Monte Carlo simulations in classical statistical physics

Anders W. Sandvik, Department of Physics, Boston University

1 Introduction

Monte Carlo simulation is a very important class of stochastic methods for calculating thermal properties of many-particle systems—arguably these are the most important numerical techniques in statistical physics. Monte Carlo simulation methods are related to the elementary Monte Carlo integration methods that we discussed earlier, but are based on more efficient non-uniform sampling schemes. By using importance sampling, the configurations (particle positions, spin directions, etc.) of a finite but large many-body system (up to millions of degrees of freedom) can be generated according to the Boltzmann distribution, so that thermal expectation values are obtained as simple arithmetic averages of functions “measured” on the configurations.

As a simple illustration of the advantages of non-uniform Monte Carlo sampling, consider a one-dimensional integral similar to a thermal expectation value in statistical physics (the discussion here can be directly generalized to multi-dimensional integrals);

\[ \langle A \rangle = \int_{-L}^{L} P(x)A(x)dx, \quad \int_{-L}^{L} P(x)dx = 1, \]  

(1)

where \( P(x) \) is an arbitrary probability distribution. By randomly sampling \( M \) points \( x_1, \ldots, x_M \) in the range \([-L, L]\), the expectation value is estimated as

\[ \langle A \rangle \approx \frac{2L}{M} \sum_{i=1}^{M} P(x_i)A(x_i). \]  

(2)

As we discussed before, if \( P(x) \) is sharply peaked in a small region, the statistical fluctuations of this estimate will be large as only a small fraction of the generated points will fall within the dominant region. If we instead sample the points according to some probability distribution \( W(x) \), i.e., the probability of picking a point in an infinitesimal range \([x, x + dx]\) is \( W(x)dx \) (we now assume that this can be done for arbitrary \( W(x) \)), and leave for later discussion how this is accomplished in practice), then the estimate for the expectation value is

\[ \langle A \rangle \approx \frac{1}{M} \sum_{i=1}^{M} P(x_i) A(x_i). \]  

(3)

This has less statistical fluctuations than the estimate (2) of the uniform sampling if \( W(x) \) is peaked in the same region as \( P(x) \) and if the function \( A(x) \) is well-behaved, in the sense of being reasonably smooth and not very small where \( P(x) \) is large and vice versa. More precisely, the fluctuation in the values sampled using the distribution \( W(x) \) is given by

\[ \sigma_W^2[A] = \int_{-L}^{L} \left( \frac{P(x)}{W(x)}A(x) - \langle A \rangle \right)^2 W(x)dx, \]  

(4)
which in principle can be minimized by choosing a particular \( W(x) \). In general (for the multidimensional integrals or sums encountered in statistical physics) it is not possible in practice to find the optimal \( W(x) \) that minimizes the fluctuations, but if \( P(x) \) has much larger variations than \( A(x) \) a very good solution is to use \( W(x) = P(x) \). The expectation value is then just the simple arithmetic average of \( A(x) \) over the sampled configurations

\[
\langle A \rangle \approx \frac{1}{M} \sum_{i=1}^{M} A(x_i), \tag{5}
\]

and the expected fluctuation of the measured values is

\[
\sigma_P^2[A] = \int_L^L [A(x) - \langle A \rangle]^2 P(x)dx. \tag{6}
\]

It should be noted that the distribution of the values \( A_i \) is typically not Gaussian, and hence to calculate the statistical errors of \( \langle A \rangle \) estimated as (5) the values should first be binned, in the same way as we discussed previously in the chapter on Monte Carlo integration.

In statistical physics, \( P \) is a sharply peaked exponential function \( e^{-E/k_BT} \) of the energy and \( A \) is typically a linear or low-order polynomial function of the system degrees of freedom. The fluctuations in \( P \) are thus very large relative to those of \( A \) and the sampling using \( P \) as the probability distribution is then close to optimal. This is what is normally meant by the term importance sampling. Using importance sampling instead of uniform random sampling is crucial when a small fraction of the configuration space dominates the partition function, which is always the case with the Boltzmann probability in statistical mechanics models at temperatures of interest. How to achieve the correct distribution in practice is the main theme of this chapter; we will discuss importance sampling schemes for both lattice and continuous-space models.

One of the primary utilities of Monte Carlo simulation is in studies of phase transitions and critical phenomena. This will be the focus of applications discussed here. Although there are analogous simulations methods available also for quantum systems (called quantum Monte Carlo methods), we will here consider only Monte Carlo simulations of classical many-body models. In addition to describing simulation algorithms, we will also discuss how simulation data is analyzed in order to locate phase transitions and extract critical exponents.

