Computational Nanoscience

NSE C242 & Phys C203

Spring, 2008

Lecture 7:

Introduction to Monte Carlo Methods

February 12, 2008

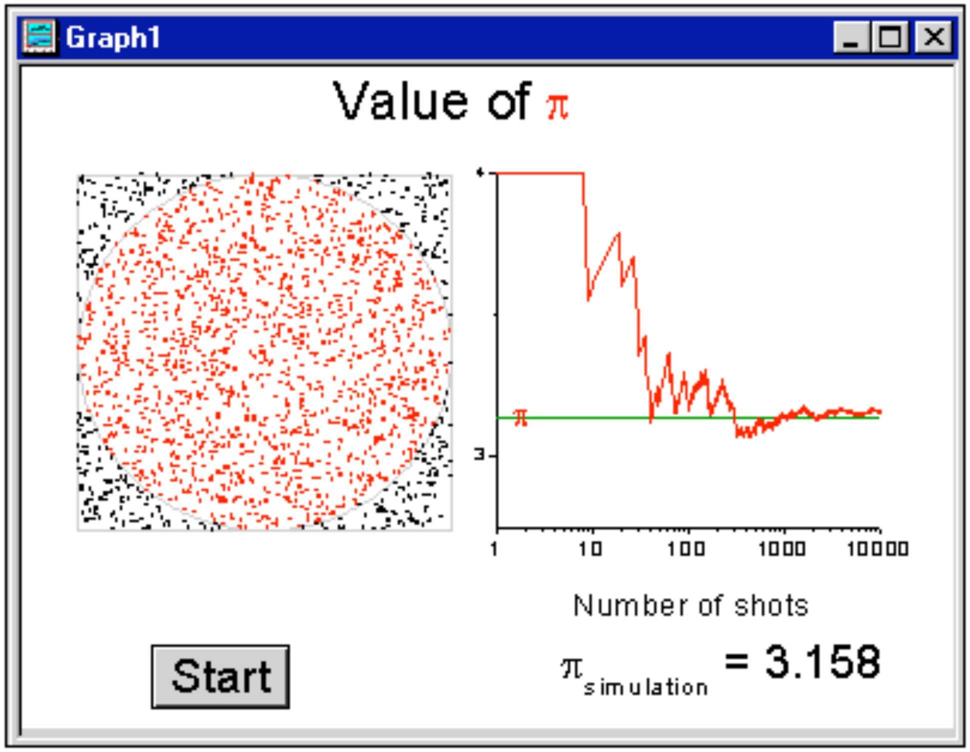
Elif Ertekin and Jeffrey C. Grossman

Monte Carlo: What and Why

Monte Carlo

- term coined by physicists at Los Alamos, 1940 (refers to gambling casinos in Monaco)
- broad term, describes an approach to solving problems that involves generating a sequence of random numbers
- Monte Carlo simulations are statistical and non-deterministic: each simulation will give a different result, but the results will be related via some statistical error
- examples: numerical integration, percolation threshold, diffusion limited aggregation, brownian motion, Ising models, radiation transport, subnuclear processes, stellar evolution, econometrics, Dow Jones forecasting … very broad, but hopefully you will have a sense for this by the time we are through with this discussion

Example: Computing Pi



from http://www.originlab.com

commercially available MC!
(why is this not common?)

How can we estimate the value of π using simple Monte Carlo methods?

Choose N points at random in the xy-plane so that -1 < x,y < 1.

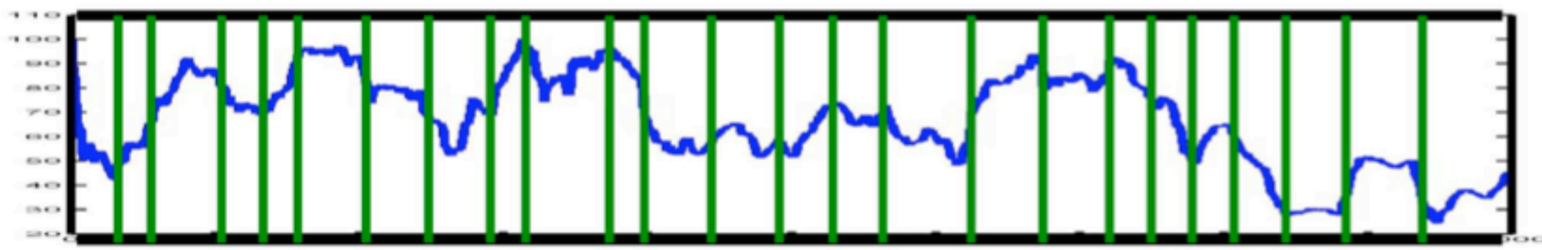
Calculate the distance from the origin for each point -- record whether it fall inside the circle or not.

Note how the estimate improves as we use more points in the simulation!

