Introduction to stochastic process and quantum Monte Carlo for electronic structure theory

I. Introduction to stochastic process

II. The Metropolis algorithm

III. The Variational Monte Carlo (VMC) approach

- A. The method
- B. The trial wavefunction

IV The Diffusion Monte Carlo (DMC) approach

- A. The method
- B. The Fixed-Node approximation for fermions

V. The trial wavefunction optimization

Some preliminaries: Notion of random variable and stochastic process

Random variable

Definition

A random variable X is a variable subject to randomness. It can take on different values, each of them with some given probability. The fundamental quantity is the probability distribution of the random variable that gives all possible values with corresponding probabilities to occur.

Discrete random variable: $X \in \{1, 2, ..., N\}$, N finite or not. The probability distribution obeys

$$\boxed{P(X = i) = P_i \geq 0 \text{ and } \sum_{i=1}^{N} P_i = 1} \tag{1}$$

Continuous random variable (typically, $X \subseteq \mathbb{R}^d$). The probability distribution density [or probability density function (PDF)] obeys

$$\boxed{P(X = \mathbf{x}) = P(\mathbf{x}) \geq 0 \text{ and } \int d\mathbf{x} P(\mathbf{x}) = 1} \tag{2}$$

The uniform distribution over $(0, 1)$

$$P(x) = \begin{cases} 1, & \text{if } x \in (0,1), \\ 0, & \text{if } x \notin (0,1). \end{cases} \tag{3}$$

In practice, the uniform distribution is realized with Random Number Generators (RNG). Most generators are based on the use of a deterministic algorithm "mimicking" randomness as best as possible (pseudo-random generators). A common one is the simple linear congruential generator

$$x_{n+1} = (ax_n + c) \bmod m \tag{4}$$

where x_0 is defined as the "seed" of the generator. Note that once the seed has been chosen, the entire series of "random" numbers can be reproduced. A vast literature is devoted to the problem of producing randomness as pure as possible (minimization of correlations between pseudo-random numbers). A popular good quality-RNG has been proposed by L'Ecuyer,[**?**] see appendix 60.

The gaussian distribution over $(-\infty, +\infty)$

As a consequence of the central-limit theorem, the gaussian distribution is ubiquitous in real applications. The one-dimensional version is defined as

$$\boxed{P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]} \tag{5}$$

where μ is the mean of the distribution

$$\mu = \langle X \rangle = \int_{-\infty}^{+\infty} dx\; x\; P(x) \tag{6}$$

and σ^2 its variance

$$\sigma^2 = \langle (X-\mu)^2 \rangle = \int_{-\infty}^{+\infty} dx\; (x-\mu)^2\; P(x) \tag{7}$$

When $\mu = 0$ and $\sigma^2 = 1$, the distribution is known as the normal distribution. Generalization to an arbitrary dimension d is as follows

$$\boxed{P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d\, detC}} \exp\left[-\frac{1}{2}\sum_{i,j}(\mathbf{x}-\mu)_i\, C_{ij}^{-1}(\mathbf{x}-\mu)_j\right]} \tag{8}$$

where μ is the mean vector

$$\mu_i = \langle X_i \rangle, \tag{9}$$

and C the $d \times d$ covariant matrix defined as

$$\boxed{C_{ij} = \langle (\mathbf{x} - \mu)_i (\mathbf{x} - \mu)_j \rangle} \tag{10}$$

A simple and practical approach to sample the $1d$-gaussian distribution is to use the Box-Muller algorithm given by

$$\begin{cases} x = \sqrt{-2 \ln u_1} \cos(2\pi u_2) \\ y = \sqrt{-2 \ln u_1} \sin(2\pi u_2) \end{cases} \tag{11}$$

where u_1, u_2 are two uniform random numbers over (0,1). The two values x and y are independent and gaussian distributed. The generalization to the d-dimensional case is trivial after diagonalization of the covariant matrix and factorization of the probability distribution using the eigenvectors of C

Stochastic process

General stochastic process

Stochastic process $X(t)$ = Series of random variables indexed by a time t.

The fundamental quantities are the *n-time probability distributions*. In the continuous case, it is written as

$$P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) \tag{12}$$

with $0 \leq t_1 \leq t_2 \leq ... \leq t_n$, $\mathbf{x}_i$ denoting the state, or configuration, of the system at time t_i [typically, $\mathbf{x} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$, N number of particles]. The interpretation of the probability distribution density is as follows.

$$P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) d\mathbf{x}_1 d\mathbf{x}_2 ... d\mathbf{x}_n \tag{13}$$

is the probability of finding the system between $\mathbf{x}_1 + d\mathbf{x}_1$ at time t_1, $\mathbf{x}_2 + d\mathbf{x}_2$ at time t_2, etc with

$$\int d\mathbf{x}_1 d\mathbf{x}_2 ... d\mathbf{x}_n P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) = 1 \tag{14}$$

By integrating the n-time distribution over all states at k first times, we can generate $(n-k)$-time probability distribution densities

$$\boxed{P_{n-k}(\mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) = \int d\mathbf{x}_1 ... d\mathbf{x}_k P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n)} \tag{15}$$

Let us now define the *conditional probability densities* as follows

$$\boxed{P_{k|(n-k)}(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) = \frac{P_n(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k; \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n)}{P_k(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k)}} \tag{16}$$

With this definition

$$P_{k|(n-k)}(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) d\mathbf{x}_{k+1} d\mathbf{x}_{k+2} ... d\mathbf{x}_n \tag{17}$$

is the probability of finding the system between $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1}, ...,$\mathbf{x}_n + d\mathbf{x}_n$ at time t_n *knowing* that the system was at $\mathbf{x}_1$ at time t_1, $\mathbf{x}_2$ at time t_2,...,$\mathbf{x}_k$ at time t_k. Stochastic process are now classified according to the nature of their n-time probability distributions.

Fully decorrelated process: The case of the branching process Fully decorrelated process are the simplest stochastic process we can think of. They describe a time series of independent random variables. The probability of being between $\mathbf{x}_{k+1}$ and $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1}, knowing that we are at $\mathbf{x}_k$ at time t_k, is independent on $\mathbf{x}_k$ (and, then, on all previous states). In terms of conditional probability densities it is written as (for all possible k)

$$\boxed{P_{k|1}(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) = P_1(\mathbf{x}_{k+1}, t_{k+1})} \tag{18}$$

where $P_1(\mathbf{x}, t)$ is the probability distribution at time t, namely

$$P_1(\mathbf{x}, t) = \int d\mathbf{x}_2 ... d\mathbf{x}_n P_n(\mathbf{x}, t; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) \tag{19}$$

Using Eqs.(16) and (18) the n-time probability distribution can be written as

$$\boxed{P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2, ...) = \prod_k P_1(\mathbf{x}_k, t_k)} \tag{20}$$

Because of their simplicity and lack of time correlations such process are usually not very useful for modelizing physical situations. As a simple example, we could use them for describing the dynamics of a Brownian particle (pollen grain in water) when observation times t_k are separated by long time intervals (say, several minutes or more). Another more interesting exemple is the so-called branching or birth-death process as it is defined in DMC simulations.

Branching process.

We describe now the so-called "branching" or "birth-death" process as it is defined in QMC. It will be used in the Diffusion Monte Carlo (DMC) algorithm presented below. Note that it is actually a very particular case of more general branching process introduced in mathematics.
Let us consider a weight $w \geq 0$ (we will see that this weight will depend on electronic configuration). The branching process is defined as

$$\boxed{X = E(w + U)} \tag{21}$$

where U is the uniform random variable over (0,1) and E the integer part. X takes on integer values. The probability of having n is denoted as

$$P_n = P(X = n) \tag{22}$$

Now, it is clear that for a given w, only two values of n with non-zero probability are possible: n_c and $n_c + 1$ where $n_c \equiv E(w)$. Now, we have

$$P_{n_c+1} = 1 - (n_c + 1 - w) \tag{23}$$

$$P_{n_c} = n_c + 1 - w \tag{24}$$

Of course, as it should be, $P_{n_c+1} + P_{n_c+1} = 1$. Let us compute the mean

$$\bar{n} = n_c(n_c + 1 - w) + (n_c + 1)(1 - (n_c + 1 - w)) = w \tag{25}$$

We thus have

$$\boxed{\langle X \rangle = w} \tag{26}$$

General Markov process

These are the key process used in the vast majority of stochastic simulations. The probability of being between $\mathbf{x}_{k+1}$ and $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1} is now dependent on the previous configurations $\mathbf{x}_k$ but not on the oldest ones $\mathbf{x}_{l<k}$. It is common to say (in a loosely way) that for a Markov process, the future (at time t_{k+1}) depends on the present (time t_k) but not on the past (times $t_{l<k}$). More precisely, **the Markov hypothesis** is written as

$$\boxed{P_{k|1}(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2, ..., \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) = P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1})} \tag{27}$$

The fundamental quantity $P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1})$ characterizing the Markov process is called the transition kernel or transition probability density. In what follows we shall use the convenient notation

$$P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) \tag{28}$$

It is easy to check that the n-time probability density can now be written as

$$\boxed{P_n(\mathbf{x}_1, t_1; ...; \mathbf{x}_n, t_n) = P_1(\mathbf{x}_1, t_1) \prod_{k=1}^{n-1} P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}).} \tag{29}$$

From Eqs.(15) and (16) we have

$$\int d\mathbf{x}_{k+1} P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = 1 \tag{30}$$

