried out for several values of the intercenter distance, quoting the minimum as the energy of the molecule. The quantity which is experimentally measured is the *dissociation energy*, which is the energy which must be provided so as to put the two atoms far away. The comparison with experiment requires two corrections: relativistic effects and zero-point vibrational motion of the atomic centers.

The results of several theoretical calculations are presented in Table 7, which shows the value of the *correlation energy*, the difference between the total energy of the molecule and the energy computed in the self-consistent field calculation. As in the previous case, the agreement between the various theoretical methods is the most important result of this comparison.

330

Table 7. The correlation energy, in millihartree, for the LiH molecule, as computed with several methods. CI is configuration interaction, CCM is coupled-cluster method. The VMC calculation quoted refers to a quite poor trial function, which was used as importance sampling for the DMC calculation.

Method	Comment	Ref.	Corr. energy
CI	Slater Orbitals	[60]	-76.5
CI	Elliptic Orbitals	[61]	-78.2
CCM	SUB(2) approx	[62]	-81.5
VMC		[38]	-54±8
DMC		[38]	-80±2
Exact	Mixed Theor. Exper.	[63]	-83.2 ±1

6.11 Fermionic Systems

There are two complications related to the study of fermionic systems. On one side, the g.s. fermionic wave function is not in general positive definite, the exception being systems of distinguishable particles, like the example of ^{4}He nucleus above presented. Then, it cannot be interpreted as a probability distribution function. The other complication arises from the spin/isospin degrees of freedom, which convert the Schrödinger equation into a set of coupled equations: relative phases between the various components may be negative or even complex, and again putting troubles to make a correspondence between the wave function and a probability distribution function.

The first case is usually solved in the so called *fixed node approximation*, followed by a *nodal relaxation*. The idea is to use the importance-sampling method and to assume that the zeros of the trial function, called *nodal surfaces*, define impenetrable barriers which cannot be crossed by the walkers: every time a walker attemps to cross one of these boundaries, is is killed, i.e., its *replication number* is made zero.

The fixed node approximation then solves the many-body Schrödinger equation in a series of non-intersecting domains subjected to Dirichlet boundary conditions. The problem is mathematically well posed, but any imaginable analytic solution will result, in general, in a different value for the energy. The stochastic MC method will generate some weighted average of these energies.

Nodal Surfaces. There are several questions difficult to answer related to the nodal surfaces. Considering the case of a polarized system, the space part of the trial wave function will be fully antisymmetric, and the equation $\Psi_T(\mathbf{R}) = 0$ may correspond to a single hypersurface or more than one. The set hypersurfaces must contain all the lower dimension *surfaces* or *hyperpoints* [64] defined by the equations $\mathbf{r}_i = \mathbf{r}_j$, $i < j$, required by Pauli principle. So,

questions like how many nodal surfaces, how many non-intersecting regions, and so on are normally very difficult to answer.

Intuition as well as simpler problems may be a bad allied. For example, one-dimensional systems have the peculiarity that the minimal Pauli conditions (now $x_i = x_j$) define actually full nodal surfaces. Also, symmetries of systems of few particles may define uniquely the nodal surfaces.

Example: The HeI Atom. Triplet states of helium atom have a fully symmetric spin wave function, and the space part must be fully antisymmetric.

The lowest energy triplet state has $L = 0$, and its most general wave function depends only on r_1, r_2 and $\cos(\theta_{12})$, as requested by the rotational invariance. Pauli principle requires the equality $\Psi(r_1, r_2, \cos(\theta_{12})) = -\Psi(r_2, r_1, \cos(\theta_{12}))$ which in turn implies that $r_1 = r_2$ must be a nodal surface. This does not exclude trial functions having other nodal surfaces, but it is reasonable to expect that the g.s. will have the smaller number of nodes compatible with the symmetries of the system.

