An Introduction to Monte Carlo Methods in Statistical Physics

D. P. Landau
Center for Simulational Physics
The University of Georgia

• Introduction and perspective
• Monte Carlo Simulations:
  “Simple” methods
  Accelerated algorithms
  Wang-Landau sampling
• Summary and overview
Introduction

Physics was first known as "natural philosophy"
⇒ research was purely theoretical

Later, experiment became accepted but was limited by problems with sample preparation and apparatus

- Modern computers allow simulations of models that are intractable using "classical" theory ⇒ can now invent new models and solve them without great simplification

- Computer simulations have become another way of doing physics research

- Sometimes simulations provide a theoretical basis for understanding experimental results, and sometimes they provide "experimental" data to compare with theory
Goal of these lectures ⇒

*to provide a basic understanding of the methods and philosophy of computer simulations*

- Emphasis on problems in statistical thermodynamics (condensed matter physics/ materials science)

- Choose the simulation algorithm and conditions with the problem to be solved in mind. The *interpretation* of the result is critical to the success of any simulational project!
What is a Monte Carlo Simulation?

... follow the "time dependence" of a model for which change is not deterministic (e.g. given by Newton's laws) but stochastic. Do this with a sequence of random numbers. Another sequence of random numbers will give results that agree to within some "statistical error".

Approach is valid for many problems, e.g:

- magnets
- percolation
- diffusion limited aggregation (DLA)

↑

discovered on the computer!
What problems can we solve with Monte Carlo?

- *Binary (AB) metallic alloys* ↔ ordering, interdiffusion or unmixing kinetics. Jump rates depend on the local environment. (Phonon “heat bath” → excitation energy for jumps ... good approximation because the characteristic times between jumps >> than atomic vibration periods.)

- *DLA growth of colloidal particles* . . . their masses >> atomic masses ⇒ motion of colloidal particles in fluids is described by Brownian motion.

- *“Micelle formation” in lattice models of microemulsions* (water-oil-surfactant fluid mixtures). Part of the “art” of simulation is the appropriate choice (or invention!) of a suitable (coarse-grained) model.

- *Interacting quantized particles* ... transform into a pseudo-classical model or consider permutation properties.

- *Polymer models*, e.g. flexible polymer as a random walk.
What difficulties will we encounter?

- Limited computer time and memory

**Statistical errors** come from finite sampling. Estimate them, then make a "policy" decision, i.e. simulate longer to reduce statistical errors or use the CPU time to study the properties of the system under other conditions.

- Developing new strategies is itself exciting!

- Computers have finite word length → limited precision.

- Systematic errors. An algorithm may not describe the physics properly, e.g. due to finite particle number, etc.

- Truncation & round-off may produce problems.
What strategy should we follow?

- **New simulations face hidden pitfalls.** Begin with a simple program, small systems, and short runs. Test the program for special values of parameters for which the answers are known. Uncover parameter ranges of interest and find unexpected difficulties. *Then*, refine the program and increase running times.

- **Use BOTH cpu time and human time effectively.** Don’t spend a month to rewrite a computer program that saves only a few minutes of cpu time.
How do Simulations Relate to Theory and Experiment?

• *Theory* might be available for a model with no good physical realization ⇔ compare with "data" from a simulation. Dramatic example: reactor meltdown . . . we want understanding but do not want to carry out experiments!

• Physical systems exist that are too complex for theory, *e.g.* in polymer blends there is a subtle interplay between competing interactions configurational entropy of flexible macromolecules, etc). If the simulation (playing the role of theory) disagrees with experiment, a new model is needed.

An important advantage of simulations ⇒ different physical effects which are simultaneously present in real systems may be isolated and through separate consideration may provide better understanding.
Ergodicity and Broken Symmetry

**Ergodicity** ⇒ all configurations are attainable. Below $T_c$ multiple ordered states may be well separated in phase space.

**Symmetry breaking** usually means *ergodicity breaking, i.e.*, the system stays in one separate region in phase space.

In an Ising model simulation, we may wish to keep the system from exploring all of phase space. If the simulation algorithm is fully ergodic, both $M > 0$ and $M < 0$ states will appear and $\langle M \rangle = 0$. A danger for simulations is that specialized algorithms may be *unintentionally* non-ergodic, thus yielding incorrect results.
Probability Distributions and the Central Limit Theorem

\[ P_G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{(x - \langle x \rangle)^2}{2\sigma^2} \right] \]

Gaussian distribution

... approximation to the binomial distribution with a large number of possible outcomes and a large number of samples.

If random variables \( x_i \) are independent and drawn from the same distribution, \( \bar{X}_N = \frac{\sum_{i=1}^{N} x_i}{N} \) as \( N \to \infty \) is always distributed as a Gaussian \( \Leftrightarrow \text{Central Limit Theorem} \)
Consider a quantity $A$ distributed according to a gaussian with mean value $<A>$ and width $\sigma$. For $n$ independent observations

$$\bar{A} = \frac{1}{n} \sum_{i=1}^{n} A_i = \text{unbiased estimator of the mean } <A>$$

and the standard error of this estimate = $\frac{\sigma}{\sqrt{n}}$

Estimate the variance $\sigma$ from the observations,

$$\delta A^2 = \frac{1}{n} \sum_{i=1}^{n} (\delta A_i)^2 = \bar{A}^2 - (\bar{A})^2$$
Relate this to $\sigma^2 = \langle A^2 \rangle - \langle A \rangle^2$

$$\left\langle \delta A^2 \right\rangle = \sigma^2 \left(1 - 1/n \right)$$

$\Rightarrow$ compute errors of averages from uncorrelated estimates:

$$\text{error} = \sqrt{\left\langle \delta A^2 \right\rangle / (n-1)} = \sqrt{\sum_{i=1}^{n} (\delta A_i)^2 / [n(n-1)]}$$
Markov Chains and Master Equations

...central to Monte Carlo simulations. Define a stochastic process at times $t_1, t_2, \ldots$ for a system with possible states $S_1, S_2, \ldots$

\[
P(X_{t_n} = S_{i_n} \mid X_{t_{n-1}} = S_{i_{n-1}}, X_{t_{n-2}} = S_{i_{n-2}}, \ldots, X_{t_1} = S_{i_1}),
\]

state of the system at time $t$

conditional probability

This is a **Markov Process** if the conditional probability depends **only** upon the immediate predecessor.
The sequence of states \( \{X_t\} \) is a Markov chain.

