Chapter 6

Classical Monte Carlo

In this chapter, we discuss one of the most important simulation methods in physics. Monte Carlo (MC) simulations are used for various purposes, e.g. to calculate material properties, to determine the behavior of systems close to phase transitions or in non-equilibrium settings, and to study the properties of QCD, providing important data for the analysis of e.g. the experiments to be performed at CERN's LHC. Moreover, MC simulations are also an important tool outside physics, e.g. for protein folding studies, the analysis of finance markets, or traffic simulations. Hence, we will discuss this method here in some detail, in particular paying attention to the estimation of the statistical errors of a MC simulation. We also discuss the problems and recent improvements of MC simulations.

In this chapter, we focus on applications of MC simulations to classical statistical physics, that is for systems with many (interacting) classical degrees of freedom. In a later chapter, we discuss MC methods for the simulation of quantum many-body systems. A useful guide for the application of MC methods in statistical physics is provided by the book "A Guide to Monte Carlo Simulations in Statistical Physics", by D. P. Landau and K. Binder.

As an example of a classical many-body system, take the monoatomic gas of Ch. 4 from the discussion of the MD method. A specific configuration of this system, i.e. a certain microstate, is specified by the positions $\vec{x}_i$, and velocities $\vec{v}_i$ of all the particles $i = 0, ..., N-1$. We call the set of all possible configurations of a system its configuration space, which in the current case equals phase space,

$$\Omega = \{C | C = (\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_{N-1}, \vec{v}_0, \vec{v}_1, \ldots, \vec{v}_{N-1})\}. \tag{6.1}$$

As discussed in Ch. 4, in statistical physics, one specifies a statistical ensemble in terms of a distribution function $P(C)$, that assigns to each configuration C the probability of finding the system (in fact an ensemble of such systems) in the configuration C. Recall for example the canonical ensemble, which applies to a system in equilibrium contact with a heat-bath of temperature T, where

$$P(C) \propto e^{-E(C)/(k_B T)} \tag{6.2}$$

depends on the energy $E(C)$ of the configuration C. The goal of the statistical description is to obtain macroscopic quantities, e.g. the system's energy, via calculating averages over the microscopic configurations, so that for a physical observable A, that takes on a value $A(C)$ in configuration C, the macroscopic expectation value is given as

$$\langle A \rangle = \frac{\int A(C)P(C)dC}{\int P(C)dC} \tag{6.3}$$

in terms of an integral over configuration space. Later in this chapter, we will also introduce statistical models, where the configuration space is discrete, and the above integral is replaced by a sum over all configurations. In both cases, relating the microscopic description to measurable macroscopic observables thus requires the calculation of high-dimensional integrals (or large sums). The Monte Carlo method is a very efficient tool to perform this task. Before developing the Monte Carlo formalism within the context of statistical physics problems, we first introduce a Monte Carlo method for the calculation of integrals, very similar to the idea behind the shooting of π, discussed in Sec. 5.1.

6.1 Monte Carlo Integration (Simple Sampling)

Monte Carlo integration is a method for using random numbers in order to calculate an integral. Consider first a one-dimensional integral,

$$I = \int_a^b dx f(x) = (b-a)\langle f \rangle, \tag{6.4}$$

where $\langle f \rangle$ is the average of the function $f(x)$ over the interval $[a, b]$. Instead of numerically calculating the integral using e.g. the methods discussed in Ch. 1, we now aim at obtaining a good statistical estimate for $\langle f \rangle$, from which I is then easily obtained. For this purpose, we generate a large number N of random points x_i inside the interval $x_i \in [a, b]$, using a uniform distribution. Then, the arithmetic mean

$$\overline{f} = \frac{1}{N}\sum_{i=1}^{N} f(x_i), \quad x_i \in [a, b] \tag{6.5}$$

over the N function values $f(x_i)$, is taken as an approximation for $\langle f \rangle$. As we will discuss below, indeed $\overline{f} \to \langle f \rangle$ for $N \to \infty$, with an expected deviation, the statistical error, that decreases $\propto \frac{1}{\sqrt{N}}$ upon increasing the number of points N. Indeed, the inverse square-root dependence of the statistical error in MC integration also holds for higher, d-dimensional integrals,

$$\langle f \rangle = \frac{1}{V}\int_V f(\vec{x}) d^d x \approx \overline{f} = \frac{1}{N}\sum_{i=1}^{N} f(\vec{x}_i), \tag{6.6}$$

where V denotes the integration volume.

Let us compare the $O(1/\sqrt{N})$ behavior of the MC integration error to discretization errors of the integration schemes discussed in Ch. 1. E.g., using the Simpson rule with a uniform mesh of N point (spacing h), the error in the one-dimensional case decreases $\propto h^4 \propto 1/N^4$. This appears to be more efficient than the one-dimensional MC integration, given a fixed number of function evaluations N. However, upon going to higher dimensions, the MC method still has an $O(1/\sqrt{N})$ error, and will eventually scale better than a higher-dimensional variant of the Simpson method. To assess the discretization error of a d-dimensional Simpson rule, consider a d-dimensional mesh of N points. Then, in each lattice direction, the lattice spacing scales $\propto 1/N^{1/d}$, so that we obtain an error of the integral that scales $\propto 1/N^{4/d}$. Hence, for $d \geq 9$, the MC method shows a better scaling than the Simpson rule. Similar arguments apply also to other integration schemes based on periodic meshes. This considerations illustrate, that MC integration will be particularly useful to calculate high-dimensional integrals.

Let us now discuss the nature and scaling of the MC integration error quoted above. For a finite sequence of random numbers, the outcome of the MC integration will obviously depend on the precise random number sequence, via the corresponding sequence of points $\{x_i\}$, at which the function f is evaluated. In particular, imagine that we repeat the calculations of $\langle f \rangle$ a number of M times, each time with a different (independent) sequence of random numbers, of length N. This way, we obtain M different approximations $\bar{f}_i, i = 1, ..., M$, which will fluctuate around the true average $\langle f \rangle$. Denoting the N points used in the i-th trial by $\vec{x}_{ij}, j = 1, ..., N$,

$$\bar{f}_i = \frac{1}{N} \sum_{j=1}^{N} f_{ij}, \quad i = 1, ..., M, \tag{6.7}$$

where $f_{ij} = f(\vec{x}_{ij})$. How are the $\bar{f}_i$ distributed around $\langle f \rangle$?

In the limit of a large number of points N, the central limit theorem of mathematical statistics states that the $\bar{f}_i$ are distributed according to a Gaussian distribution

$$P(\bar{f}_i = Y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(Y-\mu)^2}{2\sigma^2}}, \tag{6.8}$$

around the mean value $\mu = \langle f \rangle$, with variance

$$\sigma^2 = \frac{1}{N}\left(\langle f^2 \rangle - \langle f \rangle^2\right) = \frac{1}{N}\langle (f - \langle f \rangle)^2 \rangle, \tag{6.9}$$

specifying the width (the distance between the two turning points) of the Gaussian distribution as 2σ, that decreases upon increasing the number of points N, $\sigma \propto 1/\sqrt{N}$. Furthermore, to estimate μ and σ^2 from the $\bar{f}_i$, one can show that the following expressions are appropriate,

$$\overline{f} \;=\; \frac{1}{M} \sum_{i=1}^{M} \overline{f}_i, \quad \text{for} \quad \mu \tag{6.10}$$

$$\begin{aligned} S^2 &= \frac{1}{M-1}\sum_{i=1}^{M}(\overline{f}_i - \overline{f})^2, \quad \text{for} \quad \sigma^2 \\ &= \frac{1}{M-1}\sum_{i=1}^{M}(\overline{f}_i^{\,2} - 2\overline{f}_i\overline{f} + \overline{f}^{\,2}) \\ &= \frac{1}{M-1}\left[\sum_i \overline{f}_i^{\,2} - 2\overline{f}\sum_i \overline{f}_i + M\overline{f}^{\,2}\right] \\ &= \frac{1}{M-1}\left[\sum_i \overline{f}_i^{\,2} - M\overline{f}^{\,2}\right] \end{aligned} \tag{6.11}$$

in the sense that their expectation values equal μ and σ^2, respectively.

Instead of proofing here the central limit theorem, let us consider an example, revisiting therefore the shooting of π, this time put explicitly as an MC integration problem. The procedure in Sec. 5.1 corresponds to performing the two-dimensional integral

$$\frac{\pi}{4} = \int_{[0,1]^2} f(\vec{x}) d^2x, \tag{6.12}$$

over the unit square with the function $f(x_1, x_2) = \theta(1-(x_1^2+x_2^2))$, so that $f(x_1, x_2) = 1$, if $x_1^2 + x_2^2 \leq 1$, and $f(x_1, x_2) = 0$, if $x_1^2 + x_2^2 > 1$. Performing the MC integration procedure $M = 1000$ times, each time using $N = 8000$ points, uniformly distributed over the unit square, one finds a distribution of estimates $\overline{f}_i, i = 1, ..., M$ for $\pi/4$, what compares well to the Gaussian distribution with parameters taken from Eq. (6.10), and Eq. (6.11) as shown by the histogram in Fig. 6.1.

The standard deviation has a defined meaning in case the averages $\overline{f}_i$ obey a Gaussian distribution: the probability that one of the

$$\begin{aligned} \overline{f}_i &\in [\mu - \sigma, \mu + \sigma] \quad \text{is approximately } 67\%, \\ \overline{f}_i &\in [\mu - 2\sigma, \mu + 2\sigma] \quad \text{is approximately } 95\%. \end{aligned} \tag{6.13}$$

In this sense, σ denotes the statistical error, and provides an error bar on the $\overline{f}_i$, which scales as

$$\Delta\overline{f}_i = \sigma = \frac{1}{\sqrt{N}}\sqrt{\langle f^2\rangle - \langle f\rangle^2} \propto \frac{1}{\sqrt{N}}, \tag{6.14}$$

thus providing the $O(1/\sqrt{N})$ scaling of the statistical error in MC integration. Instead of performing M independent measurements to estimate the statistical error using Eq. (6.11), one can also approximate the statistical error based on a single trial as

$$\Delta\overline{f}_i \approx \frac{1}{\sqrt{N}}\sqrt{\frac{1}{N}\sum_{j=1}^{N} f_{ij}^2 - \left(\frac{1}{N}\sum_{j=1}^{N} f_{ij}\right)^2}, \quad i = 1, ..., M, \tag{6.15}$$

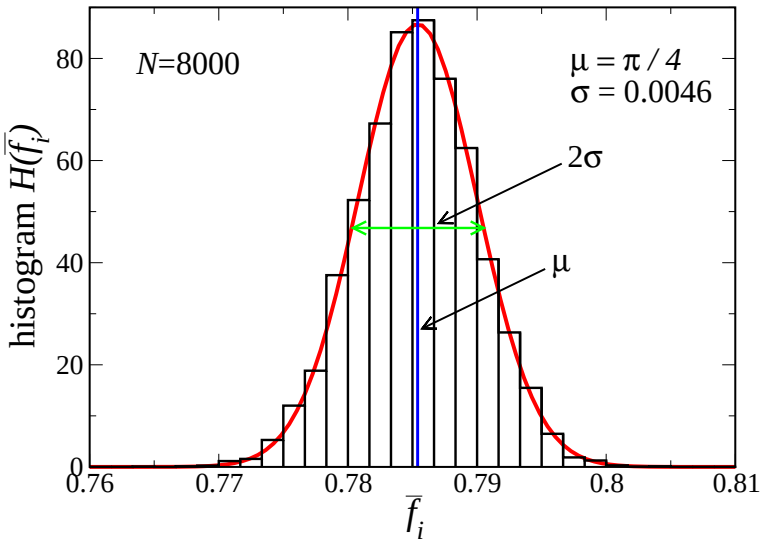

Figure 6.1: Distribution of $M = 1000$ trials on shooting π using $N = 8000$ shots in each trial. The red curve denotes a Gaussian distribution centered at $\mu = \pi/4$ of width 2σ, where $\sigma = 0.0046$.

by approximating the two averages in Eq. (6.14) based on the available data. If several trials M have been performed, then the value of $\bar{f}$ calculated from Eq. (6.10) has a smaller statistical error than each single value of $\bar{f}_i$, namely, according to the central limit theorem,

$$\Delta\bar{f} = \frac{\sigma}{\sqrt{M}}, \text{ estimated by } \frac{S}{\sqrt{M}}, \tag{6.16}$$

which again exhibits the characteristic inverse square-root dependence of the statistical MC error.