In practice, most of the Markov process used in simulations are invariant under a time shift, they are said to be *homogeneous*. In that case

$$P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = P(\mathbf{x}_k \to \mathbf{x}_{k+1}, t_{k+1} - t_k) \tag{31}$$

For simplicity, the time interval will be denoted as t and the transition probability as $P(\mathbf{x} \to \mathbf{y}, t)$. Because of the time-shift invariance, the one-body density $P_1(\mathbf{x})$ is now independent on time. Let us derive the equation obeyed by $P_1(\mathbf{x})$. We have

$$P(\mathbf{x} \to \mathbf{y}, t) = \frac{P_2(\mathbf{x}; \mathbf{y}, t)}{P_1(\mathbf{x})} \tag{32}$$

Mutiplying the equation by $P_1(\mathbf{x})$ and integrating over $\mathbf{x}$ we get

$$\int d\mathbf{x} P_1(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \int d\mathbf{x} P_2(\mathbf{x}; \mathbf{y}, t) = P_1(\mathbf{y}). \tag{33}$$

Following a popular tradition, we shall denote, here and in what follows, **the stationary distribution density** as π

$$\pi(\mathbf{x}) = P_1(\mathbf{x}) \tag{34}$$

The equation obeyed by π is thus $\boxed{\int d\mathbf{x} \pi(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \pi(\mathbf{y})}$

Starting from the distribution $\pi(\mathbf{x})$ and applying the transition kernel to all $\mathbf{x}$ leads to configurations $\mathbf{y}$ also distributed according to π. It clearly illustrates the interpretation of π as the stationary distribution of the stochastic process.

Let us now adopt an alternative point of view. As already mentioned, the transition probability density characterizes the Markov process. Considered as the kernel of a linear operator, the properties of its eigensolutions can be studied. A first remark is that the transition probability is in general not symmetric, $P(\mathbf{x} \to \mathbf{y}, t) \neq P(\mathbf{y} \to \mathbf{x}, t)$. As a consequence, it is necessary to distinguish between left- and right-eigenvectors and, in addition, the eigenvalues are not necessarily real. However, because $P(\mathbf{x} \to \mathbf{y}, t) \geq 0$ and $\int d\mathbf{y} P(\mathbf{x} \to \mathbf{y}, t) = 1$ it can be shown that the modulus of all eigenvalues ≤ 1 and that the left-eigenstate associated with the maximal eigenvalue $\lambda = 1$ is positive everywhere (Krein-Rutman theorem, a generalization of the Perron-Frobenius theorem to operators [kr] The integral equation

$$\int d\mathbf{x} \pi(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \pi(\mathbf{y}) \tag{35}$$

is thus recovered where $\pi(\mathbf{x}) \geq 0$ is the maximal eigenvector of the transition kernel which defines the stationary distribution of the stochastic process.

In the preceding section we have derived an integral equation allowing to compute the stationary density π when the transition kernel is known. Let us now consider the problem of the computation of the kernel itself. The fundamental equation for $P(\mathbf{x} \to \mathbf{y}, t)$ is a simple consequence of the Markov hypothesis. It is obtained by observing that if we introduce an arbitrary intermediate time $u \in (0, t)$ and consider the probability of going from $\mathbf{x}$ to $\mathbf{y}$ in a time t we must have

$$\boxed{P(\mathbf{x} \to \mathbf{y}, t) = \int d\mathbf{z} P(\mathbf{x} \to \mathbf{z}, u) P(\mathbf{z} \to \mathbf{y}, t - u)} \tag{36}$$

It is known under the name of **Chapman-Kolmogorov equation**. A much more interesting form is its local form relating time and space derivatives.

Let us derive such an equation in the one-dimensional case. The generalization to an arbitrary dimension is elementary. The following derivation follows closely that of [coffey] Let $h(x)$ be an arbitrary smooth function and consider the time derivative of the transition probability. We can write

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \int dy h(y) \lim_{\Delta t \to 0} \frac{P(x \to y, t + \Delta t) - P(x \to y, t)}{\Delta t}$$

Applying the Chapman-Kolmogorov equation we have

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big[\int dy h(y) \int dz P(x \to z, t) P(z \to y, \Delta\tau) - \int dy h(y) P(x \to y, t) \Big]$$

Changing the name of the dummy variable y into z in the last integral of the RHS and using $\int dy P(z \to y, \Delta t) = 1$ then

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big[\int dz P(x \to z, t) \int dy P(z \to y, \Delta\tau)[h(y) - h(z)] \Big]$$

Now, we introduce a Taylor expansion of $h(y)$ around z:

$$h(y) = h(z) + \sum_{n=1}^{\infty} h^{(n)}(z) \frac{(y - z)^n}{n!}$$

and defining the "jump moments"

$$D^{(n)}(z) = \frac{1}{n!} \lim_{\Delta t \to 0} \int dy (y - z)^n P(z \to y, \Delta\tau)$$

we get

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \int dz P(x \to z, t) \sum_{n=1}^{\infty} D^{(n)}(z) h^{(n)}(z)$$

Integrating by parts n times we get

$$\int dz h(z) \Big[\frac{\partial P(x \to z, t)}{\partial t} - \sum_{n=1}^{\infty} \Big(- \frac{\partial}{\partial z} \Big)^n [D^{(n)}(z) P(x \to z, t)] \Big] = 0$$

and finally this integral being valid for any h the equation for the transition probability can be written as

$$\boxed{\frac{\partial P(x \to y, t)}{\partial t} = \sum_{n=1}^{\infty} \Big(- \frac{\partial}{\partial y} \Big)^n [D^{(n)}(y) P(x \to y, t)]} \tag{37}$$

In its general d-dimensional version it writes

$$\boxed{\frac{\partial P(\mathbf{x}\to\mathbf{y},t)}{\partial t}=\sum_{n=1}^{\infty}(-1)^n\sum_{j_1\dots j_n}\frac{\partial^n}{\partial y_{j_1}\cdots\partial y_{j_n}}\left[D^{(n)}_{j_1,\dots,j_n}(\mathbf{y})P(\mathbf{x}\to\mathbf{y},t)\right].} \tag{38}$$

This equation is known under the name of **Kramers-Moyal expansion** (of the master equation). Here, the "jump moments" are defined as

$$\boxed{D^{(n)}_{j_1,\dots,j_m}(\mathbf{y})=\frac{1}{n!}\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\langle\prod_{\mu=1}^{n}[Y_{j_\mu}(t+\Delta t)-Y_{j_\mu}(t)]\right\rangle\Bigg|_{Y_k(t)=y_k}.} \tag{39}$$

This equation is known under the name of Kramers-Moyal expansion. Let us now discuss the Markovian process at the heart of QMC approaches presented below.

Markovian process at work in QMC

• *Free diffusion or brownian process.*
The free diffusion process is invariant by space translation and thus, $D^{(1)}=0$. It is defined by a constant diagonal diffusion matrix $D^{(2)}_{ij}=\frac{1}{2}$ and $D^{(n>2)}=0$
In one dimension the Kramers-Moyal expansion is written as

$$\boxed{\frac{\partial P(x\to y,t)}{\partial t}=\frac{1}{2}\frac{\partial^2}{\partial y^2}P(x\to y,t)} \tag{40}$$

with initial condition, $P(x \to y, t = 0) = \delta(x - y)$ This equation is known under the name of **free diffusion (or heat) equation**. By using a Fourier transform the gaussian solution of this equation is easily obtained. We have

$$p(x \to y, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \tag{41}$$

In d dimensions the solution is a product of independent one-dimensional gaussian distributions for each coordinate

$$p(\mathbf{x} \to \mathbf{y}, t) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y_i - x_i)^2}{2t}} = \frac{1}{\sqrt{2\pi t}^d} e^{-\frac{(\mathbf{y}-\mathbf{x})^2}{2t}} \tag{42}$$

Using the gaussian transition probability density, brownian trajectories can be generated step-by-step. From Eq.(42) it is seen that the quantities $\frac{(y_i - x_i)}{\sqrt{t}}$ are independent and normally distributed. $\mathbf{y}$ can thus be obtained from $\mathbf{x}$ by drawing a gaussian number for each coordinate

$$\frac{(y_i - x_i)}{\sqrt{t}} = \eta_i \quad i = 1, d \tag{43}$$

where η is a normal random vector. The previous expression can be rewritten as

$$\boxed{y_i = x_i + \sqrt{t}\eta_i \quad i = 1, d} \tag{44}$$

This last equation is the simplest example of a discretized form of the so-called **Stochastic Differential Equation (SDE)** associated with a diffusion process.