Transition probability to move from state \( i \) to state \( j \),

\[
P(X_t = S_j | X_{t-1} = S_i) = W_{ij} P(X_{t-1} = S_i).
\]

Importance sampling Monte Carlo \( \Leftrightarrow \) a Markov process.

The principle of detailed balance must be satisfied:

Equilibrium probability

\[
W_{ji} P_{eq}(S_j) = W_{ij} P_{eq}(S_i), \quad \text{i.e.,} \quad \frac{d P_{eq}(S_j, t)}{dt} \equiv 0
\]
The “Art” of Random Number Generation

Monte Carlo methods need fast, efficient production of random numbers. Physical processes, e.g. electrical circuit white noise, pre-calculated tables, etc. are too slow!

Software algorithms are actually deterministic, producing "pseudo-random" numbers...this may be helpful, e.g. to test a program, want to compare the results with a previous run made using identical random numbers.

- Poor quality random numbers → systematic errors!!

In general, the random number sequences should be uniform, uncorrelated, and have a long period.
Congruential Method

A fixed multiplier \( c \) and seed \( X_0 \) are chosen. Then,

\[
X_n = (c \times X_{n-1} + a_o) \text{MOD } N_{\text{max}}
\]

where \( X_n \) is an integer between 1 and \( N_{\text{max}} \). \( c \) must have “good” properties. Performance is best for \( X_0 \) odd. A “good” generator is the 32-bit linear congruential algorithm (CONG)

\[
X_n = (16807 \times X_{n-1}) \text{MOD } (2^{31} - 1)
\]

(Another, popular congruential generator had noticeable correlation between consecutive triplets of random numbers.)
Shift Register (Tausworthe) Algorithm

A new random number is produced by combining two different existing numbers from a pre-existing table:

\[ X_n = X_{n-p} \oplus X_{n-q} \]

where \( p \) and \( q \) must be properly chosen. The \( \oplus \) operator is the bitwise exclusive-OR operator. The best choices of the pairs \((p,q)\) are given by primitive Mersenne trinomials:

\[ x^p + x^q + 1 \]
Examples of pairs which satisfy this condition are:

- \( p = 250 \quad q = 103 \quad \leftarrow \text{popular, known as R250} \)
- \( p = 1279 \quad q = 216, 418 \)
- \( p = 9689 \quad q = 84, 471, 1836, 2444, 4187 \)

Large \( p \) and \( q \) yield better sequences. The “table” must be properly initialized. Use a good congruential generator to generate a different random number for each bit in turn.

**Lagged Fibonacci Generators**

Replace the exclusive-or by another operator, e.g.,

\[ X_n = X_{n-p} \ast X_{n-q} \]
Tests for Quality (some examples)

**Uniformity test:** Break up the internal $0 < x < 1$ into many small bins... check for uniformity in the population of the bins.

**Overlapping M-tuple test:** Check the statistical properties of the occurrence of M-tuples of digits.

**Parking lot test:** Plot points in an $m$-dimensional space using coordinates of each point determined by $m$-successive random numbers. Look for regular structures.

**These mathematical tests are not sufficient!** A “practical” test... a Monte Carlo study of a small Ising model (exactly solvable) found “local” and “non-local” sampling yielded different systematic errors with different generators.
Example:
Plot points using consecutive pairs of random numbers for $x$ and $y$. 

“bad” generator

“good” generator
Non-uniform Distributions of Random Numbers

Look at the integrated distribution function $F(x)$ of the desired distribution $f(x)$, generate a uniform distribution of random numbers $y_i$ and take the inverse function with the uniformly chosen random number as the variable, i.e.

$$y = F(y) = \int_{0}^{y} f(x) \, dx \implies x = F^{-1}(y)$$

**Examples**

Generate random numbers distributed according to $f(x) = x$.

$$y = F(x) = \int_{0}^{x} x' \, dx' = 0.5x^2.$$  

If a random number $y$ is chosen from a uniform distribution, then $x = 2.0y^{1/2}$. 
Non-uniform Distributions of Random Numbers

Gaussian distribution (Box-Muller method):

Draw $x_1$ and $x_2$ from a uniform distribution, and the desired random numbers are

$$y_1 = (-2 \ln x_1)^{1/2} \cos(2\pi x_2)$$
$$y_2 = (-2 \ln x_1)^{1/2} \sin(2\pi x_2)$$
Using Monte Carlo for Numerical Integration

• Our goal is to estimate:

\[ y = \int_{a}^{b} f(x) \, dx \]
Using Monte Carlo for Numerical Integration

• Our goal is to estimate:

$$y = \int_{a}^{b} f(x) \, dx$$

**Acceptance - rejection ("hit and miss") method**

Draw a box from $a$ to $b$ and from 0 to $y_o$ where $y_o > f(x)$
Our goal is to estimate:

\[ y = \int_{a}^{b} f(x) \, dx \]

**Acceptance - rejection ("hit and miss") method**

Draw a box from \(a\) to \(b\) and from 0 to \(y_o\) where \(y_o > f(x)\)

Drop \(N\) points randomly into the box; \(N_o\) fall below \(f(x)\).
Using Monte Carlo for Numerical Integration

- Our goal is to estimate:

\[ y = \int_{a}^{b} f(x) \, dx \]

**Acceptance - rejection ("hit and miss") method**

Draw a box from \( a \) to \( b \) and from 0 to \( y_o \) where \( y_o > f(x) \)

Drop \( N \) points randomly into the box; \( N_o \) fall below \( f(x) \).