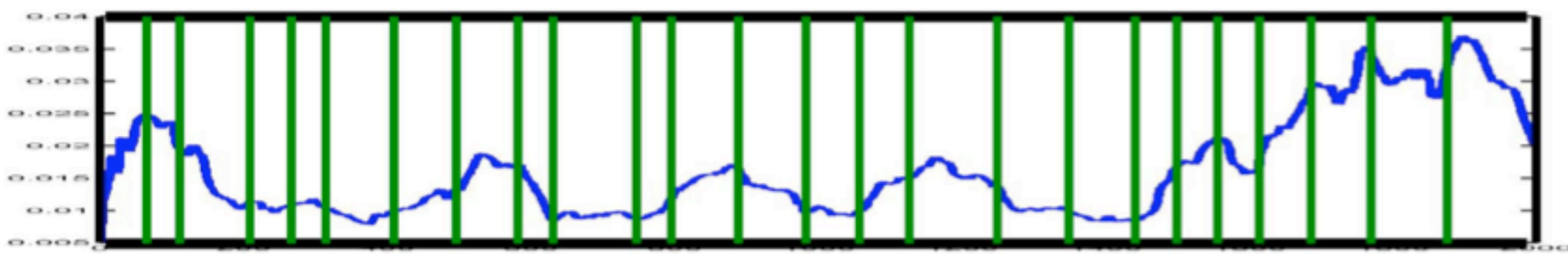
Example: Video Scheme Segmentation

Zhai & Shah. IEEE Transactions on Multimedia, Vol. 8, No. 4, August 2006.

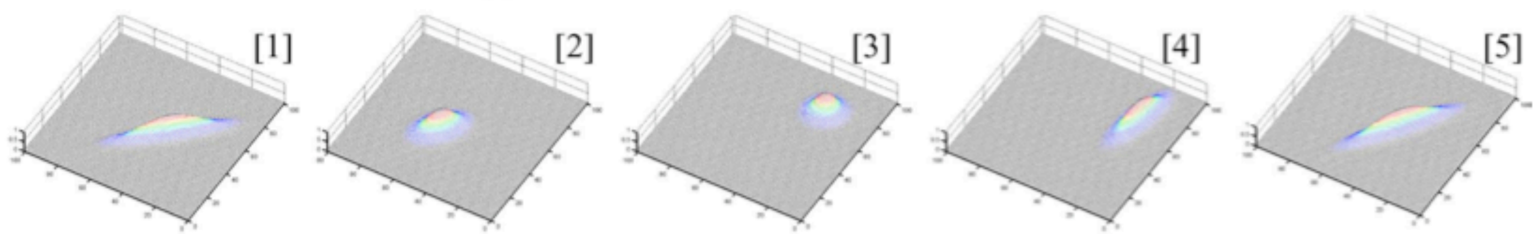
(a) Representative frames of example scenes in the movie.



(b) Plot of the shot length in the movie.



(c) Plot of the visual content in the movie.



(d) PDFs of the 2D normal distributions of the five scenes.

Example: Video Scheme Segmentation

Zhai & Shah. IEEE Transactions on Multimedia, Vol. 8, No. 4, August 2006.

TABLE III
ACCURACY MEASURES FOR THREE FEATURE MOVIES

Measures	*Gone in 60 Seconds*	*Dr. No - 007*	*Mummy Returns*
Length	01:46:09	01:30:55	01:45:33
Num. of Frames	152665	130811	151802
Num. of Shot	2237	677	1600
Num. of Scenes	29	17	18
Detected Scenes	25	20	18
Match	24	14	15
Insertion	1	3	3
Deletion	5	6	3
Precision	0.960	0.700	0.833
Recall	0.828	0.824	0.833

Example: Percolation of Permeable, Hard-Spheres

Rottereau et al. Eur. Phys. J. E 11, 61-64 (2003).

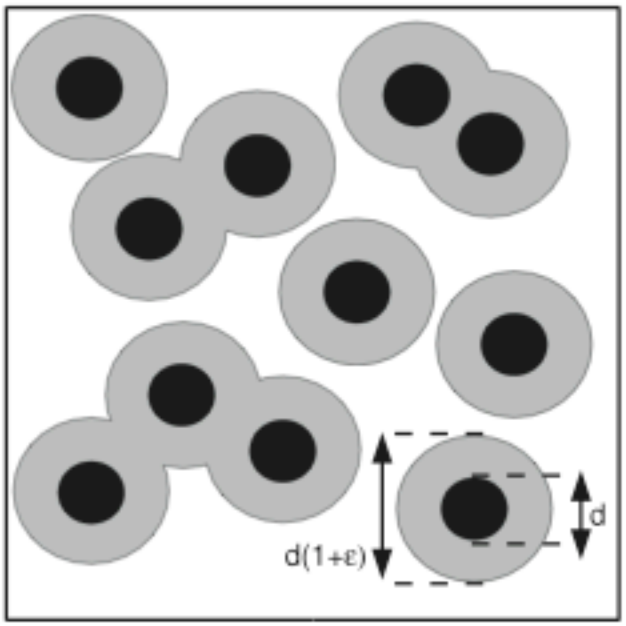


Fig. 1. Schematic representation of hard spheres with permeable shell. The black areas represent ϕ and the sum of black and grey areas represents ϕ_e.