6.2 Importance Sampling

In the simple sampling MC integration scheme, one uses uniformly distributed random points to sample a function f over an integration domain in order to estimate its mean value. Such a procedure performs badly (in the sense that the statistical error will be large), in case the function f has its dominating values in only a small subregion of the integration domain, because in that case only a small fraction of the uniformly distributed random points fall inside that subregion. The function f is not sampled efficiently, and thus many more points are needed in order to arrive at a low statistical error. More formally, using Eq. (6.14), we see that for a function f showing larger differences between $\langle f^2\rangle$ and $\langle f\rangle^2$, a larger number of points N are needed in order to assure a given accuracy.
In such a case, one would prefer to generate random points inside the relevant subdomain more frequently, and take that bias into account in the final averaging. This

idea underlies the concept of importance sampling, which we develop in this section.

To give an example, consider the following modification of the shooting of π problem: Assume we want to integrate the function

$$f(x_1, x_2) = (1 - x_1)^2(1 - x_2)^2\theta(1 - (x_1^2 + x_2^2)), \quad x_1, x_2 \in [0, 1]^2 \tag{6.17}$$

on the unit square. This function has larger function values close to the origin than further out on the unit circle. Simple sampling, where

$$I = \int_V f(x)d^2x \approx \frac{V}{N}\sum_{i=1}^{N} f(\vec{x}_i), \tag{6.18}$$

will not sample the function according to the important regions. It would be better to increase the number of points towards the origin, which is what importance sampling does. But, which distribution of random points should we use, and how do we construct such a distribution?

In order to specify the importance sampling strategy further, we now relax uniformity of the random points, and instead consider them to be distributed according to a possibly non-uniform distribution $W(\vec{x})$, where $W(\vec{x}) > 0$ on the support of f, and $\int W(\vec{x})d^dx = 1$ (normalization). Then, the integral I can be expressed as

$$I = \int_V f(\vec{x})\, d^dx = \int_V \frac{f(\vec{x})}{W(\vec{x})}W(\vec{x})\, d^dx := \left\langle \frac{f}{W} \right\rangle_W, \tag{6.19}$$

where the subscript W denotes averaging with respect to the W-distribution measure. This is approximated as

$$I \approx \frac{1}{N}\sum_{i=1}^{N} \frac{f(\vec{x}_i)}{W(\vec{x}_i)}, \tag{6.20}$$

where the $\vec{x}_i$ are generated according to the distribution W (this reduces to Eq. (6.18) in case of a uniform distribution on V, $W = 1/V$). The variance on the integral in this case becomes

$$\begin{aligned}
\sigma_W^2 &= \frac{1}{N}\left(\left\langle \left(\frac{f}{W}\right)^2 \right\rangle_W - \left\langle \frac{f}{W} \right\rangle_W^2\right) \\
&= \frac{1}{N}\left(\int \left(\frac{f}{W}\right)^2 W d^dx - \left(\int \frac{f}{W} W d^dx\right)^2\right) \\
&= \frac{1}{N}\left(\int \frac{f^2}{W} d^dx - \left(\int f d^dx\right)^2\right).
\end{aligned} \tag{6.21}$$

We can find the optimal W by minimizing σ_W^2 subject to the constraint of normalization by a functional variation

$$0 = \frac{\delta}{\delta W}\left(\int \frac{f^2}{W} d^d x - \left(\int f d^d x\right)^2 + \lambda \int W d^d x\right), \tag{6.22}$$

with a Lagrange multiplier λ. As the middle term does not depend on W, the variation gives

$$0 = -\frac{f^2}{W^2} + \lambda \tag{6.23}$$

$$\Longrightarrow \qquad W(\vec{x}) = \frac{|f(\vec{x})|}{\sqrt{\lambda}} = \frac{|f(\vec{x})|}{\int |f(\vec{x}\,')|\, d^d x'}, \tag{6.24}$$

where λ has been chosen to enforce normalization. In case $f \geq 0$, this choice will in fact result in a vanishing variance, as now $f(\vec{x})/W(\vec{x}) = I$ is constant. Implementing this optimal choice for W apparently requires knowledge of the integral of $|f|$, which usually is as difficult as calculating I. So, we would basically need to know the integral in order to generate optimally distributed random numbers for an optimal importance sampling of the integral. While this optimal choice for W is thus of little practical use, one can nevertheless make use of the importance sampling idea. Namely, in case we can separate

$$f(\vec{x}) = \underbrace{P(\vec{x})}_{\text{peaked}} \times \underbrace{A(\vec{x})}_{\text{smooth}} \tag{6.25}$$

into a product of a positive definite function $P(\vec{x})$, which to a large extend follows the peaks and dominating structures in f, and a less so varying part $A(\vec{x})$. Assuming that it is possible to implement a random number generator that provides random points distributed according to $P(\vec{x})$ (we can also assume that P is normalized), we use $W = P$ in Eq. (6.20), and then approximate

$$I \approx \frac{1}{N}\sum_{i=1}^{N} A(\vec{x}_i) = \langle A \rangle_P \tag{6.26}$$

based on random points distributed according to $W = P$, with variance

$$\sigma_W^2 = \sigma_P^2 = \frac{1}{N}\left(\langle A^2 \rangle_P - \langle A \rangle_P^2\right). \tag{6.27}$$

In our example, we could use

$$P(x_1, x_2) = 9(1-x_1)^2(1-x_2)^2 = P_s(x_1)P_s(x_2), \quad P_s(x_i) = 3(1-x_i)^2 \tag{6.28}$$

which follows the increase of f near the origin, and is normalized,

$$\int P(x_1, x_2) d^2 x = 1 \tag{6.29}$$

on the unit square. Furthermore, using the transformation method, $P_s(x_i)$ can be implemented based on a uniform random number generator. With this choice of P, we then average

$$A(x_1, x_2) = \frac{1}{9}\,\theta(1 - (x^2 + y^2)) \tag{6.30}$$

in the importance sampling scheme.

Other improved sampling strategies for high-dimensional integrals, such as stratified sampling and adaptive MC (the VEGAS algorithm) are discussed in Ch. 7.8 of Numerical Recipes. Here, we move on to applications of MC integration in statistical physics.

In statistical mechanics, one needs to calculate high-dimensional integrals over configuration space,

$$\langle A \rangle = \frac{\int A(C)P(C)dC}{\int P(C)dC} \tag{6.31}$$

which in case of the canonical ensemble are dominated by the Boltzmann distribution function

$$P(C) \propto e^{-E(C)/(k_B T)}, \tag{6.32}$$

with $E(C)$ being the energy of the system at given configuration C. To apply the idea of importance sampling to such configuration space integrals, we should thus generate configurations C distributed according to $P(C) \propto e^{-E(C)/(k_B T)}$. A possible way to generate a sequence of configurations according to such a distribution is to use a Markov chain, as we will discuss in Sec. 6.4. Before doing so, we introduce one of the most important models of statistical physics: the Ising model. Then, we discuss importance sampling via Markov chains for this model.

6.3 Ising Model

The Ising model is one of the most important models of classical statistical physics. It is a simple model of a magnetic system, and shows a second order phase transition, as we will discuss later on. The Ising model describes coupled magnetic moments ("spins") on a lattice. At each lattice site $i = 0, ..., N_s - 1$, a binary variable $\sigma_i = \pm 1$ is defined, corresponding to two possible orientations of the magnetic moment at site i: "up" ($\sigma_i = 1$) or "down" ($\sigma_i = -1$). The energy for a given configuration

$$\sigma = (\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{N_s - 1})\,, \tag{6.33}$$

of the σ_i is defined as

$$E(\sigma) = \sum_{i<j} J_{ij}\sigma_i\sigma_j \tag{6.34}$$

Here, the J_{ij} describe couplings between the spins. If a coupling J_{ij} is positive (negative), the system can lower its energy by making the two spins opposite (equal, or parallel). We will discuss this in more detail, below.
The Ising model has also been considered as an effective model for other, non-magnetic systems, e.g. in a binary alloy, $\sigma_i = +1$ might correspond to the site i being occupied by an atom of type A, and $\sigma_i = -1$ to an atom of type B. Or, as a model of particle absorption on a substrate, the σ_i correspond to the presence $(+1)$ or absence (-1) of a particle at lattice site i.

Here, we take the Ising model as a simple model of a magnetic material. Considering only uniform nearest-neighbor interactions, the energy is

$$E(\sigma) = J \sum_{<i,j>} \sigma_i \sigma_j, \tag{6.35}$$

where the notation $< i, j >$ means that each nearest neighbor bond is taken only once in the summation. We consider the system coupled to a heat-bath at a finite temperature T. In this case, the statistical weight of a configuration is given by the Boltzmann weight

$$P(\sigma) \propto e^{-\beta E(\sigma)}, \tag{6.36}$$

where $\beta = 1/(k_B T)$ denotes the inverse temperature. The partition function of the Ising model

$$Z = \sum_{\{\sigma\}} e^{-\beta E(\sigma)}, \tag{6.37}$$

involves a discrete sum over the configuration space, which in this case is the discrete set of 2^{N_s} Ising spin configurations. Using the partition function,

$$P(\sigma) = \frac{1}{Z} e^{-\beta E(\sigma)}, \tag{6.38}$$

is seen to be normalized,

$$\sum_{\{\sigma\}} P(\sigma) = 1. \tag{6.39}$$

The thermal expectation value of an observable, that takes the value $A(\sigma)$ for the system in configuration σ, (e.g. the energy $E(\sigma)$) is

$$\langle A \rangle = \sum_{\{\sigma\}} P(\sigma) A(\sigma) = \frac{1}{Z} \sum_{\{\sigma\}} A(\sigma) e^{-\beta E(\sigma)}. \tag{6.40}$$

We will discuss other observables of the Ising model below. Let us first try to understand the behavior of the Ising model in two limits of temperature: (i) $T \to 0$, and (ii) $T \to \infty$.