An estimate for the integral is given by the fraction of points which fall below the curve times the area of the box, \( i.e. \)

\[ y_{est} = \left( \frac{N_o}{N} \right) \times [y_o (b-a)] \]
Example:  *Estimate “π” using simple sampling Monte Carlo*

- Choose \( N \) points randomly in the xy-plane so that \( 0 < x < 1, 0 < y < 1 \).
- Count points \( N_O \) that are less than a distance = 1 from the origin.
- \( N_O/N \) ⇒ estimate for the area of one-quarter of a circle = \( 4N_O/N \).
- Repeat multiple times ⇒ use the variance to estimate the error.

**Sample results:** After 700 points estimates appear to have converged to the **wrong answer**. The apparent difficulty is due to the use of too few points. *This lesson should not be forgotten!*

<table>
<thead>
<tr>
<th>( N )</th>
<th>Result</th>
<th>( 100 )</th>
<th>3.1600</th>
<th>( 200 )</th>
<th>3.0400</th>
<th>( 300 )</th>
<th>3.1067</th>
<th>( 400 )</th>
<th>3.0800</th>
<th>( 500 )</th>
<th>3.0560</th>
<th>( 600 )</th>
<th>3.0800</th>
<th>( 700 )</th>
<th>3.0743</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>3.0800</td>
<td>200</td>
<td>3.0400</td>
<td>300</td>
<td>3.1067</td>
<td>400</td>
<td>3.0800</td>
<td>500</td>
<td>3.0560</td>
<td>600</td>
<td>3.0800</td>
<td>700</td>
<td>3.0743</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simple "crude method"

- Choose $N$ values of $x$ randomly, and then

\[ y_{est} = \frac{1}{N} \sum_{i} f(x_i) \]

- As $N \to \infty$, the estimate converges to the correct result.

- Simple variation of this method... divide the interval into unequal sub-intervals and perform a separate Monte Carlo integration for each sub-interval.

- Less effort can be expended on those sub-intervals over which the function is small.
Intelligent methods for Monte Carlo Integration

**Control variate method**

Select a known, integrable function and only integrate the difference \[ f'(x) - f(x) \], i.e.

\[
y_{est} = F' + \int_{a}^{b} [f'(x) - f(x)] \, dx
\]

where \( F' = \int_{a}^{b} f'(x) \, dx \)

( Note: can be improved without additional numerical effort by an intelligent choice of \( f'(x) \) ! )
Intelligent methods for Monte Carlo Integration

**Importance sampling**

Choose points by the expected importance $p(x)$ of the value of the function at that point to the integral. Weight the contribution by the inverse of the probability of choice:

$$y_{est} = \sum_{i} p^{-1}(x_i)f(x_i)$$

Values of $x$ are no longer chosen with equal probability $\Rightarrow$ need random numbers that are *not* drawn from a uniform sequence. Some expertise is often needed in choosing $p(x) \rightarrow$ Iterate...

carry out a rough study, get an improved $p(x)$, and repeat.

**Intelligent importance sampling is far more effective in improving convergence than the brute force method of simply generating many more points!**

**Homework:** Suppose $f(x) = x^{10}-1$. Use a "hit or miss" Monte Carlo simulation to estimate the integral between $x=1$ and $x=2$. 
Monte Carlo Solution of Boundary Value Problems

Often have a differential equation with a boundary condition

\[ \nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{Laplace's equation} \]

where \( u(r) = f \) on the boundary. Re-express as a finite difference equation using a small increment \( \Delta \),

\[
\nabla^2 u = \frac{u(x + \Delta, y) + u(x - \Delta, y) + u(x, y + \Delta) + u(x, y - \Delta) - 4u(x, y)}{\Delta^2} = 0
\]

\[ \Rightarrow u(x, y) = \frac{u(x + \Delta, y) + u(x - \Delta, y) + u(x, y + \Delta) + u(x, y - \Delta)}{4} \]
Probabilistic interpretation: Consider a grid of points in the $x$-$y$ plane with a lattice spacing $\Delta$. The probability of a random walk returning to point $(x, y)$ from a nearest neighbor site is $1/4$. Place the boundary on the grid, a random walk will end at a boundary point $(x', y')$ where $u(x', y') = f(x', y')$.

Execute many walks beginning at point $(x, y)$, then

$$u(x, y) \approx \frac{1}{N} \sum_i f(x'_i, y'_i), \quad \text{let } N \to \infty.$$
Consider two concentric, circular conductors in a plane which are placed into the center of a square box which is 20cm on a side. The inner conductor has a radius of 4cm and carries a potential of 2V.; the outer conductor has a radius of 16cm and has a potential of 4V. What is the potential halfway between the two conductors? Consider a square box with an $L \times L$ grid.

\[
\begin{array}{cccccc}
N & L=10 & L=20 & L=40 & L=80 & L=160 & L=320 \\
500 & 3.6560 & 3.3000 & 3.2880 & 3.3240 & 3.3400 & 3.3760 \\
\end{array}
\]

**exact value** = 3.3219 V.
Phase transitions and critical phenomena are of great interest! How can we learn about them?

The *Partition function* contains all thermodynamic information:

*Lattice models* are simple and amenable to high resolution study via computer simulation (*magnets, lattice gauge models, polymers, etc.*)

How do we get to long times, *i.e.* sample much of phase space?
Monte Carlo and Stat. Mech.- An Introduction

Phase transitions and critical phenomena are of great interest! How can we learn about them?

The Partition function contains all thermodynamic information:

\[ Z = \sum_{\text{all states}} e^{-\frac{H}{k_B}T} \]

Lattice models are simple and amenable to high resolution study via computer simulation (magnets, lattice gauge models, polymers, etc.)

How do we get to long times, i.e. sample much of phase space?
The Percolation problem

. . . a geometric problem in which random addition of objects can create a contiguous path which spans the entire system.