• *Drifted diffusion or drifted Brownian motion*. As we shall see later, QMC methods are based on a more general version of the free Brownian motion where a drift part is introduced to enhance the Monte Carlo convergence (importance sampling). In this case, both $D^{(1)}$ and $D^{(2)}$ are non-vanishing. The first jump moment is known as the drift vector

$$\mathbf{b}(\mathbf{x}) = \mathbf{D}^{(1)}(\mathbf{x}) \tag{45}$$

In this case, the equation of evolution (KM expansion) is known as the **Fokker-Planck equation**. It is written as

$$\boxed{\frac{\partial P(\mathbf{x} \to \mathbf{y}, t)}{\partial t} = \frac{1}{2}\nabla_y^2 P(\mathbf{x} \to \mathbf{y}, t) - \nabla_y[\mathbf{b}(\mathbf{y})P(\mathbf{x} \to \mathbf{y}, t)]} \tag{46}$$

In the case of a constant drift vector **b** this equation can still be solved using a Fourier transform, we get

$$P(\mathbf{x} \to \mathbf{y}, t) = \frac{1}{\sqrt{2\pi t}^d} e^{-\frac{(\mathbf{y}-\mathbf{x}-\mathbf{b}\ t)^2}{2t}} \tag{47}$$

Stochastic trajectories are generated using the discretized SDE

$$y_i = x_i + b_i(x_1, ..., x_d)t + \sqrt{t}\eta_i \quad i = 1 \ \ to \ \ d \tag{48}$$

In the case of a general drift $\mathbf{b}(\mathbf{x})$, no analytical solution exists. However, it is still possible to generate trajectories by using a small enough time-step τ instead of an arbitrary time t as above. For that, we need to introduce a short-time approximation of the transition probability. When the time-step is sufficiently small, the variation of position is small and at leading order the drift vector can be considered as constant. The transition probability density is thus approximated as

$$\boxed{P(\mathbf{x} \to \mathbf{y}, \tau) = \frac{1}{\sqrt{2\pi t}^d} \exp - \frac{(\mathbf{y} - \mathbf{x} - \mathbf{b}(\mathbf{x})\tau)^2}{2\tau}} \tag{49}$$

This qualitative statement can be made more rigorous by looking at the small time-step limit of the exact solution of the Fokker-Planck equation, Eq.(46). Having a short-time gaussian expression for the transition probability, stochastic trajectories can be generated according to

$$\boxed{y_i = x_i + b_i(\mathbf{x})\tau + \sqrt{\tau}\eta_i \quad i = 1, d} \tag{50}$$

Note that the equations for each component are now coupled through the drift vector. The stationary density π of the process can be obtained by solving $\frac{\partial P(\mathbf{x} \to \mathbf{y}, t)}{\partial t} = 0$ that is

$$\frac{1}{2}\nabla^2 \pi - \nabla(\mathbf{b}\pi) = 0$$

It is easily seen that this equality is fulfilled when

$$\boxed{\mathbf{b}(\mathbf{x}) = \frac{1}{2}\frac{\nabla \pi(\mathbf{x})}{\pi(\mathbf{x})}} \tag{51}$$

Markov process with drift can thus be used to sample a given distribution $\pi(\mathbf{x})$ (for example, the Boltzmann distribution $\pi(\mathbf{x}) = \frac{e^{-\beta E(\mathbf{x})}}{Z}$). For that, we choose a drift vector according to Eq.(51) (here, $\mathbf{b} = -\frac{\beta}{2}\nabla E(\mathbf{x})$) and we generate trajectories using the stochastic differential equation, Eq.(48). Note that with such a scheme a (small) bias on the stationary distribution related to the use of a small but finite time-step is present. In contrast, it is not the case with the Metropolis algorithm presented in the next section.

• *Other Markov process*. There exist a great variety of Markovian process. Let us just say a few words about two important examples.
i) *The Lévy flight: A generalization of the browian motion allowing large moves*
Probability distribution:

$$f(x;\mu,c) = \sqrt{\frac{c}{2\pi}} \quad \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

where $x > \mu$, $\mu =$ location parameter, and $c =$scale parameter.
"Heavy-tailed" probability distribution (large values of x have non-negligible probability to occur). Note that $< x^2 >= \infty$ (mean), $< x^2 >= \infty$ (variance)!!
Kramers-Moyal equation derived above

$$\frac{\partial P(x \to y, t)}{\partial t} = \sum_{n=1}^{\infty} \left(- \frac{\partial}{\partial y} \right)^n [D^{(n)}(y) P(x \to y, t)]$$

becomes here

$$\frac{\partial P(x \to y, t)}{\partial t} = -(-\frac{\partial^\alpha}{\partial y^\alpha})[D^{(2)}(y)P(x \to y, t)] - \frac{\partial}{\partial y}[D^{(1)}P(x \to y, t)]$$

with fractional derivative ($0 < \alpha \leq 2$).
An intense activity aout the modelization of the paths followed by animals or humans when searching for food, hunting, (or even searching for lost keys on the beach...) has been developed. See, for example, the influential work by H. Eugene Stanley and collaborators of 1999 ("Optimizing the success of random searches" [**?**]).

ii) *The Poisson process: A simple example of discrete Markov process*
Poisson process of intensity λ ($\lambda > 0$. Equation of evolution of discrete variable X

$$\frac{p(X = n)(t + \Delta t) - p(X = n)(t)}{\Delta t} = p(X = n - 1)(t) - p(X = n)(t)$$

when Δt goes to zero, the probability distribution is given by

$$\mathbb{P}(X = n, t) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n \ \text{integer}$$

• *Stochastic process with memory effects (beyond Markov ones)*. Being almost never used in realistic simulations, they will not discussed here.

The Metropolis algorithm

Sampling a general density in high dimension

The purpose of the Metropolis algorithm,[?][?] is to sample a general density π in arbitrary dimension. Two remarkable features of the algorithm are that :

i) It can be used for (very) large-dimensional spaces and

ii) Only the ratio of probability densities $\frac{\pi(\mathbf{x})}{\pi(\mathbf{y})}$ are to be evaluated, not the probability density π alone.

i.) The first property is remarkable and make in practice the Metropolis algorithm the only practical choice for treating problemes in high dimensions. This is the reason why the algorithm is so widely used and is in the list proposed in 2000 by Dongarra and Sullivan of the "10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century" [?] Note that applications including dimensions as large as several thousands are routine, and much larger dimensions can be successfully treated.

ii) The second important feature is that there is no need to know the normalization of π. It is an important practical point since the normalization is usually a physically relevant quantity (for example, the partition function in statistical physics) and is in general not known.

The basic idea of the Metropolis algorithm is to generate by a step-by-step procedure configurations in space distributed according to π. The fundamental quantity of the algorithm is the trial transition probability density denoted here as $P^T(\mathbf{x} \to \mathbf{y})$. The algorithm is as follows.
m

METROPOLIS ALGORITHM

At each Monte Carlo step a new state $\mathbf{x}_{i+1}$ is generated from the current state $\mathbf{x}_i$ by a two-step procedure:
1) Draw a "trial" state denoted as $\mathbf{x}^T$ using some trial transition probability $P^T(\mathbf{x} \to \mathbf{y})$
2) Accept the trial state as the new state ($\mathbf{x}_{i+1} = \mathbf{x}^T$) or reject it ($\mathbf{x}_{i+1} = \mathbf{x}_i$) with probability $q(\mathbf{x}_i, \mathbf{x}_T)$ ($0 \leq q \leq 1$) given by

$$q = Min\Big[1, \frac{\pi(\mathbf{x}^T)P^T(\mathbf{x}^T \to \mathbf{x}_i)}{\pi(\mathbf{x}_i)P^T(\mathbf{x}_i \to \mathbf{x}^T)}\Big] \tag{52}$$

At this point, several remarks are in order.

• A necessary condition that the algorithm be valid (sample the density π) is that the transition probability is ergodic. Ergodicity means that for any initial state $\mathbf{x}_0$ and final state $\mathbf{x}$, and any neighborhood of $\mathbf{x}$ (for example, neighborhood= set of all states $\mathbf{y}$ such as $||\mathbf{x} - \mathbf{y}|| \leq \epsilon$) there is a *finite* probability starting from $\mathbf{x}_0$ to reach the neighborhood of $\mathbf{x}$ in a *finite* number of moves.

• If the ergodicity property is fulfilled, the Metropolis algorithm converges to π **independently** on the choice of the trial transition probability and/or the initial conditions $\mathbf{x}_0$. Such quantities only determines the rate of convergence of the Markov chain towards π.

• To have a practical scheme, the trial transition density must be chosen easy to sample (see, below).

• To accept a change with probability q means: Draw a uniform random number u over (0,1), if $u \leq q$ the change is accepted, if not it is rejected.

For a derivation of the Metropolis algorithm in the discrete case, see appendix 62.

Computing multi-dimensional integrals with the Metropolis algorithm

Integrals as probabilistic averages

In a Monte Carlo calculation, the integrals to be computed are of the form

$$I(f) = \int d\mathbf{x}\pi(\mathbf{x})f(\mathbf{x}) \tag{53}$$

where π is some probability density defined over $\mathbb{R}^d$ and f some integrand.
Note that the most general form for a d-dimensional integral (without π), namely $I = \int d\mathbf{x}g(\mathbf{x})$, can always be rewritten under the form given in Eq.(53) by introducing some arbitrary *positive* function g_0 with

$$\pi = \frac{g_0}{\int d\mathbf{x}g_0(\mathbf{x})}$$

and

$$f = \frac{g}{\pi}$$

However, to be able to do that, we need to know the normalization of the function g_0 since it enters now the integrand f, a constraint which severely reduces the possible choices for g_0. In practice, a reasonable strategy is to search where g is maximal and choose g_0 as some gausian approximation around its maximum. However, this strategy will work in high dimension only if f does not vary too much in region where $\pi(= g_0)$ is large.

VERY IMPORTANT PHYSICAL REMARK

Actually, a fundamental point to realize is that for *all* physical problems defined in (very) high dimension some density π is always present in the integrands, the density giving the weight of the state (configuration) with respect to all other possible states. In practice, this density is non-zero only for a *very tiny fraction* of all possible states, such states corresponding to the so-called "physically accessible" states. If it would not be the case, the situation would just be desesperate since sampling a huge number of states with a limited number of Monte Carlo steps (say, up to about a few billions) is not possible.

