Ensemble Average by Monte Carlo Simulations

Introduction
In molecular dynamics simulations to get quantity of interest such as temperature or pressure we average "instantaneous" quantities over the simulation time;

$$A_{obs} = \langle A \rangle_{time} = \frac{1}{\tau_{obs}} \sum_{\tau=1}^{\tau_{obs}} A(\tau), p(\tau)).$$

Instead of following real trajectories of particles we can randomly sample configuration in (r, p) space and this way calculate desired thermodynamic averages. This is typically done in the canonical ensemble where the ensemble average is given by;

$$A_{obs} = \langle A \rangle = \frac{\sum_{N_{conf}} A(r, p) e^{-E(r, p)/k_B T}}{\sum_{N_{conf}} e^{-E(r, p)/k_B T}}.$$  (2)

Since we do not intend to follow trajectories with velocities we focus on quantities that depend only on positions A=A(r). With E(r,p) = K(p) + V(r) we can write

$$A_{obs} = \langle A \rangle = \frac{\sum_{N_{conf}} A(r) e^{-V(r)/k_B T}}{\sum_{N_{conf}} e^{-V(r)/k_B T}}.$$  (3)

Rigorously speaking to calculate the average given by Eq. 3 we need to sum over all possible configurations. However, we expect that with sufficient fraction of configurations sampled, the average value will be close to the exact result.

Uniform and Importance Sampling
The most straightforward way of evaluating Eq. 3 is to randomly generate sufficient number of configurations and calculate averages. This procedure is called uniform sampling. As a example consider 100 particles interacting with LJ potential placed in the cubic box of volume V. A random configuration can be generated by randomly placing each of 100 particles in the box. For each configuration its energy has to be calculated to evaluate Eq. 3.

The uniform sampling can be computationally very inefficient. For example, assume that to sample "enough" the fist particle has to have ~100 random coordinates, than the second one needs its own 100 random coordinates etc. Thus with 100 particles we need 100^100 configuration which is beyond capabilities of any computer. Furthermore, for a dense system, such as liquid, if we pick particle positions at random, the particles can easily overlap with each other and than V(r) →∞, meaning that this configuration has zero probability it does not contribute to the average given by Eq. 3.
This leads to the key question: Can we design a sampling procedure such that we do not pick configurations that do not contribute to the averages and pick only configurations that are important?

In fact, for the most efficient sampling, we would like to pick configurations with probability proportional to Boltzmann distribution:

$$P_{\text{Boltz}}(r) = \frac{e^{-V(r)/k_B T}}{\sum_{\text{conf}} e^{-V(r)/k_B T}}, \quad \text{(4)}$$

and instead of uniform sampling averages given by Eq. 3, we simply get:

$$\langle A \rangle = \frac{\sum_{\text{conf}\{r\}} A(r)}{N_{\text{conf}}} \quad \text{(5)}$$

where in Eq. 5 configurations are chosen with probability equal to $P_{\text{Boltz}}(r)$. This not only eliminates all time wasted for configurations with very high energies (e.g. overlapping particles), but also focus the averaging process on configurations that have the highest probability of occurring. This procedure is called the importance sampling, since most of the time only important (high probability) configurations are sampled.

**Markov Chain**

The next question is how to actually pick configurations with probabilities proportional to Boltzmann distribution? To address this problem we first introduce the concept of fictitious dynamics, in which we will start from a given configuration and attempt to "move" from one configuration to the other according to such rules that the average "time" spent in each configuration will be proportional to Boltzmann probability. The "time" $t$ is a computer time (marking the number of iterations of the procedure), NOT real time -- our statistical system is considered to be in equilibrium, and thus time is invariant. Also "moves" can be very unphysical, such as particles jumping over large distances.

To formalize the discussion, let $P\{r_1, t\}$ be the probability of being in configuration $\{r_1\}$ at time $t$, and $W(1 \to 2)$ be the (transitional) probability of moving from configuration $\{r_1\}$ to configurations $\{r_2\}$ per unit time. Then the conservation of probabilities $P$ requires that:

$$P\{r_{1,t} + \delta t\} = P\{r_{1,t}\} + \sum_2 W(2 \to 1) P\{r_{2,t}\} - W(1 \to 2) P\{r_{1,t}\} \quad \text{(6)}$$

If we consider system in equilibrium, probability $P\{r_{1,t}\}$ is time independent, $P\{r_{1,t}\} = P\{r_1\}$. To satisfy this condition we can impose detailed balance;

$$W(1 \to 2) P\{r_{1,t}\} = W(2 \to 1) P\{r_{2,t}\} \quad \text{(7)}$$
By inserting Eq. 7 to Eq. 6 it can be easily verified that \( P\{r_1, t\} = P\{r_1, t+\delta t\} = P\{r_1\} \). This method can be used for any probability distribution. With our choice of the Boltzmann distribution, combining Eqs. 4 and 7 allows us to calculate the ratio of transitional probabilities, \( W \).

\[
\frac{W\{1 \rightarrow 2\}}{W\{2 \rightarrow 1\}} = \frac{P_{\text{Boltz}}\{r_2\}}{P_{\text{Boltz}}\{r_1\}} = \frac{e^{-V(r_2)/k_B T}}{e^{-V(r_1)/k_B T}} = e^{-\Delta E/k_B T} \tag{8}
\]

where \( \Delta E = V\{r_2\} - V\{r_1\} \) is the energy difference between configurations \{\( r_1 \)\} and \{\( r_2 \)\}. This energy difference is easy to calculate.