Site Percolation

- Lattice sites are randomly occupied with probability $p$.
- Connect occupied nearest neighbor sites to form clusters.
- Generate many (independent) realizations $P_{\text{span}} = \text{probability of having a spanning (infinite) cluster}$.
- The order parameter $M$ is the fraction of occupied sites in the lattice which belong to the infinite cluster.
- $p = p_c$ (percolation threshold); $(p-p_c) \Leftrightarrow (T_c-T)$ for a thermal transition.

Simplest “infinite” cluster  random infinite cluster
Some Brief Background

The *Partition function* contains all information about a system

\[ Z = \sum_{\text{all states}} e^{-\frac{H}{k_B} T} \]

**Example**

*N* non-interacting Ising spin in an external magnetic field *H*

\[ H = -H \sum_{i} \sigma_i, \quad \sigma_i = \pm 1 \]

then \[ Z = \left( e^{-\frac{H}{k_B} T} + e^{\frac{H}{k_B} T} \right)^N \]

Usually *Z* cannot be evaluated exactly. For an *N*=10,000 Ising model *Z* has \(2^{10,000}\) terms!
Probability that the system is in state $\mu$:

$$P_\mu = e^{-\frac{H(\mu)}{k_BT}} / Z$$

**Partition function ⇔ thermodynamic quantities**

$$F = -k_BT \ln Z \quad \text{Free energy}$$

$$U = -T^2 \frac{\partial (F/T)}{\partial T}$$

Can also estimate Free Energy differences by integration, e.g.,

$$\Delta(F/T) = \int d(1/T)d(U/T).$$
Fluctuations

Mean values, *e.g.*:

\[
\overline{U}(\beta) = \sum_{\mu} P_\mu H(\mu) = \sum_{\mu} H(\mu) e^{-\beta H(\mu)} / \sum_{\mu} e^{-\beta H(\mu)}
\]

Fluctuation relation:

\[
k_B T^2 C_v = \left\langle H^2 \right\rangle - \left\langle H \right\rangle^2 = \left\langle (\Delta U)^2 \right\rangle_{NVT}, \quad \Delta U \equiv H - \left\langle H \right\rangle
\]

In experiment \(N \approx 10^{22}\) and fluctuations are too small to be detectable. *In simulations they are readily observable!*

Other fluctuation relations exist, *e.g.* isothermal susceptibility

\[
\chi = \left( \frac{\partial M}{\partial H} \right)_T = N \left[ \left\langle M^2 \right\rangle - \left\langle M \right\rangle^2 \right] / k_B T
\]
Single Spin-Flip Monte Carlo Methods

Typical spin configurations for the Ising square lattice

$T \ll T_c$  $T \sim T_c$  $T \gg T_c$
Simple Sampling Monte Carlo

The *Partition function* contains all information about a system

\[ Z = \sum_{\text{all states}} e^{-\frac{H}{k_B} T} \approx \sum_{M \text{ states}} e^{-\frac{H}{k_B} T} \]

**Example:** \(N\) Ising spins on a square lattice

\[ H = -J \sum_{i,j} \sigma_i \sigma_j, \quad \sigma_i = \pm 1 \]

At low temperature, only two states contribute very much (i.e. all spins up or all spins down). *Simple sampling is very inefficient since it is very unlikely to generate these two states!* Use importance sampling instead.

Remember, for an \(N=10,000\) Ising model \(Z\) has \(2^{10,000}\) terms!
Single spin-flip sampling for the Ising model

Produce the $n^{th}$ state from the $m^{th}$ state ... relative probability is $P_n / P_m \rightarrow$ need only the energy difference, i.e. $\Delta E = (E_n - E_m)$ between the states.

Any transition rate that satisfies detailed balance is acceptable, usually the Metropolis form (Metropolis et al, 1953).

$$W(m \rightarrow n) = \tau_o^{-1} \exp(-\Delta E/k_B T), \quad \Delta E > 0$$

$$= \tau_o^{-1}, \quad \Delta E < 0$$

where $\tau_o$ is the time required to attempt a spin-flip.
Metropolis Recipe:

(1.) Choose an initial state
(2.) Choose a site $i$
(3.) Calculate the energy change $\Delta E$ which results if the spin at site $i$ is overturned
(4.) Generate a random number $r$ such that $0 < r < 1$
(5.) If $r < \exp(-\Delta E/k_B T)$, flip the spin
(6.) Go to the next site and go to (2)

Critical Slowing Down!
Some Practical Advice

1. *In the very beginning, think!*  
   What problem do you really want to solve? What method and strategy are best suited to the study? (A little thought may reduce the number of false starts.)

2. *In the beginning think small!*  
   Use small lattices and short runs... gives fast “turnaround” and aids debugging. Can rapidly search through parameter space to find ranges with physically interesting behavior.

3. *Test the random number generator!*  
   Find limiting cases where accurate, or exact values of some properties can be compared with your results with different random number sequences and/or generators.
Some Practical Advice (continued)

4. Look at systematic variations with system size and run length!

5. Calculate error bars!

- Use a wide range of sizes and runs lengths and then use scaling forms to analyze data.

6. Make a few very long runs!

- This enables the evaluation of the correctness of conclusions which are drawn from the data.
Montmorency Royce Sebastian Carlow (1878-1927). The son of a Buckinghamshire peat-digger, he was born in the sedate village of Gossoon two years after Alexander Graham Bell's invention of the telephone, and died the year Lindberg flew the Atlantic.

Montmorency Carlow's contribution to science is no less striking. He displayed no noticeable mathematical talent - indeed no talent at all worth speaking of - until the age of forty-eight, when he succeeded in destroying the entire village of Gossoon, and much of the surrounding countryside, in a singlehanded air-raid mounted from a Handley Page 0/0400 Night Bomber. At his subsequent trial he offered the defence that he was attempting to estimate the area of Gossoon's village pond by calculating the proportion of hits from a random bombing pattern; adding that since the total area of Gossoon was 946.32 acres and he had hit the pond exactly once using 143 bombs, the area of the pond was approximately 6.6176224 and a bit acres.