Coming back to the definition of $I(f)$, Eq.(53), it can be interpreted as the probabilistic mean of f with respect to π writting

$$I(f) = \langle f \rangle$$

and the integral can be expressed as the average of f over an infinite number of configurations sampled with the Metropolis algorithm (ergodic property)

$$I(f) = \lim_{K \to +\infty} \frac{1}{K} \sum_{i=1}^{K} f(\mathbf{x}_i). \tag{54}$$

Of course, in practical simulations, a large but finite number of points is used. The integral is thus written as

$$I_K(f) = I(f) + \epsilon(K) \tag{55}$$

where $I_K(f)$ is the Monte Carlo value obtained with K configurations,

$$I_K(f) = \frac{1}{K} \sum_{i=1}^{K} f(\mathbf{x}_i) \tag{56}$$

and $\epsilon(K)$ some residual statistal error. This error is discussed in the next section.

Optimizing the sampling.

As seen, in the Metropolis algorithm the configurations are changed using the the trial transition probability density. Although the values of the integrals do not depend on the transition density, it determines the quality of the sampling and thus the rate of convergence to the density π and then to the exact values for the integrals.

• A first natural choice is the "historical" one made by Metropolis and collaborators in their original work where $P^T(\mathbf{x} \to \mathbf{y})$ is taken to be a uniform transition density in some small region around $\mathbf{x}$. Mathematically, it is written as

$$P^T(\mathbf{x} \to \mathbf{y}) = \begin{cases} \frac{1}{\Delta^d}, & \text{if } \mathbf{y} \in [x_i - \frac{\Delta}{2}, x_i + \frac{\Delta}{2}]^d \\ 0, & \text{otherwise} \end{cases} \tag{57}$$

where Δ is some positive constant defining the magnitude of the proposed trial move around the current position. The acceptance probability q, as defined in Eq.(52), is given by

$$q(\mathbf{x}, \mathbf{y}) = Min\big[1, \frac{\pi(\mathbf{x})}{\pi(\mathbf{y})}\big] \tag{58}$$

To make the simulation converge rapidly, it is desirable to take large values of Δ, leading to a better sampling of the configuration space. Unfortunately, when using large values of Δ the trial configuration, which is chosen randomly and far from the physically-accpetable state $\mathbf{x}$, has almost no chance to be accepted. On the opposite case where Δ is chosen very small, the trial configuration is almost systematically accepted since $q \simeq 1$. However, the new state is now very close of $\mathbf{x}$ and the sampling of the configuration space is very inefficient.
In actual simulations, some estimator allowing to determine the optimal value of Δ is introduced. A standard solution consists in defining the average acceptance probability $\eta = \langle q \rangle$ as

$$\eta = \frac{\#\text{ of accepted moves}}{\#\text{ of moves}} \tag{59}$$

and to adjust Δ so that the acceptation ratio is about one half.

•Optimal choice. The optimal choice of P^T consists in drawing trial configurations according to $\pi(\mathbf{y})$, independently on the current position $\mathbf{x}$, that is, $P^T(\mathbf{x} \to \mathbf{y}) = \pi(\mathbf{y})$. In that case, successive drawings are independent, large moves in the configuration space can be done and the acceptance probability is equal to one:

$$q(\mathbf{x}, \mathbf{y}) = \frac{\pi(\mathbf{y}) P^T(\mathbf{y} \to \mathbf{x})}{\pi(\mathbf{x}) P^T(\mathbf{x} \to \mathbf{y})} = \frac{\pi(\mathbf{y})\pi(\mathbf{x})}{\pi(\mathbf{x})\pi(\mathbf{y})} = 1$$

Unfortunately, there is no efficient algorithm known to draw directly a general density in a high-dimensional case. Actually, it is the very reason why the Metropolis algorithm has been introduced!

• Trial transition density of the Fokker-Planck equation To go beyond the standard uniform transition density, it is very desirable to include some information about the shape of the distribution π to be sampled. Indeed, instead of making "blind moves" in random directions as in the historical algorithm it is much better to propose moves into directions where π increases significantly and avoiding moves toward region where π decreases stiffly.

This can be beautifully realized with the transition probability density of the drifted random walk introduced above, Eq.(49). It is the transition probability used when Variational Monte Carlo (VMC) calculations for electronic structure, see the next section.

• *Statistical error.*

The Metropolis algorithm is a simple and efficient algorithm for generating states distributed according to an arbitrary density. However, the price to pay for such a simplicity is the fact that the successive states produced are correlated. Accordingly, some care is needed when estimating the statistical error associated with the arithmetic averages computed. First of all, it is important to check that we are not in the transient regime associated with the initial configuration used. Second we have to estimate the correlation time of the Markov chain.

Let $f(x)$ a quantity whose expectation value is to be computed, $I(f) = \int dx \pi(x) f(x)$

A unbiased estimator of the expectation value is the arithmetic sum

$$\bar{f}_n = \frac{1}{n} \sum_{i=1}^{n} f(x_i) \tag{60}$$

where n is a finite number of configurations drawn with the Metropolis algorithm. Note that $\bar{f}_n$ is a random variable and that its value depends on the series of random numbers used to generate the successive states of the sum. Unbiased means here that if the finite sum is computed an infinite number of times with different random realizations, then

$$\langle \bar{f}_n \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle f(x_i) \rangle = I(f) \tag{61}$$

Due to the central limit theorem valid for Markov process, we know that for sufficiently large n the distribution of the random variable $\bar{f}_n$ becomes gaussian

$$P(\bar{f}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(\bar{f}_n - \langle f_n \rangle)^2}{2\sigma_n^2}}$$

where

$$\sigma_n^2 = \langle \bar{f}_n{}^2 \rangle - \langle \bar{f}_n \rangle^2 \tag{62}$$

Now, a practical way to compute the error bar is to realize a certain number of independent calculations of $\bar{f}_n$ and to estimate the variance of the distribution $P(\bar{f})$. Let N_b the number of independent calculations, we denote $\bar{f}_n{}^k \quad k = 1, N_b$, the values obtained for each calculation. Unbiased estimates of the mean and variance are

$$\langle \bar{f}_n \rangle = \frac{1}{N_b} \sum_{k=1}^{N_b} \bar{f}_n{}^k$$

and

$$\sigma_n^2 = \frac{1}{N_b - 1} \sum_{k=1}^{N_b} (\bar{f}_n{}^k - \langle \bar{f}_n \rangle)^2$$

An estimate of the statistical error δf on the estimate of $I(f)$ is then $\delta f = \frac{\sqrt{\sigma_n^2}}{\sqrt{N_b}}$, that is

$$\boxed{\delta f = \frac{1}{\sqrt{N_b(N_b-1)}} \sqrt{\sum_{k=1}^{N_b} \left(\bar{f_n}^k - \langle \bar{f_n} \rangle\right)^2}} \tag{63}$$

In practical calculations, the N_b calculations are never fully independent and some correlation are introduced. Such correlations can be explicited as follows. By inserting (61) into (62) we get

$$\sigma^2 = \frac{1}{n}[c_0 + 2\sum_{i=1}^{n-1}(1 - \frac{i}{n})c_i]$$

where

$$c_i = \langle f_k f_{k+i} \rangle - \langle f_k \rangle \langle f_{k+i} \rangle$$

(time translation implies independence on k). Calculation of the c_i can be performed by estimating the various correlators from the N_b realizations. Formula (63) can be easily generalized using such correlators. For a discussion of such aspects, see for example [**?**].

Computing the quantum-mechanical properties associated with some trial wavefunction: The Variational Monte Carlo (VMC) method

The basic idea

• Consider a trial wavefunction Ψ_T (in our applications: [$\mathbf{x} = (\mathbf{r}_1, ..., \mathbf{r}_N)$, N number of particles (electrons)] chosen to be a good representation of the unknown wavefunction

• Use the Metropolis algorithm for sampling the quantum-mechanical probability density associated with Ψ_T, namely

$$\pi(\mathbf{x}) = \frac{|\Psi_T(\mathbf{x})|^2}{\int d\mathbf{x}|\Psi_T(\mathbf{x})|^2}$$

• Compute properties as probabilistic averages over sampled configurations.
In the case of the energy, the variational energy E_v is obtained as

$$E_v = \frac{\langle\Psi_T|H|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \frac{\int d\mathbf{x}|\Psi_T|^2 \frac{H\Psi_T}{\Psi_T}}{\int d\mathbf{x}|\Psi_T|^2}$$

that is

$$\boxed{E_v = \int d\mathbf{x}\pi(\mathbf{x})E_L(\mathbf{x})}$$

where $E_L(\mathbf{x})$ is the so-called local energy.

$$\boxed{E_L(\mathbf{x}) = \frac{H\Psi_T}{\Psi_T}}$$

The probabilistic average is then evaluated as follows

$$E_{var} = \langle E_L \rangle = \lim_{K\to\infty} \frac{1}{K} \sum_{i=1}^{K} E_L(\mathbf{x}_i)$$

where $\mathbf{x}^i$ denotes the configurations drawn with the Metropolis algorithm.
Other properties can be computed in a similar way

$$\frac{\langle \Psi_T | O | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \int d\mathbf{x} O(\mathbf{x}) \pi(\mathbf{x}) = \lim_{K\to\infty} \frac{1}{K} \sum_{i=1}^{K} O(\mathbf{x}_i)$$

The trial transition probability density is chosen to be the short-time drifted gaussian transition probability density, Eq.(49)
Zero-variance property for the energy. As seen above, the statistical error on probabilistic averages is proportional to the square root of the variance of the integrand, that is here, of the local energy. Now, the "closest" the trial wave function is of the exact solution, the smaller the fluctuations of E_L are. In the limit of an exact wavefunction, fluctuations vanish. This property is referred to as the *zero-variance property*.
Zero-variance property for general observables O. Using the Hellman-Feynman theorem expressing $\langle O \rangle$ as the derivative of the energy with respect to the magnitude of the operator considered as an external field and using the ZV property for the energy, it is possible to construct new estimators for O having also a zero-variace property. For more details, see [**?**, **?**].

