Random Walks

\[ \int dR e^{-V(R)/k_BT} \]

- It difficult to \textit{directly} sample the Boltzmann distribution for many particles.
  - If we sample from another distribution, \textbf{the overlap will be order \( \exp(-aN) \)}, where \( N \) is the number of variables. The error bars will get exponentially larger as \( N \) increases.

- \textbf{Markov chains} (random walks), \textit{detailed balance} and transition rules.
  - This method was introduced by \textit{Metropolis et al.} in 1953 who applied it to a hard sphere liquid.
  - It is one of the most powerful and used algorithms in all of computer science.
Equation of State Calculations by Fast Computing Machines

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

Edward Teller,* Department of Physics, University of Chicago, Chicago, Illinois
(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of
state for substances consisting of interacting individual molecules is described. The method consists of a
modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere
system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared
to the free volume equation of state and to a four-term virial coefficient expansion.

I. INTRODUCTION

The purpose of this paper is to describe a general
method, suitable for fast electronic computing
machines, of calculating the properties of any substance
which may be considered as composed of interacting
individual molecules. Classical statistics is assumed,
only two-body forces are considered, and the potential
field of a molecule is assumed spherically symmetric.
These are the usual assumptions made in theories of
liquids. Subject to the above assumptions, the method
is not restricted to any range of temperature or density.
This paper will also present results of a preliminary two-
dimensional calculation for the rigid-sphere system.
Work on the two-dimensional case with a Lennard-
Jones potential is in progress and will be reported in a
later paper. Also, the problem in three dimensions is
being investigated.

II. THE GENERAL METHOD FOR AN ARBITRARY
POTENTIAL BETWEEN THE PARTICLES

In order to reduce the problem to a feasible size for
numerical work, we can, of course, consider only a finite
number of particles. This number $N$ may be as high as
several hundred. Our system consists of a square† cont-
aining $N$ particles. In order to minimize the surface
effects we suppose the complete substance to be periodic,
consisting of many such squares, each square contain-
ing $N$ particles in the same configuration. Thus we
define $d_{AB}$, the minimum distance between particles $A$
and $B$, as the shortest distance between $A$ and any of
the particles $B$, of which there is one in each of the
squares which comprise the complete substance. If we
have a potential which falls off rapidly with distance,
there will be at most one of the distances $AB$ which
can make a substantial contribution; hence we need
consider only the minimum distance $d_{AB}$.

* Now at the Radiation Laboratory of the University of Cali-
forinia, Livermore, California.

† We will use the two-dimensional nomenclature here since it
is easier to visualize. The extension to three dimensions is obvious.
Markov chain or Random Walk

- Markov chain is a random walk through phase space:
  \[ s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \rightarrow \ldots \]
  Here “s” is the state of the system.

- The transition probability is \( P(s_n \rightarrow s_{n+1}) \) a stochastic matrix:
  \[ P(s \rightarrow s') \geq 0 \quad \sum_{s'} P(s \rightarrow s') = 1 \]

- In a Markov chain, the distribution of \( s_{n+1} \) depends only on \( s_n \) (by definition).
- Let \( f_n(s) \) be the probability after “n” steps. It evolves according to a “master equation.”
  \[ f_{n+1}(s') = \sum_s f_n(s) P(s \rightarrow s') \quad \text{or} \quad f_{n+1} = P f_n \]

- The stationary states are eigenfunctions of \( P \):
  \[ P \phi = \varepsilon \phi \]
Properties of Random Walk

- Because $P$ is positive, the eigenvalues have $\varepsilon \leq 1$. 
  
  *An equilibrium state must have $\varepsilon = 1$.*

- How many equilibrium states are there?
  - If $P$ is *ergodic*, then $f$ will converge to a unique stationary distribution. Only one eigenfunction has $\varepsilon = 1$.

- In contrast to molecular dynamics, ergodicity can be proven if:
  - One can move everywhere in a finite number of steps with non-zero probability. *No barriers!*
  - Non-periodic transition rules (e.g. hopping on bi-partite lattice).
  - Average return time is finite. (No expanding universe.) This is guaranteed for a finite system.