In conclusion, the triplet $L = 0$ g.s. will have a single nodal surface in a three-dimensional space $\{r_1 \in [0, \infty], r_2 \in [0, \infty], \cos(\theta_{12}) \in [-1, 1]\}$ which is a plane along the $\cos(\theta_{12})$ axis with $r_1 = r_2$. The space will be divided in two regions, $r_1 < r_2$ and $r_1 > r_2$ and for a given point in one region there is a corresponding point in the other region obtained by applying the permutation (12), i.e., by exchanging r_1 and r_2. Any reasonable trial function will posess only this nodal surface.

A different situation happens with the triplet $L = 1$ states. Considering the $M = 1$ state, a trial wave function could be the one provided by the central field model

$$\Psi_T(L = 1, M = 1) = \det \begin{bmatrix} \psi_s(r_1) & \psi_s(r_2) \\ \psi_p(r_1)z_1 & \psi_p(r_2)z_2 \end{bmatrix} .$$

The description of the system involves now four coordinates, $\{r_1 \in [0, \infty], z_1 \in [-r_1, r_1], r_2 \in [0, \infty], z_2 \in [-r2, r2]\}$. If $\psi_s = \psi_p$ the nodal surface will be determined by the equation $z_1 = z_2$. However, a different nodal surface will result for ψ functions not satisfying the above equality.

These two examples have shown the two kinds of situations to face up: the nodal surfaces are fully determined by the symmetries of the system, or there may be families of nodal surfaces compatible with the symmetries of the system. Using the fixed node approximation in systems like the former will ensure to obtain exactly its energy. On the contrary, in systems where the nodal surfaces are not uniquely defined the fixed node approximation will not provide the exact value of the energy.

Ceperley Theorem. Ceperley [65, 38] and almost simultaneously Moskowitz and collaborators [66] proved that the energy obtained with the mixed estimator in the fixed node approximation is an upper bound to the fermion

g.s. energy. If the nodal surfaces of the trial function were the exact surfaces, then this upper bound will coincide with the g.s. energy.

The proof of the theorem is as follows. Assume that the nodal surfaces have divided the space in domains $\{D_k, \; k = 1, M\}$. The fixed node approximation solves the Schrödinger equation in each of these domains, with Dirichlet boundary conditions at the edges, providing values for the energy E_k and wave function Φ_k in each domain (actually, the QMC algorithm will only give a weighted average of the domain energies). From any of the domain solutions one may construct a fully antisymmetric wave function by applying the antisymmetrizer

$$\bar{\Phi}_k(\mathbf{R}) = \sum_P \epsilon_P \Phi_k(P\mathbf{R}) \; ,$$

having in this form an adequate variational trial function. The expectation value of the Hamiltonian with respect to $\bar{\Phi}_k$, which has the value E_k, will be an upper bound to the fermion g.s. energy.

A Counter Example. Just searching for an example to explain the meaning of Ceperley's theorem I found a counter example. Imagine a system of two particles in a one-dimensional space, with a translationally invariant interaction which, in addition, commutes with the parity operator. Concentrating on the relative wave function the exchange of the two particles is equivalent to the parity operation. Assume that we look for solutions *with a wrong nodal surface*, which in our case will just be a wrong node. Figure 8 shows, in the lhs part, the two solutions which are obtained, one of them (labelled R) with energy *higher* than the energy of the first excited state (which will be the *fermionic* gs), but the other (labelled L) will have an energy *lower* than the fermionic g.s. energy. This means that Ceperley's theorem cannot be satisfied, or, in other words, that the expectation value of the Hamiltonian with respect to the antisymmetrized wave functions cannot be the same than the domain energies.

The reason lies in the presence of discontinuities of the derivative of function R at the imposed nodal point and its image. The presence of these discontinuities, and the corresponding effect on the kinetic energy, was already realized in Ref. [38], but there it was assumed that the wave function will be null at those points. This is not the case for the trial function L of our example, but it is the case of the other wave function R.

The reason of having non-null values of the antisymmetrized wave function at the discontinuities is that the permutations may overlap a given domain with itself. It should be said that normally importance sampling functions with such wrong nodal surfaces will not be used, if one is able to recognize how unappropriate they are, as in our simple example.