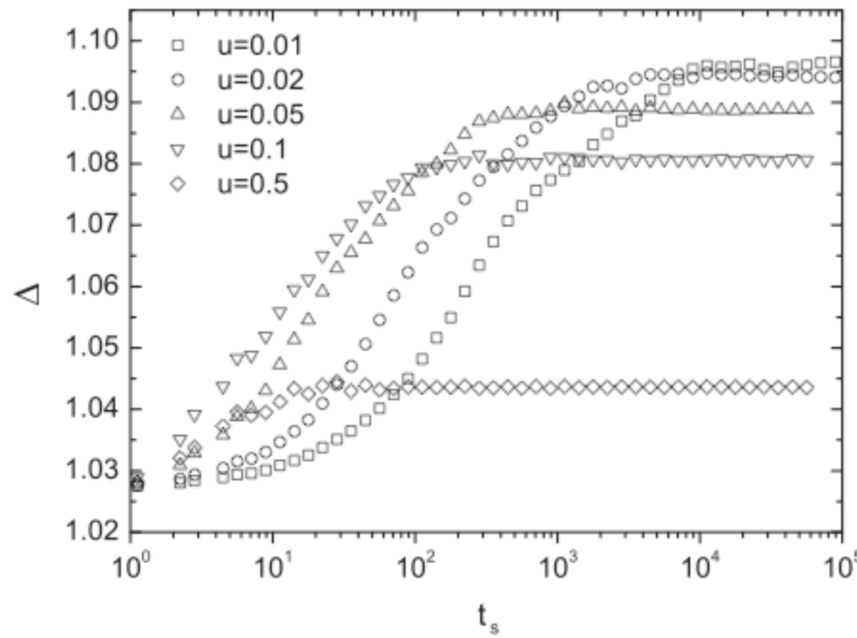


Fig. 2. Example of the time dependence of the mean distance between nearest-neighboring spheres for $\phi = 0.2$ with different step lengths for $L = 40$.

Example: Percolation of Permeable, Hard-Spheres

Rottereau et al. Eur. Phys. J. E 11, 61-64 (2003).

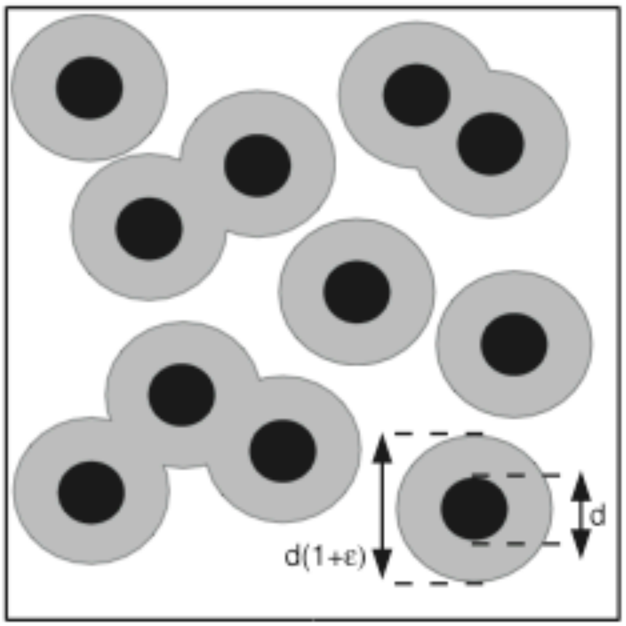
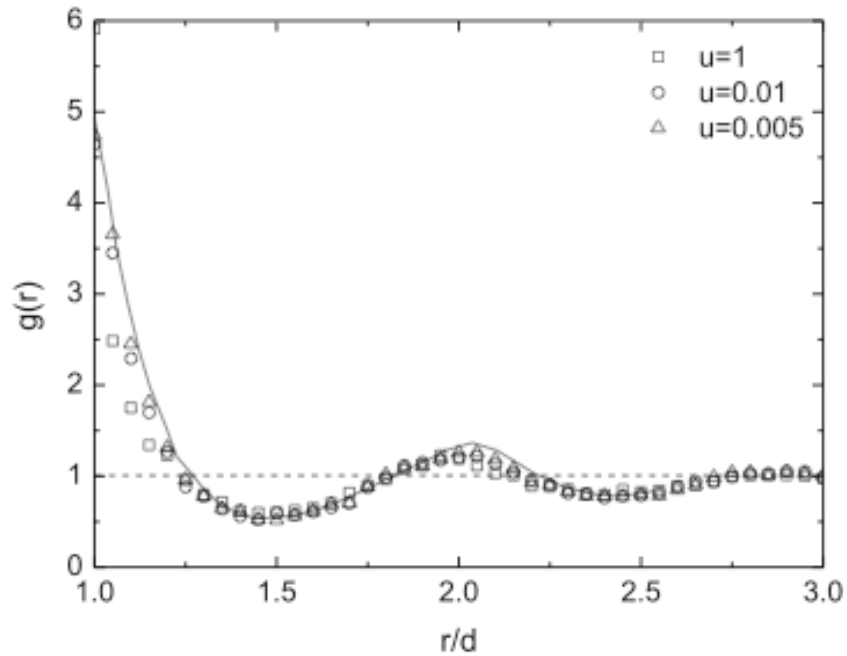


Fig. 1. Schematic representation of hard spheres with permeable shell. The black areas represent ϕ and the sum of black and grey areas represents ϕ_e.