In the following sections we will first briefly review the the expressions for thermal expectation values in systems of particles identified by coordinates in continuous space. We then consider models with discrete degrees of freedom on a lattice, focusing on spin models, the Ising model in particular. We will discuss the general form of the detailed balance condition that can be used to sample configurations according to any desired probability distribution. Monte Carlo simulation algorithms are for instructional purposes often developed in the context of the Ising model, and we will follow this path here as well (it should also be noted that Ising models are of continued importance in research). We will develop a standard program for simulations using the Metropolis algorithm, which is based on evolving (updating) configurations by flipping individual spins. We will study the performance of this method using autocorrelation functions, which characterize the way in which generated configurations gradually become statistically independent of the past configurations. This leads us to the problem of critical slowing down, which makes accurate studies close to phase transitions difficult. Critical slowing down can be often be greatly reduced, in some cases completely
eliminated, using *cluster algorithms*, where clusters of spins are flipped collectively (using cluster-building rules that satisfy detailed balance). We will develop a cluster Monte Carlo program and use it to study the ferromagnetic phase transition in the two-dimensional Ising model. For this purpose, we will also discuss *finite size scaling* methods used to extract critical points and exponents. Finally, we will return to problems involving particles in continuous space; we will develop a program for simulating a mono-atomic gas and its phase transition into the liquid state.

2 Statistical mechanics of many-body systems

We here briefly review the mathematical formalism used for evaluating thermal expectation values in classical many-body physics, considering first particles in continuous space and after that focusing on the lattice spin models for which we will develop Monte Carlo simulation algorithms initially. We will also discuss the magnetic phase transitions that can be studied using Monte Carlo simulations.

2.1 Particles in continuous space

For a system of $N$ particles, with position coordinates $\vec{x}_i$ and momenta $\vec{p}_i$ in a $d$-dimensional space, the thermal expectation value of a quantity $A$ at temperature $T$ is given by

$$\langle A \rangle = \frac{1}{Z} \int \prod_{i=1}^{N} dx_i^d \int \prod_{i=1}^{N} dp_i^d A(\{\vec{x}_i, \vec{p}_i\}) e^{H(\{\vec{x}_i, \vec{p}_i\})/k_B T},$$

(7)

where $Z$ is the partition function

$$Z = \int \prod_{i=1}^{N} dx_i^d \int \prod_{i=1}^{N} dp_i^d e^{H(\{\vec{x}_i, \vec{p}_i\})/k_B T},$$

(8)

$k_B$ is Bolzmann’s constant, and $H$ is the Hamiltonian. For identical particles of mass $m$ in a potential $U(\vec{x}_i)$ and a two-particle interaction $V(\vec{x}_i, x_j)$, the Hamiltonian is

$$H(\{\vec{x}_i, \vec{p}_i\}) = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \sum_{i=1}^{N} U(\vec{x}_i) + \sum_{i \neq j} V(\vec{x}_i, x_j).$$

(9)

If the observable $A$ is velocity independent (i.e., a function only of the positions $x_i$), the momentum integrals cancel in (7), leading to

$$\langle A \rangle = \frac{1}{Z} \int \prod_{i=1}^{N} dx_i^d A(\{\vec{x}_i\}) e^{E(\{\vec{x}_i\})/k_B T},$$

(10)

$$Z = \int \prod_{i=1}^{N} dx_i^d e^{E(\{\vec{x}_i\})/k_B T},$$

(11)

i.e., only the potential energy,

$$E(\{\vec{x}_i\}) = \sum_{i=1}^{N} U(\vec{x}_i) + \sum_{i \neq j} V(\vec{x}_i, x_j),$$

(12)
is relevant for the static properties [the density \( \rho(\vec{x}) \), density fluctuations, equal-time correlation functions, etc.] of the system. Often the only velocity dependent quantity considered in equilibrium statistical mechanics is the kinetic energy, which for a single particle is given by

\[
K_i = \left\langle \frac{\vec{p}_i^2}{2m} \right\rangle = \frac{1}{Z_p} \int dp_i^d \frac{P_i^2}{2m} e^{-\vec{p}_i^2/2mk_B T}, \tag{13}
\]

\[
Z_p = \frac{1}{Z_p} \int dp_i^d e^{-\vec{p}_i^2/2mk_B T}, \tag{14}
\]

since all integrals except those over \( \vec{p}_i \) cancel. This gives the equipartition theorem;

\[
K_i = \frac{d}{2} k_B T. \tag{15}
\]

In general it is not possible to analytically calculate expectation value of more complicated functions of the particle momenta or positions, except in one dimension. In a Monte Carlo simulation, real-space expectation values are evaluated by importance sampling of a finite number of the configurations \( \{\vec{x}_i\} \). Before discussing how this is done for particles in continuous space, we will consider the slightly simpler case of lattice models.