The trial wavefunction

In QMC there is a great freedom in choosing the functional form of the trial wavefunction (no computation of one- or bi-electronic integrals, just first and second derivatives of Ψ_T). A great variety of functional forms has thus been considered.

Spin-free formalism In constrast with most electronic structure methods where spin variables are introduced, in QMC the trial wavefunctions are spin-free, that is they depend only on the space coordinates of the electrons, $\mathbf{x} = (\mathbf{r}_1, ..., \mathbf{r}_N)$. This is possible since the Schrödinger equation to be solved is spin-variable independent. For a discussion of the use of a spin-free formalism in quantum chemistry, see for example[**?**]. Without entering into the details, let us just say that the matrix elements of a fully symmetric and spin-free operator between two determinants $|I\rangle$ and $|J\rangle$ can be obtained as

$$\langle D_I|O|D_J\rangle_{\mathbf{x},\sigma} = \langle D_I^\alpha D_I^\beta|O|D_J^\alpha D_J^\beta\rangle_{\mathbf{x}} \tag{64}$$

where D^σ $(\sigma = \alpha, \beta)$ are space-only determinants built from the space orbitals corresponding to spin σ. The subscript over brackets indicates the variables of integration used.

To give an example, the following spin-space determinants describing a set of doubly occupied orbitals

$$\begin{vmatrix} \phi_1(\mathbf{r}_1)\alpha & \phi_1(\mathbf{r}_2)\alpha & \cdots & \phi_1(\mathbf{r}_N)\alpha \\ \phi_1(\mathbf{r}_1)\beta & \phi_1(\mathbf{r}_2)\beta & \cdots & \phi_1(\mathbf{r}_N)\beta \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N/2}(\mathbf{r}_1)\alpha & \phi_{N/2}(\mathbf{r}_2)\alpha & \cdots & \phi_{N/2}(\mathbf{r}_N)\alpha \\ \phi_{N/2}(\mathbf{r}_1)\beta & \phi_{N/2}(\mathbf{r}_2)\beta & \cdots & \phi_{N/2}(\mathbf{r}_N)\beta \end{vmatrix}$$

has the same averages over spin-free operators as the pure space product of determinants

$$\begin{vmatrix} \phi_1(\mathbf{r}_1) & \cdots & \phi_1(\mathbf{r}_{N/2}) \\ \vdots & \vdots & \vdots \\ \phi_{N/2}(\mathbf{r}_1) & \cdots & \phi_{N/2}(\mathbf{r}_{N/2}) \end{vmatrix} \begin{vmatrix} \phi_1(\mathbf{r}_{N/2+1}) & \cdots & \phi_1(\mathbf{r}_N) \\ \vdots & \vdots & \vdots \\ \phi_{N/2}(\mathbf{r}_{N/2+1}) & \cdots & \phi_{N/2}(\mathbf{r}_N) \end{vmatrix} \tag{65}$$

where α-electrons have been arbitrarily chosen to have particle labels $\{1, ..., N/2\}$ and β-electrons particle labels $\{N/2 + 1, ..., N\}$

Different types of wavefunction used

• *Multi-determinant Slater-Jastrow*. The most popular form is the multi-determinant Slater Jastrow form written as

$$\Psi_T = e^{J(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \sum_{k=1}^{N_{det}} c_k Det_k(\{\Phi_i^\alpha\}) Det_k(\{\Phi_i^\beta\}), \tag{66}$$

where $\{\Phi_i^\sigma\}(\sigma = \alpha, \beta)$ is a set of molecular orbitals and e^J is the Jastrow factor. The role of the Jastrow factor is to impose the exact behavior of the wavefunction in the $[r_{ij} \to 0]$-limit (electron-electron cusp condition) and, also, to incorporate some two-body (electron-electron and electron-nucleus) and three-body (electron-electron-nucleus) correlations (to describe the best as possible the shape of the Coulomb hole[**?**]). Many different forms for the Jastrow factor have been introduced. Typically,

$$J = \sum_{i<j} u(r_{ij}) + \sum_i \sum_\alpha v(r_{i\alpha}) + \sum_{i<j} \sum_\alpha w(r_{ij}, r_{i\alpha}, r_{j\alpha})$$

where $r_{ij} = |r_i - R_\alpha|$, and $r_{i\alpha} = |r_i - R_\alpha|$. Various forms for the functions u,v, and w have been tested. For example, the minimal Padé form for u

$$u(r_{ij}) = \frac{a r_{ij}}{1 + b r_{ij}}.$$

• *Use of a backflow term.* In trial wavefunctions including backflow, the electron coordinate $\mathbf{r}_i$ is replaced by a quasi-particle (dressed) coordinate $\bar{\mathbf{r}}_i = \mathbf{r}_i + \sum_{j \neq i} \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j)$ and is introduced in Slater forms. Physically, this backflow displacement is supposed to reproduce the characteristic "flow pattern" where the quantum fluid is pushed out of the way in front of a moving particle and fills in the space behind it. For more details, see Ref.[**?**]

• *Resonating VB form and geminal forms*. Let Φ be the pairing function (geminal) which takes into account the correlation between two electrons with opposite spin. If the system is unpolarized and the state is a spin singlet, the antisymmetrized geminal product (AGP) wavefunction is

$$\mathbf{\Psi}_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \hat{A}[\Phi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_2^{\downarrow})\Phi(\mathbf{r}_3^{\uparrow}, \mathbf{r}_4^{\downarrow}) \cdots \Phi(\mathbf{r}_{N-1}^{\uparrow}, \mathbf{r}_N^{\downarrow})], \tag{67}$$

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

$$\Phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right), \tag{68}$$

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

$$\Psi_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = det(A_{ij}), \tag{69}$$

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

$$A_{ij} = \phi(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}). \tag{70}$$

For more details, see Ref.[**?**]

• *Perturbatively selected Configuration Interaction expansion*. In quantum chemistry Configuration Interaction (CI) expansions are widely used. They allow a systematic improvement of the wavefunction through increase of the number of determinants and of the basis set used. In QMC the use of CI expansions is problematic due to the very large number of determinants. Indeed, at each Monte Carlo iteration -and there can be as many as one billion of such elementary steps- the first and second derivatives (Laplacian) must be computed for the current electronic configuration. However, despite these drawbacks, CI expansions have nevertheless been recently employed in QMC. It is possible only because 1) the CI expansion is reduced by a suitable selection of the most important determinants[**?**, **?**] 2) efficient techniques have been developed to make the CI expansion computable in a reasonable time.[**?**, **?**, **?**]. Some applications can be found in Ref.[**?**],[**?**].

• *Valence Bond trial wavefunction*. The use of Valance Bond (VB) wavefunctions is very attractive in quantum chemistry. Indeed, VB forms give a simple and very appealing interpretation of the electronic structure in terms of Lewis pairs (bound pairs, lone pair, etc.). Unfortunately, from a technical point of view VB wavefunctions are made of non-orthogonal determinants, a point which dramatically increases the computational effort (passing from a standard N^3 law to a $N!$ law). A number of QMC works using VB wavefunctions have been presented, see Ref.[**?**, **?**, **?**]

• *Multi-Jastrow form* The so-called Multi-Jastrow is obtained by replacing the global Jastrow form into local Jastrows attached to one-particle molecular orbitals. Using such local forms allows to describe the electron-electron correlation in a more specific way (electron correlation is different into a 1s orbitals, 3d orbitals, polarizable lone pairs, etc.) See [**?**].

Computing the exact ground-state energy: The Diffusion Monte Carlo (DMC) method

Diffusion Monte Carlo Let us start with the time-dependent Schrödinger equation (atomic units)

$$i\frac{\partial \Psi(\mathbf{x},t)}{\partial t} = -\frac{1}{2}\nabla^2\Psi(\mathbf{x},t) + (V(\mathbf{x}) - E_T)\Psi(\mathbf{x},t)$$

where E_T is some arbitrary reference energy. Let us make the transformation to imaginary time (Wick's rotation)

$$t \to -it$$

$$\frac{\partial \Psi(\mathbf{x},t)}{\partial t} = \frac{1}{2}\nabla^2\Psi(\mathbf{x},t) - (V(\mathbf{x}) - E_T)\Psi(\mathbf{x},t) \tag{71}$$

Important: As far as time-independent properties are considered, this transformation has no consequences. In particular, the eigensolutions of the Hamiltonian are not modified.