Example: Percolation of Permeable, Hard-Spheres

Rottereau et al. Eur. Phys. J. E 11, 61-64 (2003).

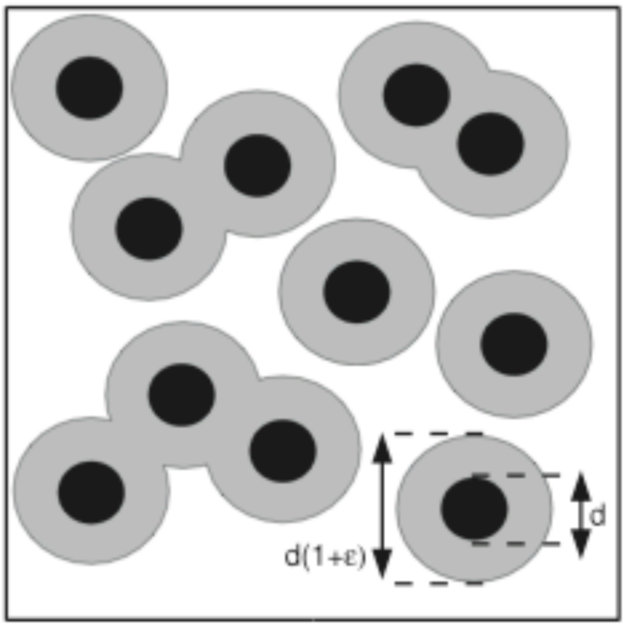


Fig. 1. Schematic representation of hard spheres with permeable shell. The black areas represent ϕ and the sum of black and grey areas represents ϕ_e.

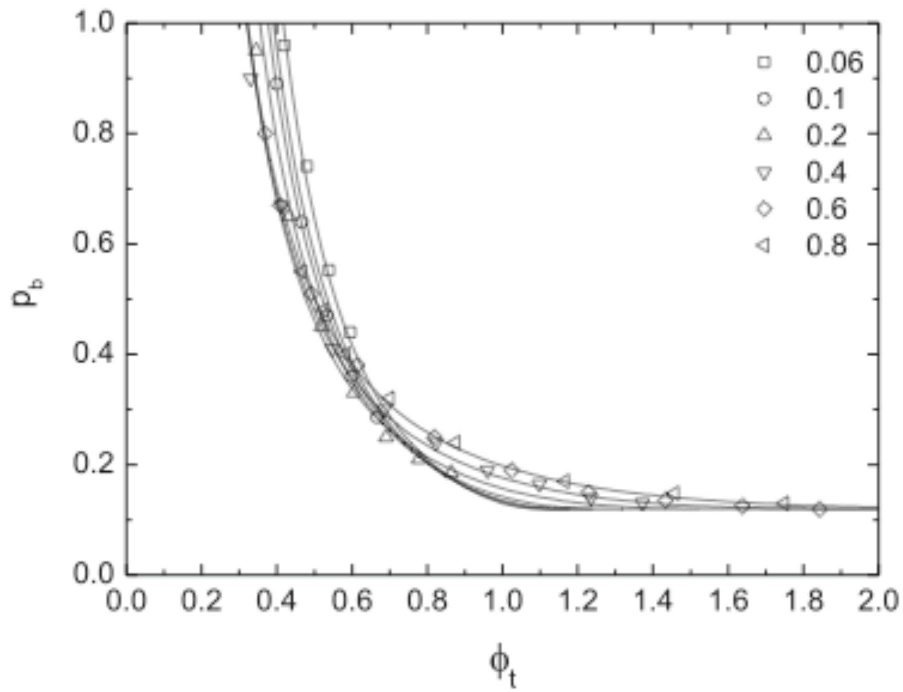


Fig. 8. The bond percolation threshold as a function of the total volume fraction for different values of ε indicated in the figure. The solid lines represent fits to equation (2) after transformation of ϕ_e into ϕ_t.

Historical Perspective

- Prior to computer simulations, liquids were modeled mechanically - large assemblies of macroscopic spheres or ball bearings (crude, and what about thermal motion?)

CHEMICAL PHYSICS

VOLUME 12, NUMBER 1 JANUARY, 1944

On the Statistical Mechanics of Liquids, and the Gas of Hard Elastic Spheres

O. K. RICE
University of North Carolina, Chapel Hill, North Carolina
(Received September 4, 1943)

Because of the extreme complications arising from a direct deductive approach, the theory of the liquid state requires the use of a model involving simplifying assumptions. In this paper an attempt is made to formulate general principles which any such model must follow. The first step is a general discussion of "communal" entropy, arising from the sharing of available space by all the atoms. Arguments are advanced to support the contention that for a gas of hard elastic spheres the communal entropy is fully excited in each direction of space, and amounts in all to $3R$ per mole. The communal entropy of assemblages of atoms exerting normal attractive and repulsive forces (in particular, the Debye solid) is considered. The geometry, the equation of state, and the partition function for an assemblage of hard elastic spheres are considered in detail. By extension of these ideas, allowing for the type of force actually exerted on each other by real atoms, a general form of partition function for a monatomic liquid is set up. This partition function involves a sum of two parts, one corresponding to a vibrational motion expressed in terms of the Debye characteristic temperature of the solid, and the other being a translational term, each part carrying with it its own communal entropy.