In conclusion, either the theorem is wrong or some extra conditions must be added for its fulfillment. This is a point that should be studied more in depth.
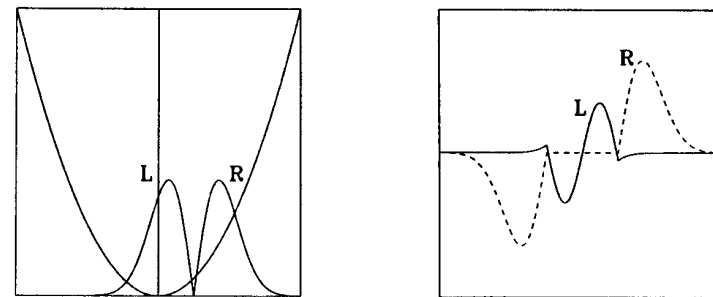

Fig. 8. Solution of a one-dimensional problem with a wrong nodal surface. The lhs shows the potential and the solutions confined in the left (L) and right (R) domain. The rhs shows the fermionic trial functions constructed from the domain solution by antisymmetrization.

Nodal Relaxation. The fixed-node QMC may be followed by a *nodal relaxation* process. Once the fixed-node MC walk has stabilized, one allows walkers to cross the nodal surfaces, adding a *signature* which indicates if their contribution to the energy should be added or substracted, accordingly with the number of crosses (even or odd) undergone.

The danger with crossing nodes is that the boson g.s. may start to grow up, basically because of the statistical noise. Then it may well happen that the variance of the result (nodes relaxed) is too high to ascertain if there has been a real gain in energy, so the extra work will be useless. Sooner or later, the boson g.s. will start to be the most important piece of the set of walkers, so the relaxation may only be continued for a limited number of time steps. Such fact is the reason for adding the adjective *transient* to the energy estimator. A qualitative discussion of the releasing of nodes may be fund in [67].

7 Closing Remarks

This is the end of these quite incomplete notes on Monte Carlo methods applied to many-body systems. The objective was to show the basic ideas, and adding examples was considered as a complementary didactic tool which, hopefully, will be appreciated by the reader.

Specific applications require some extra work. Even if MC methods are very powerful, they require our help (in other words, a good importance sampling driving function) to become really effective. This means that one must know in depth the physical system to be described, and this knowledge may be more complex than the mere MC techniques.

A last comment regarding to QMC methods, of special interest for new-commers to this technique: most of the effort in MC is actually devoted to have a variance as low as possible. Brute force is not the best way of diminishing the variance.

Acknowledgements

References

1. J. R. Newman, *The World of Mathematics* (Simon and Schuster, New York, 1956)
2. W. Feller, *An Introduction to Probability Theory and its Applications* (J. Wiley, New York, 1957)
3. B. Gnedenko, *The theory of Probability* (Mir Publ., Moscow, 1975)
4. J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods* (Methuen, Norwich, 1965)
5. M.H. Kalos and P.A. Whitlock, *Monte Carlo Methods* (J. Wiley, New York, 1986)
6. D.E. Knuth, *The Art of Computer Programmer*, Vol. II *Seminumerical Algorithms*, Chap. 3 (Addison Wesley, New York, 1971)
7. K. Binder and Stauffer, in *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder, (Springer–Verlag, Berlin 1984)
8. J.A. Carlson and R.B. Wiringa, in *Computational Nuclear Physics 1*, edited by K. Langanke, J.A. Maruhn and S.E. Koonin (Springer–Verlag, Berlin, 1991). The algorithm appears in the accompanying diskette, and is coded in FORTRAN.
9. G. Marsaglia and A. Zaman, Computers in Phys. **8**, 117 (1994)
10. W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge U. Press, Cambridge Mass., 1992), Chap. 7
11. I.R. Afnan and Y.C. Tang, Phys. Rev. **175**, 1337 (1968)
12. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. W. Teller , J. Chem. Phys. **21**, 1087 (1953)
13. Martin Gardner, *Mathematical Circus*, spanish translation *Circo Matemático* (Alianza, Madrid, 1985)
14. R. Jastrow, Phys. Rev. **98**, 1479 (1955)
15. R.F. Bishop, E. Buendía, M.F. Flynn and R. Guardiola, J. Phys. **G18**, L21 (1992)
16. R. Guardiola, P.I. Moliner and J. Navarro, Phys. Lett. **B383**, 243 (1996)
17. R. Malfliet and J.A. Tjon, Nucl. Phys **A127**, 161 (1969)
18. Y. Akaishi, Nucl. Phys. **A416**, 409 (1984)