In summary, we demonstrated that the problem of selecting given configuration with the Boltzmann probability was is equivalent to a procedure in which we start with some initial configuration and transition to new configurations with the transitional probabilities given by \( W_s \). This sequence of configurations generated in the procedure is called the Markov chain. As far as \( W_s \) satisfy Eq. 8 we indeed sample configurations according to the Boltzmann distribution.

**The Metropolis Algorithm**

There are many possible choices of the \( W_s \)'s which will satisfy detailed balance. The most popular one selected by the inventors of the method: Metropolis, Teller, and Rosenbluth in 1953 (supposedly at a Los Alamos dinner party) is;

\[
\begin{align*}
W\{1 \rightarrow 2\} &= e^{-\Delta E/k_B T} & \text{if } \Delta E > 0 \\
W\{1 \rightarrow 2\} &= 1 & \text{if } \Delta E < 0 
\end{align*} \tag{9}
\]

Eq. 9 states that the transition probability to a configuration with higher energy is given by \( e^{-\Delta E/k_B T} \), and probability of transition to a configuration with lower energy is equal to unity. By inspection we can verify that if \( E\{2\} > E\{1\} \);

\[
\frac{W\{1 \rightarrow 2\}}{W\{2 \rightarrow 1\}} = \frac{e^{-(E(r_2) - E(r_1))/k_B T}}{1} = e^{-\Delta E/k_B T} \tag{10}
\]

and if \( E\{2\} < E\{1\} \);

\[
\frac{W\{1 \rightarrow 2\}}{W\{2 \rightarrow 1\}} = \frac{1}{e^{-(E(r_1) - E(r_2))/k_B T}} = e^{-\Delta E/k_B T}. \tag{11}
\]

In both cases the condition given by Eq. 8 is satisfied.

The computation procedure implementing the above Markov chain based sampling process is remarkably simple. It consists of the following steps:

1. Generate initial configuration, \{\( r_1 \)\}.
2. Generate a new configuration \{\( r_2 \)\}. For example move a particle to a new position.
3. Calculate the energy difference $\Delta E = E(r_2) - E(r_1)$.
4. If $\Delta E < 0$, accept new configuration.
5. If $\Delta E > 0$, accept new configuration with probability $p = e^{-\Delta E/k_BT}$. This can be done by generating a random number $0 < z < 1$. If $p > z$, accept, otherwise reject.
6. Use the current configuration (new if accepted, old if rejected) to calculate all sort of averages.
7. Go to point 2 and repeat the procedure 2-7 for $n$ times.

After $n$ steps your average will be given by

$$\langle A \rangle = \frac{1}{n} \sum A_{\text{conf}}$$

(12)

Often, your starting configuration can be far away for equilibrium. In this case do not average until the system is in equilibrium. Depending on the choice of your “move” i.e., way of generating a new configuration you might get stuck in a local energy minimum, particularly at low temperatures, and thus do not sample the configuration space correctly.

**Lattice Models**

The Monte-Carlo importance sampling procedures according to the Metropolis algorithm are particularly useful to simulate so-called lattice models. In a lattice model rather than moving particles in a continuous space, particles or model objects exist on a fixed lattice and the only thing that can change is the "state of the lattice site". For example, in Fig. 1 we show a square lattice in which each square can be occupied by a particle of type A, or type B, or not occupied at all. Depending of the choice and on the problem, Monte-Carlo moves can involve, moving an existing particle to an empty square, exchanging particle A with B, or creating a new particle (A or B) in an empty square. The last “move” will be allowed in simulations of chemically open systems in which the total number of particles can change.

**The Ising Model**

In order to study a given physical phenomena and implement the Metropolis algorithm it is necessary to describe physical interactions and thus be able to calculate the energy of the system for a given configuration. The simplest possibility is to consider only the nearest neighbor interactions. In an example given in Fig. 1, only two particles residing in adjacent (via an edge) squares will interact. Other pairs of particles will not interact.
The most celebrated nearest neighbor interaction model is the Ising model. The Ising model consists of spins, $S$, arranged on a lattice. Each spin can have value one ($S=1$, up spin) or negative one ($S=-1$, down spin). The total energy of the system is given by:

$$E = -J \sum_{i,j} S_i S_j - h \sum_i S_i$$  \hspace{1cm} (13)$$

where the double sum is over all pairs of spins that are nearest neighbors, $J$ is the nearest neighbor interaction energy and $h$ represent external fields, such as magnetic field, but also can represent chemical potential for an open system. The original Ising model was constructed to describe magnetism, but, as we will see, it can also describe surface adsorption of a gas or binary alloys.