- If ergodic, convergence is *geometrical and monotonic*.

$$f_n(s) = n(s) + \sum_{\nu} \varepsilon^n c_{\nu} \phi_{\nu}(s)$$

If $\varepsilon < 1$, then after “$n$” iterations, this is 0! Hence, $\varepsilon = 1$ in the stationary state.
Metropolis algorithm

Three key concepts:
1. Sample by using an ergodic random walk.
2. Determine equilibrium state by using detailed balance.
3. Achieve detailed balance by using rejections.

**Detailed balance:** \( \pi(s) P(s \to s') = \pi(s') P(s' \to s) \).

*Ratel balance from s to s’.*

Put \( \pi(s) \) into the master equation. (*Or sum above Eq. on s.*)

\[
\sum_s \pi(s) P(s \to s') = \pi(s') \sum_s P(s' \to s) = \pi(s')
\]

- Hence, \( \pi(s) \) is an eigenfunction of \( P \).
- Then if \( P(s \to s') \) is ergodic, then \( \pi(s) \) is the unique steady state solution.
- Possible to stay in same state: \( P(s \to s) = 1 - \sum_{s' \neq s} P(s' \to s) \)
Rejection Method

Metropolis achieves detailed balance by rejecting moves.

General Approach:
1. Choose distribution to sample, e.g., \( n(s) = \exp[-\beta H(s)]/Z \)
2. Impose detailed balance on transition: \( K(s \rightarrow s') = K(s' \rightarrow s) \)
   where \( K(s \rightarrow s') = n(s) P(s \rightarrow s') \)
   \((\text{probability of being at } s) \times (\text{transition probability of going to } s').\)
3. Break up transition probability into sampling and acceptance:
   \( P(s \rightarrow s') = T(s \rightarrow s') A(s \rightarrow s') \)
   \((\text{probability of generating } s' \text{ from } s) \times (\text{probability of accepting move}).\)

The optimal acceptance probability that gives detailed balance is:

\[
A(s \rightarrow s') = \min[1, \frac{T(s' \rightarrow s)\pi(s')}{T(s \rightarrow s')\pi(s)}] = \min[1, \frac{\pi(s')}{\pi(s)}]
\]

Normalization of \( \pi(s) \) is not needed or used!

If \( T \) is constant!
Quantum Monte Carlo

We need to go beyond classical simulations since the microscopic world is quantum mechanical.

The methods I will describe are the simplest generalization of classical molecular dynamics and Monte Carlo.
There are many types of QMC

- Lattice models: DetMC, DiagMC, Lattice Gauge MC
- Second quantized methods: FCI-QMC, AF-QMC,….

I will only introduce continuum methods for strongly correlated systems:

- Non-zero temperature: path integral MC
- Zero temperature: projector quantum MC
  - Variational MC
  - Diffusion MC

- We can treat $10^5$ bosons, $10^3$ fermions.
Atomic Theory of the \( \alpha \) Transition in Helium

R. P. Feynman
California Institute of Technology, Pasadena, California
(Received May 15, 1953)

It is shown from first principles that, in spite of the large interatomic forces, liquid He\(^\text{4}\) should exhibit a transition analogous to the transition in an ideal Bose-Einstein gas. The exact partition function is written as an integral over trajectories, using the space-time approach to quantum mechanics. It is next argued that the motion of one atom through the others is not opposed by a potential barrier because the others may move out of the way. This just increases the effective inertia of the moving atom. This permits a simpler form to be written for the partition function. A rough analysis of this form shows the existence of a transition, but of the third order. It is possible that a more complete analysis would show that the transition implied by the simplified partition function is actually like the experimental one.
Quantum Simulations

• We do Classical Monte Carlo simulations to evaluate averages such as:

\[ \langle V \rangle = \frac{1}{Z} \int dRV(R)e^{-\beta V(R)} \quad \text{with} \quad \beta = 1/(k_BT) \]

• Quantum mechanically for T>0, we need both to generate the distribution and do the average:

\[ \langle V \rangle = \frac{1}{Z} \int dRV(R)\rho(R;\beta) \]

\[ \rho(R;\beta) = \text{diagonal density matrix} \]

