UNITED NATIONS EDUCATIONAL SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

**SMR.998a - 5**

Research Workshop on Condensed Matter Physics
30 June - 22 August 1997
**MINIWORKSHOP ON**
**QUANTUM MONTE CARLO SIMULATIONS OF LIQUIDS AND SOLIDS**
**30 JUNE - 11 JULY 1997**
**and**
**CONFERENCE ON**
**QUANTUM SOLIDS AND POLARIZED SYSTEMS**
**3 - 5 JULY 1997**

---

# "Introduction to Quantum Monte Carlo Methods"

**G. SENATORE**
**Dipartimento di Fisica Teorica**
**Universita' di Trieste**
**Strada Costiera 11**
**Trieste**
**Italy**

---

**These are preliminary lecture notes, intended only for distribution to participants.**

# Introduction to Quantum Monte Carlo Methods

Gaetano Senatore
Dipartimento di Fisica Teorica,
Università di Trieste, Italy
(senatore@ts.infn.it)

*Miniworkshop on*
*Quantum Monte Carlo Simulation of Liquids and Solids*
ICTP, Trieste, 30 June – 11 July 1997



## Outline

- Quantum Averages
- The Monte Carlo Method
- Variational Monte Carlo (VMC)
- Green Function Monte Carlo (DMC & GFMC )
- Path Integral Monte Carlo (PIMC)
- Path Integral Ground State
- Auxiliary Field Monte Carlo (?)

## Hamiltonian and notation

- The N-body Hamiltonian (for a one–component system!) is

$$H = -D\nabla^2 + V(R),$$

with

$$D = \frac{\hbar^2}{2m}, \quad R = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N),$$

and

$$\nabla \equiv \nabla_R = (\nabla_1, \nabla_2, \ldots, \nabla_N).$$

- $R$ is a $d$·N–dimensional vector

- $\mathcal{O}$ is a generic Hermitian operator ( $I$, $H$, $\hat{n}(\mathbf{r})$, $\hat{n}(\mathbf{r},\mathbf{r}')$, $\hat{\gamma}(\mathbf{r},\mathbf{r}')$, . . . )

- $\tau$ and $\beta$ are respectively the imaginary time ($\tau = -it$) and the inverse temperature ($\beta = 1/(K_B T)$).



## Quantum Averages

Quantum averages involve multidimensional integrations:

- (Ground) state average

$$\langle \mathcal{O} \rangle_\phi = \frac{\langle \phi|\mathcal{O}|\phi\rangle}{\langle \phi|\phi\rangle} = \frac{\int dR\, \phi(R)^* \mathcal{O}\phi(R)}{\int dR |\phi(R)|^2}$$

$$= \int dR \left[\frac{|\phi(R)|^2}{\langle \phi|\phi\rangle}\right]\left[\frac{\mathcal{O}\phi(R)}{\phi(R)}\right] \equiv \int dR\, \pi(R)\mathcal{O}(R).$$

- *Temperature* average

$$\langle \mathcal{O} \rangle_\beta = \frac{Tr\, e^{-\beta H}\mathcal{O}}{Z(\beta)} = \frac{\int dR \langle R|e^{-\beta H}\mathcal{O}|R\rangle}{Z(\beta)}$$

$$= \frac{\int dR dR' \rho(R,R';\beta)\mathcal{O}(R,R')}{Z(\beta)},$$

$$Z(\beta) = Tr\, e^{-\beta H} = \int dR \rho(R,R;\beta)$$

with $Z(\beta)$ the canonical partition function and $\rho(R,R';\beta)$ the N–body *temperature* density matrix.

## Multidimensional Integration

- Quadrature (Simpson–like) schemes are unfeasible!

  A regular grid with 10 mesh point per axis would require $10^{d \cdot N}$ evaluations of the integrand, for N particles in $d$ dimensions, i.e., $10^{30}$ operations for 10 particles in 3 dimensions!

  In other words, if the error behaves like $h^l$, with $h$ the mesh size*, its scaling with the number M of evaluations is

$$\text{error} \propto 1/M^{-l/(d \cdot N)}$$

  As $l$ is of order unity, the error decays exceedingly slowly with $M$. In fact the higher is the dimension $d \cdot N$ the slower it decays.

  For 20 particles in 2 dimensions and $l = 4$, halving the error of an evaluation with M points requires going to $2^{2 \cdot 20/4} \cdot M = 1024 \cdot M$ points; to reduce the error by a factor 4 requires $10^6 \cdot M$ points, and so on!

  * If we integrate over a hypercube with side $L$, the number of grid points is $M = (L/h)^{d \cdot N}$, i.e. $h \propto M^{-1/(d \cdot N)}$.