Historical Perspective

- Prior to computer simulations, liquids were modeled mechanically - large assemblies of macroscopic spheres (crude, and what about thermal motion?)

J.R. Bernal, Bakerian Lecture, Proc. Roy. Soc. 280:299, 1964.

In the end I fell back on the study of a model of a large number of ball-bearings. Precautions had to be taken particularly to prevent any plane surfaces around the block even if it consisted of more than a thousand balls because, as will be shown later, any regular two-dimensional array produces an effect of regular packing which goes far into the mass, figure 16, plate 17. The problem of fixing such an arrangement so that it could be measured was achieved very simply by adding black paint and letting it harden. This provided, as figure 11, plate 14 shows, marks indicating contacts and near contacts. The counting of these contacts and near contacts for a large number of balls, of the order of several hundred, was carried out by Mr J. Mason (Bernal & Mason 1960). It provided one further clue of great importance, namely, that the numbers of contacts were arranged in some definite statistical order, that is, the number of balls having five, six, seven, etc., up to eleven contacts formed a determinate curve and was absolutely distinct from the regular arrangement, where every ball must have twelve contacts. It was evident that this *variation of contact numbers* or *co-ordination* was one of the most significant features, possibly the most significant feature of the irregular liquid

Historical Perspective

- Prior to computer simulations, liquids were modeled mechanically - large assemblies of macroscopic spheres (crude, and what about thermal motion?)

J.R. Bernal, Bakerian Lecture, Proc. Roy. Soc. 280:299, 1964.

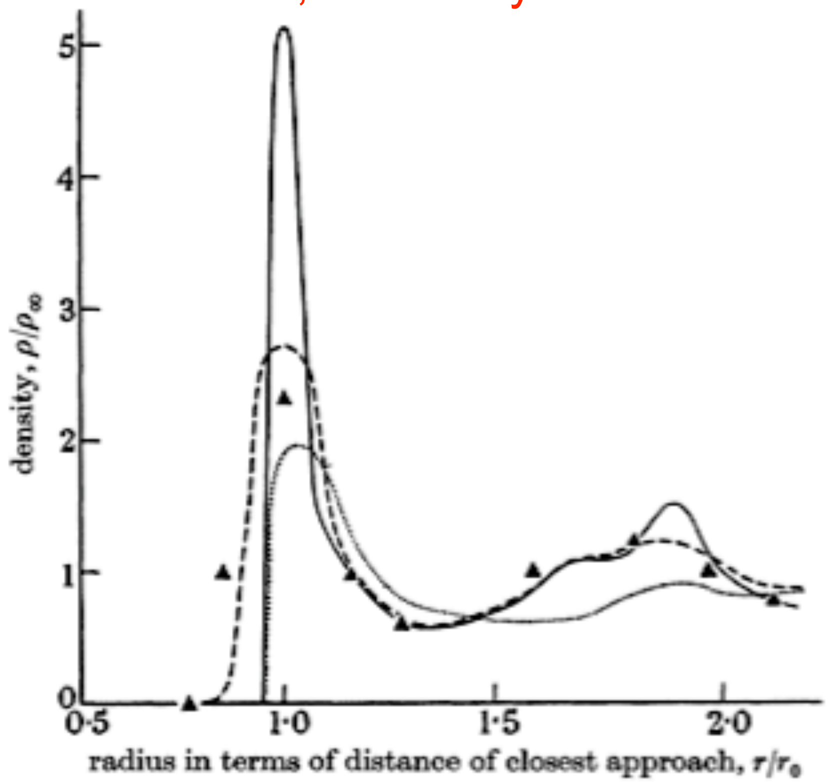


FIGURE 1. Radial distribution functions: variation of mean particle density as a function of r/r_0. ……, Derived from calculated random model; ——, derived from squeezed random model; — — — —, calculated for liquid lead by Furukawa (1960) with X-rays calculated for liquid argon by Henshaw, D. G. (*Phys. Rev.* 105, 976; 1957) with neutron diffraction.

Historical Perspective

• 1953: Electronic computers made available for non-classified research. Numerical simulation of these liquids was one of the first problems tackled:

–Metropolis et al. (Los Alamos) introduces Monte Carlo method.
J. Chem. Phys. 21: 1088-1092, 1953.

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

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.

The Problem We Want to Solve

Our goal is the evaluation of integrals such as:

$$\langle A \rangle = \frac{\int dr^N dp^N A(r^N, p^N) \exp\left[-E(r^N, p^N)/kT\right]}{\int dr^N dp^N \exp\left[-E(r^N, p^N)/kT\right]}$$

Where do such integrals arise? From statistical mechanics, the classical expression for partition function looks like this (c is normalization):

$$Z = c \int dr^N dp^N \exp\left[-E(r^N, p^N)/kT\right]$$

So computing any physical property of the system of interest will involve integral expressions like the one above.