### 2.2 Lattice and spin models

In a lattice model the degrees of freedom of the system "live" on the vertices of a lattice; these degrees of freedom can be continuous or discrete. Spin models constitute an important class of lattice models in which the degrees of freedom correspond to magnetic moments of fixed magnitude \( S \) and variable orientation; the energy is a function of the orientation angles. In nature, spin models have direct realizations in crystals of atoms with unpaired electronic spins that are localized at the atomic sites, i.e., in insulators where the spins are not carried by delocalized conduction electrons but can be associated with individual atoms (\( S = 1/2 \) for single unpaired electrons; higher spins can result from Hund’s rule and/or electronic states with non-zero angular momentum). If the spin quantum number \( S \) is relatively large, quantum fluctuations can be neglected to a good approximation and the spins can be described with classical vectors (however, for small spin, \( S = 1/2 \) and 1 in particular, quantum effects do play a big role and classical models may give quantitatively, and often even qualitatively, wrong results). One of the most important models of this kind is the Heisenberg model, where the interaction between spins at sites \( i \) and \( j \) is proportional to their scalar product;

\[
E = \sum_{i,j} J_{ij} \vec{S}_i \cdot \vec{S}_j. \tag{16}
\]
The coupling constants $J_{ij}$ are often restricted to be non-zero only for lattice sites $i,j$ that are nearest neighbors. Here the spin vectors are three dimensional, but anisotropies can lead to effective spin models in which the spin orientations are confined to within a plane, as illustrated in Fig. 1, or along a single axis.

The simplest spin model is the Ising model, in which the spins have only two possible orientations along a chosen axis, "up" or "down". Denoting the degrees of freedom $\sigma_i = \pm 1$, the energy is

$$E = \sum_{i,j} J_{ij} \sigma_i \sigma_j - h \sum_i \sigma_i,$$

(17)

where we have also included an external magnetic field. The interaction $J_{ij}$ is again often (but not always) non-zero only between nearest neighbors. Ising couplings can arise in a system of $S = 1/2$ quantum spins when anisotropies make the interactions in one spin direction dominant, e.g., only $S_i^x S_j^x$ may have to be considered. There is also a plethora of other physical situations that can be mapped onto Ising models with various forms of the interaction $J_{ij}$ and the field $h$ in Eq. (17), e.g., binary alloys (where $\sigma_i$ correspond to the two species of atoms) and atoms adsorbed on surfaces (where $\sigma_i$ correspond to the presence of absence of an atom on a surface).

Considering nearest-neighbor interactions only and zero external field, the energy is

$$E = J \sum_{\langle i,j \rangle} \sigma_i \sigma_j,$$

(18)

where $\langle i, j \rangle$ denotes a pair of nearest-neighbor sites $i,j$. In sums like these one normally counts each interacting spin pair only once, i.e., if the term $\langle i, j \rangle$ is included in the sum, the term $\langle j, i \rangle$ is not. Denoting by $\sigma$ the whole set of spin configurations $\{\sigma_1, \ldots, \sigma_N\}$, where $N$ is the total number of spins in the system, the thermal expectation value of a function $A(\sigma)$ is

$$\langle A \rangle = \frac{1}{Z} \sum_{\sigma} A(\sigma) e^{-E(\sigma)/T},$$

(19)

$$Z = \sum_{\sigma} e^{-E(\sigma)/T}.$$

(20)

For ferromagnetic interactions (i.e., $J < 0$) when $T \to 0$ there are only two contributing spin configurations; those with all spins pointing up or down. For antiferromagnetic interactions ($J > 0$) there
are also two lowest-energy configurations if the lattice is bipartite, i.e., if the system can be subdivided into two sublattices such that all interacting pairs $\langle i, j \rangle$ have one member on each sublattice. For example, on a two-dimensional square lattice the lowest-energy configurations have alternating up and down spins in a checkerboard pattern (the up and down spins form the two sublattices). The two-dimensional ferromagnetic and antiferromagnetic ground states are illustrated in Fig. 2. In this case (and for all other bipartite lattices) there is a simple mapping between the antiferromagnet and the ferromagnet; by performing a spin rotation $\sigma_i \rightarrow -\sigma_i$ on one of the sublattices we effectively achieve $J \rightarrow -J$, and hence all the properties of these two models at any $T/J$ are simply related to each other. This is not the case for a non-bipartite lattice, e.g., a triangular lattice, in which case antiferromagnetic interactions are said to be geometrically frustrated (meaning that the individual energies of the interacting spin pairs cannot all be simultaneously minimized). The lowest-energy states of the antiferromagnet are then non-trivial, while the completely polarized states remain the lowest-energy states of the ferromagnet. Frustrated antiferromagnets are of great interest in current magnetism research, but we will here for simplicity consider mainly ferromagnetic interactions.