(i) At low temperatures, i.e. for $\beta \to \infty$, those configurations will dominate the Boltzmann distribution, that have the lowest energy. The nature of these ground state configurations of the Ising model depends on the sign of J.

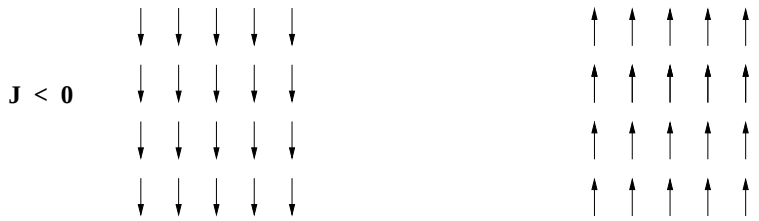

Figure 6.2: Ground state configurations of the Ising model on a square lattice with ferromagnetic nearest-neighbor interactions.

We first consider $J < 0$: In this case, the energy contribution from a given bond is minimized, if the two neighboring spins point in the same direction (such that $\sigma_i \sigma_j = 1$). Therefore, in the ground state all spins point in the same direction, either all up, or all down. The ground state is thus two-fold degenerate. This fully polarized, ferromagnetic arrangement is shown in Fig. 6.2 for a square lattice geometry.

In case $J > 0$, the contribution of a given bond to the total energy will be lowest, if the two neighboring spins point in opposite directions (such that $\sigma_i \sigma_j = -1$). Such an alternating arrangement of neighboring spins can be realized globally, only if the lattice is bipartite: meaning that it can be split into two sub-lattices, such that only sites from different sub-lattices are coupled by a bond. For example, the square lattice is bipartite (think of a checkerboard) and thus this antiferromagnetic arrangement of spins in what is also called a Néel state becomes possible for the Ising model on the square lattice. The two-fold degenerate ground state of the Ising antiferromagnet on the square lattice is illustrated in Fig. 6.3. Maybe more interesting is the case of a non-bipartite lattice, such as a triangular lattice. For simplicity, consider only a single triangle. In this case, the three spins cannot be arranged such that all three bonds contribute a minimal value to the total energy. There will be at least one bond, for which the energy contribution is $+J$. In a sense, the system is frustrated, as not all energy contributions can be minimized. In such geometrically frustrated systems, the nature of the ground state is more complicated, and in fact understanding the properties of complex, frustrated antiferromagnets is a fascinating, active area of current solid state physics research.

In both the ferromagnetic ($J < 0$) and the antiferromagnetic ($J > 0$) case on a bipartite lattice, the Ising model thus exhibits a magnetically ordered ground state. We focus here on the ferromagnetic case, $J < 0$, and typically have a square lattice geometry in mind. Note, that in this ferromagnetic case, the fully polarized states form the ground state even on a non-bipartite lattice. Furthermore, on a bipartite lattice, the ferromagnetic and the antiferromagnetic case can be mapped onto each other, and are thus equivalent.

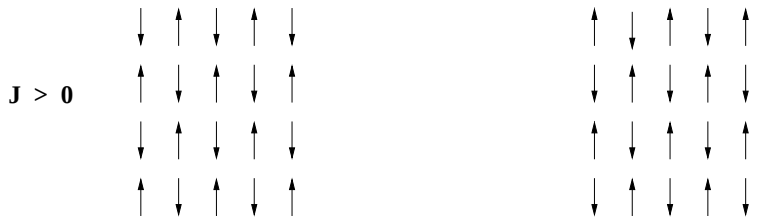

Figure 6.3: Ground state configurations of the Ising model on a square lattice with antiferromagnetic nearest-neighbor interactions.

(ii) Next, we consider the high temperature regime. For $T \to \infty$, $\beta \to 0$, and

$$P(\sigma) \to \frac{1}{Z} = \frac{1}{2^N}, \tag{6.41}$$

so that all configurations contribute equally well to thermal averages. For large temperatures, the magnetic order is destroyed, and the system is magnetically disordered, i.e. in a paramagnetic state.

From the above considerations we conclude, that there must be a transition from the magnetically ordered low temperature to the magnetically disordered, high temperature phase. In fact, in case the spatial dimension $d > 1$, there is a second order phase transition in the Ising model at a finite critical temperature $T_c > 0$. For $T < T_c$ the system is magnetically ordered, with the majority of the spins pointing in the same direction. How can we quantify this order below T_c? Based on the ideas of Ginzburg and Landau, one uses an order parameter for this purpose. In case of the Ising model, the order parameter is given by the expectation value $\langle |m| \rangle$ of the magnetization

$$m = \frac{1}{N_s} \sum_{i=1}^{N_s} \sigma_i. \tag{6.42}$$

For the thermodynamic limit (TDL), i.e. for $N_s \to \infty$, the behavior of the order parameter $\langle |m| \rangle$ is sketched in Fig. 6.4.

In the ordered phase, for $T < T_c$, the system exhibits spontaneous symmetry breaking: the energy functional $E(\sigma)$, defining the model, is invariant under the global transformation $\sigma \to -\sigma$ (i.e. for all i, $\sigma_i \to -\sigma_i$). This transformation, together with the identity mapping $\sigma \to \sigma$, forms the $\mathbb{Z}_2$ symmetry group of the Ising model. However, for a configuration in the ordered phase, this $\mathbb{Z}_2$ symmetry is broken, as the system picks a preference for the majority spin.

At high temperatures, the symmetry is restored, as thermal fluctuations destroy the magnetic order for $T > T_c$. The case $d = 1$ is special, in the sense that only the ground state at $T = 0$ is ordered, and any finite thermal fluctuations immediately destroy the magnetic order. Thus, for $d = 1$ the critical temperature $T_c = 0$.

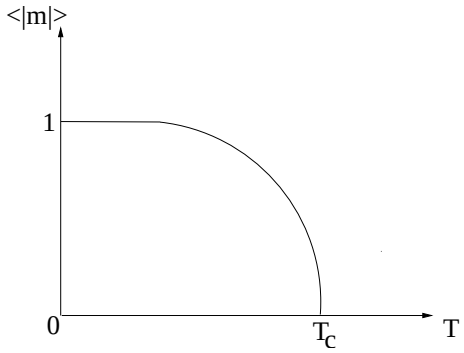

Figure 6.4: Schematic temperature dependence of the order parameter of the Ising model for $d > 1$ in the TDL.

In MC simulations on a computer, one usually considers a finite system, for which spontaneous symmetry breaking does not occur, as the fluctuations can turn the majority spins from "up" to "down" in a finite time on a finite system, and only in the TDL this time scale diverges. In fact, it is easy to show that on a finite system $\langle m \rangle = 0$.
Furthermore, on a finite system, as those accessible by MC simulations, the order parameter does not exhibit a sharp transition as in Fig. 6.4, but instead exhibits a smooth behavior, which sharpens when the system size is increased. In the TDL, the order parameter shows a characteristic behavior

$$\langle |m| \rangle \propto (T_c - T)^{\beta}, \quad T \to T_c^-, \tag{6.43}$$

where β is an example of a critical exponent. These describe characteristic properties of the phase transition, as we will discuss later in this chapter.
For example, in case of the Ising model on the square lattice, for which an exact solution was found by Onsager, one finds

$$\beta = \frac{1}{8}, \tag{6.44}$$

$$k_B T_c = \frac{2|J|}{\ln(1+\sqrt{2})} \approx 2.269\,|J|. \tag{6.45}$$

You can find more details on the Ising model and Onsager's solution in the book "Statistical Mechanics" by K. Huang. The exact solution of the one-dimensional model based on the transfer matrix method, is also nicely described in Ch. 11.2 of Thijssen's book "Computational Physics".

Our goal here will be to use MC methods to calculate the properties of the Ising model, in particular in order to extract its critical properties such as critical exponents at the magnetic phase transition. Such methods are applicable beyond the exactly solvable cases, and allow the study of various other statistical mechanics

models, as discussed later. Based on the Ising model, we will also discuss problems of such MC methods, ways to extract TDL behavior from finite system simulations, and recent improvements of classical MC methods (the cluster methods).

6.4 Markov Process and Detailed Balance

We now discuss MC methods for the simulation of statistical physics models, such as the Ising model. At a temperature T, the expectation value of an observable A in the canonical ensemble is given as

$$\langle A \rangle = \sum_{\{C\}} P(C)A(C), \quad P(C) = \frac{1}{Z} e^{-\beta E(C)}, \tag{6.46}$$

involving a sum over configuration space $\Omega = \{C\}$. In this context, simple sampling MC corresponds to an unbiased, random generation of configurations C, and then calculating $\langle A \rangle$ based on this set of generated configurations. Obviously, such a procedure will be quite inefficient, if P is not uniform. In particular at low temperatures, one mainly generates configurations that are not important, in the way that they do not exhibit ordered domains. Important sampling corresponds to generating configurations C according to the distribution $P(C)$. But, how can we do so? We do not have such a random configuration generator at hand.

The idea is to generate such a sequence of configurations using a stochastic process: Starting from an initial configuration $C_{i(0)} \in \Omega$, we generate a sequence

$$C_{i(0)} \to C_{i(1)} \to \ldots \to C_{i(N)} \tag{6.47}$$

where in each step a random modification (update) is performed on $C_{i(k)}$, resulting in $C_{i(k+1)}$ in the k-th step of the sequence. Such a stochastic process is thus defined in terms of the probabilities for the various possible updates to occur. If furthermore, the probability for an update $C_{i(k)} \to C_{i(k+1)}$ in the k-th step depends only on $C_{i(k)}$, but not on the previous configurations, the stochastic process is also called a Markov process, named after A. Markov, who introduced it around 1907. The sequence of configurations is then also called a Markov chain. In order to specify the Markov process, we thus need to specify the transition probabilities

$$P(C_i \to C_j) \tag{6.48}$$

for an update from configuration C_i to configuration C_j. What conditions do these transition probabilities need to fulfill, in order for the Markov chain to follow the desired distribution $P(C_i)$ (i.e. to have the occurrence of a configuration C_i in the Markov chain proportional to $P(C_i)$)?