• Simulation is possible since the density matrix is positive.
The thermal density matrix

\[ \hat{H} \phi_\alpha = E_\alpha \phi_\alpha \]

\[ \rho(R; \beta) = \sum_\alpha |\phi_\alpha(R)|^2 e^{-\beta E_\alpha} \]

\[ \hat{\rho}_\beta = e^{-\beta \hat{H}} \quad \text{operator notation} \]

off-diagonal density matrix:

\[ \rho(R, R'; \beta) = \sum_\alpha \phi_\alpha^*(R') \phi_\alpha(R) e^{-\beta E_\alpha} \]

\[ \rho(R, R'; \beta) \geq 0 \quad \text{(without statistics)} \]

\[ \rho(R_1, R_2; \beta_1 + \beta_2) = \]

\[ = \int dR' \rho(R_1, R'; \beta_1) \rho(R', R_2; \beta_2) \]

or with operators: \[ e^{-(\beta_1+\beta_2)\hat{H}} = e^{-\beta_1\hat{H}} e^{-\beta_2\hat{H}} \]
Trotter’s formula (1959)

- We can use the effects of operators separately as long as we take small enough time steps.
  \[ \hat{\rho} = e^{-\beta(\hat{T} + \hat{V})} \]
  \[ \hat{\rho} = \lim_{M \to \infty} \left[ e^{-\tau \hat{T}} e^{-\tau \hat{V}} \right]^M \]
  \[ \tau = \beta / M \]

- M is number of time slices.
- \( \tau \) is the “time-step”

- We now have to evaluate the density matrix for potential and kinetic matrices by themselves:
  - Do by FT’s
    \[ \langle r | e^{-\tau \hat{T}} | r' \rangle = \left( 4\pi \lambda \tau \right)^{-3/2} e^{-(r-r')^2/4\lambda \tau} \]
  - V is “diagonal”
    \[ \langle r | e^{-\tau \hat{V}} | r' \rangle = \delta(r - r') e^{-\tau V(r)} \]
  - Error at finite n is roughly:
    \[ e^{-\tau^2 [\hat{T},\hat{V}]/2} \]
Using this for the density matrix.

• We sample the distribution:

\[
- \sum_{i=1}^{M} S(R_i, R_{i+1}; \tau) \frac{e^{-S(R_i, R_{i+1}; \tau)}}{Z} \quad \text{with} \quad Z = \int dR_1 \ldots dR_M e^{-\sum_{i=1}^{M} S(R_i, R_{i+1}; \tau)}
\]

Where the “primitive” link action is:

\[
S(R_0, R_1) = \frac{(R_0 - R_1)^2}{4\lambda \tau} + \frac{\tau}{2} \left[ V(R_0) + V(R_1) \right]
\]

• Similar to a classical integrand where each particle turns into a “polymer.”
  – K.E. is spring term holding the polymer together.
  – P.E. is the inter-polymer potential.
• Trace implies \(R_1 = R_{m+1}\) \(\Rightarrow\) closed or ring polymers
“Distinguishable” particles

- Each atom is a ring polymer; an exact representation of a quantum wavepacket in imaginary time.
- The dots represent the “start” of the path (but all points are equivalent).
- The lower the real temperature, the longer the “string” and the more spread out the wavepacket.
Quantum statistics

- For quantum many-body problems, not all states are allowed: only totally symmetric or antisymmetric ones. Statistics are the origin of BEC, superfluidity, the lambda transition.
- Use permutation operator to project out the correct states:

\[ \hat{P}f(R) = \sum_{P=1}^{N!} \frac{1}{N!} f(PR) \]

\[ Z = \sum_{P=1}^{N!} \frac{1}{N!} \int dR_1 ... dR_M e^{-\sum_{i=1}^{M} S(R_i, R_{i+1})} \]

- Means the path closes on itself with a permutation. \( R_1 = PR_{M+1} \)
- Too many permutations to sum over; we sample them.
- PIMC task: sample path \( \{ R_1, R_2, ... R_M \} \) and \( P \) with Metropolis Monte Carlo (MCMC) using “action”, \( S \), to accept/reject.
Bosonic Path Integrals