19. H. Kameyama, M. Kamimura and Y. Fukushima, Phys. Rev. C **40**, 974 (1989)
20. A. Kievsky, M. Viviani and S. Rosati, *Condensed Matter Theories* 6, edited by S. Fantoni and S. Rosati (New York, Plenum, 1991) p. 391
21. J.G. Zabolitzky, Phys. Lett. **B100**, 5 (1981)
22. H. Kamada and W. Glöckle, Nucl. Phys. **A584**, 205 (1992)
23. W.W. Wood and F.R. Parker, J. Chem. Phys. **27**, 720 (1957)
24. R. Guardiola, in *First International Course on Condensed Matter*, edited by D. Prosperi, S. Rosati and G. Violini (World Sci. Singapore, 1988) p. 155
25. W.L. McMillan, Phys. Rev A **138**, 442 (1965)
26. S. Schiff and L. Verlet, Phys. Rev. A **160**, 208 (1967)
27. D.M. Ceperley and M.H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer Verlag, Berlin, 1979) p. 145
28. M.A. Preston, *Physics of the Nucleus*, chap. 5, Addison Wesley (Reading, Mass 1965)
29. B.H. Bransden and C.J. Joachain, *Physics of Atoms and Molecules* (Longman Sci. Tech., Essex, 1983)
30. F.M. Fernández and R. Guardiola, J. Phys **A30**, 5825 (1997)
31. N. Metropolis and S. Ulam, J.Am. Stat. Assoc. **44**, 335 (1949)
32. M.D. Donskar and M. Kac, J. Res. Nat. Bur. Standards **44**, 551 (1950)
33. M.H. Kalos, Phys. Rev. **128**, 1791 (1962)
34. J.B. Anderson, J. Chem. Phys. **63**, 1499 (1975)
35. J.B. Anderson, J. Chem. Phys. **65**, 4121 (1976)
36. J.B. Anderson, J. Chem. Phys. **73**, 3897 (1980)
37. D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980)
38. P.J. Reynolds, D.M. Ceperley, B.J. Alder and W.A. Lester Jr., J. Chem. Phys. **77**, 5593 (1982)
39. B.J. Hammond, W.A. Lester Jr. and P.J. Reynolds *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994)
40. M.H. Kalos, J. Comput. Phys. **2**, 257 (1967)
41. M.H. Kalos, Phys. Rev. A **2**, 250 (1970)
42. M.H. Kalos, D. Levesque and L. Verlet, Phys. Rev. A **9**, 2178 (1974)
43. D.M. Ceperley, G.V. Chester and M.H. Kalos, Phys. Rev. B **16**, 3081 (1977)
44. M.H. Kalos, M.A. Lee, P.A. Whitlock and G.V. Chester, Phys. Rev. B **24**, 115 (1981)
45. K. Schmidt and M.H. Kalos, in *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer–Verlag, Berlin, 1984), p. 125
46. S. Caracciolo and A. Fabrocini Eds., *Monte Carlo Methods in Theoretical Physics*, ETS Editrice (Pisa, 1991)
47. S.A. Chin, Phys. Rev. A **42**, 6991 (1990)
48. J. Carlson, Phys. Rev. C **36**, 2026 (1987)
49. J. Carlson, Phys. Rev. C **38**, 1879 (1988)
50. D.M. Ceperley and R. Bernu, J. Chem. Phys **89**, 6316 (1988)
51. S.A. Chin and E. Krotscheck, Phys. Rev. B **45**, 852 (1992)
52. J. Vrbik and S.M. Rothstein, J. Comput. Phys. **63**, 130 (1986)
53. E.P. Wigner, Phys. Rev. **46**, 1002 (1934)
54. R.F. Bishop and K.H. Lührmann, Phys. Rev. B **17**, 3757 (1978)
55. R.F. Bishop and K.H. Lührmann, Phys. Rev. B **26**, 5523 (1982)
56. K. Emrich and J.G. Zabolitzky, Phys. Rev. B **30**, 2049 (1984)