To derive such a condition, consider an ensemble of configurations, which is already distributed according to $P(C)$, such that the number of times the configuration C_i occurs within the ensemble,

$$N_0(C_i) \propto P(C_i). \tag{6.49}$$

After performing an update step on this ensemble of configurations, we find for the number of times the configuration C_i occurs within the ensemble,

$$N_1(C_i) = N_0(C_i) + \sum_{j \neq i} [N_0(C_j)P(C_j \to C_i) - N_0(C_i)P(C_i \to C_j)], \tag{6.50}$$

taking the configurations C_j that were changed into and out of C_i. This equation defines the master equation of the Markov process. If the process is to preserve the desired distribution, the sum in the above equation needs to vanish for all C_i. Since, initially $N_0(C_i) \propto P(C_i)$, we thus arrive at the condition

$$\sum_{j \neq i} [P(C_j)P(C_j \to C_i) - P(C_i)P(C_i \to C_j)] = 0, \tag{6.51}$$

which must be fulfilled for all i.

A special solution which satisfies all the above conditions term-by-term is

$$\forall_{j \neq i}: \quad P(C_j)P(C_j \to C_i) - P(C_i)P(C_i \to C_j) = 0. \tag{6.52}$$

This particular solution of the master equation for the transition probabilities is called detailed balance, and is commonly written

$$\frac{P(C_i \to C_j)}{P(C_j \to C_i)} = \frac{P(C_j)}{P(C_i)}. \tag{6.53}$$

In statistical physics the configuration probability $P(C_i)$ is given as

$$P(C_i) = \frac{1}{Z} W(C_i), \quad W(C_i) = e^{-\beta E(C_i)}, \tag{6.54}$$

so that the detailed balance condition Eq. (6.53) can be written as

$$\frac{P(C_i \to C_j)}{P(C_j \to C_i)} = \frac{W(C_j)}{W(C_i)} = e^{-\beta(E(C_j) - E(C_i))}. \tag{6.55}$$

If the detailed balance condition is fulfilled for the transition probabilities, an ensemble of configurations will remain distributed according to the desired distribution. However, in a MC simulation, we consider the temporal evolution of a single Markov chain. In order for this single Markov chain to follow the desired distribution, an additional requirement must be fulfilled: Namely, the transition probabilities must ensure ergodicity. This means, that starting from any initial configuration, any other configurations can in principle be obtained in a finite number of update steps. Thus,

a Markov chain reflects the desired distribution, if detailed balance and ergodicity are both guaranteed.
In practice, the initial configuration $C_{i(0)}$ will be chosen randomly (e.g. in the Ising model by assigning random values to each spin), and thus usually will not be one of the important configurations. The Markov chain will thus initially not be distributed according to the desired distribution, and one needs to allow for some time for it to reach this distribution.
In order to see, that the equilibrium distribution is indeed eventually reached, let us analyze the dynamics of the Markov chain in more detail. For this purpose, we denote the transition matrix T, with elements

$$T_{ij} = P(C_j \to C_i) \geq 0, \tag{6.56}$$

so that in an ensemble of configurations, that evolves in time according to our transition probabilities, the number of times a configuration C_i occurs within the ensemble after t update steps, $N_t(C_i)$, obeys:

$$N_{t+1}(C_i) = \sum_j T_{ij} N_t(C_j). \tag{6.57}$$

Since the probability of landing anywhere in Ω from a given C_j equals one, T satisfies

$$\sum_i T_{ij} = 1, \tag{6.58}$$

whereas ergodicity implies that there is an integer n_s, such that all elements of T^{n_s} are positive, $(T^{n_s})_{ij} > 0$. For such stochastic matrices, the Perron Frobenius theorem states that (i) the matrix T has 1 as an eigenvalue of multiplicity one, for which the corresponding eigenvector can be chosen to be positive in all its elements, and (iii) all other eigenvalues λ of T have $|\lambda| < 1$. The eigenvector P corresponding to the eigenvalue 1 obviously fulfills the requirement of stationarity:

$$T_{ij} P_j = P_i \tag{6.59}$$

and thus corresponds to the desired (equilibrium) distribution, $P_j = P(C_j)$. That N_t indeed approaches this equilibrium distribution P may be seen as follows: Let

$$P_t(C_i) = \frac{N_t(C_i)}{\sum_j N_t(C_j)} \tag{6.60}$$

be the fraction in the ensemble in state C_i after t update steps, then

$$\begin{aligned} ||P_{t+1} - P|| &:= \sum_i |P_{t+1}(C_i) - P(C_i)| \\ &= \sum_i |\sum_j T_{ij} P_t(C_j) - \sum_j T_{ij} P(C_j)| \end{aligned}$$

Figure 6.5: In a MC simulation one starts the Markov chain from an arbitrary configuration $C_{i(0)}$. After a sufficient number of equilibration steps, the system evolved towards the desired distribution. After that the configurations generated can be used to measure observables.

$$\begin{aligned} &\leq \sum_i \sum_j T_{ij} |P_t(C_j) - P(C_j)| \\ &= \sum_j \left(\sum_i T_{ij} \right) |P_t(C_j) - P(C_j)| \\ &= \sum_j |P_t(C_j) - P(C_j)| = ||P_t - P||. \end{aligned} \tag{6.61}$$

In fact, for an arbitrary initial ensemble distribution, the convergence to P is determined by the eigenvalue of T with second largest absolute value, λ_2. The rate of convergence will then scale as $\exp(-t/\tau)$, with the "mixing-time" $\tau = -1/\log(|\lambda_2|) > 0$. This can be seen from a similar analysis as above, performed in the basis of eigenvectors of T.
This initial phase of a MC simulation, in which the system tends towards the equilibrium distribution is also called equilibration or thermalization phase. It is an important part of a MC simulation, and care has to be taken in order to assure, that sufficient equilibration has been allowed for before taking measurements. In Fig. 6.5, the resulting sequence of phases of a MC simulation is shown. In practice, one might monitor the time evolution of the relevant observables to assure, that sufficient equilibration has been achieved.

What type of updates, i.e. transitions $C_i \to C_j$ does one consider in a MC simulation? This question concerning the update method, is what we discuss next. The answer to this question obviously depends on the specific model under consideration. Traditionally, one would employ so-called local updates, where only a small part in the current configuration is changed in an MC update step:

For example, in the Ising model, such a local update corresponds to changing the orientation of a single, randomly picked spin. Such updates allow to reach any configuration from any initial configuration, and are thus ergodic.

In an MC simulation of a gas or fluid, a local update corresponds to e.g. moving a randomly picked particle to a new position inside a cube of a given size around its current position. Apparently, such updates are also ergodic.

Here, we have to make a very important remark: In contrast to MD, where we

followed the real time dynamics of the system, in a MC simulation, we do not trace any dynamics of the model! The dynamics seen in a MC simulation is set by the random updates we perform and not by the equations of motion. Also note, that for the Ising model, there are no equations of motions defined, but only the energy for a given configuration. Therefore, the dynamics of a MC simulations is fully in our hands, and merely has to satisfy detailed balance and ergodicity. In fact, as we will discuss later, one can increase the efficiency of a MC simulation by modifying the MC dynamics appropriately.

Let us consider local updates in more detail: such an update involves two steps. First, a randomly picked spin (particle) is attempted to be flipped (shifted). Then, this change is eventually accepted with some probability. The total transition probability $P(C_i \to C_j)$ is thus a product of two probabilities

$$P(C_i \to C_j) = P_{attempt}(C_i \to C_j) \times P_{accept}(C_i \to C_j) \tag{6.62}$$

Often, such as in the two examples above, the attempt probability $P_{attempt}(C_i \to C_j)$ is uniform, $P_{attempt}(C_i \to C_j) = \text{const}$, independent of i and j. Then, the detailed balance condition Eq. (6.55) becomes

$$\frac{P_{accept}(C_i \to C_j)}{P_{accept}(C_j \to C_i)} = \frac{W(C_j)}{W(C_i)} = e^{-\beta(E(C_j)-E(C_i))}. \tag{6.63}$$

There are various solutions to this equation. The most commonly used is the Metropolis acceptance probability:

$$P_{accept}(C_i \to C_j) = \min\left[\frac{W(C_j)}{W(C_i)}, 1\right] = \min\left[e^{-\beta(E(C_j)-E(C_i))}, 1\right] \tag{6.64}$$

If the new configuration has lower or equal energy $E(C_j) \leq E(C_i)$ (corresponding to a larger or equal weight $W(C_j) \geq W(C_i)$), the update is always accepted ($C_{i(k+1)} = C_j$); while if the new configuration has higher energy $E(C_j) > E(C_i)$ (corresponding to smaller weight $W(C_j) < W(C_i)$), it will be accepted only with a probability, equal to the ratio of the new and old weight $\frac{W(C_j)}{W(C_i)} = \frac{e^{-\beta E(C_j)}}{e^{-\beta E(C_i)}} = e^{-\beta(E(C_j)-E(C_i))} \in (0,1)$. To determine whether to accept the update when $P_{accept}(C_i \to C_j) < 1$, one can generate a uniformly distributed random number $z \in (0,1)$; if $z < P_{accept}(C_i \to C_j) = e^{-\beta(E(C_j)-E(C_i))}$ the update is accepted, and otherwise rejected.

The resulting MC algorithm is called the Metropolis algorithm, and was developed by N. Metropolis et al. in J. Chem. Phys. **21**, 1087 (1953). It can be applied very generally to classical many-body systems, once appropriate local updates have been identified, such that the calculation of the energy difference $E(C_j) - E(C_i)$ becomes feasible. In the next section, we discuss the application of the Metropolis algorithm to the Ising model in more detail.

Another important solution of Eq. (6.63) is the heat bath solution, where

$$P_{accept}(C_i \to C_j) = \frac{W(C_j)}{W(C_i) + W(C_j)}, \tag{6.65}$$

is set by the relative weight of the two alternatives, C_j vs. C_i.
In general, the heat bath acceptance probability involves a selection among p possible options $C_{j(l)}, l = 1, ..., p$, where option $C_{j(k)}$ is accepted with probability

$$P_{accept}(C_{j(k)}) = \frac{W(C_{j(k)})}{\sum_{l=1}^{p} W(C_{j(l)})}, \quad k = 1, ..., p. \tag{6.66}$$

For lattice models where each lattice site can be in $p > 2$ different states (such as e.g. in the p-states Potts model, for which the σ_i can take p different integer values from 1 to p, with an energy

$$E(\sigma) = \sum_{i,j} J_{ij} \delta_{\sigma_i, \sigma_j} \tag{6.67}$$

involving a Kronecker δ-function), the heat bath acceptance probability is commonly used, and Eq. (6.65) is the special case for $p = 2$ (also, the 2-states Potts model is equivalent to an Ising model).