Lord Justice Milnesshawe-Ffreebes, failing to appreciate the revolutionary nature of the method, commented that since the bomb crater had obliterated all trace of the pond, the calculation left something to be desired. Before the conclusion of the trial, Carlow attempted to estimate the probability of surviving a fall by repeatedly jumping from a high window, fell on his head at the first attempt, and broke his neck. His name lingers on, however: the techniques of estimation that he pioneered are known throughout the world as Monty Carlow Methods.
Types of Computer Simulations

Deterministic methods . . . (Molecular dynamics)

Stochastic methods . . . (Monte Carlo)
Boundary Conditions

Different boundary conditions ⇒ different physics
Different boundary conditions $\Rightarrow$ different physics

Extrapolate size to $\infty$
$\rightarrow$ bulk behavior
Different boundary conditions $\Rightarrow$ different physics

Introduce an interface parallel to the top “surface”
$\rightarrow$ study interfacial phenomena
Boundary Conditions

Different boundary conditions $\Rightarrow$ different physics

Study nanoparticles
Boundary Conditions

Different boundary conditions $\Rightarrow$ different physics

AND, free edges on top, pbc on the sides
$\Rightarrow$ Study surface critical behavior
Typical spin configurations for the Ising square lattice with pbc

\[ T << T_c \]
\[ T \sim T_c \]
\[ T >> T_c \]
Correlation times

Define an equilibrium relaxation function $\phi(t)$

$$\varphi_{MM}(t) = \frac{<(M(0)M(t))> - <M>^2}{<M^2> - <M>^2}$$

$$t \to \infty \implies e^{-t/\tau}$$

and

$$\tau \propto |T - T_c|^{-\nu Z}$$

i.e. $\tau$ diverges at $T_c$!

“Critical slowing down”
Finite Sampling Time Effects

Consider the expectation of the square of the statistical error

\[
\langle (\delta A)^2 \rangle = \left\langle \left[ \frac{1}{N} \sum_{\mu=1}^{N} (A_\mu - \langle A \rangle) \right]^2 \right\rangle
\]

\[
= \frac{1}{N^2} \sum_{\mu=1}^{N} \langle (A_\mu - \langle A \rangle)^2 \rangle + \frac{2}{N^2} \sum_{\mu_1=1}^{N} \sum_{\mu_2=\mu_1+1}^{N} \left( \langle A_{\mu_1} A_{\mu_2} \rangle - \langle A \rangle^2 \right)
\]

Define

\[\tau_A \equiv \int_{0}^{\infty} \phi_A(t)dt\]

Correlation time

\[
\langle (\delta A)^2 \rangle = \frac{1}{N} \left[ \langle A^2 \rangle - \langle A \rangle^2 \right] \left( 1 + \frac{2\tau_A}{\delta t} \right)
\]

No. of measurements, made time $\delta t$ apart

True “error”
Biased sampling: a systematic error

Consider the magnetic susceptibility

\[ \chi \propto (\langle m^2 \rangle - \langle m \rangle^2) \]

No. of independent measurements

\[ E(s^2) = \sigma^2 (1 - 1/n) \]
Finite Size Effects

A scaling ansatz for the singular part of the free energy:

\[ F(L, T) = L^{(2-\alpha/\nu)} F(\varepsilon L^{1/\nu}) \]

where \( \varepsilon = (T - T_c)/T_c \).

Choose the scaling variable \( x = \varepsilon L^{1/\nu} \) because \( \xi \sim \varepsilon^\nu \sim L \) as \( T \to T_c \). (\( L \) “scales” with \( \xi \); but can use \( L/\xi \propto \varepsilon^\nu L \) or \( \varepsilon L^{1/\nu} \) as the argument of \( F \).) Differentiate \( \Rightarrow \)

\[ M = L^{-\beta/\nu} M_o(\varepsilon L^{1/\nu}) \quad \text{magnetization} \]
\[ \chi = L^{-\gamma/\nu} \chi_o(\varepsilon L^{1/\nu}) \quad \text{susceptibility} \]
\[ C = L^{-\alpha/\nu} C_o(\varepsilon L^{1/\nu}) \quad \text{specific heat} \]

\( M_o(x), \chi_o(x), \) and \( C_o(x) \) are scaling functions. Corrections to scaling and finite size scaling appear for small \( L \) and large \( \varepsilon \).
A case study: $L \times L$ Ising square lattice with p.b.c. (from 1976)

The large scatter in the data is characteristic of early Monte Carlo work--the effort is within the reach of a PC today!
Dynamic Finite Size Scaling

\[ \tau = L^z F(\varepsilon L^{1/\nu}) \]

so at \( T_c \),

\[ \tau \propto L^z \quad \text{Critical Slowing Down} \]

This is valid only as long as \( L \) is sufficiently large that corrections to finite size scaling do not become important.

For the Metropolis method, \( z \sim 2.1 \)
Consider the magnetic susceptibility:

\[ \chi \propto (\langle m^2 \rangle - \langle m \rangle^2) \]
Cluster Flipping Methods

*Fortuin-Kasteleyn theorem.* . . . map a ferromagnetic Potts model onto a corresponding percolation model for which successive states are uncorrelated ⇒ **no critical slowing down!**

For the *q*-state Potts model

\[ Z = \sum_{\{\text{all states}\}} e^{-K \sum_{i,j} (\delta_{\sigma_i \sigma_j} - 1)} \]

where \( K = J / k_B T. \)

- Replace each pair of interacting spins by a bond with probability

\[ p = 1 - e^{-K \delta_{\sigma_i \sigma_j}} \]

- Repeat for all pairs → lattice with bonds that connect some sites to form clusters with diverse sizes and shapes. *Note: all spins in each cluster have the same value.*
The spins may then be integrated out (leaving a factor of $q$ behind for each cluster), and for the $N_c$ remaining clusters

$$Z = \sum_{\text{bonds}} p^b (1 - p)^{(N_b - b)} q^{N_c}$$

where $b = \# \text{ of bonds}$, $N_b = \text{total} \# \text{ of possible bonds}$, $(1-p)$ is the probability that no bond exists between a pair of sites.
Swendsen-Wang method