336

57. L. Lantto and P. J. Siemens, Nucl. Phys. **A317**, 55 (1979)
58. J.G. Zabolitzky, Phys. Rev. B **22**, 2353 (1980)
59. D.M. Ceperley, Phys. Rev. B **18**, 3126 (1978)
60. H. Partridge and S.R. Laughoff, J. Chem. Phys **74**, 2361 (1981)
61. D.M. Bishop and L.M. Cheung, J. Chem. Phys **78**, 1396 (1983)
62. B. Jeziorski, H.J. Monkhorst, K. Szalevicz and J.G. Zabolitzky, J. Chem. Phys **81**, 368 (1984)
63. K. Szalevicz, B. Jeziorski, H.J. Monkhorst, and J.G. Zabolitzky, J. Chem. Phys **79**, 5543 (1983)
64. D.J. Klein and H.M. Pickett, J. Chem. Phys **64**, 4811 (1976)
65. D.M. Ceperley, in *Recent Progress in Many–Body Theories*, edited by J.G. Zabolitzky, M. de Llano, M. Fortes and J.W. Clark, Lecture Notes in Physics **142** 262 (Springer Verlag, Berlin, 1981)
66. J. Moskowitz, K.E. Schmidt, M.A. Lee and M.H. Kalos, J. Chem. Phys. **77**, 349 (1982)
67. D. Ceperley and B.J. Alder, Science **231**, 555 (1986)

Monte Carlo Calculations of Nuclei

Steven C. Pieper

Physics Division, Argonne National Laboratory,
Argonne, IL 60439, USA

Abstract. Nuclear many-body calculations have the complication of strong spin- and isospin-dependent potentials. In these lectures I discuss the variational and Green's function Monte Carlo techniques that have been developed to address this complication, and present a few results.

1 Introduction

A major goal in nuclear physics is to understand how nuclear binding, stability, and structure arise from the underlying interactions between individual nucleons. To achieve this goal, we must both determine the Hamiltonian to be used, and devise reliable methods for many-body calculations with it. In principle quantum chromodynamics can prescribe the nuclear Hamiltonian, but it will be a long time before this will be done with useful accuracy. Thus the nuclear Hamiltonian is determined phenomenologically, and our knowledge of it is refined, in part, by the many-body calculations we make with it. A large amount of empirical information about the nucleon-nucleon scattering problem has been accumulated over time, resulting in ever more sophisticated *NN* potential models. These models have strong spin and isospin dependence, and spin and orbital angular momentum are mixed by a strong tensor interaction. In addition the three-nucleon interaction must be considered in realistic calculations, however there is very little experimental knowledge of it.

Thus the nuclear many-body Hamiltonian is significantly more complicated than those encountered in most atomic and condensed-matter problems, and progress in accurate nuclear ground state calculations has been quite slow. As Table 1 shows, it took 30 years to progress from the two-body ground state to the three-body one; the four-body problem was solved in a few more years. It then took some time until ongoing advances in computational resources, particularly the advent of massively parallel computers, allowed us to apply sophisticated quantum Monte Carlo methods to the study of light *p*-shell nuclei, up to 8-body nuclei.

Quantum Monte Carlo methods for central interactions are discussed in R. Guardiola's contribution to this volume; here I will concentrate on the complications due to the state-dependence of the nuclear forces. The nuclear Hamiltonian is presented in the next section, variational Monte Carlo (VMC) for nuclei is presented in Sect. 3 and Green's function Monte Carlo (GFMC) in Sect. 4. Finally a very few recent results are given in the last section. The