6.5 Metropolis Algorithm for the Ising Model

In this section, we consider the Metropolis algorithm as applied to the Ising model in more detail. In this case, a local update means to select a spin σ_k at random and flip it with the acceptance probability Eq. (6.64). Acceptance of such an update takes a configuration C to a new configuration C', where all spins remain constant, except for $\sigma_k \to \sigma'_k = -\sigma_k$. The weight ratio $W(C')/W(C)$ in the acceptance probability Eq. (6.64) is thus explicitly given by

$$\begin{aligned}
\frac{W(C')}{W(C)} &= \exp\Big[-\beta(E(C') - E(C))\Big] \\
&= \exp\left[-\beta J \sum_{<i,j>} (\sigma'_i \sigma'_j - \sigma_i \sigma_j)\right] \\
&= \exp\left[2\beta J \sum_{j \in n(k)} \sigma_k \sigma_j\right] \\
&= \exp\left[2\beta J \sigma_k \sum_{j \in n(k)} \sigma_j\right],
\end{aligned} \tag{6.68}$$

where $n(k)$ denotes the set of nearest neighbor sites of site k. For a d-dimensional hyper-cubic lattice the coordination number $|n(k)| = z = 2d$.
In an efficient implementation, one saves the possible values for these ratios in a table before starting the MC simulation, instead of recalculating the exponential function over and over again. Such a table could be realized using a two-dimensional array, that is indexed by the values of σ_k and $\sum_{j\in n(k)} \sigma_j$. For example, on a square lattice, where $z = 4$, $\sum_{j\in n(k)} \sigma_j$ can take on the possible values $-4, -2, 0, 2$, and 4.
Furthermore, note that if the ratio $W(C')/W(C) \geq 1$, no random number needs to be generated, as the update is definitely accepted in that case.

For the initial configuration $C_{i(0)}$ one could randomly assign each $\sigma_i = \pm 1$ with equal probability. Otherwise, at low temperatures, one could consider the fully polarized state as the initial configuration. For a lattice with N_s spins, we define one MC step as N_s trials of flipping a randomly chosen spin.
After performing a series of N_{eq} thermalization steps, one performs a number N of MC steps, each step $k = 1, ..., N$ followed by a measurement of various observables, such as the energy $E_k = E(C_{i(k)})$, the magnetization $m_k = m(C_{i(k)})$, and the ferromagnetic order parameter $|m|_k = |m(C_{i(k)})|$. From these N values of an observable A, one can finally calculate the average

$$\overline{A} = \frac{1}{N} \sum_{k=1}^{N} A_k. \tag{6.69}$$

If one repeats this procedure with M independent simulations (i.e. independent random number sequences), each simulation containing N Monte Carlo measurements, one obtains M such averages

$$\overline{A}_i = \frac{1}{N} \sum_{k=1}^{N} A_k, \quad i = 1, ..., M \tag{6.70}$$

and the overall mean

$$\overline{A} = \frac{1}{M} \sum_{i=1}^{M} \overline{A}_i. \tag{6.71}$$

As the $\overline{A}_i$ are statistically independent, we can estimate the statistical error $\Delta\overline{A}$ of the overall mean $\overline{A}$ using (c.f. Eq. (6.16) and Eq. (6.11))

$$\Delta\overline{A} = \frac{1}{\sqrt{M(M-1)}} \left(\sum_{i=1}^{M} \left(\overline{A}_i - \overline{A}\right)^2 \right)^{\frac{1}{2}} \tag{6.72}$$

$$= \frac{1}{\sqrt{M(M-1)}} \left(\sum_{i=1}^{M} \overline{A}_i^{\,2} - M\overline{A}^{\,2} \right)^{\frac{1}{2}}$$

$$= \frac{1}{\sqrt{M(M-1)}} \left(\sum_{i=1}^{M} \overline{A}_i^{\,2} - \frac{1}{M} \left(\sum_{i=1}^{M} \overline{A}_i \right)^2 \right)^{\frac{1}{2}}. \tag{6.73}$$

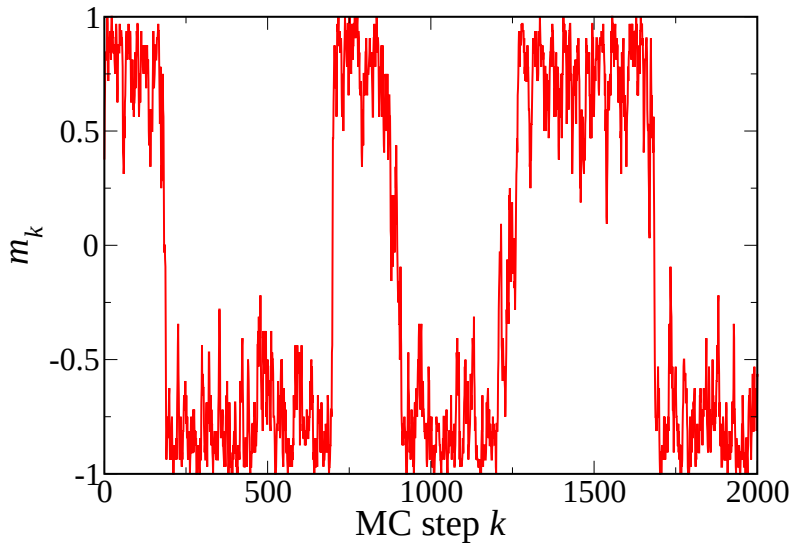

Figure 6.6: Time-series of the magnetization measurements from a simulation of the Ising model on a 8×8 square lattice at $k_B T = 2.4J$, employing the Metropolis algorithm.

In the next section, we discuss how to calculate the error of the mean from a single MC simulation ($M = 1$), where we will find that correlations in the Markov chain are important to be accounted for.

6.6 Correlations and Error Analysis

The Metropolis algorithm applied to the Ising model generates a sequence of configurations $C_{i(k)} \in \Omega$, $k = 0, 1, ..., N$, where we denote from now on by $C_{i(k)}$ the configuration after the k-th MC step, that results after attempting N_s local spin flips to the previous configuration $C_{i(k-1)}$. Here, we consider only those configurations, for which measurements are taken, i.e. those generated after the thermalization phase. Let us look at the time-series of measurements of an observable A, performed on such a Markov chain, $A_k = A(C_{i(k)})$. In Fig. 6.6 we show as an example the values for the magnetization $m_k = m(C_{i(k)})$ vs. k from the simulation of the Ising model on a 8×8 square lattice at $k_B T = 2.4J$. The data shown in the figure clearly show, that the consecutive measurements are highly correlated: e.g., if $m_k > 0$, then most likely also $m_{k'} > 0$ for k' close to k. This indicates that in the Markov chain, consecutive configurations cannot be considered to be statistically independent; instead they are strongly correlated. Only upon considering configurations, that are sufficiently separated (in terms of the MC step index k), will these correlations be reduced. How many MC steps are required between two configurations, before they can be considered statistically independent?
In order to answer this important question, we study the autocorrelation function,

which for a quantity A is defined for an integer separation τ as

$$C_A(\tau) = \frac{\langle A_k A_{k+\tau}\rangle - \langle A_k\rangle^2}{\langle A_k^2\rangle - \langle A_k\rangle^2}, \tag{6.74}$$

where Eq. (6.74) is normalized such that

$$C_A(0) = 1. \tag{6.75}$$

If the configurations $C_{i(k)}$ and $C_{i(k+1)}$ for all k were statistically independent, then $\langle A_k A_{k+\tau}\rangle = \langle A_k\rangle \cdot \langle A_{k+\tau}\rangle = \langle A_k\rangle^2$, and thus $C_A(\tau) = 0$ for $\tau > 0$. Finite values of $C_A(\tau)$ for $\tau > 0$ thus indicate the presence of correlations in the Markov chain. Considering the limit of large separation, we expect statistical independence for an ergodic simulation, and so the autocorrelation function approaches 0 as $\tau \to \infty$. For large τ the autocorrelation function Eq. (6.74) decays exponentially,

$$C_A(\tau) \longrightarrow e^{-\tau/\Theta_A}, \tag{6.76}$$

where Θ_A is called the autocorrelation time of observable A. If the autocorrelation function would be a pure exponential in a continuous τ, then

$$\int_0^\infty C_A(\tau)\, d\tau = \int_0^\infty e^{-\tau/\Theta_A} d\tau = \Theta_A. \tag{6.77}$$

This motivates the general definition of the integrated autocorrelation time

$$\Theta_A^{(int)} = \frac{1}{2} + \sum_{\tau=1}^{\infty} C_A(\tau), \tag{6.78}$$

which corresponds to the trapezoidal integration formula for $\int_0^\infty C_A(\tau)d\tau$ with a unit step size, and where the first term, $\frac{1}{2} = \frac{C_A(0)}{2}$. Here, we formally consider an infinitely long Markov chain.

Both Θ_A and $\Theta_A^{(int)}$ provide a measure for the number of the MC steps required between two configurations, before they can be considered statistically independent. In order to make this statement more precise, let us consider the effects of autocorrelations in the MC error analysis. Assume that A_k, $k = 1, 2, \ldots, N$ are the consecutive values of the observable A, as obtained from a MC simulation. Then, we can estimate $\langle A\rangle$ by the average

$$\overline{A} = \frac{1}{N}\sum_{k=1}^{N} A_k. \tag{6.79}$$

Next, we estimate its error from the mean square deviation:

$$(\Delta\overline{A})^2 \;=\; \left\langle \left(\overline{A} - \langle A\rangle\right)^2 \right\rangle$$

$$
\begin{aligned}
&= \left\langle \left(\frac{1}{N} \sum_{k=1}^{N} A_k - \langle A \rangle \right)^2 \right\rangle \\
&= \left\langle \left(\frac{1}{N} \sum_{k=1}^{N} A_k - \langle A \rangle \right) \cdot \left(\frac{1}{N} \sum_{k'=1}^{N} A_{k'} - \langle A \rangle \right) \right\rangle \\
&= \underbrace{\left\langle \frac{1}{N^2} \sum_{k=1}^{N} (A_k - \langle A \rangle)^2 \right\rangle}_{k=k'} + \underbrace{\frac{2}{N^2} \sum_{k=1}^{N} \sum_{\tau=1}^{N-k} (\langle A_k A_{k+\tau} \rangle - \underbrace{\langle \langle A \rangle^2 \rangle}_{\langle A \rangle^2})}_{k'=k+\tau \neq k} \\
&= \frac{1}{N} \left(\langle A^2 \rangle - \langle A \rangle^2 \right) + \frac{2}{N^2} \sum_{k=1}^{N} \sum_{\tau=1}^{N-k} C_A(\tau) \left(\langle A^2 \rangle - \langle A \rangle^2 \right) \\
&\approx \frac{1}{N} \left(\langle A^2 \rangle - \langle A \rangle^2 \right) \cdot \left(1 + 2 \sum_{\tau=1}^{\infty} C_A(\tau) \right) \quad ((N-k) \to \infty \text{ in the last sum}) \\
&= \frac{1}{N} \left(\langle A^2 \rangle - \langle A \rangle^2 \right) \cdot 2\Theta_A^{(int)} \qquad (6.80)
\end{aligned}
$$

In case the measurements are uncorrelated, such that $\Theta_A^{(int)} = 1/2$, this formula reproduces the result obtained from the central limit theorem, c.f. Eq. (6.9). However, we see that the naive application of the error formula corresponding to Eq. (6.72),

$$
(\Delta \overline{A})_{\text{naive}} = \frac{1}{\sqrt{N(N-1)}} \left(\sum_{k=1}^{N} \left(A_k - \overline{A} \right)^2 \right)^{\frac{1}{2}}, \qquad (6.81)
$$

which assumes statistical independence of the A_k, underestimates the error of $\overline{A}$, which is in fact larger, due to the correlations in the Markov chain, namely

$$
\Delta \overline{A} = (\Delta \overline{A})_{\text{naive}} \cdot \sqrt{2\Theta_A^{(int)}}. \qquad (6.82)
$$

Eq. (6.80) shows, that the presence of correlations in the Markov chain corresponds to a reduction of the effective number of measurements, as if only $N/2\Theta_A^{(int)}$ statistically independent measurements were actually performed.