- Choose an initial spin state.
- Place bonds between each pair of spins with probability $p$.
- Find all clusters i.e. connected networks of bonds.
- Randomly assign a spin value to all sites in each cluster.
- Erase bonds $\rightarrow$ new spin state.

original spins  clusters formed  “decorated” clusters
At high $T$, clusters are small. At low $T$, almost all sites with nearest neighbors in the same state are in the same cluster and the system oscillates back and forth between similar structures. Near $T_c$, a rich array of clusters is produced and each configuration differs substantially from its predecessor

⇒ *critical slowing down is reduced!*

$z \sim 2.1$ for Metropolis
$\sim 0$ in 2-dim and $\sim 0.5$ in 3-dim for SW *(Wang, 1990)*

Code is more complex than for single spin-flip methods.
⇒ For small lattices the SW technique may be slower, but for sufficiently large lattices it will be more efficient.
**Wolff method**

The Swendsen-Wang approach spends great effort on small clusters that do not contribute to the critical slowing down.

*Instead, grow single clusters and flip them sequentially:*

- Randomly choose a site.
- Draw bonds to all nearest neighbors that are in the same state with probability $p = 1 - e^{-K}$. Repeat, iteratively, to form a cluster of connected sites.
- Flip the entire cluster of connected sites.
- Choose another initial site and repeat the process.

*Wolff kinetics has a smaller prefactor and smaller dynamic exponent than does the Swendsen-Wang method.*
“Improved estimators”
It may be possible to calculate a thermodynamic property using clusters . . . for some quantities “noise reduction” occurs, e.g. the susceptibility for O(N) models is given by the mean cluster size, i.e.,

\[ \chi = \beta \langle |C| \rangle \]

where \(|C|\) is the size of a cluster. The statistical error in this definition is less than that obtained from fluctuations in the order parameter (fluctuations due to small clusters cancel).
**N-fold way and extensions**

- The above methods are *time-step driven* → **low T problems**!
- In *event driven* algorithms (e.g., “N-fold Way” Bortz et al, 1975) a flip occurs at each step:

*For discrete spin models* ⇒ only a few flipping probabilities

- Collect spins into lists; spins in each have equivalent local environments (*Ising square lattice has N=10 classes*).
- Total probability of *some* class \( l \) spin flipping in a step is

\[
p_l = n_l e^{-\Delta E_l / k_B T}
\]

\( n_l \) = number of spins in class \( l \)

\[
Q_M = \sum_{l=1}^{M} p_l = \text{total for all classes with } l \leq M
\]
• Generate a random number $0 < rn < Q_N \Rightarrow$ class for the next flip, i.e., class $M$ is chosen if $Q_{M-1} < rn < Q_M$.

• Chose another $rn$ to pick a class $M$ spin.

• $3^{rd}$ random number $\Rightarrow$ time elapsed before flipping

$$\Delta t = -\frac{\tau}{Q_N} \ln R$$

Find properties from lifetime weighted averages.

$\Rightarrow$ At low $T$, the gain in performance can be huge!

**Generalize:** “absorbing Markov chains” (Novotny, 1995)
Classical spin models
e.g., the classical Heisenberg model with spin vectors \( \vec{S}_i \)

\[
H = -J \sum_{nn} \vec{S}_i \cdot \vec{S}_j, \quad |\vec{S}_i| = 1
\]

The Metropolis method involves “spin-tilts” instead of “spin-flips”

Over-relaxation method (Brown and Woch, 1987; Creutz, 1987)
Precess the spin about the effective interaction field (due to neighbors) by an angle \( \theta \) using the equation of motion

\[
\dot{\vec{S}} = -\vec{S} \times \vec{H}_{eff}
\]

This is microcanonical (and also deterministic) but it decorrelates successive states. Combine with Metropolis \( \Rightarrow \) becomes canonical.

What about cluster-flipping?
Wolff embedding trick and cluster-flipping

The “embedding trick” turns the classical spin model into an inhomogeneous Ising model (Wolff, 1989):

- Choose a direction \( \hat{n} \) randomly in space.
- Project spins onto that direction to form an Ising model with interactions that depend on the spin projections.
- Then, use cluster flipping, e.g. for Wolff flipping, bonds are added between nearest neighbors with probability

\[
p = 1 - \exp\{\min[0, 2\beta J(\hat{n} \cdot S_i)(\hat{n} \cdot S_j)]\}
\]

- Reverse the components parallel to \( \hat{n} \) for all spins in a cluster to yield a new spin configuration.
- Choose a new (random) direction and repeat.
Hybrid methods

Combine updating types ⇒ a single scheme, e.g., a study of the 2-dim classical, XY-model used a mixture of Metropolis, over-relaxation and embedding trick Wolff flips, together with finite size scaling, to determine the Kosterlitz-Thouless temperature to higher precision than previously possible (Evertz and Landau, 1996), $J/k_B T_{KT} = 0.700(5)$.  
Improvements in Performance (Ising model):

- Computer speed
- Algorithmic advances - cluster flipping, reweighting . . .
Variation on a Theme...

Probability-Changing Cluster Algorithm *(Tomita and Okabe, 2001)*

**Goal:** Tune the critical point automatically

Extend the SW algorithm by increasing or decreasing the probability of forming bonds depending on whether or not the clusters are percolating.

**Recipe:**
1. Choose an initial configuration and a value of $p$
2. Construct the Kasteleyn-Fortuin clusters using probability $p$.
   
   Check to see if the system is percolating.
3. Update the spins using the SW rule.
4. If the system was percolating, decrease $p$ by $\Delta p$. If the system was not percolating, increase $p$ by $\Delta p$
5. Go to 2.

**Note:** Start with some value of $\Delta p$ and decrease it as the simulation proceeds.
Test for the 2-dim Ising model
What is Multicanonical sampling?