I will present a quick introduction to why we are interested in integrals of this form in ….

Fastest Introduction to Stat Mech Ever

1 cm^3 of Silicon = 5×10^{22} atoms.

Let's say we know the differential equations that govern the motion of the atoms and wish to solve them directly. Then:

- 6 coordinates per atom (x, y, z, px, py, pz)
- 8 bytes per coordinate
- memory requirements = 8 bytes * 6 * 5 * 10^{22} = 2.4 * 10^{24} bytes

(1 GB of memory = 10^9 bytes)

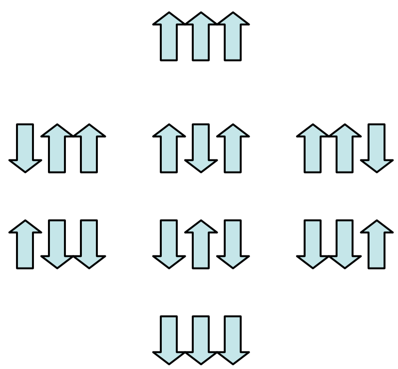
Even the memory requirements alone are absurd, even by today's standards!!

(Aside: I played a little "Moore's Law" game. If we assume that the memory available doubles every year, we are still ~ 50 years away from meeting these memory requirements.)

Statistical Mechanics - Bare Minimum

In Stat Mech, we solve this problem by coarse-graining our system:

state of a system is characterized not by 10^{23} coordinates, but instead by a small set of parameters: T, P, magnetization, etc.



3m

2m

-2m

-3m

microstates **macrostates**

Ideally, these coarse-grained parameters are sufficient to describe the properties with which we are concerned.

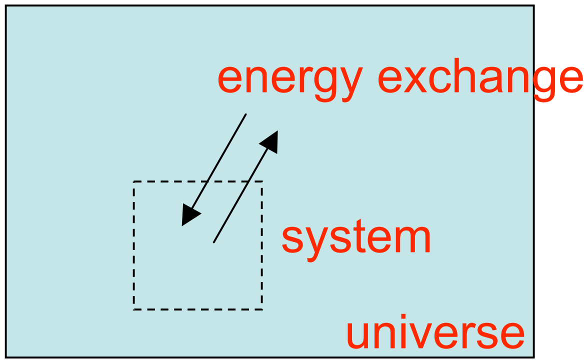
Note how different microstates can give us the same macrostate.

Underlying assumption of Stat Mech: given a particular macrostate, all corresponding microstates are equally likely

Statistical Mechanics - Bare Minimum

For a system, in contact (thermal equilibrium) with a heat bath, what is the probability $p(\varepsilon_{sys})$ of finding the system at a given microstate with energy ε_{sys}?

We get this information from the canonical distribution. (i.e., NVT ensemble)



Energy conservation:

$$\varepsilon_{tot} = \varepsilon_{bath} + \varepsilon_{sys}$$

Let $\Omega_{sys}(\varepsilon_{sys})$ = # of microstates of system with energy ε_{sys}

From Boltzmann statistics, the relative probability of finding this system at a microstate with energy ε_{sys} scales as $\exp(-\beta\varepsilon_{sys})$.

Normalizing, the probability of finding the system at a given microstate is:

$$p(\varepsilon_{sys}) \;\alpha\; e^{-\beta\varepsilon_{sys}} \quad \Rightarrow \quad p(\varepsilon_{sys}) = \frac{e^{-\beta\varepsilon_{sys}}}{\sum_{all\ microstates} e^{-\beta\varepsilon}}$$