In dimensions $d > 1$, the Ising model exhibits a phase transition between a disordered (paramagnetic) state at high temperatures and an ordered (ferromagnetic) state at low temperatures (in one dimension, thermal fluctuations prohibit order at $T > 0$, and the system then exhibits true long-range order only exactly at $T = 0$). The order parameter of this phase transition is the magnetization,

$$m = \frac{1}{N} \sum_{i=1}^{N} \sigma_i.$$  \hfill (21)

Fig. 3 shows the expectation value of the magnetization versus the temperature for an infinitely large ferromagnet, in which the spin-reversal symmetry can be broken, i.e., if the system has been prepared with the spins predominantly in one direction below $T < T_c$ it cannot spontaneously (in a finite time) fluctuate through a series of local spin flips into the phase with the opposite magnetization. In a finite system such fluctuations are possible and then $\langle m \rangle = 0$ for all $T$. In simulations of finite lattices one can instead define the order parameter as $\langle |m| \rangle$ or $\sqrt{\langle m^2 \rangle}$; such magnetization curves will be smooth and non-zero for all $T$. As the system size is increased the

---

1The symmetry breaking can be achieved by applying a weak magnetic field, which is removed once equilibrium has been reached. The system will order also spontaneously, without an external symmetry-breaking field, through fluctuations into a state with excess magnetization in one direction, but then large domains with different magnetizations will typically form.
magnetization sharpens close to the critical temperature and approaches the infinite-size symmetry-
broken \( \langle m(T) \rangle \), which has the form \( \langle m(T) \rangle \sim (T_c - T)^\beta \) for \( T \to T_c \) from below. Here \( \beta \) is an example of a critical exponent. In subsequent sections we will discuss Monte Carlo simulation algorithms for the Ising model and learn how to extract transition temperatures and critical exponents from simulation data. Before that, we will consider the general conditions for importance sampling according to a desired distribution.

3 Importance sampling and detailed balance

We will here consider a discrete configuration space \( \{ C \} = C_1, C_2, \ldots, C_N \) (where \( N \) can be finite or infinite), but the discussion can be directly generalized to a continuum of configurations as well (we will mention an example of this as well). For a system at temperature \( T \), an expectation value is given by

\[
\langle A \rangle = \sum_i P(C_i) A(C_i), \quad P(C_i) = \frac{1}{Z} e^{-E(C_i)/T},
\]

where we work in units such that \( k_B = 1 \) (i.e., we measure energies in degrees Kelvin). In a simulation we start with some arbitrary configuration \( C_{i(0)} \) and from it stochastically generate a sequence \( C_{i(1)}, C_{i(2)}, \ldots, C_{i(M)} \), which we use to approximate various expectation values of interest. Our goal is for the configurations to be distributed according to \( P \).

We use some stochastic process in which a configuration \( C_{i(k+1)} \) is obtained from the previous configuration \( C_{i(k)} \) by making some kind of random change in the latter. We consider a sequence of configurations constituting a Markov chain, i.e., the probability of making a transition from \( C_{i(k)} \) to \( C_{i(k+1)} \) is not dependent on how we arrived at \( C_{i(k)} \) (its history). We will discuss conditions on the transition probabilities \( P(C_i \to C_j) \) for the desired distribution \( P(C) \) to be achieved. It should be noted that \( P \) can be any probability distribution; not necessarily the Bolzmann probability that we are interested in here.