Now, we present a systematic method to calculate the above error estimate, and thus $\Theta_A^{(int)}$ from the MC data for A from a single simulation: the binning analysis. The basic idea of the binning analysis is to group (bin) consecutive measurements into bins of data, such as indicated in Fig 6.7:
If we divide the data into M_B bins, each bin contains $N_B = N/M_B$ original MC measurement values. N_B is also called the bin length. Now calculate for each bin $i = 1, ..., M_B$ the bin-mean value:

$$
\overline{A}_i^B = \frac{1}{N_B} \sum_{A_j \in \text{bin } i} A_j, \quad i = 1, \ldots, M_B. \qquad (6.83)
$$

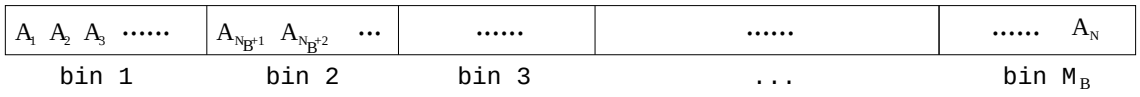

Figure 6.7: Binning N MC measurements into M_B bins, each of $N_B = N/M_B$ values.

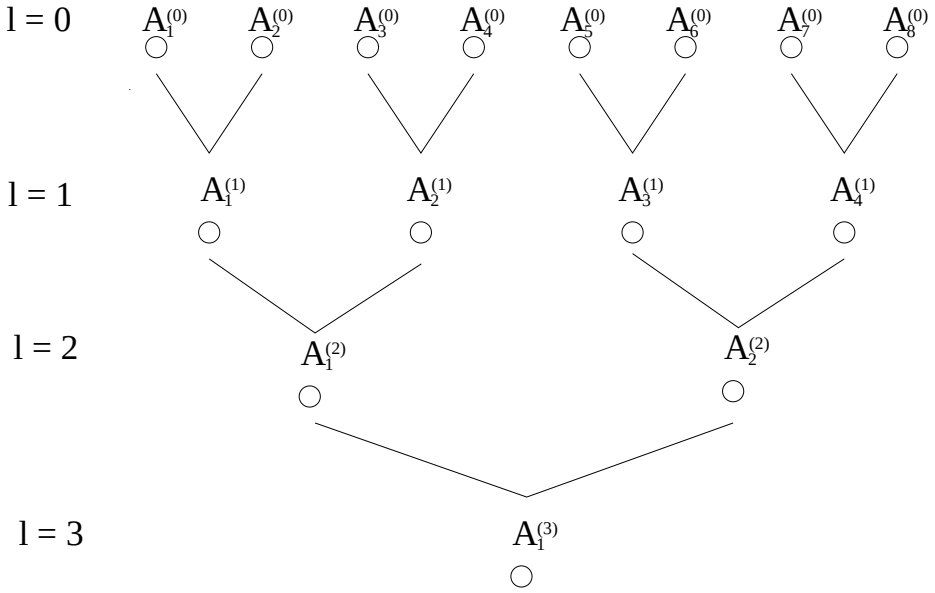

Figure 6.8: Binning analysis for $N = 8$ MC data $A_i^{(0)} = A_i, i = 1, ..., N$ up to the max. possible binning level $l = 3$ in this case.

If the bin length $N_B \gg \Theta_A^{(int)}$, the bin-mean values $\overline{A}_i^B$ will be statistically independent, and thus the naive error estimate, applied to the bin-mean values, converges to the error $\Delta\overline{A}$ of the correlated original MC data for sufficiently large values of N_B.

In practice, one often increases the bin length by powers of two, thus considering $N_B = 2^l$, where $l = 0, 1, ...$ denotes the binning level. Taking $l = 0$, the bin length is 1, and the bins correspond to the original MC data. At binning level l, the number of bins $M_B = \frac{N}{2^l} =: M_l$. We denote the M_l bin-mean values of the observable A at binning level l by $A_i^{(l)}$, $i = 1, ..., M_l$. Note, that $A_i^{(0)} = A_i$. Instead of recalculating the bin-mean values for each new level, one can use a simple iteration formula:

$$A_i^{(l)} = \frac{1}{2}\left(A_{2i-1}^{(l-1)} + A_{2i}^{(l-1)}\right), \tag{6.84}$$

This binning procedure is illustrated in Fig. 6.8 for $N = 8$ up do the max. possible binning level $l = 3$ in this case. For $M_l > 1$, we calculate the naive error of the

bin-mean values at binning level l,

$$\Delta\overline{A}^{(l)} = \frac{1}{\sqrt{M_l(M_l-1)}} \left(\sum_{i=1}^{M_l} (A_i^{(l)} - \overline{A})^2 \right)^{1/2}, \tag{6.85}$$

and obtain an estimate of the mean error $\Delta\overline{A}$ of the MC measurement values from

$$\Delta\overline{A} = \lim_{l\to\infty} \Delta\overline{A}^{(l)}. \tag{6.86}$$

In practice, only a finite number of binning steps can be performed. If one plots $\Delta\overline{A}^{(l)}$ vs. l, one obtains an increasing curve, which eventually becomes flat, showing that beyond a certain binning level, the bin-mean values become uncorrelated. However, for large values of l, the number of bins $M_l = N/2^l$ drastically decreases, and the curve starts to show large fluctuations, indicating that more data would be needed to reliably estimate $\Delta\overline{A}^{(l)}$ for such values of l. One thus needs to carefully monitor the behavior of $\Delta\overline{A}^{(l)}$ in order to reliably estimate the MC error from the binning analysis.
Since $(\Delta\overline{A})_{\text{naive}} = \Delta\overline{A}^{(0)}$, we can estimate the integrated autocorrelation time from

$$\Theta_A^{(int)} = \frac{1}{2} \left(\frac{\Delta\overline{A}^{(\infty)}}{\Delta\overline{A}^{(0)}} \right)^2, \tag{6.87}$$

where in practice one approximates $\Delta\overline{A}^{(\infty)}$ by $\Delta\overline{A}^{(l)}$ from the range of l-values, where $\Delta\overline{A}^{(l)}$ shows no more significant l-dependence.

6.7 Critical Slowing Down

We discussed already, that the Ising model in $d > 1$ exhibits a second order phase transition, which in case of a square lattice geometry takes place at

$$k_B T_c/|J| = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269. \tag{6.88}$$

If one uses the Metropolis algorithm to perform MC simulations of the model for different temperatures, and studies the behavior of the integrated autocorrelation time, e.g. of the order parameter $A = |m|$ as a function of the (linear) system size L, one finds that for $T > T_c$, $\Theta_A^{(int)}$ converges for sufficiently large systems (e.g., at $k_B T = 3|J|$ for $L > 30$). However, for T close to T_c, $\Theta_A^{(int)}$ steadily increases with L. Close to criticality, the dynamics of the Markov chain thus slows down, and the algorithm becomes inefficient in generating statistically independent configurations. This phenomenon is called critical slowing down. Quite generally, MC simulations using local updates suffer from critical slowing down close to second order phase

transitions. The reason for this behavior lies in the increasingly growing domains of alike spins, that form when the system approaches T_c. This formation of increasingly growing domains in the Ising model close to T_c can be studied by measuring the correlation between distant spins, i.e. the spin-spin correlation function $C(i,j) = \langle \sigma_i \cdot \sigma_j \rangle$, which shows an exponential decay for $T > Tc$, indicating the absence of long range order (correlations),

$$C(i,j) = \langle \sigma_i \cdot \sigma_j \rangle \propto e^{-r_{i,j}/\xi}, \quad T > T_c \tag{6.89}$$

where the correlation length ξ corresponds to the typical size of the ordered domains in the system. It quantifies the characteristic length scale of the thermal fluctuations in the system. Similar, for $T < T_c$, the correlation length quantifies the exponential decay of the spin-spin correlations induced by fluctuations atop the ordered state. As $T \to T_C$ the correlation length diverges,

$$\xi \propto \left(\frac{|T - T_C|}{T_C} \right)^{-\nu}, \tag{6.90}$$

where ν is another example of a critical exponent. For example, from the exact solution, one finds that for the two-dimensional Ising model, $\nu = 1$.

Close to T_c local updates, such as employed in the Metropolis algorithm, can only slowly change the structure of configurations including large domains. Away from T_c, the domains are smaller, and their shape and position can much more easily be changed by the Metropolis algorithm. Once the linear system size $L \gg \xi$, the system covers the relevant clusters, and the properties of the system and the autocorrelation times converge. However, the closer T approaches T_c, the more correlated become the MC configurations generated by the Markov process. Right at the critical point, where ξ diverges, the spin-spin correlations show an algebraic, power-law decay,

$$C(r) \propto \frac{1}{r^{d-2+\eta}}, \tag{6.91}$$

with yet another critical exponent η ($\eta = 1/4$ for the two-dimensional Ising model). The infinite system at T_c is thus scale-invariant, in that no characteristic length scale exists, and the domains exhibit a self-similar structure on all length scales larger than the lattice spacing. This relevance of fluctuations on all length scales is at the basis of the renormalization group theory of critical phenomena, which is a powerful, predictive theoretical framework to understand and quantitatively describe critical phenomena near second order phase transitions.
For a finite system close to T_c, the only relevant length scale is thus set by the system size L. The autocorrelation times then diverge as a power of L,

$$\Theta_A, \Theta_A^{(int)} \propto L^z, \tag{6.92}$$

where z is called the dynamical exponent. When using the Metropolis algorithm for the two-dimensional Ising model, $z \approx 2.2$, which quantifies the problem of critical

slowing down. Note, that z, in contrast to the other exponents introduced above is not intrinsic to the model, but instead depends strongly on the MC update method. We will next present MC methods for the Ising model, that overcome this drastic critical slowing down, via improving the simulation dynamics. In particular, they allow for the dynamical exponent to be reduced down to $z \approx 0.2$. This was a major improvement for performing large-scale MC simulation of statistical physics models close to second order phase transitions.

6.8 Cluster-Methods

In cluster algorithms, clusters of connected spins are flipped simultaneously in order to achieve an efficient MC dynamics, with reduced autocorrelations. The main goal is to construct these clusters of spins based on the relevant domains formed in the system. The algorithms described in this section are fundamental to an understanding of many recent advances in modern MC simulations.