$P(E)$ may contain multiple maxima that are widely spaced in configuration space (1st order phase transitions, spin glasses, etc.) ⇒ Standard methods become “trapped” near one maximum; infrequent transitions between maxima leads to ill determined relative weights of the maxima and the minima of $P(E)$.

∴ modify the single spin flip probability to enhance the probability of the “unlikely” states between the maxima ⇒ accelerates effective sampling!
Reformulate the problem $\Rightarrow$ an effective Hamiltonian

$$H_{\text{eff}}(\sigma) = H_{\text{eff}}(\beta H(\sigma))$$

The probability distribution for the energy is then:

$$P(E) = \frac{\exp(S(E) - H_{\text{eff}})}{\sum_{E} \exp(S(E) - H_{\text{eff}})}$$


$P(E)$ from one simulation is an estimate for a 2nd simulation, etc.

This can be very time consuming!
Compare ensembles:

Then,

$$\langle A \rangle_\beta = \frac{\langle A \exp(H_{\text{eff}} - H) \rangle}{\langle \exp(H_{\text{eff}} - H) \rangle}$$

Thermodynamic variable
How to get $H_{\text{eff}}$:

Find where it is easy to measure (via standard Monte Carlo) and use it as an estimate for a 2nd run made closer to the interesting region. Continue to the “unknown” region where standard sampling fails.

Example: A $q=7$ Potts model \textit{(After Janke, 1992).}
Phase Switch Monte Carlo

In some cases, e.g. the freezing transition, a crystal state forms that is defect ridden…these do not anneal out in reasonable time scales → the system is trapped

**Strategy:** Make a global coordinate transformation that maps one pure phase onto another

*Wilding and Bruce (2000)*
{R}^{\gamma} is the representative configuration for phase $\gamma$

Define particle positions:

$$r_i^{\gamma} = R_i^{\gamma} + u_i$$

**Phase switch:**

Replace $\{R\}^{\gamma}$ by $\{R\}^{\gamma'}$

Results for the order parameter of a 256 particle system
A New Approach to Monte Carlo Simulations in Statistical Physics

Fugao Wang and DPL

• Introduction
• Method
• Applications (how well does it work?)
  - 1st Order transition – 2d Potts Models
  - 2nd order transitions – 2d Ising model
  - Spin glass transition – 3d EA model
  - Critical endpoint - triangular Ising model
• Conclusions
A New Approach to Monte Carlo Simulations in Statistical Physics: The “Random Walk in Energy Space with a Flat Histogram” method

or

“Wang-Landau sampling”
Problems and Challenges

**Statics:** Monte Carlo methods are valuable, but near $T_c$

$\Rightarrow$ **critical slowing down** for 2\textsuperscript{nd} order transitions

$\Rightarrow$ **metastability** for 1\textsuperscript{st} order transitions

\therefore Try to reduce characteristic time scales or circumvent them

\textit{“Dynamics”}: stochastic vs deterministic
Metropolis Recipe:

1. Choose an initial state
2. Choose a site \( i \)
3. Calculate the energy change \( \Delta E \) that results if the spin at site \( i \) is overturned
4. Generate a random number \( r \) such that \( 0 < r < 1 \)
5. If \( r < \exp(-\Delta E/k_B T) \), flip the spin
6. Go to 2.

This is not a unique solution. An alternative \((Glauber, 1963)\):

\[
W_{n \rightarrow m} = \tau_0^{-1} [1 + \sigma_i \tanh (E_i /k_B T)],
\]

where \( \sigma_i E_i \) is the energy of the \( i^{th} \) spin in state \( n \). Both Glauber and Metropolis algorithms are special cases of a general transition rate \((Müller-Krumbhaar and Binder, 1973)\)
Correlation times (reminder)

Define an equilibrium relaxation function \( \phi(t) \)

\[
\phi_{MM}(t) = \frac{\langle M(0)M(t) \rangle - \langle M \rangle^2}{\langle M^2 \rangle - \langle M \rangle^2}
\]

\( t \to \infty \Rightarrow e^{-t/\tau} \)

and

\( \tau \propto |T - T_c|^{-\nu} \) i.e. \( \tau \) diverges at \( T_c \)!
A Quite Different Approach

Random Walk in Energy Space with a Flat Histogram

Reminder:

$$Z = \sum_{\text{all states}} e^{-\frac{H}{k_B}T} = \sum_{\text{all energies}} g(E) e^{-\frac{H}{k_B}T}$$

Estimate the density of states $g(E)$ directly — how?

Solution: Perform a random walk in energy space:

1. Set $g(E)=1$; choose a modification factor (e.g. $f_0 = e^1$)

2. Randomly flip a spin with probability:

$$p(E_1 \rightarrow E_2) = \min\left(\frac{g(E_1)}{g(E_2)}, 1\right)$$

3. Set $g(E_i) \rightarrow g(E_i) \cdot f$

4. Continue until the histogram is “flat”; decrease $f$, e.g. $f_{i+1} = f^{1/2}$

5. Repeat steps 2 - 4 until $f = f_{\text{min}} \sim \exp(10^{-8})$

6. Calculate properties using final density of states $g(E)$
How can we test the method?

For a 2<sup>nd</sup> order transition, study the 2-dim Ising model:

\[
H = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j , \quad \sigma_i = \pm 1
\]

- \( T_c \) is known for an infinite system \((\text{Onsager})\)
- Bulk properties are known
- \( g(E) \) is known exactly for small systems

For a 1<sup>st</sup> order transition, study the 2-dim Potts model:

\[
H = -J \sum_{\langle i, j \rangle} \delta_{\sigma_i \sigma_j} , \quad \sigma_i = 1, \ldots, q \quad \text{for } q=10
\]

- \( T_c \) is known for an infinite system \((\text{duality})\)
- Good numerical values exist for many quantities
Density of States for the 2-dim Ising model

Compare exact results with data from random walks in energy space: $L \times L$ lattices with periodic boundaries
Density of States for the 2-dim Ising model

Compare exact results with data from random walks in energy space: \( L \times L \) lattices with periodic boundaries

\( \varepsilon = \text{relative error} \) (exact solution is known for \( L \leq 64 \))
Density of States: Large 2-dim Ising Model

- use a parallel, multi-range random walk

**Question:**
- Need to perform a random walk over ALL energies?

NO exact solution is available for comparison!
How Does $f_i$ Affect the Accuracy?