Instead of considering a single sequence of configurations, it is useful to first imagine an ensemble of a large number of configurations. If this ensemble is distributed according to \( P \), then the number \( N_0(C_i) \) of configurations \( C_i \) in the ensemble is proportional to \( P(C_i) \). At a given time (step) we apply some scheme to change (update) the configurations, with the probability of changing \( C_i \) to \( C_j \) denoted \( P(C_i \to C_j) \). The number of configurations \( C_i \) after updating all the configurations is

\[
N_1(C_i) = N_0(C_i) + \sum_{j \neq i} [N_0(C_j)P(C_j \to C_i) - N_0(C_i)P(C_i \to C_j)],
\]

where the two terms for each \( j \) in the sum correspond to the number of configurations that were changed into and out of \( C_i \), respectively. This is called the master equation. If we want the ensemble to remain distributed according to \( P \), we clearly must have, for all \( i = 1, \ldots, N \),

\[
\sum_{j \neq i} [N_0(C_j)P(C_j \to C_i) - N_0(C_i)P(C_i \to C_j)] = 0,
\]

or, since \( N_0(C_i) \propto P(C_i) \),

\[
\sum_{j \neq i} [P(C_j)P(C_j \to C_i) - P(C_i)P(C_i \to C_j)] = 0.
\]
This equation may have many solutions, and in general it would be very difficult to find all solutions. However, we can obtain a particular solution by satisfying the above condition term-by-term;

\[ P(C_j)P(C_j \rightarrow C_i) - P(C_i)P(C_i \rightarrow C_j) = 0, \] (26)

which gives a condition, called *detailed balance*, on the ratio of the transition probabilities;

\[ \frac{P(C_i \rightarrow C_j)}{P(C_j \rightarrow C_i)} = \frac{P(C_j)}{P(C_i)}. \] (27)

In statistical mechanics the configuration probabilities \( P(C_i) \) is given by

\[ P(C_i) = \frac{1}{Z} W(C_i), \quad W(C_i) = e^{-E(C_i)/T}, \] (28)

where \( W(C_i) \) is referred to as the configuration weight. Since the partition function cancels in the ratio on the right hand side of Eq. (27) we can also write

\[ \frac{P(C_i \rightarrow C_j)}{P(C_j \rightarrow C_i)} = \frac{W(C_j)}{W(C_i)}, \] (29)

which is the way the detailed balance condition normally is written.

Although we have derived the detailed balance condition starting from an ensemble of configurations, it is valid for a single Markov chain as well. This statement would clearly be true if the Markov chains formed by the time evolution of all of the individual configurations in the ensemble would have the same distribution over time, in which case they clearly all would be distributed according to \( P \). For this to hold the sampling has to be *ergodic*, i.e., the types of updates made must be such that from an arbitrary configuration any configuration can be reached by a series of updates. Most Monte Carlo simulation schemes are based on the principles of detailed balance and ergodicity.

We have shown that detailed balance maintains the desired distribution of configurations if we start from an ensemble that is already in that distribution. In practice one starts a Markov chain from some arbitrary state, which may be an improbable state of the target distribution. It will then take some time before the generated configurations are correctly distributed, but with detailed balance and ergodicity fulfilled we are guaranteed to reach the correct distribution after some time. This can be seen in the master equation (23), where it is clear that if we have an ensemble with an excess of configurations \( C_i \) (implying an over-all deficit of other configurations), then after one updating step the excess is reduced because when \( N_0(C_i) \) is large there are also more configurations changing out of \( C_i \) than into it. The time needed for *equilibration* depends on the system under study and one should make sure that a sufficient number of updates are carried out before the configurations are used to measure observables.

In a simulation, one typically does not consider all possible transitions \( C_i \rightarrow C_j \), but only a subset of all transitions corresponding to making certain small changes in \( C_i \). For example, for an Ising configuration with \( N \) spins one can consider flipping a randomly selected spin, in which case \( N \) new configurations can be reached. For a system of particles in continuous space, one can consider moving a randomly chosen particle by a displacement vector \( \delta \), with \( \delta \) chosen randomly within a sphere of radius \( \Delta \). These updates are illustrated in Fig. 4; these clearly constitute ergodic processes as we can create any configuration by repeating the update many times.
The transition probability $P(C_i \rightarrow C_j)$ in the examples given above can be written as a product of two probabilities; one for attempting a certain update (selection of the spin to be flipped, or the particle to be moved and the displacement vector $\vec{\delta}$) and one for actually carrying out the change (accepting it). We hence write

$$P(C_i \rightarrow C_j) = P_{\text{attempt}}(C_i \rightarrow C_j)P_{\text{accept}}(C_i \rightarrow C_j).$$

(30)

It is often the case, as it is in the examples mentioned above, that the probability of attempting each of the possible updates is trivially uniform, i.e., $P_{\text{attempt}}(C_i \rightarrow C_j) = \text{constant}$, independent of $i,j$. This part of the transition probability then drops out of the detailed balance condition (29) and we are left with a detailed-balance condition for the acceptance probabilities;

$$P_{\text{accept}}(C_i \rightarrow C_j)P_{\text{accept}}(C_j \rightarrow C_i) = \frac{W(C_j)}{W(C_i)}.$$  