Let us note $\Psi_T(\mathbf{x})$ a (time-independent) trial wavefunction and introduce a "mixed" density

$$\boxed{f(\mathbf{x},t) \equiv \Psi_T(\mathbf{x})\Psi(\mathbf{x},t)} \tag{72}$$

Multiplying each side of Eq.(81) by Ψ_T, we get

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = \frac{1}{2}\Psi_T(\mathbf{x})\nabla^2[\frac{f(\mathbf{x},t)}{\Psi_T(\mathbf{x})}] - (V(\mathbf{x}) - E_T)f(\mathbf{x},t)$$

With simple algebra we get

$$\frac{1}{2}\Psi_T\nabla^2[\frac{f}{\Psi_T}] = \frac{1}{2}\nabla^2 f - \mathbf{b}\nabla f - \frac{1}{2}\frac{\nabla^2\Psi_T}{\Psi_T} + \mathbf{b}^2 f$$

where the drift vector is given by

$$\boxed{\mathbf{b} = \frac{\nabla\Psi_T}{\Psi_T}} \tag{73}$$

Remarking that

$$E_L = \frac{H\Psi_T}{\Psi_T} = -\frac{1}{2}\frac{\nabla^2\Psi_T}{\Psi_T} + V$$

finally, we have

$$\boxed{\frac{\partial f(\mathbf{x},t)}{\partial t} = \frac{1}{2}\nabla^2 f(\mathbf{x},t) - \nabla[\mathbf{b}(\mathbf{x})f(\mathbf{x},t)] - (E_L(\mathbf{x}) - E_T)f(\mathbf{x},t)} \tag{74}$$

or

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = \big(L - (E_L - E_T)\big)f(\mathbf{x},t)$$

where L is the **Fokker-Planck operator**

$$\boxed{L = \frac{1}{2}\nabla^2 - \nabla[\mathbf{b}.]} \tag{75}$$

Eq.(74) determining the evolution of the mixed density f can be considered as the fundamental equation of diffusion Monte Carlo.
The time evolution of the density results from two coupled contributions:
(1) A first term describing a diffusion process associated with a constant diffusion $D = \frac{1}{2}$ and a drift term, $\mathbf{b} = \frac{\nabla \Psi_T}{\Psi_T}$. Note that the stationary density is given by $\pi = \Psi_T^2$.
(2) A potential part given by the local energy. Considered alone, the equation of evolution is

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = -(E_L(\mathbf{x}) - E_T) f(\mathbf{x}, t)$$

whose solution is

$$f(\mathbf{x}, t) = f(\mathbf{x}, t = 0) e^{-t(E_L(\mathbf{x}) - E_T)}$$

This part describes a so-called birth-death process or branching process. At point $\mathbf{x}$ the density increases/decreases in time according to the variation of the local energy around the trial energy. Denoting τ the small time-step used in the simulation we have

$$f(\mathbf{x}, t + \tau) = w(\mathbf{x}, \tau) f(\mathbf{x}, t) \tag{76}$$

where the weight w is defined as

$$w(\mathbf{x}, \tau) = \mathrm{e}^{-\tau(E_L(\mathbf{x}) - E_T)}$$

Diffusion Monte Carlo combines both process. The resulting stationary distribution can be obtained by writing

$$L - (E_L - E_T) = 0$$

It is easy to check that π fulfilling this equation is given by

$$\boxed{\pi_{DMC} = \Psi_T \Phi_0} \tag{77}$$

where Φ_0 is the unknown exact ground-state and E_T has ben taken equal to E_0.
An unbiased estimator of the ground-state energy is the expectation value of the local energy over the stationary distribution Indeed, because the operator H is a hermitian (self-adjoint) operator we can write

$$E_0 = \frac{\int \Phi_0 H \Psi_T}{\int \Phi_0 \Psi_T} = \frac{\int \Phi_0 \Psi_T \frac{H\Psi_T}{\Psi_T}}{\int \Phi_0 \Psi_T}$$

and then

$$\boxed{E_0 = \int d\mathbf{x} \pi_{DMC}(\mathbf{x}) E_L(\mathbf{x})} \tag{78}$$

A schematic DMC algorithm is thus

- Start from a population of walkers (a set of configurations $\mathbf{x}_i^k$ with $k = 1, N_w$)
- Move independently each walker according to Eq.(79)
- For each walker compute the branching weight w. From w build an integer whose expectation value gives w, for example $m = E(w + u)$, u random number and E=integer part.
- Remove ($m = 0$) or duplicate reach walker a certain number of times ($m \geq 0$). In average, this step reproduces the evolution of the density as given in Eq.(76)
- Modify the reference energy E_T to keep the number of walkers approximately constant (population control step).
- Add contribution of the new walkers to each average (for the energy, Eq.(78)) and iterate.

Population control step. As seen the number of walkers can varied in time. The total number of walkers at time t is given by

$$M(t) = \int d\mathbf{x} f(\mathbf{x}, t)$$

and its time variation by

$$\frac{dM(t)}{dt} = \int d\mathbf{x} \frac{\partial f(\mathbf{x}, t)}{\partial t}$$

In the case where only the diffusion part is considered, we have

$$\frac{dM(t)}{dt} = \int d\mathbf{x} L f(\mathbf{x}, t) = 0$$

The norm of the density is conserved and the number of walkers can be kept constant. It is no longer the case when adding the branching term

$$\frac{dM(t)}{dt} = -\int d\mathbf{x} (E_L(x) - E_T) f(\mathbf{x}, t) = -\sum_{k=1}^{M(t)} (E_L(k) - E_T)$$

Since nothing prevents the population to increase or decrease indefinitely a population control step must be introduced. A standard solution consists in modifying smoothly the reference energy such that to keep in average the population constant.

$$E_T(t + \tau) = E_T(t) + \frac{K}{\tau} \ln[\frac{M(t + \tau)}{M(t)}]$$

DMC for fermions: The sign problem and the fixed-node approximation

As just presented the DMC algorithm is exact only if the trial wavefunction Ψ_T never vanishes (at finite distances), say $\Psi_T > 0$. It can be directly employed for quantum systems with no Fermi constraints (bosonic systems, quantum oscillators, ensemble of distinguishable particles, etc.). Indeed, in such cases the ground-state eigenfunction Φ_0 is nodeless (say, positive).

Unfortunately, for fermionic systems such an eigenstate is physically forbidden by the Pauli exclusion principle [wigner], and the fermionic ground-state has now a sign.

Let us see what happens if the DMC algorithm is used as it is.
Let us recall that the walkers are moved move according to

$$\boxed{y_i = x_i + b_i(\mathbf{x})\tau + \sqrt{\tau}\eta_i \quad i = 1, 3N} \tag{79}$$

with

$$\boxed{\mathbf{b}(\mathbf{x}) = \frac{\nabla \Psi_T(\mathbf{x})}{\Psi_T(\mathbf{x})}} \tag{80}$$

• The values of $\mathbf{x}$ where Ψ_T vanishes are called the **zeros (or nodes)** of Ψ_T. It can be shown that the nodes of the exact wavefunctions are a variety of dimension $(3N-1)$ (the nodes "cut" the configuration space). It is the same for the trial wavefunctions used.

- Nodal pockets are the subdomains of constant sign for the wavefunction
- The union of nodal pockets is the entire configuration space
- The **nodes of Ψ_T are infinitely repulsive barriers for the walkers**, and thus each walker is trapped for ever into the nodal pocket where it starts from.
- The nodes of Ψ_T being not exact, the Schrödinger equation is solved with the approximation that the solution vanishes wherever Ψ_T vanishes. It is the **fixed-node approximation.** We can easily show the variational property

$$E_0^{FN} \geq E_0$$

Alternative point of view
The basic idea of DMC is to transform the (imaginary) time-dependent Schrödinger equation

$$\boxed{\frac{\partial \Psi(\mathbf{x}, t)}{\partial t} = -(H - E_T)\Psi(\mathbf{x}, t)} \tag{81}$$

into a generalized diffusion equation by introducing a mixed density f as

$$f(\mathbf{x}, t) \equiv \Psi_T(\mathbf{x})\Psi(\mathbf{x}, t) \tag{82}$$

It can be viewed as applying a similarity transformation to the SE so that

$$\boxed{\frac{\partial f(\mathbf{x},t)}{\partial t} = L^* f(\mathbf{x},t)}$$

with

$$L^* = L - (E_L - E_T) = \Psi_T (H - E_T) \frac{1}{\Psi_T}$$

The eigensolutions of L^* and $\Psi_T (H - E_T) \frac{1}{\Psi_T}$ are related via

$$L^* u_i = -(E_i - E_T) u_i$$

with $u_i = \Psi_T \Phi_i$.
If Ψ_T vanishes, some new boundary conditions depending on Ψ_T are put on the operator L^*. The energy obtained by simulating L^* is now the ground-state energy of the Hamiltonian with these new boundary-conditions which are not exact,**this is the fixed-node approximation.**

Mathematical digression For fermions the functional space of wave functions is divided into two orthogonal spaces

$$L^2(\mathbb{R}^{dN}) = B \oplus F \tag{83}$$

where F is the vector space of "fermionic" wavefunctions defined as follows:

$$\Psi \in F \text{ if and only if } \Psi[\sigma(\mathbf{x})] = \mathrm{sgn}(\sigma)\Psi[(\mathbf{x})] \tag{84}$$

where σ ranges in some permutation subgroup of the symmetric group S_N leaving invariant some 2-subsets partition of $\{1, \ldots, N\}$ (corresponding to "spin up" or "spin down" electrons). In particular, all totally skew-symmetric functions are in this case. B, the vector space of "bosonic" wavefunctions, is then simply the orthogonal of F. In particular, all totally symmetric functions are in B.
The Pauli principle can then be summarized by saying that the "fermionic" eigensolutions of H physically admissible are those obtained by restricting the Hamiltonian to the vector space F. In particular, the totally symmetric nodeless lowest eigenstate of H is forbidden for fermions (the so-called "bosonic" ground-state).
Note that in contrast with standard presentations of the Pauli exclusion principle, no spin coordinates have been introduced here. Actually, at the non-relativistic level such coordinates are not needed, see *e.g.*[wigner,matsen]. However, they are of common use since within a spin-space representation the Pauli exclusion principle is particularly simple to express. The eigenstates are written as a combination of space *and* spin functions and only those that are totally antisymmetric under the exchange of space-spin coordinates of any pair of particles are physically allowed. In a spin-free (space-only) formalism as used here, the spatial wavefunctions $\Psi(\mathbf{x})$ just need to be antisymmetric under permutations within two subsets of particles that can be formally associated with spin "up" and "down" particles.
Because the Schrödinger Hamiltonian is spin-independent and the diffusion processes introduced are defined in a pure space representation, the use of spin coordinates is not adapted and is thus avoided in QMC.