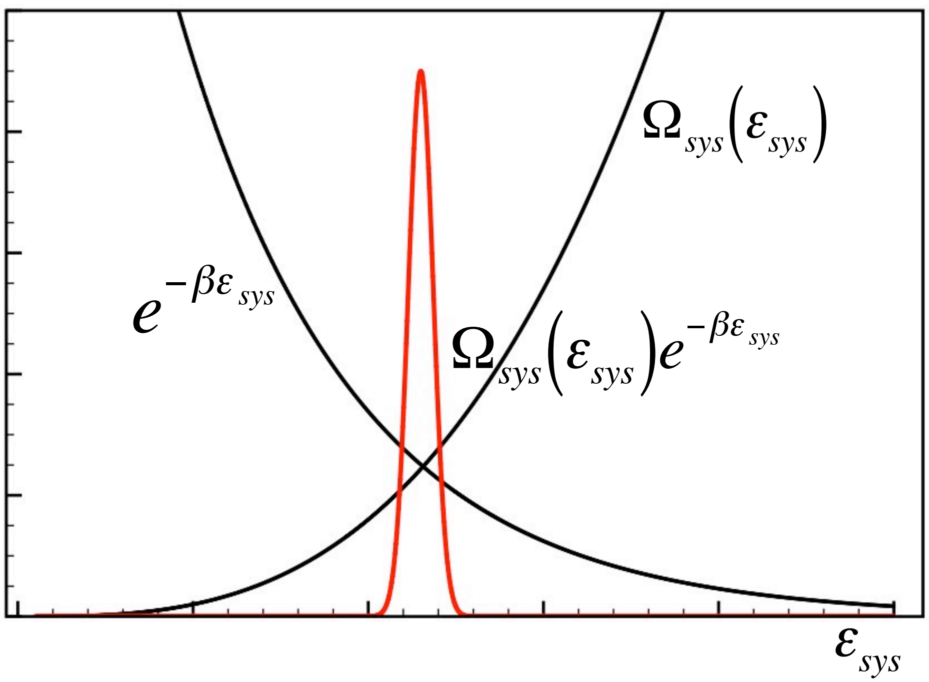
We define the partition function Z

$$Z = \sum_{all\ microstates} e^{-\beta\varepsilon_{sys}} = \sum_{all\ energies} \Omega_{sys}(\varepsilon_{sys}) e^{-\beta\varepsilon_{sys}}$$

Statistical Mechanics - Bare Minimum

What about the probability of finding the system at a particular macrostate, with energy ε_{sys}?

$$p = \frac{\Omega_{sys}\left(\varepsilon_{sys}\right)e^{-\beta\varepsilon_{sys}}}{\sum_{all\ energies}\Omega_{sys}\left(\varepsilon\right)e^{-\beta\varepsilon}}$$



Fluctuations in measured properties arise because the system is constantly exchanging energy with the universe.

Such fluctuations are inherent to the system, not just an experimental feature!

The average properties, however, are well-defined even though instantaneous properties fluctuate.

Known as the canonical distribution.

Back to Monte Carlo

Now we understand why we are likely to run into integrals of the form:

$$\langle A \rangle = \frac{\int dr^N dp^N A(r^N, p^N) \exp\left[-E(r^N, p^N)/kT\right]}{\int dr^N dp^N \exp\left[-E(r^N, p^N)/kT\right]}$$

How to evaluate? Consider direct integration, first. Let's consider even the simplest case. For a classical system, the total energy can be written as:

$$E(\mathbf{R}, \mathbf{V}) = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{v}_i^2 + V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

where **V** is a specified interaction potential.

The great majority of observables of interest do not explicitly depend on the velocities; in that case, the integration over the velocities factors out and the equation for <A> becomes:

$$\langle A \rangle = \frac{\int d\mathbf{R} A(\mathbf{R}) \exp\left[-\beta V(\mathbf{R})\right]}{\int d\mathbf{R} \exp\left[-\beta V(\mathbf{R})\right]}$$

Why Monte Carlo?

$$\langle A \rangle = \frac{\int d\mathbf{R} A(\mathbf{R}) \exp\left[-\beta V(\mathbf{R})\right]}{\int d\mathbf{R} \exp\left[-\beta V(\mathbf{R})\right]}$$

Consider numerical integration (quadrature, Simpson's rule, etc). Choose m equidistant points along each axis for the integration, in a D-dimensional space with N particles.

We have m^{DN} points at which to evaluate the function. For N=100 particles in a D=3 dimensional space, say we choose m=5. We must evaluate 10^{210} points! This is clearly impossible.

Additionally, even if we could do it, the statistical error would be huge because numerical integration works well only for functions that are smooth relative to the mesh size.

For most interatomic potentials, the Boltzmann factor is a rapidly varying function of particle coordinates. In fact, for the overwhelming majority of points, the Boltzmann factor is vanishingly small.

For example, for a fluid of 100 hard spheres at the freezing point, the Boltzmann factor is nonzero for only 1 out of every 10^{260} configurations.

Clearly, we need another approach … and this is where Monte Carlo comes in.

Metropolis Monte Carlo

Metropolis, in 1953, showed that it is possible to evaluate this ratio of integrals using statistical methods. His approach is known as "Metropolis Monte Carlo", and it is the most common form of Monte Carlo methods.

We will start by describing it, and then moving on to some generalizations of the Metropolis method to other forms of Monte Carlo.

In the Metropolis method, we randomly generate points in configuration space according to the Boltzmann distribution. This is called importance sampling.

We do this via a random walk in configuration space, however, the next configuration is rejected or accepted in such a way that we sample the configuration according to the probability distribution.

Thus, we generate a sequence of points in configuration space with a relative probability proportional to the Boltzmann factor.

Metropolis Monte Carlo

• This is why we use Monte Carlo simulations to generate sequences of configurations of the system weighted according to the canonical distribution.

• If we measure the property of interest (energy, magnetization) at each configuration, we can compute the average value of that property over the course of the simulation.

• Of course, there will be some statistical error in our calculated average (more later…)

Note: we have lost something along the way!

- There is no inherent sense of “time” in a pure Monte Carlo simulation - each step of the simulation corresponds to a different microstate or configuration that is physically accessible.
- We will get to “time” when we discuss Kinetic Monte Carlo later on.

Metropolis Monte Carlo

How do we sample the configuration according to the Boltzmann distribution?