- Average by sampling over all paths and over connections.
- Trial moves involve reconnecting paths differently.
- At the superfluid transition a “macroscopic” permutation appears.
- This is reflection of bose condensation within PIMC.
“Direct” Fermion Path Integrals

\[ Z = \sum_{p=1}^{N!} \frac{(-1)^p}{N!} \int dR_1 \ldots dR_M e^{-\sum_{i=1}^{M} S(R_i, R_{i+1})} \]

- Bosons are easy: simply sample \( P \).
- Fermions: sample the “action” and carry \((-1)^p\) as a weight.
- Observable is even \( P \) - odd \( P \). Scales exponentially in \( N \) and \( T^{-1} \)!
- THIS IS THE SIGN PROBLEM

\[ \text{CPUtime} \sim \varepsilon^{-2} e^{2N[F_F - F_B]/k_B T} \]

\( \varepsilon = \) statistical error

F = free energy
First Major QMC Calculation

- Variational Monte Carlo calculation of ground state of liquid $^4$He.
- Applied MC techniques from classical liquid theory.
- Ceperley, Chester and Kalos (1976) generalized to fermions.
Variational Monte Carlo (VMC)

- Variational Principle. Given an appropriate trial function:
  - Continuous
  - Proper symmetry
  - Normalizable
  - Finite variance
- Quantum chemistry uses a product of single particle functions, e.g. Hartree-Fock
- With MC we can use any “computable” function.

  - Sample R from $|\psi|^2$ using MCMC.
  - Take average of local energy:
  - Optimize $\psi$ to get the best upper bound

$$E_V = \frac{\int dR \langle \psi | H | \psi \rangle}{\int dR \langle \psi \psi \rangle} \geq E_0$$

$$\sigma^2 = \frac{\int dR \langle \psi | H^2 | \psi \rangle}{\int dR \langle \psi \psi \rangle} - E_V^2$$

$$E_L(R) = \Re \left[ \psi^{-1}(R) H \psi(R) \right]$$

$$E_V = \langle E_L(R) \rangle_{\psi^2} \geq E_0$$

- Better wavefunction, lower variance! “Zero variance” principle. (non-classical)
Jastrow trial function for liquid $^4$He

- Whenever 2 atoms get close together wavefunction should vanish.
- The function $u(r)$ is similar to classical potential. The Jastrow
- Local energy has the form:

$$\psi(R) = \prod_{i<j} e^{-u(r_{ij})}$$

$$E_\psi(R) = \sum_{i<j} v(r_{ij}) - 2\lambda \nabla^2 u(r_{ij}) - \lambda \sum_i G_i^2$$

$$G_i = \sum \nabla_i u(r_{ij})$$

- We want finite variance of the local energy.

If $v(r)$ diverges as $r^{-n}$ how should $u(r)$ diverge? Assume:

$$U(r) = \alpha r^{-m}$$

Gives a cusp condition on $u$ and value of $m$. 

Ceperley Tutorial IPAM 51
Fermions: antisymmetric trial function

- At mean field level the wavefunction is a Slater determinant.
- Orbitals for homogenous systems are a filled set of plane waves.
- We can compute this energy analytically (HF).
- To include correlation we multiply by Jastrow. We need MC to evaluate properties.
- New feature: how to compute the derivatives of a determinant and sample the determinant. Use tricks from linear algebra.
- Reduces complexity to $O(N^2)$.

$$\Psi_s (R) = \text{Det} \left\{ e^{ik_i r_j} \eta_i (\sigma_j) \right\}$$

PBC: \quad \mathbf{k} \cdot \mathbf{L} = 2\pi n + \{\theta\}

$$\Psi_{SJ} (R) = \text{Det} \left\{ e^{ik_i r_j} \right\} e^{-\sum_{i<j}^{} u(r_{ij})}$$

Slater-Jastrow trial function.

$$\det (\phi_k (r_j^T)) = \det (\phi_k (r_j)) \sum_k \phi_k (r_j^T) M_{k,i}^{-1}$$

$$\frac{1}{\det (M)} \frac{\partial \det (M)}{\partial a} = \text{Tr} \left\{ M^{-1} \frac{\partial M}{\partial a} \right\}$$
Dependence of energy on wavefunction