Finally, the problem to solve in QMC is to design an efficient algorithm allowing to converge to the ground-state fermionic eigenfunction (lowest eigenstate of H restricted to vector space F). Unfortunately, up to now it has not been possible to define a computationally tractable (polynomial) algorithm implementing exactly such a property for a general fermionic system. This problem -known under the name of "sign problem" is of uttermost practical importance and is viewed as one of the most important problems to be solved in computational many-body physics [signproblem1,signproblem2,signproblem3,signproblem4]
Ψ_0^{FN} denotes the Fixed-Node (FN) ground-state eigenfunction obtained by imposing the nodal boundaries to Ψ_T. Due to its very construction the fixed-node solution has the same sign as the trial wavefunction ($\Psi_T(\mathbf{x})\Psi_0^{FN}(\mathbf{x}) \geq 0$). The fermionic problem defined over the entire configuration space $\mathbb{R}^{dN}$ is thus recast in a sum of *independent* bosonic-type problems defined in each nodal volume cut by the nodes of the approximate trial wavefunction. Instead of defining a unique Fokker-Planck operator with a non-divergent drift vector over all space, a set of independent FP operators restricted to each nodal cell domain is considered. Transposed into the original Hamiltonian problem, it means that the Schrödinger equation is solved independently in each nodal cell (mathematically, the N-body Schrödinger ground-state is computed with additional Dirichlet boundary condition on the nodal set $\mathcal{N}$ where Ψ_T^F vanishes, $\mathcal{N} = \{\mathbf{x} \in \mathbb{R}^{dN} : \Psi_T^F(\mathbf{x}) = 0\}$. In the general case, the zeroes of the trial wavefunction do not coincide with those of the unknown fermionic eigensolution and we are thus left with a systematic bias, the fixed-node error.
At this point, several important theoretical and practical aspects of the fixed-node approximation must be discussed.

Mathematical foundation of the fixed-node approach.

A mathematical analysis of the fixed-node approach and the justification of the statements given above can be found in Cancès *et al.* [cances] and Rousset [rousset]. A convenient framework to analyze the fixed-node approach is to express it as a variational problem in the functional space of anti(skew)-symmetric functions with Dirichlet-type boundary conditions.

The tiling theorem. By solving the Schrödinger equation as a juxtaposition of independent problems, there is no reason why ground-state energies computed separately in each domain should be identical. The fixed-node energy is defined as the minimum of such energies. Unfortunately, in QMC calculations for non-trivial systems, the minimum found may depend on the initial conditions in the case where not all nodal domains are sampled, a situation that may arise since the number and localization of such domains in high dimension is in general not known. Hopefully, for fermionic ground-states Ceperley [ceperley-nodes] has proved under physically reasonable conditions the existence of a tiling theorem for the exact ground-state: There is only one distinct kind of nodal regions. All others are related to it by permutational symmetry (with same energy). Unfortunately, in practice we need that Ψ_T satisfies the tiling property, not just the unknown ground-state. In actual simulations, it is generally assumed that Hartree-Fock or Kohn-Sham-type wavefunctions satisfy the tiling property. Results seem to validate such a statement. However, some (mathematical) work is needed to clarify this point.

When Ψ_T is chosen to be positive and does not vanish (except at infinity) the DMC algorithm just presented will converge to the stationary density corresponding to the lowest (positive) eigenstate of H. For bosonic systems this latter state is the physical ground-state and DMC is an exact method for solving the Schrödinger equation.

When we are dealing with fermions (electrons) the situation is different. The fermionic ground-state is antisymmetric and has a non-constant sign. The algorithm presented can also be used using a fermionic Ψ_T (for example, a Hartree-Fock determinant). However, the ground-state properties obtained are no longer exact.

- Wherever Ψ_T vanishes, the drift vector diverges. The walkers are trapped for ever within the nodal cells. The problem is recast into a set of independent bosonic calculations in each nodal region.
- The nodes of Ψ_T are of two types: exchange nodes and other nodes. Exchange nodes are $(3N-3)$-dimensional, exact nodes $(3N-1)$-dimensional. The nodes are not known and there is a fixed-node bias.
- Fixed-node energy has a variational property

$$E_0^{FN} \geq E_0$$

the equality occuring when the nodes of Ψ_T are exact.

- *A priori* each simulation performed in each nodal cell leads to a different energy. The DMC energy is the minimum of them. However, Ceperley has shown that there exists a tiling property,[**?**].
- The nodes being a $(3N-1)$-dimensional object, their structure is not trivial and to decrease of fixed-node error in a systematic way is not a simple problem.

Trial wavefunction optimization

The problem

When accurate results are searched for, we need to reduce the two following errors:
(1) The statistical fluctuations related to the finite number of Monte Carlo steps
(2) The fixed-node bias related to the use of an approximate nodal hypersurface.

Both errors can be decreased by optimizing the parameters of the trial wavefunction. Different criteria can be used to define the "quality" of a trial wavefunction. The two most employed:

- Minimization of the variational energy

$$E(\mathbf{p}) = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

where $\mathbf{p}$ denotes the set of parameters of $\Psi_T(\mathbf{x}, \mathbf{p})$

- Minimization of the variance of the Hamiltonian

$$\sigma^2(\mathbf{p}) = \frac{\langle \Psi_T | [H - E(\mathbf{p})]^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

The correlated approach

The most natural idea to optimize the trial wavefunction is to minimize the total energy evaluated for a finite number of configurations N_c drawn in a preliminary Variational Monte Carlo step:

$$E(\mathbf{p}) \simeq \frac{1}{N_c} \sum_{i=1}^{N_c} E_L(\mathbf{x}_i)$$

In practice, such a idea is difficult to realize for two reasons:
(1) For a finite number of walkers $E(\mathbf{p})$ is not bounded from below and the minimizer can change parameters in a wird way so that to concentrate the wavefunction around one or a few points having a very low local energy.
(2) The stationary distribution, $\Psi_T^2(\mathbf{x}, \mathbf{p})$ depends on parameters $\mathbf{p}$, and thus new configurations must be redrawn at each change of parameters. The variational energy being calculated for a finite number of points, the energy curve $E(\mathbf{p})$ is then noisy and it is a tricky situation for the minimizer.

Practical solution:
(1) When a not too large number of configurations is used (a few thousand's) it is much preferable to minimize the variance since it is a quantity bounded from below ($\sigma^2 \geq 0$) for *any finite number* of configurations.
(2) To avoid the noisy character of $E(\mathbf{p})$ or $\sigma^2(\mathbf{p})$ a fixed set of configurations can be used and a correlated approach introduced[**?**]

$$\sigma^2(\mathbf{p}) = \frac{\frac{1}{N_c}\sum_{i=1}^{N_c} w_i (E_L - E)^2(\mathbf{x}_i, \mathbf{p})}{\frac{1}{N_c}\sum_{i=1}^{N_c} w_i}$$

where N_c number of configurations and $w_i = \frac{\Psi_T^2(\mathbf{x}_i,\mathbf{p})}{\Psi_T^2(\mathbf{x}_i,\mathbf{p}_0)}$. The configurations are drawn once for all according to $\Psi_T^2(\mathbf{x}_i, \mathbf{p}_0)$. In such conditions the energy curve is no longer noisy and standard minimizers (for example, quasi-Newton) can be employed.
Note that $\sigma^2(\mathbf{p})$ is a reasonable estimate of

$$\frac{\langle \Psi_T | (H-E)^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

only if the weights remain all close to one. It is thus important to quantify this aspect in some way, for example by introducing

$$\eta = \frac{1}{N_c} \frac{(\sum_i w_i)^2}{\sum_i {w_i}^2}$$

When η is close to one, the number of configurations playing a role is close to N_c and the estimation of the energy/variance is reasonable. In contrast, when only a few configurations contribute, η is close to zero and a new set of reference points must absolutely be drawn.