(31)

This condition can be fulfilled in a number of ways, among which the most commonly used is the Metropolis acceptance probability;

$$P_{\text{accept}}(C_i \rightarrow C_j) = \min \left[ \frac{W(C_j)}{W(C_i)}, 1 \right].$$

(32)

In other words, if the new configuration weight is higher (corresponding to lowering the energy of the system) we always accept the update, whereas if it is lower we accept it with a probability equal to the ratio of the new and old weights. It can easily be checked that this Metropolis acceptance probability indeed satisfies the detailed balance condition (31). To determine whether or not to accept the update when $P_{\text{accept}}(C_j) < 1$, the acceptance probability can be compared with a random number $r \in [0,1)$; if $r < P_{\text{accept}}(C_i \rightarrow C_j)$ the update is accepted, and otherwise it is rejected. If an update is rejected, the old configuration $C_i$ should be considered the next configuration in the sequence. The whole procedure of attempting updates and accepting or rejecting them using the above scheme goes under the name of the Metropolis algorithm, after the first author of the paper where this method was first introduced.\(^2\)

Another often used acceptance probability that can be used with the Metropolis algorithm is

\[ P_{\text{accept}}(C_i \rightarrow C_j) = \frac{W(C_j)}{W(C_i) + W(C_j)}, \quad (33) \]

which is a special case of a *heat bath* probability involving selection among a number \( m \) of choices

\[ P_{\text{select}}(C_{j(k)}) = \frac{W(C_{j(k)})}{\sum_{l=1}^{m} W(C_{j(l)})}. \quad (34) \]

Here the current configuration is \( C_i = C_{j(l)} \) for some \( l \in \{1, \ldots, m\} \) and there is no explicitly rejected update, i.e., one of the options \( l = 1, \ldots, m \) is always chosen, according to the above probabilities. Since one of the choices equals the old configuration the update may still lead to no change. This approach is useful, e.g., for lattice models where each lattice site can be in \( m > 2 \) different states. The acceptance probability (33) can be considered a special case of the heat-bath approach when there are just two states to be selected among.

### 4 Metropolis algorithm for the Ising model

It was already indicated above how the Metropolis algorithm works in the case of the Ising model, for which the energy in the presence of a magnetic field is given by Eq. (17); a configuration update amounts to selecting a spin at random and flipping it with probability (32). When updating an Ising configuration; \( C \rightarrow C' \), by flipping any number of spins, the weight ratio \( W(C')/W(C) \) in the acceptance probability is given explicitly by

\[ \frac{W(C')}{W(C)} = \exp \left[ -\frac{J}{T} \sum_{(i,j)} (\sigma'_i \sigma'_j - \sigma_i \sigma_j) + \frac{h}{T} \sum_i (\sigma'_i - \sigma_i) \right], \quad (35) \]

where \( \{\sigma'_i\} \) are the spins of the updated configuration. Flipping a single spin \( j \), we get

\[ \frac{W(C')}{W(C)} = \exp \left[ \frac{2J}{T} \sigma_j \left( \sum_{\delta[j]} \sigma_{\delta[j]} - \frac{h}{J} \right) \right], \quad (36) \]

where \( \delta[j] \) denotes a nearest neighbor of site \( j \) (of which there are \( 2d \) on a \( d \)-dimensional cubic lattice). Since the accept/reject criterion in practice amounts to comparing the above ratio with a random number \( 0 \leq r < 1 \), these ratios can be used directly without taking the minimum with 1, which required in the actual probability (32). In order to avoid repeated time-consuming evaluations of exponential functions, the weight ratios should be precalculated and stored in a table.

It should be pointed out that it is actually not necessary to select the spin to be flipped at random; one can also go through all spins one-by-one. In this case detailed balance is not fulfilled for each step, but with some more effort one can show that the correct distribution is nevertheless obtained. It is likely, however, that the random spin selection makes the simulation less sensitive to flaws in the random number generator (whether or not this statement really holds clearly depends on the random number generator used) and hence this is the preferred way to do it.
4.1 Program implementation

The ising spins \( \sigma_i = \pm 1 \) can be stored in a one-dimensional vector; \( \text{spin}[1:n] \). For a simple cubic lattice with dimensionality \( d = 1, 2, \) or \( 3 \), and the number of spins \( N = L^d \), the correspondence between the index \( i \) and the coordinates of the lattice sites can then be conveniently chosen as \( x=i \) for \( d = 1 \), \( x=\text{mod}(i,1), y=\text{div}(i,1) \) for \( d = 2 \), and \( x=\text{mod}(i,1), y=\text{mod}(i,1^2)/1, z=\text{div}(i,1^2) \) for \( d = 3 \). The neighbors \( \delta[i] \) can be easily calculated on the run using these coordinates. For more complicated non-cubic lattices it may be better to precalculate the neighbors and store them in an array.