Let "c" denote the current configuration, and "n" denote a possible new configuration. Let N(c) give the relative probability of sampling state c, and let $\pi(c \rightarrow n)$ denote the transition probability from state c to state n.

The transition probabilities must satisfy one important rule. Once an equilibrium distribution is reached, the transition probabilities must maintain that equilibrium.

The average number of accepted trial moves that result in the system leaving state c must be exactly equal to the number of accepted trial moves from all other states n into c.

In practice, we impose much stricter condition for convenience: the average number of trial moves from c to a specific state n is exactly canceled by the number of moves from n to c.

This is the condition of detailed balance. Mathematically,

$$N(c)\pi(c \rightarrow n) = N(n)\pi(n \rightarrow c)$$

Metropolis Algorithm

$$N(c)\pi(c \rightarrow n) = N(n)\pi(n \rightarrow c)$$

Theorists' Warning: be extremely wary of simulations that violate detailed balance! Detailed balance implies configurations consistent w/ equilibrium …

We can implement detailed balance in many ways. Let's break it down further…

$$\pi(c \rightarrow n) = \alpha(c \rightarrow n) \times acc(c \rightarrow n)$$

Probability of selecting state n as the next trial configuration given that you are currently in state c

Probability of accepting the move from c to n, given that you have selected n.

Metropolis Algorithm

In the original Metropolis scheme, the factor α is symmetric:

$$\alpha(c \rightarrow n) = \alpha(n \rightarrow c)$$

Thus, our detailed balance condition becomes:

$$N(c) \times acc(c \rightarrow n) = N(n) \times acc(n \rightarrow c)$$

To correctly sample the canonical distribution, this means that the relative acceptance rates must satisfy:

$$\frac{acc(c \rightarrow n)}{acc(n \rightarrow c)} = \frac{N(n)}{N(c)} = \frac{\exp[-\beta\varepsilon_n]}{\exp[-\beta\varepsilon_c]} = \exp[-\beta(\varepsilon_n - \varepsilon_c)]$$

Many ways to do this, but Metropolis approach is:

$$acc(c \rightarrow n) = \begin{cases} \exp[-\beta(\varepsilon_n - \varepsilon_c)] & \text{if } \varepsilon_n > \varepsilon_c \\ 1 & \text{if } \varepsilon_n < \varepsilon_c \end{cases}$$

Metropolis Algorithm - Implementation

1. Assign an initial configuration (non-trivial!!) for your system, record it's energy.
2. Choose a second possible configuration for the system (e.g. move an atom by a diffusion step in a random direction, flip a spin, move all of the atoms, etc), and compute the change in energy $\Delta\varepsilon$ associated with the change in configuration
3. If $\Delta\varepsilon < 0$: accept the change

   If $\Delta\varepsilon > 0$: accept the change with probability $e^{-\beta(\Delta\varepsilon)}$

   - i.e. generate a random number r uniformly distributed between 0 and 1.

   - If $r < e^{-\beta(\Delta\varepsilon)}$, accept the change. Otherwise, reject it.

   This accepting and rejecting ensures that we choose our configurations consistently with the Boltzmann population distribution.

4. Repeat steps 2 and 3 as long as reasonable. We can use these successive configurations to obtain an estimate for the desired average.

Basic idea: The Metropolis algorithm obeys detailed balance and exhibits the dynamics of a canonical distribution.

How do we generate trial moves?

First, we have to devise a way to choose a new trial configuration $\mathbf{R}^n$ from a current one, $\mathbf{R}^c$

Generally, we use Markov chains to do this. The probability of choosing a trial state $\mathbf{R}^n$ depends only on the current state which you are in $\mathbf{R}^c$.

There are many ways to choose $\mathbf{R}^n$ from $\mathbf{R}^c$. For instance,

- move all particles by a random 3D vector
- move one particle by a random 3D vector
- move all particles by a diffusion step of fixed length in a random direction
- move one particle by a diffusion step of fixed length in a random direction

Of course, after choosing the trial configuration, the next step will be an acceptance test.

That is, we have only proposed a move, but we need to determine whether that attempt should actually be taken or whether it should be discarded and a new trial move computed.

Efficient Sampling

We want an efficient sampling procedure -- we want the lowest statistical error for a given amount of computing time.

We can assume that the mean-square error is inversely proportional to the number of uncorrelated configurations visited. But the number of independent configurations depends on how much of phase space we can cover.

Maximizing efficiency is a balancing game:

- if we use too large a step size, it is likely that the new configuration will be high energy and thus rejected

- if we use too small a step size, then successive measurements are correlated and we have less independent configurations.

Efficient Sampling

Is there a single optimal acceptance ratio?

Often, 50% is cited as a target acceptance ratio.

In truth, not really - it depends on the specifics of your system.

For instance, does the amount of computing required to test whether a trial move is accepted depend on the magnitude of the move?

Not for continuous potentials, but it does for hard spheres -- a move can be rejected as soon as neighbor overlap is detected.

Thus, for hard spheres, rejection is cheap, and we can accommodate lower acceptance ratios (20%).