The linear method

The linear method has been recently introduced by Umrigar et al.[?] and is presently one of the most efficient approach to optimize a large number of parameters (both linear and non-linear).
The method is based on the minimization of the variational energy. Let us call N_p the number of parameters. The method consists in introducing a linear Taylor expansion around the current parameters p_0.

$$\Psi_T(\mathbf{x}, \mathbf{p}) = \Psi_T(\mathbf{x}, \mathbf{p}_0) + \sum_{i=1}^{N_p} (\mathbf{p} - \mathbf{p}_0)_i \Psi_i \tag{85}$$

where the functions Ψ_i are defined as

$$\Psi_i = \frac{\partial \Psi_T(\mathbf{x}, \mathbf{p}_0)}{\partial p_i}$$

Functions Ψ_i are now considered as a basis for the trial wavefunction and the energy is minimized in this basis set. Remarking that the Ψ_i are not orthogonal, the problem to solve is thus a generalized eigenvalue problem

$$H \Delta \mathbf{p} = E S \Delta \mathbf{p} \tag{86}$$

where H and S are the Hamiltonian and overlap matrices, respectively.

$$H_{ij} = \langle \Psi_i | H | \Psi_j \rangle \text{ and } S_{ij} = \langle \Psi_i | \Psi_j \rangle$$

These quantities can be calculated in a VMC calculation using Ψ_0^2 as stationary distribution

$$H_{ij} = \langle \frac{\Psi_i}{\Psi_0} \frac{H\Psi_j}{\Psi_0} \rangle_{\Psi_0^2}$$

and

$$S_{ij} = \langle \frac{\Psi_i}{\Psi_0} \frac{\Psi_j}{\Psi_0} \rangle_{\Psi_0^2}$$

A few numerical applications: Exploring atomic, molecular and solid-state systems

Let us now present some (very) recent applications of QMC for a variety of systems.

- **G2 benchmark**

Benchmark sets are useful in electronic structure theory. They allow to compare the results obtained by various methods against some reference data. The so-called G2 set (actually G1 set) has been introduced by Curtiss and collaborators and has been extensively used as benchmark in quantum chemistry. The benchmark consists in comparing the (corrected) experimental values for the atomization energies of a set of $N_{mol} = 55$ simple molecules with those obtained with the method to be evaluated. The criterium used is the mean absolute deviation (MAD) defined as

$$MAD = \frac{1}{N_{mol}} \sum_{i=1}^{N_{mol}} |E_i^{at} - E_i^{at;expt}|$$

where E_i^{at} is the atomization energy of molecule i. The smaller the MAD is the better the method reproduces the experimental values.

DFT and post-HF methods
- LDA: MAD $\sim$ 40 kcal/mol
- B3LYP and B3PW91: MAD $\sim$ 2.5 kcal/mol
- CCSDT/aug-cc-pVQZ MAD $\sim$ 2.8 kcal/mol
- CCSDT Complete Basis Set limit, MAD $\sim$ 1.3 kcal/mol

QMC
- Grossman (2002) HF nodes, use of pseudo-potientials, MAD $\sim$ 2.9 kcal/mol
- Nemec et al. (2010) HF nodes, all-electron, MAD $\sim$ 3.2 kcal/mol.
- Petruziello et al. (2012) MAD: 5z basis set $\sim$ 1.2 kcal/mol.

The best MAD obtained with QMC is comparable to that obtained with CCSDT in the infinite basis set limit.

• **Non-covalent interactions**

See Table 1 of the paper of Dubecky *et al.* (2016) "Noncovalent Interactions by Quantum Monte Carlo".[**?**] A long list of references (in chronological order) presenting QMC calculations of noncovalent interactions is given.

• **Water nano-droplets** Reference: [**?**]

• **Barrier heights** Krongchon *et al.* "Accurate barrier heights using diffusion Monte Carlo" (2017)[**?**] Benchmark calculations of the barrier heights of 19 non-hydrogen-transfer chemical reactions.

- **3d-metal containing molecules**

K. Doblhoff-Dier et al. "Diffusion Monte Carlo for Accurate Dissociation Energies of 3d Transition Metal Containing Molecules" (2016)[**?**]. Benchmark calculations of for 20 transition metal containing dimers. Set introduced by Truhlar *et al.* (2015)[**?**]

- **H and He under very high pressure** Reference: [**?**]
- **Cuprates** Reference: [**?**],[**?**],[**?**]
- **Solids** References: [**?**],[**?**]

APPENDIX: L'Ecuyer pseudo-random generator

The L'Ecuyer pseudo-random generator is a combined multiple recursive generator

$$z_n = (x_n - y_n) \bmod m_1$$

where x_n and y_n are
$x_n = (a_1 x_{n-1} + a_2 x_{n-2} + a_3 x_{n-3}) \bmod m_1$
$y_n = (b_1 y_{n-1} + b_2 y_{n-2} + b_3 y_{n-3}) \bmod m_2$
with coefficients
$a_1 = 0, a_2 = 63308, a_3 = -183326, b_1 = 86098, b_2 = 0, b_3 = -539608$, and moduli
$m_1 = 2^{31} - 1 = 2147483647$ and $m_2 = 2145483479$.
The period is approximately 2^{185} (about 10^{56}).

APPENDIX: Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is associated with a linear drift vector

$$b(x) = -kx$$

, where k some positive constant. The transition probability density is

$$P(x \to y, t) = \frac{1}{\sqrt{1-\gamma^2}} \exp -\frac{(y-\gamma x)^2}{\sqrt{1-\gamma^2}}$$

where $\gamma = e^{-kt}$.

APPENDIX: Derivation of the Metropolis algorithm in the discrete case

- Def. 1 Probability distribution $\pi_i \geq 0$ i=1,N and $\sum_i \pi_i = 1$
- Def. 2 transition probability (or stochastic matrix) $P_{i\to j}$:
  i. $P_{i\to j} \geq 0$
  ii. $\sum_{j=1}^{N} P_{i\to j} = 1$ (independent on i)
- Def. 3 *Ergodic* transition probability
  $\forall i_0 \ \ \forall i$ there exist a non-zero probability that after a *finite* number of steps starting from i_0 we end at i.
- Def. 4 Stationary (or invariant) distribution π:

$$\sum_i \pi_i P_{i\to j} = \pi_j$$

Metropolis algorithm
Let $P^T_{i\to j}$ being a trial ergodic transition probability, then $P_{i\to j}$ defined as follows

$$\begin{cases} P_{i\to j} = P^T_{i\to j}\mathrm{Min}(1, R_{ij}) & \mathrm{j} \neq \mathrm{i} \\ P_{i\to i} = P^T_{i\to i} + \sum_{k\neq i} P^T_{i\to k}(1 - \mathrm{Min}(1, R_{ik})) & \mathrm{j} = \mathrm{i} \\ \text{with} \quad R_{ij} = \frac{\pi_j P^T_{j\to i}}{\pi_i P^T_{i\to j}} \end{cases}$$

is an ergodic transition probability admitting π_i as stationary distribution.
Proof:

- $P_{i\to j}$ is a transition probability
- $P_{i\to j} \geq 0$ obvious
- $\sum_{j=1}^{N} P_{i\to j} = \sum_{j\neq i} P_{i\to j} + P_{i\to i}$
  $= \sum_{j\neq i} P^T_{i\to j} + P^T_{i\to i}$
  $= 1$

Stationary distribution
We have to show: $\sum_i \pi_i P_{i\to j} = \pi_j$
For that we first show that $\{P_{i\to j}; \pi_i\}$ obeys detailed balance

$$\pi_i P_{i\to j} = \pi_j P_{j\to i} \quad \forall(i, j)$$

Proof:

- i=j obvious
- i≠j: the ratio of the two sides of the previous equality is

$$\frac{\pi_j P_{j\to i}}{\pi_i P_{i\to j}} = \frac{R_{ij}\mathrm{Min}(1, R_{ji})}{\mathrm{Min}(1, R_{ij})}.$$

  Remarking that $R_{ij} = 1/R_{ji}$ and distinguising between the two cases corresponding to $R_{ij} \geq 1$ and $R_{ij} < 1$, we easily verify that this ratio is equal to 1.

Finally, using the detailed balance relation we get

$$\sum_i \pi_i P_{i\to j} = \sum_i \pi_j P_{j\to i} = \pi_j$$

thus, π_i is the stationary distrbution.

APPENDIX: Convergence of the Metropolis algorithm

Let us precise the way the distribution converges to the stationary one.
Let $f^{(k)}$ be a distribution, that is a set of N positive real numbers. The application of the stochastoc matrix to this distribution is written as

$$f_i^{(k+1)} = \sum_j f_j^{(k)} P_{j\to i} \equiv P f_i^{(k)}$$

We have the following property

$$\lim_{n\to\infty} f_i^{(n)} \sim P^n f_i^{(0)} = \pi_i \qquad \forall f^{(0)}$$

The different steps of the proof are as follows.

- Let us associate to $P_{i\to j}$ a *symmetric* real matrix defined as follows

$$M_{ij} = \sqrt{\pi_i} P_{i\to j} \frac{1}{\sqrt{\pi_j}}$$

  Let us insist that the stochastic matrix is in general *not symmetric*.

- It is easy to check that $\sqrt{\pi}$ is eigenstate of M with eigenvalue 1

$$\sum_j M_{ij} \sqrt{\pi_j} = \sqrt{\pi_i}$$

- We also see that

$$P^n f^{(0)} = \sqrt{\pi} M^n \frac{f^{(0)}}{\sqrt{\pi}}$$

- Let us now use the spectral decomposition of M. For large n, M^n becomes the projector in the eigenspace associated with the largest eigenvalue. Due to its particular structure, it can be shown that M has eigenvalues λ_i such that $0 \leq |\lambda_i| \leq 1$ and in the case where π does not vanish, the associated eigenspace is not degenerate. As a consequence

$$P^n f^{(0)} = c\ \pi$$

where c is the overlap between the initial distribution $f^{(0)}/\sqrt{\pi}$ and the eigenstate $\sqrt{\pi}$ of matrix M.