6.8.1 Swendsen-Wang Algorithm

We first derive the cluster algorithm presented by Swendsen and Wang in Phys. Rev. Lett. **58**, 86 (1987), as applied to the ferromagnetic Ising model ($J < 0$). We start by rewriting the energy Eq. (6.35) of an Ising configuration in terms of an explicit summation over the bonds of the lattice,

$$E(\sigma) = -|J| \sum_{b=1}^{N_b} [\sigma_{i(b)}\sigma_{j(b)} + 1] = -\sum_{b=1}^{N_b} E_B(\sigma|_b), \tag{6.93}$$

where for a d-dimensional hypercubic lattice with periodic boundary conditions, the number of bonds is $N_b = d \cdot N_s$. In the equation above, we added a constant $|J|$ to the energy for each bond, so that the contribution from a bond with two anti-parallel spins vanishes (we will see later, why this is useful), and denote by $i(b)$ and $j(b)$ the two sites connected by bond b. Of course, adding a constant to the energy does not change the physics of the model. Furthermore, we denote by $E_B(\sigma|_b)$ the contribution of bond b,

$$E_B(\sigma|_b) = |J|(\sigma_{i(b)}\sigma_{j(b)} + 1). \tag{6.94}$$

Then, we can express the partition function as

$$Z = \sum_\sigma e^{-\beta E(\sigma)} = \sum_\sigma \prod_{b=1}^{N_b} e^{\beta E_B(\sigma|_b)} = \sum_\sigma \prod_{b=1}^{N_b} [\underbrace{1}_{F_B(0,\sigma|_b)} + \underbrace{(e^{\beta E_B(\sigma|_b)} - 1)}_{F_B(1,\sigma|_b)}]. \tag{6.95}$$

where $F_B(0)$ and $F_B(1)$ are the <u>bond functions</u>

$$\begin{aligned} F_B(0, \sigma|_b) &= 1, \\ F_B(1, \sigma|_b) &= e^{\beta E_B(\sigma|_b)} - 1. \end{aligned} \tag{6.96}$$

Thus, the partition function can be rewritten as

$$Z = \sum_{\sigma} \prod_{b=1}^{N_b} [F_B(0, \sigma|_b) + F_B(1, \sigma|_b)]. \tag{6.97}$$

We next introduce a set of auxiliary bond variables $\tau_b = 0, 1$ which are used as arguments in the bond function Eq. (6.96) for each bond b, and use the notation $\tau = (\tau_1, \tau_2, \ldots, \tau_{N_b})$ to refer to a whole bond configuration, in analogy to a spin configuration σ. We can then express Z as sum over both spins and bond variables:

$$Z = \sum_{\sigma} \sum_{\tau} \prod_{b=1}^{N_b} F_B(\tau_b, \sigma|_b). \tag{6.98}$$

This representation of the partition function of the Ising model was already introduced by Fortuin and Kasteleyn in 1969. In the combined configuration space of spins and bonds, the configuration weight in the partition function Eq. (6.97) is thus

$$W(\sigma, \tau) = \prod_{b=1}^{N_b} F_B(\tau|_b, \sigma|_b) \tag{6.99}$$

The bond function F_B depends as follows on the spins connected by bond b :

$$F_B(0, \sigma|_b) = 1, \quad \text{independent of } \sigma \tag{6.100}$$

$$F_B(1, \sigma|_b) = e^{\beta E_B(\sigma|_b)} - 1 = \begin{cases} e^{2\beta|J|} - 1, & \text{if } \sigma_{i(b)} = \sigma_{j(b)}, \\ 0, & \text{if } \sigma_{i(b)} \neq \sigma_{j(b)}. \end{cases} \tag{6.101}$$

We refer to a bond b with $\tau_b = 0$ as an empty bond, contributing 1 to the weight in Eq. (6.99). A bond b with $\tau_b = 1$ we call a filled bond, which contributes $e^{2\beta|J|} - 1$ to the weight only, when the bond is between parallel spins, otherwise, it contributes 0. The two important features of the weight $W(\sigma, \tau)$ in Eq. (6.102) are thus:

(i) $W(\sigma, \tau)$ is zero, if a filled bond is placed between antiparallel spins,
(ii) if no such "illegal" bonds are present, $W(\sigma, \tau)$ is independent of σ. In this case, the configuration weight is

$$W(\sigma, \tau) = (e^{2\beta|J|} - 1)^{n_f}, \tag{6.102}$$

where n_f is the number of filled bonds.

The spin configuration σ affects the weight only via imposing a restriction on where filled bonds can be placed. This property relies on the added constant $-|J|$ in each bond energy in Eq. (6.93); without this term there could be filled bonds also between antiparallel spins, and the weight function would have a more complex σ-dependence.

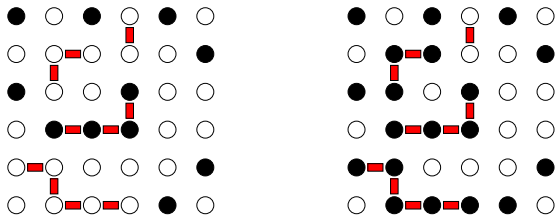

Figure 6.9: A spin configuration (circles) with filled bonds (red retangles between two circles) inserted between equal spins according to the appropriate probability Eq. (6.104) (left panel) . Two out of the four clusters are then flipped (right panel).

We can now use the above properties of the weight in order to construct an MC update method, that performs a spin configuration update by using an intermediate step of generating a "legal" bond configuration, as follows:

$$\sigma \to (\sigma, \tau) \to (\sigma', \tau) \to \sigma' \tag{6.103}$$

Let us describe the three involved sub-steps:

$\underline{\sigma \to (\sigma, \tau)}$: For each bond b, assign $\tau_b = 0, 1$, with heat-bath probabilities

$$P_B(\tau_b) = \frac{F_B(\tau|_b, \sigma|_b)}{F_B(0, \sigma|_b) + F_B(1, \sigma|_b)}, \tag{6.104}$$

so that the probability to fill the bond is

$$P_B(\tau_b = 1) = 1 - e^{-2\beta|J|}, \text{ if } \sigma_{i(b)} = \sigma_{j(b)}, \tag{6.105}$$

$$P_B(\tau_b = 1) = 0, \qquad \text{if } \sigma_{i(b)} \neq \sigma_{j(b)}. \tag{6.106}$$

$\underline{(\sigma, \tau) \to (\sigma', \tau)}$: After τ has been assigned, determine the clusters of spins, that are connected by filled bonds. If all spins in such a cluster are flipped collectively, the weight Eq. (6.102) is unchanged, c.f. Fig. 6.9. We flip each cluster with probability 1/2. After that a new spin configuration σ' is obtained, which is also consistent with the bond configuration τ.

$\underline{(\sigma', \tau) \to \sigma'}$: Use σ' as the new spin configuration, and forget τ.

Upon repeating the above procedure, we obtain the Swendson-Wang algorithm:

- (1): Start with an arbitrary spin configuration.
- (2): Cast filled bonds between equal spins according to the probabilities Eq. (6.105) and Eq. (6.106).
- (3): Identify all spin clusters; flip each of them with probability 1/2.

- (4): Start a new update from step (2).

The most demanding part of implementing the Swendsen-Wang algorithm is the identification of all clusters. There are several cluster identification algorithms, the most efficient one being the Hoshen-Kopelmann algorithm, scaling linear in system size N_s.

The dynamical exponent z is low for the Swendsen-Wang algorithm ($z \approx 0.2$ for the two-dimensional Ising model), although there is still a mild critical slowing down problem. But compared with the Metropolis algorithm ($z \approx 2.2$ for the two-dimensional Ising model), the improvement is very significant.

6.8.2 Wolff Algorithm and Continuous Symmetries

In Phys. Rev. Lett. **62**, 361 (1989), Wolff introduced a simple modification of the Swendsen-Wang algorithm, that makes it much easier to implement. He also provided an extension to models with a continuous symmetry.
The difference between Wolff's algorithm and the Swendsen-Wang algorithm, as applied to the Ising model, is that in the former at each step one spin is selected randomly and only a single cluster is constructed, starting from this seed spin, and the resulting cluster is flipped with probability one. The filled bonds are cast with the same probabilities, Eq. (6.105) and Eq. (6.106), as in the Swendsen-Wang algorithm. Therefore, in each step of the Wolff algorithm one constructs only one of the clusters that would be constructed in the same step of the Swendsen-Wang algorithm. There is thus no need to identify all the clusters. Furthermore, since the seed spin is selected randomly, the probability of selecting large clusters will be high.

There is a simple means to reconstruct the Wolff algorithm: Consider two spins on a randomly chosen bond of the lattice. If one considers in the Metropolis algorithm to flip one of the two spins, then the energy change on that bond would be $\Delta E = -2|J| < 0$ in case the spins are antiparallel. Thus the probability to accept the flip (just considering these two spins) is 1. In case the two spins are initially parallel, the energy change on that bond would be $\Delta E = +2|J| > 0$, and the probability to accept the flip (just considering these two spins) $e^{-2\beta|J|}$; with probability $P = 1 - e^{-2\beta|J|}$ no relative spin flip will be introduced. Now, instead of flipping one of the spins against the other, one might consider, with the same probability P, to flip *both* spins simultaneously, preserving their parallel alignment. Continuing in a similar fashion with all the other spins connected to the two spins, etc., one obtains a cluster of parallel spins that are finally flipped together. This way, one precisely constructs and flips a single cluster as in the Wolff algorithm.
Wolff also showed how the idea of cluster updates can be applied to models which have a continuous symmetry. Let us first discuss a few important examples of such models:

Classical Heisenberg Model

In the classical Heisenberg model, one considers instead of a binary variable a continuous three-dimensional unit vector degree of freedom at each lattice site:

$$\sigma_i \to \vec{S}_i \in S^2 = \{\vec{x} \in \mathbb{R}^3 \mid |\vec{x}| = 1\}. \tag{6.107}$$

The energy of a given configuration $\vec{S} = (\vec{S}_1, ..., \vec{S}_{N_s})$ is defined in terms of scalar products,

$$E(\vec{S}) = \sum_{i,j} J_{ij} \vec{S}_i \cdot \vec{S}_j, \tag{6.108}$$

so that this model exhibits a continuous global $O(3)$ symmetry. Compare this to the Ising model, that has a discrete global $\mathbb{Z}_2$ symmetry. In fact, the difference in symmetry has drastic effects on the behavior of the model, even on the same lattice. Namely, for a two-dimensional lattice, the Heisenberg model does not have a finite temperature phase transition. This is indeed a very deep theoretical result, the Mermin-Wagner theorem: There can be no spontaneous symmetry breaking of a continuous symmetry in one- and two-dimensional systems at finite temperature. So, e.g. the ferromagnetic Heisenberg model on the square lattice orders only in the ground state ($T_c = 0$). Any finite thermal fluctuations immediately destroy the long range order.