It is useful to define a size-normalized "time" unit of a simulation, so that the probability of carrying out an update of a spin during a time unit is independent of the system size. Hence we define a Monte Carlo step as \( N \) attempts to flip randomly selected spins.

As an example we here consider the two-dimensional ferromagnetic Ising model with \( n=lx*ly \) spins and \( J = -1 \). We discuss the main elements of the program 'ising2d.f90' which is available on the course web site. The spins \( \text{spin}[s] \) can be initially set to arbitrary values \( \pm 1 \), e.g., chosen at random. The precalculated weight ratios (36) are stored in a matrix \( \text{pflip}[s0,ss] \) where the indexes \( s0 \) and \( ss \) correspond to the values of \( \sigma_j = \pm 1 \) and \( \sum \delta[j] \sigma[\delta[j]] = -4, -2, 0, 2, 4 \), respectively (since we work on a periodic lattice in which each spin has four neighbors, the sum is always even).

We use a simple array (which has more elements than needed); with \( S0=-1,0,1 \) and \( ss=-4,-3,\ldots,4 \). With the temperature stored in a floating-point variable \( \text{temp} \) the probability matrix can be constructed by

```plaintext
for ss=-4:2:4
    pflip[-1,ss]=exp(+((ss+h)*2/temp))
    pflip[+1,ss]=exp(-(ss+h)*2/temp)
end
```

Using the Julia function \( \text{rand()} \) with two methods to generate either integer random numbers between 1 and \( n \) or floating-point values between 0 and 1 , a Monte Carlo step can be carried out as follows;

```plaintext
for i=1:n
    s=rand(1:n)
    x=mod(s-1,lx)
    y=div(s-1,lx)
    s1=spin[1+mod(x+1,lx)+y*lx]
    s2=spin[1+mod(x+1,ly)*lx]
    s3=spin[1+mod(x-1,lx)+y*lx]
    s4=spin[1+mod(y-1,ly)*lx]
    if \( \text{rand()}<\text{pflip[spin[s],s1+s2+s3+s4]} \)
        spin[s]=-spin[s]
    end
end
```

A full simulation consists of a number of equilibration steps and a much larger number of steps.
after each of which measurements of physical quantities of interest are carried out (however, in some cases it may not be statistically worthwhile to measure after each step, as we will discuss further below in connection with autocorrelation functions). The main part of the simulation should be subdivided into bins, for which separate averages are calculated that can be subsequently used for calculating statistical errors (in the same way as we discussed in the chapter on Monte Carlo integration). The statistical analysis can be carried out using a separate program; hence all the bin averages are stored on disk (this is useful if one later wants to increase the number of bins without having to redo the bins already completed). The main part of a simulation program can then look like this:

```plaintext
for i=1:binsteps
    mcstep()
end
for j=1:bins
    resetdatasums()
    for i=1:binsteps
        mcstep()
        measure()
    end
    writebindata()
end
```

where equilibration is done for a number of steps corresponding to one bin (which may or may not be sufficient) and the subroutine names are self-explanatory (with the exception perhaps of `resetdatasums()`, which sets to zero all the variables used to accumulate measurements). No arguments have been indicated for the functions, as those will depend on exactly how the program is written.

As a general rule, the number of bins should be at least 10 in order to calculate the statistical errors reliably. In order not to generate unnecessarily large data files, one should adapt the number of steps per bin so that the number of bins required in order to reach a satisfactory statistical error is not too large (how many bins are too many is somewhat a matter of taste, but to have more than 1000 bins would generally be considered excessive, unless one has some specific reason for this, e.g. to construct high-precision histograms of bin averages).

We discussed data binning also in the context of Monte Carlo integration, where its purpose was to obtain data following a normal distribution (which is the case in the limit of a large number samples per bin). In Monte Carlo simulations based on a Markov chain, where by construction the configurations are time-correlated, an additional purpose of binning is to achieve statistically independent data; this will also be the case if the bins are sufficiently long; a statement which can be made more precise in terms of autocorrelation functions, which we will discussed below in Sec. ??.
4.2 Evaluating physical observables