Data for $L=32$:

$$\varepsilon[\ln(g_i)] = \frac{1}{N_E} \sum_E \ln[g_i(E)] - \ln[g^{\text{exact}}(E)]$$

![Graph showing error vs. $f_{i-1}$]
Specific Heat of the 2-dim Ising Model

\[ \varepsilon = \text{relative error} \]
Free Energy of the 2-dim Ising Model

\[ \varepsilon = \text{relative error} \]
Wang-Landau vs Metropolis Sampling in the 2-dim Ising model

Compare states generated using Wang-Landau sampling and using the Metropolis method at $T_c$ in $L \times L$ Ising square lattices with periodic boundaries.
What About a 1st Order Transition?

Look at the $q=10$ Potts model in 2-dim

At $T_c$ coexisting states are separated by an energy barrier
$q=10$ Potts Model: Determine $T_c$
$q=10$ Potts Model: Internal Energy

![Graph showing the internal energy $U(L, T)/N$ as a function of temperature $T$. The graph includes a inset showing the behavior for different system sizes $L$. There are markers indicating the latent heat $\Delta Q$ and the critical temperature $T_c$.](image)
What About a System with Complex “Order”?

The EA (Edwards-Anderson) spin glass model in 3 dim:

\[ H = - \sum_{<i,j>} J_{ij} \sigma_i \sigma_j, \quad \sigma_i = \pm 1 \quad \text{and} \quad J_{ij} = \pm J. \]

At \( T_c \) (if it exists) a spin glass state forms \( \Rightarrow \) get a “rough” energy landscape where multiple minima are separated by high energy barriers

Define an Order Parameter

First choose a finite lattice groundstate \( \{\sigma^\alpha\} \), then

\[ q = \frac{1}{n} \sum_{\alpha} < \frac{1}{N} \sum_{i} \sigma_i^\alpha \sigma_i > \quad \text{"EA order parameter"} \]

\[ \uparrow \]

number of bond configurations

Extend random walk \( \Rightarrow \) multi-dimensional parameter space
Energy-order parameter histogram ($L=6$) performed for different random walk sequences for each bond distribution.
Distribution of States: 3d EA Spin Glass

... For larger L, $P(q,T)$ becomes even more complex!
Energy landscape – 3d EA Spin Glass Model
3d EA Spin Glass Model

Look for a phase transition (Preliminary results)
A "new" old problem: A Critical Endpoint

Theory predicts new singularities at \((g_e, T_e)\) (Fisher and Upton, 1990)

A schematic view: \(T\) is temperature; \(g\) is a non-ordering field

A "new" old problem: A Critical Endpoint

Theory predicts new singularities at \((g_e, T_e)\) (Fisher and Upton, 1990)

A schematic view: \(T\) is temperature; \(g\) is a non-ordering field
Triangular Ising Model with Two-Body and Three-Body Interactions

\[ H = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j - J_3 \sum_{\langle i, j, k \rangle} \sigma_i \sigma_j \sigma_k - H \sum_i \sigma_i, \quad \sigma_i = \pm 1 \]

\( J = 1, \quad J_3 = 2 \)

Must search for the critical endpoint in \((H, T)\) space
Critical Endpoint: Singularity in the Curvature of the Spectator Phase Boundary

\[
\frac{d^2H_\alpha(T,L)}{dT^2} \propto L^\frac{\alpha}{\nu} \quad \alpha/\nu = 2/5
\]
Critical Endpoint: Singularity in the Order Parameter

\[ d = 2 \]

\[ L = 6.27 \]

\[ \frac{dp}{dT} \]

\[ \frac{d\rho}{dT} \]
Postscript: (incomplete)

F. Wang and DPL, PRL 86, 2050 (2001); PRE 64, 056101 (2001)

Improved sampling:
- C. Yamaguchi and N. Kawashima, PRE 65, 0556710 (2002)

Proteins:

Polymer Films:

Continuum (fluid) simulations:

AF Potts model groundstate:

Reaction coordinates:

Quantum Monte Carlo:

Kondo problem:

Biological circuits:
- H.-B. Schüttler, et al.

Combinatorial number theory:
- V. Mustonen and R. Rajesh

Performance studies:
- Dayal et al. cond-mat/0306108
- C. Zhou and R. H. Bhatt, cond-matt/0306711
Overview: Wang-Landau sampling

The random walk in energy space method is powerful and flexible

- It provides direct access to the density of states
- It is easily parallelizable
- It is effective for the study of critical phenomena
- It eliminates the problem with energy barriers at 1st order phase transitions
- It can be used to study systems with complex energy landscapes and multi-dimensional parameter spaces
Critical “Dynamics”, a few comments

Time dependent behavior is important because:
- correlations reduce precision of static properties
- “dynamic” critical behavior inherently interesting

**Stochastic models** ⇒ critical relaxation
  e.g. Monte Carlo sampling
  *note: “acceleration” methods alter the critical relaxation!*

**Deterministic models** ⇒ true critical dynamics
Summary and Overview

Static behavior of many models in statistical physics can be studied by Monte Carlo sampling methods

- The “standard” is the Metropolis method which is flexible and easy to implement
- Long time-scales may be reduced with inventive algorithms (cluster flipping, N-fold way, multicanonical, etc.)
- Wang-Landau sampling provides all information from a single simulation

Time dependent behavior is more difficult . . . Spin dynamics simulations allow the study of true dynamic properties . . . more on this to come
Learn more about Monte Carlo methods

Published by Cambridge University Press

ISBN 0 521 65314 2
(hardbound)
ISBN 0 521 65366 5
(paperback)