- Monte Carlo Integration is the only choice:

$$\int dR\, \pi(R) \mathcal{O}(R) \simeq \frac{1}{M} \sum_{i=1}^{M} \mathcal{O}(R_i),$$

  with an

$$\text{error} \propto 1/\sqrt{M},$$

  provided that the configurations or *walkers* $\{R_i\}$ are distributed with the probability $\pi(R)$.

  To halve the error only $4 \cdot M$ points are required; $16 \cdot M$ points are sufficient to reduce the error by a factor 4; and so on. Also, there is no dependence on the dimensionality of the configuration space.

  Configurations distributed with a given probability can be generated with a variety of algorithms:

  - (generalized) Metropolis algorithm,
  - Molecular Dynamics,
  - Langevin Dynamics,
  - combination of the above,
  - other.

  In the following we shall restrict to the (generalized) Metropolis algorithm.

– G. Senatore –

# Random Walks (Markov Chains)

The state $s$ of the system is changed randomly according to a transition probability $p(s, s') = p(s \to s')$ satisfying

$$\sum_{s'} p(s, s') = 1$$

and

$$p(s, s') \geq 0,$$

thus generating a random walk (or sample) $(s_0, s_1, s_2, \ldots)$.
If $p(s, s')$ is ergodic there exists a (unique) probability measure $\pi(s)$ satisfying at equilibrium the stationarity condition:

$$\sum_{s} \pi(s) p(s, s') = \pi(s').$$

Moreover if $p^n(s, s')$ is the probability to reach $s'$ from $s$ in $n$ steps then

$$\lim_{n \to \infty} p^n(s, s') = \pi(s') :$$

the random walk converges to the equilibrium distribution irrispective of the initial distribution.



## Ergodicity

1. Irreducibility: for each $(s, s')$ there exists an $n \geq 0$ such that $p^n(s, s') > 0$;

2. Aperiodicity: $p(s, s) > 0$;

3. The average return time is finite: it exists $N_{s,s'} < \infty$ such that, for $n > N_{s,s'}$, $p^n(s, s') > 0$.

## Detailed balance

A sufficient condition to obtain $\pi(s)$ as stationary distribution is to chose the transition probability to satisfy

$$\pi(s) p(s, s') = \pi(s') p(s', s).$$

In fact summing the above over $s$ one gets

$$\sum_{s} \pi(s) p(s, s') = \pi(s') \sum_{s} p(s', s) = \pi(s')$$

– G. Senatore –

## Generalized Metropolis Algorithm

The transition probability may be conveniently decomposed into the product of an irreducible proposal or sampling matrix $T(s,s')$ and an acceptance matrix $A(s,s')$

$$p(s,s') = T(s,s')A(s,s').$$

Imposing the detailed balance yields

$$\frac{A(s,s')}{A(s',s)} = \frac{\pi(s')T(s',s)}{\pi(s)T(s,s')} \equiv q(s,s'),$$

which can be satisfied quite generally by chosing

$$A(s,s') = F[q(s,s')],$$

where the function $F : [0,\infty] \to [0,1]$ satisfies

$$\frac{F[z]}{F[1/z]} = z, \quad \text{for} \quad \text{all} \quad z.$$

- Metropolis choice:

$$F[z] = min[1,z]$$

- An altyernative choice could be:

$$F[z] = \frac{z}{1+z}$$



## Implementation of Metropolis algorithm

Given a probability $\pi(s)$ to sample (it may be not known in closed form, see, e.g., DMC, GFMC):

1. Chose the proposal matrix $T(s,s')$ ;

2. Initialize the system in the state $s_0$;

3. To advance from $s_n$ to $s_{n+1}$:
   - sample $s'$ from $T(s_n,s')$,
   - calculate
   $$q(s_n,s') = \frac{\pi(s')T(s',s_n)}{\pi(s_n)T(s_n,s')}$$
   ,
   - generate a random number $r_n$ and compare it with $q(s_n,s')$:
     – if $q(s_n,s') > r_n$: $s_{n+1} = s'$
     – else $s_{n+1} = s_n$.

4. Throw away the first $k$ *states* as being out of equilibrium;

5. Collect averages using the configurations with $n > k$ and block them to calculate error bars (???).

6. Example: T a constant in a cube, $\pi(s) \propto exp(-\beta V(s))$.

Some facts about Metropolis:

- The normalization of the probability, $\int ds\, \pi(s)$, is never needed and in fact cannot be calculated (... easily).
- Particles can be moved one at time ( hard spheres!);
- For the generalized algorithm ($T(s, s')$ is not a constant) one has to sample both forward and reverse transition;
- An optimal acceptance is

$$\mathcal{A} = \frac{\text{moves accepted}}{\text{total moves}} \simeq 1/2.$$

  In fact the overall efficiency may dictate different choices (see, e.g., DMC).