A quantity of natural interest in the context of the ferromagnetic Ising model is the magnetization, which is the order parameter of the phase transition occurring at a temperature $T_c > 0$ in two and three dimensions. We denote by $M$ the full magnetization and by $m$ the corresponding size-normalized quantity:

$$M = \sum_{i=1}^{N} \sigma_i, \quad m = \frac{M}{N}. \quad (37)$$

As we discussed in Sec. 2.2, on a finite lattice the spin-reversal symmetry is not broken in a simulation running for a long time and hence $\langle m \rangle = 0$ (although in practice the time to “tunnel” between states with positive and negative magnetization becomes very long in Metropolis simulations of large systems when $T < T_c$; in practice one will then “measure” (a term often used on the field for accumulating expectation values in simulations) $\langle m \rangle \neq 0$). One can instead measure $\langle |m| \rangle$ or $\sqrt{\langle m^2 \rangle}$, since in the thermodynamic limit the become equal to the symmetry-broken $\langle m \rangle$. Another quantity of great interest is the magnetic susceptibility, defined as

$$\chi = \frac{d\langle m \rangle}{dh}, \quad (38)$$

i.e., the linear response of the magnetization to a uniform magnetic field. In the thermodynamic limit, the susceptibility of a system in zero field ($h = 0$) diverges at the ferromagnetic transition.

We can take the derivative in the statistical expression for $\langle m \rangle$ to obtain an estimator for $\chi$. Writing the energy (17) as $E = E_0 - hM$, we have

$$\langle m \rangle = \frac{1}{Z} \sum S m e^{-(E_0-hM)/T}, \quad Z = \sum S e^{-(E_0-hM)/T}. \quad (39)$$

The susceptibility (38) is then

$$\chi = -\frac{dZ/dh}{Z^2} \sum S m e^{-(E_0-hM)/T} + \frac{1}{Z T} \sum S m M e^{-(E_0-hM)/T}, \quad (40)$$

and with

$$\frac{dZ}{dh} = \frac{1}{T} \sum S M e^{-(E_0-hM)/T}, \quad (41)$$

we obtain

$$\chi = \frac{1}{N T} \left( \langle M^2 \rangle - \langle M \rangle^2 \right). \quad (42)$$

This expression can be used for $h \neq 0$ at any $T$. In the zero field case $\langle m \rangle = 0$ for $T \geq T_c$ and thus

$$\chi = \frac{1}{N T} \langle M^2 \rangle, \quad (h = 0, T \geq T_c). \quad (43)$$

For $T < T_c$ we can again substitute $|M|$ for $M$ in (42) and use

$$\chi = \frac{1}{N T} \left( \langle M^2 \rangle - \langle |M| \rangle^2 \right). \quad (44)$$

This formula can also be used for $T > T_c$, since there $\langle |M| \rangle \to 0$ as $N \to \infty$, and hence the two expressions (43) and (44) are equivalent in the thermodynamic limit.
The specific heat can be derived in a similar way;

\[ C = \frac{1}{N} \frac{dE}{dT} = \frac{1}{N} \frac{d}{dT} \sum_C \frac{E(C) e^{-E(C)/T}}{Z} = \frac{1}{NT} \left( \langle E^2 \rangle - \langle E \rangle^2 \right). \] (45)

Like the susceptibility, the specific heat is singular at the ferromagnetic phase transition. Another quantity of interest is the spin-spin correlation function

\[ C(\vec{r}_j - \vec{r}_i) = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle^2, \] (46)

which can be averaged over pairs of sites with the same separation $|\vec{r}| = r \rightarrow \infty$; $C(\vec{r}) \sim e^{-r/\xi}$, where $\xi$ is the correlation length. As $T \rightarrow T_c$ the correlation length diverges, and exactly at $T_c$ the correlation function decays to zero as a power-law. We will discuss these issues further in Sec. ??.

In the example program ising2d.f90 the energy and the magnetization, as well as their squares, are measured, and their bin averages are stored in a file on disk for later processing. This is the main part of the measurement routine:

```fortran
  e=0
  for y=0:ly
    for x=0:lx
      s=1+x+y*lx
      e=e+spin[s]*(spin[1+mod(x+1,lx)+y*lx]+spin[1+x+mod(y+1,ly)*lx])
    end
  end
  m=sum(spin)
  enrg1=enrg1+e
  enrg2=enrg2+e**2
  magn1=magn1+m
  magn2=magn2+m**2
```

After each completed bin, the averages $\langle E \rangle/N$, $\langle E^2 \rangle/N$, $\langle |M| \rangle/N$, and $\langle M^2 \rangle/N^2$ are written to the file bindata.dat. This file should be processed by the program average.f90, which calculates the final averages and statistical errors. The susceptibility and the specific heat are computed based on bin averages according to Eqs. (44) and (45).