The three-dimensional Heisenberg model on a simple cubic lattice has a finite temperature second order phase transition at $T_C \approx 2.05|J|$. Below T_c, the system has long-range order.

XY Model

Here, one assigns to each lattice site a planar vector (or angular variable),

$$\vec{S}_i \in S^1 = \{\vec{x} \in \mathbb{R}^2 \mid |\vec{x}| = 1\}, \tag{6.109}$$

with the same expression Eq. (6.108) for the energy. The XY model thus has a continuous, global $O(2)$ symmetry. So, you would probably expect, based on the Mermin-Wagner theorem, that is does not develop long-range order at finite temperatures in two dimensions, which would correspond to spontaneous breaking of the continuous $O(2)$ symmetry. Indeed, this is the case, and the model shows long-range order only at $T = 0$. However, there still is a transition at a finite temperature. This transition is named Kosterlitz-Thouless transition, and the critical temperature $T_{KT} \approx 0.887|J|$ on the square lattice. For $T > T_{KT}$, the spin-spin correlation function decays exponentially with distance, similar to the Ising model above T_c. However, for $T < T_{KT}$ the spin-spin correlation function decays algebraically with distance, according to a power law, with a temperature dependent exponent. This transition derives from a binding of topological excitations (vortices) in the spin configuration below T_{KT}, where the system behaves like at a critical point. At larger T, these vortices unbind, and fully disorder the system.

For more on these models, and their critical properties, c.f. e.g. the book by J. Cardy, Scaling and Renormalization in Statistical Physics, Cambridge Lecture Notes in Physics 5, 1997, Cambridge Univ. Press.

The main idea of Wolff for constructing a cluster update scheme for models with an $O(n)$ symmetry, is to generalize the concept of simultaneous spin-flips from the Ising model to the general case. In each update step, one first chooses a random vector $\vec{u} \in S^{n-1}$. A spin-flip of a local spin $\vec{S}_i$ is then defined as a reflection with respect to the hypersurface perpendicular to $\vec{u}$:

$$\vec{S}_i \longrightarrow \vec{S}_i - 2(\vec{S}_i \cdot \vec{u})\vec{u}, \tag{6.110}$$

such that the component of $\vec{S}_i$ parallel to $\vec{u}$ is flipped. After picking a random $\vec{u}$, one selects a random spin and flips it according to the above specification. This defines the seed of the cluster. Then, one visits each neighbor of that spin, and attemps to flip it with a probability

$$P = 1 - \exp[\min(0, 2\beta|J|(\vec{S}_i \cdot \vec{u})(\vec{S}_j \cdot \vec{u}))] \tag{6.111}$$

where j denotes the site of the considered neighbor (note that the cluster spin $\vec{S}_i$ has already been flipped, in contrast to $\vec{S}_j$). If flipped, that spin is added to the cluster of flipped spins. As in the Wolff algorithm for the Ising model, one continues with the neighbors of any new site that has been added to the cluster, and which have not been visited already. Note, that in this formulation, flipping is done on-the-fly. For the case of the Ising model (i.e. $S_i, u \in \pm 1$), the above algorithm reduces to the Wolff algorithm as introduced above, with on-the-fly flipping. See the article by Wolff for more details and performance studies of the algorithm.

6.9 Critical Phenomena and Finite-Size Scaling

In this last section on MC simulations, after having introduced efficient methods to study second order phase transitions in statistical physics models, we now provide a short introduction to the main ideas behind critical phenomena and the method of finite-size scaling.

Consider, again, the Ising model on the square lattice, with a finite temperature second order phase transition at T_c. We already mentioned, that close to T_c increasingly large domains of equally oriented spins form (for $T < T_c$ atop the oriented, ordered background). The characteristic size of these domains relates to the correlation length ξ, which diverges when the system approaches the critical point. In the thermodynamic limit,

$$\xi \propto t^{-\nu}, \quad t = \frac{|T - T_c|}{T_c}, \tag{6.112}$$

in terms of the critical exponent ν, which is the same for approaching T_c from above or below.

Right at the critical point T_c, the correlation function decays in a power-law fashion,

$$\langle \sigma_i \sigma_j \rangle \propto \frac{1}{r_{ij}^{d-2+\eta}}, \tag{6.113}$$

with another critical exponent η. The order parameter is zero above T_c, and below T_c it behaves as

$$\langle |m| \rangle \propto (T_c - T)^{\beta}, \tag{6.114}$$

for T close to T_c. The magnetic susceptibility also exhibits a divergence at T_c,

$$\chi = \left. \frac{dm}{dh} \right|_{h=0} = \frac{N_S}{T} \langle m^2 \rangle \propto t^{-\gamma}, \tag{6.115}$$

where h denotes a magnetic field, that adds a term $-h \sum_i \sigma_i$ to the energy $E(\sigma)$. For $h > 0$, the field tends to align the spins to point "up". The specific heat also has a singularity at T_c,

$$C = \frac{dE}{dT} = \frac{1}{N_S T^2} \left(\langle E^2 \rangle - \langle E \rangle^2 \right) \propto t^{-\alpha}. \tag{6.116}$$

When the exponent $\alpha > 0$, C diverges at T_c; when $\alpha < 0$, the specific heat exhibits a cusp. In case of the two-dimensional Ising model, $\alpha = 0$, and the specific heat shows a logarithmic singularity $C \propto -\ln(t)$ close to T_c.
There are various relations between the exponents, such as

$$\gamma = \nu(2 - \eta), \tag{6.117}$$

$$\alpha + 2\beta + \gamma = 2, \tag{6.118}$$

$$\nu d = 2 - \alpha. \tag{6.119}$$

The last relation is called hyperscaling, and involves the dimensionality d of the system. You can check these relations for example for the two-dimensional Ising model, where $\nu = 1$, $\eta = \frac{1}{4}$, $\beta = \frac{1}{8}$, $\gamma = \frac{7}{4}$, $\alpha = 0$ are known exactly.

A fascinating fact about critical phenomena is that the values of the critical exponents, such as $\nu, \eta, \beta, \gamma, \delta, \alpha$ are exactly the same for different models, if these models fall into the same universality class. The universality class depends only on the dimensionality of the system and the system's symmetry, as described by the order parameter.

There are various methods to calculate critical properties of the various universality classes. These include series expansion techniques, the renormalization group, and MC simulations. As we are concerned here with the latter, let us consider them in

more detail.

First assume, that we could obtain MC data for such large system sizes, that they are representative for the TDL. For a given reduced temperature $t \neq 0$, there is a certain system size L above which measured expectation values converge exponentially to their TDL values. Typically, data for system sizes $L > 4\xi$ is considered indistinguishable from the TDL values. Then, consider an observable A with a scaling $A \propto t^{-\lambda}$ close to T_c, or

$$\ln(A) = const. - \lambda \ln(t). \tag{6.120}$$

In case we know T_c, we could extract the critical exponent λ by fitting a straight line between $\ln(A)$ and $\ln(t)$. If T_c is not known, we could treat it as an additional parameter, and adjust it to give the best such linear fit. However, since the pure power-law critical scaling behavior is valid only asymptotically as $t \to 0$, one may need to go to very large system sizes in order to obtain high-precision estimates for T_c and the exponents. Data obtained on finite systems does not exhibit singular behavior and thus there will be finite-size corrections to the exact TDL behavior, when performing the above analysis.

The method of finite-size scaling is a way to overcome this limitation, and extract TDL critical behavior from studies on finite systems, by employing the regularity of these deviations to extract information. It is based on the finite-size scaling hypothesis, which states that an observable close to T_c behaves as a power of L multiplied by a non-divergent function of L/ξ i.e., the ratio of the two relevant length scales in a finite system. The finite-size scaling hypothesis can be proven using the renormalization group theory. Applied to an observable A, the finite-size scaling ansatz is

$$A_L(t) = L^{\sigma_A} f_A(L/\xi(t)), \tag{6.121}$$

where $A_L(t)$ denotes the value of A in a system of size L at a reduced temperature t. As $\xi(t) \propto t^{-\nu}$, we also get (since $L/\xi(t) \propto L/t^{-\nu} = (L^{1/\nu}t)^{\nu}$)

$$A_L(t) = L^{\sigma_A} g_A(L^{1/\nu}t), \tag{6.122}$$

with the scaling function g_A. For a fixed value of t close to 0, the infinite size scaling form has to be recovered, i.e.

$$A_L(t) \propto t^{-\lambda}, \quad \text{for } L \to \infty, \tag{6.123}$$

where λ is the corresponding critical exponent for the quantity A in the TDL. In order for Eq. (6.123) to reproduce this limiting behavior, the scaling function must obey

$$g_A(x) \propto x^{-\lambda}, \quad \text{for } x \to \infty, \tag{6.124}$$

and $\sigma_A = \lambda/\nu$, since then

$$A_L(t) \propto L^{\sigma_A}(L^{1/\nu}t)^{-\lambda} = t^{-\lambda}, \quad \text{for } L \to \infty. \tag{6.125}$$

Thus, the finite-size scaling form of A is

$$A_L(t) = L^{\lambda/\nu} g_A(L^{1/\nu} t). \tag{6.126}$$

Consider the magnetic susceptibility χ as an example, so that $A = \chi$ and $\lambda = \gamma$, and thus

$$\chi_L(t) = L^{\gamma/\nu} g_\chi(L^{1/\nu} t). \tag{6.127}$$

If one plots $\chi_L / L^{\gamma/\nu}$ vs. $x = L^{1/\nu} t$ for different system sizes, all curves corresponding to different system sizes collapse onto each other (data collapse), and this curve is the scaling function $g_\chi(x)$. The scaling exactly at the critical point ($x = t = 0$) can be used to extract γ/ν, namely,

$$\chi_L(T_c) \propto L^{\gamma/\nu}. \tag{6.128}$$

If T_c is not known, one can adjust it until a linear behavior is seen in a log-log plot of $\chi_L(T_c)$ vs. L. However, such two-parameter scalings are easily affected by corrections to the asymptotic scaling form, and lead to systematic errors in the extracted T_c and exponents.
A more robust means of estimating T_c, without knowledge of exponents is to consider the so-called Binder ratio. For example, one can show that for the Ising model, the quantity

$$U_L(T) = \frac{3}{2}\left(1 - \frac{\langle m^4 \rangle}{3\langle m^2 \rangle^2}\right). \tag{6.129}$$

for different system sizes crosses at $T = T_c$. At low T, $m \to \pm 1$, so that $U \to 1$, while at large T, a Gaussian distribution in m gives $U = 0$.

You can find more details about finite size scaling, critical phenomena and their MC study in the books D. P. Landau and K. Binder, "A Guide to Monte Carlo Simulations in Statistical Physics", 2^{nd} Edition, 2005, Cambridge Univ. Press, and J. Cardy, "Scaling and Renormalization in Statistical Physics", Cambridge Lecture Notes in Physics 5, 1997, Cambridge Univ. Press.