3d Electron fluid at a density $r_s=10$

*Kwon, Ceperley, Martin, Phys. Rev. B58, 6800, 1998*

- Wavefunctions
  - Slater-Jastrow (SJ)
  - three-body (3)
  - backflow (BF)
- Energy $\langle \psi | H | \psi \rangle$ converges to ground state
- Variance $\langle \psi [H-E]^2 \psi \rangle$ to zero.
- Using 3B-BF gains a factor of 4.
Advantages and problems with variational methods

- Powerful method since you can use any trial function
- Scaling (computational effort vs. size) is almost classical
- Learn directly about what works in wavefunctions
- No sign problem
- Optimization is time consuming
- Energy is insensitive to order parameter
- Non-energetic properties are less accurate. $O(1)$ vs. $O(2)$ for energy.
- Difficult to find out how accurate results are.
- Favors simple states over more complicated states, e.g.
  - Solid over liquid
  - Polarized over unpolarized

What goes into the trial wave function comes out!

We need a more automatic method!

Projector Monte Carlo \ OR \ machine learning techniques
Projector Monte Carlo

• Originally suggested by Fermi and implemented in 1950 by Donsker and Kac for H atom.

• Practical methods and applications developed by Kalos:

PHYSICAL REVIEW A  VOLUME 9, NUMBER 5  MAY 1974

Helium at zero temperature with hard-sphere and other forces

M. H. Kalos
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

D. Levesque and L. Verlet
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France
(Received 22 August 1973)

Various theoretical and numerical problems relating to helium-like systems in their ground states are treated. New developments in the numerical solution of the Schrödinger equation permit the solution of 256-body systems with hard-sphere forces. Using periodic boundary conditions, fluid and crystal states can be described; results for the energy and radial-distribution functions are given. A new method of correcting for low-lying phonon excitations
Diffusion Monte Carlo (DMC)

- Expand into exact eigenstates of $H$.
- Then the evolution is simple in this basis.
- Long time limit is lowest energy state that overlaps with the initial state, usually the ground state.
- How to carry out on the computer?

\[
\psi(R, t) = e^{-(H-E_T)t} \psi(R, 0)
\]

\[
H \phi_\alpha = E_\alpha \phi_\alpha
\]

\[
\psi(R, 0) = \sum_\alpha \phi_\alpha (R) \langle \phi_\alpha | \psi(0) \rangle
\]

\[
\psi(R, t) = \sum_\alpha \phi_\alpha (R) e^{-t(E_\alpha-E_T)} \langle \phi_\alpha | \psi(0) \rangle
\]

\[
\lim_{t \to \infty} \psi(R, t) = \phi_0 (R) e^{-t(E_0-E_T)} \langle \phi_0 | \psi(0) \rangle
\]

$E_0 \approx E_T \Rightarrow \text{normalization fixed}$
Monte Carlo process

- Now consider the variable “t” as a continuous time (it is really imaginary time).
- Take derivative with respect to time to get evolution.
- This is a diffusion + branching process.
- Justify in terms of Trotter’s formula.

\[ -\frac{\partial \psi(R,t)}{\partial t} = (H - E_T)\psi(R,t) \]

\[ H = -\sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 + V(R) \]

\[ -\frac{\partial \psi(R,t)}{\partial t} = \sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 \psi(R,t) \]

\[ -\frac{\partial \psi(R,t)}{\partial t} = (V(R) - E_T)\psi(R,t) \]

Requires interpretation of the wavefunction as a probability density.

But is it? Only in the boson ground state. Otherwise there are nodes.
Basic DMC algorithm

- Construct an ensemble (population $P(0)$) sampled from the trial wavefunction. \{R_1, R_2, \ldots, R_P\}
- Go through ensemble and diffuse each one (timestep $\tau$)
  \[ R'_k = R_k + \sqrt{2\lambda\tau}\zeta(t) \]
- number of copies = $\frac{e^{-\tau(V(R)-E_T)}}{u}$
- Trial energy $E_T$ adjusted to keep population fixed.

\[
E_0 = \lim_{t \to \infty} \frac{\int dRH\phi(R,t)}{\int dR\phi(R,t)} \approx \langle V(R)\rangle_{\phi(\infty)}
\]

- Problems:
  - Branching is uncontrolled
  - Population unstable
  - What do we do about fermi statistics?
Importance Sampling

Kalos 1970, Ceperley 1979

• Why should we sample the wavefunction? The physically correct pdf is \(|\psi|^2\).