For two dimensional square lattice the phase diagram of the Ising model was calculated analytically, while in 3D only numerical or approximate solutions exist. In general, at high temperatures, $J/k_B T = 1$ or higher, spins will randomly point up or down and with $h=0$, the average spin is zero. Tuning magnetic field to positive or negative values will continuously change the average spin to positive or negative values. At low temperatures $J/k_B T <<1$, the spins align themselves and with positive $h$ the average spin is $-1$, while with negative is $-1$. Thus in this case, changing the magnetic field from the positive to negative value will lead to discontinuous change in the average spin. This behavior represents a phase transition between a single (random) phase at high temperatures and two ordered phases at low temperatures. We can simulate the thermodynamic behavior of the Ising model using the Metropolis algorithm in which moves will involve changing values of spins from one to minus one and vice versa.

**Ising Model and Binary Mixtures**

Let us consider a copper – zinc 50/50 alloy. We can represent a site with copper atom as $S_i = 1$ and site with zinc atom as $S_i = -1$. Defining $J_{\text{CuCu}}$, $J_{\text{ZnZn}}$, and $J_{\text{CuZn}}$ as the interaction between two nearest neighbor copper atoms, zinc atoms and a copper and a zinc atom, respectively we may write down the total energy of the system as;

$$E = \frac{1}{4} \sum_{i,j} J_{\text{CuCu}} (1 + S_i)(1 + S_j) + \frac{1}{4} \sum_{i,j} J_{\text{ZnZn}} (1 - S_i)(1 - S_j) + \frac{1}{4} \sum_{i,j} J_{\text{CuZn}} \{ (1 + S_i)(1 - S_j) + (1 - S_i)(1 + S_j) \}$$  \hspace{1cm} (14)$$

and by collecting terms in the above equation we can recover the original Ising model;

$$E = -J \sum_{i,j} S_i S_j - h \sum_i S_i + C$$  \hspace{1cm} (15)$$

with $J = 1/4(2J_{\text{CuZn}} - J_{\text{CuCu}} - J_{\text{ZnZn}})$, $C$ being just a constant and $h$ not relevant since the sum over $S_i$ is equal to zero due to equal number of zinc and copper atoms. If we wish
to simulate 50/50 alloy with the MC simulations we can only perform moves that preserve composition, such as exchanging position of Cu atom with Zn atom.

Due to the above demonstrated equivalence, the phase diagram of model CuZn alloy is exactly the same as that of the Ising model. In particular at high temperatures spins will be random, i.e., we will have a random distribution of zinc and copper atoms in a single-phase material (disordered alloy). At low temperatures, if on average Cu-Cu and Zn-Zn binding energies are stronger than Zn-Cu binding energy (positive $J$), the system will separate into two phases Cu ($S=1$) and Zn ($S=-1$) rich, respectively. If the Zn-Cu bonding is stronger that average of Cu-Cu and Zn-Zn bonding (negative $J$), the Zn atoms will tend to be neighbors on Cu atoms and we will have a single phase material – an ordered alloy. In the original Ising model it corresponds to antiferromagnetic phase with neighboring spins having opposite values.

**Ising Model and Surface Adsorption**

Adsorption of a gas onto a solid surface can be modeled by the lattice gas model. The lattice gas model on a 2 dimensional lattice specifies the state of each lattice site as either occupied, $t_i = 1$, or not occupied, $t_i = 0$. With variables $t_i$, the energy of the system is a given configuration is;

$$E = -J_L \sum_{i,j} t_i t_j - \mu_L \sum_i t_i$$  \hspace{1cm} (16)

where $J_L$ is the nearest neighbor attraction between occupied sites and $\mu_L$ is a chemical potential which controls the adsorption process: a large positive $\mu_L$ lead to most sites being occupied and a large negative $\mu_L$ will favor low coverage.

By a simple transformation;

$$t_i = (1-S_i)/2$$  \hspace{1cm} (17)

we can recover the usual nearest neighbor Ising model, with the magnetic field $h$ related to the chemical potential, $\mu_L$.

From the Ising model behavior we already know that at high temperature surface coverage (magnetization) will vary continuously as a function of chemical potential (magnetic field). For example, coverage of gas on a solid surface will continuously increase with increasing gas pressure. At low temperature there will a discontinuous jump in surface coverage, at a characteristic to the system gas pressure.

**Problems**

MC.1. Insert Eq. 17 into Eq. 16 and rearrange the terms to recover the original Ising model form (Eq. 13). Find $J$ and $h$ in terms of $J_L$ and $\mu_L$. At low temperatures at which combination of $J_L$ and $\mu_L$ you will expect an abrupt change of surface coverage to be observed.