- The length of the necessary *thermalization* (deciding the number $k$ of initial moves to discard) can be investigated monitoring cumulative averages of physically relevant quantities (energy, density profile, ...).
- But ... what are averages anyway?



## Monte Carlo Estimates and Averages

One would like to evaluate the *true* mean

$$\langle \mathcal{O} \rangle = \int ds\, \pi(s) \mathcal{O}(s),$$

whereas MC yield a sample $(s_1, s_2, ..., s_M)$ of lenght $\simeq M$ of states distributed according to $\pi(s)$. Evidently, one can define a sample mean

$$\overline{\mathcal{O}} = \frac{1}{M} \sum_{i=1}^{M} \mathcal{O}_i,$$

with $\mathcal{O}_i \equiv \mathcal{O}(s_i)$.

The sample mean is an unbiased estimator of the true mean, i.e., $\langle \overline{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle$ independently of $M$.Also, it is possible to prove:

- the law of large numbers, $\lim_{M \to \infty} \overline{\mathcal{O}} = \langle \mathcal{O} \rangle$;
- The central limit theorem, which states that $\overline{\mathcal{O}}$ is *normally* distributed around $\langle \mathcal{O} \rangle$.

Therefore we need to evaluate the variance

$$\sigma^2(\overline{\mathcal{O}}) = \langle (\overline{\mathcal{O}} - \langle \mathcal{O} \rangle)^2 \rangle,$$

whose root we may interpret as statistical error on $\overline{\mathcal{O}}$.

## The statistical error

Using $\overline{\mathcal{O}} = (1/M)\sum_{i=1}^{M} \mathcal{O}_i$, one obtaines for the variance

$$\begin{aligned} \sigma^2(\overline{\mathcal{O}}) &= \frac{1}{M}\sum_{t=-(M-1)}^{t=M-1}\left(1-\frac{|t|}{M}\right)C(t) \\ &\approx \frac{\tau}{M}C(0) = \frac{\tau}{M}\sigma^2(\mathcal{O}). \end{aligned}$$

Here

$$C(t) = \langle \mathcal{O}_s \mathcal{O}_{s+t}\rangle - \langle \mathcal{O}\rangle^2$$

is the unnormalized *time* autocorrelation function, which evidently reduces to the variance of $\mathcal{O}$ at time 0, $C(0) = \sigma^2(\mathcal{O})$, and the integrated correlation time

$$\tau = 1 + 2\sum_{t=1}^{\infty}\frac{C(t)}{C(0)},$$

accounts for the correlation existing between walkers in the Markov chain. In general $\tau > 1$.



A sample estimate of $C(t)$, with a bias of order $1/M$ is given by

$$\tilde{C}(t) = \frac{1}{M-|t|}\sum_{i=1}^{M-|t|}(\mathcal{O}_i - \overline{\mathcal{O}})(\mathcal{O}_{i+|t|} - \overline{\mathcal{O}}).$$

Thus one has an estimate for $\sigma^2(\mathcal{O}) = C(0) \approx \tilde{C}(0)$,

$$\tilde{\sigma}^2(\mathcal{O}) = \frac{1}{M}\sum_{i=1}^{M}(\mathcal{O}_i - \overline{\mathcal{O}})^2,$$

and the correlation time can also be calculated from $\tilde{C}(t)$.

– G. Senatore –

## Blocking

The precise estimate of the error bar require the calculation of time correlation functions, which one would rather avoid. An alternative is provided by the blocking procedure. The sample is broken in a number of blocks $M = N_b n_b$, with $N_B$ the number of blocks and and $n_b$ the length of each block. New variable are constructed as block averages

$$\mathcal{O}_{b,I} = \frac{1}{n_b} \sum_{i=1}^{n_b} \mathcal{O}_{(I-1)n_b+i},$$

and clearly have a mean equal to the run mean $\overline{\mathcal{O}}$. Intuitively, if $n_b \gg \tau$, this new variables should become statistically independent and therefore have a a variance around their mean $\overline{\mathcal{O}}$ given by

$$\sigma^2(\mathcal{O}_b) = \frac{1}{N_b(N_b-1)} \sum_{I=1}^{N_b} (\mathcal{O}_{b,I} - \overline{\mathcal{O}})^2$$

One can indeed show that provided $n_b \gg \tau$ and yet $n_b \ll M$ or equivalently $N_b$ large

$$\sigma^2(\mathcal{O}_b) \simeq \sigma^2(\overline{\mathcal{O}}).$$

A plot of $\sigma^2(\mathcal{O}_b)$ versus $n_b$ will reveal a plateau, where in fact the above relation holds, and therefore it also yields an estimate of the correlation time.