• Importance sample (multiply) by trial wave function.

\[ f(R,t) \equiv \psi_T(R)\phi(R,t) \quad \lim_{t \to \infty} f(R,t) \equiv \psi_T(R)\phi_0(R) \]

\[ -\frac{\partial f(R,t)}{\partial t} = \psi_T(R)H \left[ \frac{f(R,t)}{\psi_T(R)} \right] \quad \text{Commut e } \psi \text{ through } H \]

\[ -\frac{\partial f(R,t)}{\partial t} = -\lambda \nabla^2 f - \lambda \nabla (2f\nabla \ln \psi_T(R)) + \left( \psi^{-1}_T H \psi_T \right) f(R,t) \]

Evolution = diffusion + drift + branching

We have three terms in the evolution equation. Trotter’s formula still applies.
Fermions in DMC

- How can we do fermion simulations? The initial condition can be made real but not positive (for more than 1 electron in the same spin state)

- In *transient estimate* or *released-node* methods one carries along the sign as a weight and samples the modulus.

\[
\phi(t) = e^{-(\hat{H}-E_t)t} \text{sign}(\phi(R,0)) |\phi(R,0)|
\]

- Do not forbid crossing of the nodes, but carry along sign when walks cross.
- What’s wrong with node release:
  - Because walks don’t die at the nodes, the computational effort increases (bosonic noise)
  - The signal/noise ratio goes to zero exponentially in \( t \).
General statement of the “fermion problem”

- Given a system with $N$ fermions and a known Hamiltonian and a property $O$. (usually the energy).
- How much time $T$ will it take to estimate $O$ to an accuracy $\varepsilon$? How does $T$ scale with $N$ and $\varepsilon$?
- If you can map the quantum system onto an equivalent problem in classical statistical mechanics then:

$$T \propto N^\alpha \varepsilon^{-2} \quad \text{With } 0 < \alpha < 4$$

This would be a “solved” quantum problem!

- All approximations must be controlled!
- Algebraic scaling in $N$!
  
  e.g. properties of Boltzmann or Bose systems in equilibrium.
Fixed-node DMC

• Initial distribution from VMC. \( f(R,0) = |\psi_T(R)|^2 \)

• Impose the fixed-node BC. \( \phi(R) = 0 \) when \( \psi_T(R) = 0 \).

• Will give the best upper bound to the exact energy

\[
E_{FN} \geq E_0 \\
E_{FN} = E_0 \quad \text{if} \quad \phi_0(R)\psi(R) \geq 0 \quad \text{all } R
\]

• \( f(R,t) \) has a discontinuous gradient at the nodal location.
• Accurate method because Bose correlations are done exactly.
• Scales well, like the VMC method, as \( N^3 \). Classical complexity.
• Can be generalized from the continuum to lattice, finite temperature, magnetic fields, ...
• One needs trial functions with accurate nodes.
Dependence of energy on wavefunction

3d Electron fluid at a density $r_s=10$

*Kwon, Ceperley, Martin, Phys. Rev. B58, 6800, 1998*

- Wavefunctions
  - Slater-Jastrow (SJ)
  - three-body (3)
  - backflow (BF)
  - fixed-node (FN)
- Energy $\langle \psi | H | \psi \rangle$ converges to ground state
- Variance $\langle \psi [H - E]^2 \psi \rangle$ to zero.
- Using 3B-BF gains a factor of 4.
- Using DMC gains a factor of 4.
Summary of $T=0$ methods:

Variational (VMC), Fixed-node (FN), Released-node (RN)
• Methods of classical simulation carry over to quantum simulation.
• BUT they are only **exact** for thermodynamic properties of some quantum systems because of the sign problem.
• Otherwise they are variational
• See *Interacting Electrons (Cambridge)* Chapters 22-25 for details and examples: