Institut für Theoretische Physik
Fakultät für Physik und Geowissenschaften
Universität Leipzig

Diplomarbeit

Directional Ordering in the Classical Compass Model in Two and Three Dimensions

vorgelegt von

Max Henner Gerlach

Leipzig, den 24.02.2012

Betreuer: Prof. Dr. Wolfhard Janke

Gutachter: Prof. Dr. Wolfhard Janke
Prof. Dr. Ulrich Behn

Contents

Introduction and overview

In condensed matter physics there has been a long-standing interest in the properties of transition metal oxide compounds. In these systems the $3d$-shell of the transition metal ion is only partially filled. Since the $3d$-orbitals are nearly degenerate, this can give rise to an additional electronic degree of freedom next to charge and spin, where an electron can occupy one of the different orbitals. Different mechanisms of interaction can then lead to global, cooperative effects of orbital ordering. Since orbitals extend in real space, the Hamiltonians describing orbital-orbital interactions are highly anisotropic and frustrated. In the case of Kugel-Khomskii superexchange interactions, the orbital degrees of freedom appear as quantum pseudo-spin $1/2$ operators in the Hamiltonian. If the phonon-mediated Jahn-Teller effect dominates, the operators are basically classical pseudo-spins [1, 2].

In the case of three-fold degenerate t_{2g}-orbitals, which are present for instance in compounds with Ti^{3+} or V^{3+} ions, the orbital-interaction can be described by the compass model [3]. While the compass model is closely related to the well-studied $O(n)$ and Heisenberg lattice spin models of statistical physics, it does not have an ordered phase with a finite magnetization [4]. Competing interactions in different directions, however, allow for long-ranged directional ordering. The angular dependence of the interaction is qualitatively similar to a set of magnetic dipoles—or compass needles—arranged in a square or cubic lattice.

The quantum compass model has furthermore received increased attention in recent years because in the field of topological quantum computing it may serve to protect qubits from decoherence [5, 6]. The model can be realized in the form of arrays of superconducting Josephson junctions, which have already been implemented successfully in experiments [7].

Earlier Monte Carlo studies have established the existence of a continuous thermal phase transition in both the classical and the quantum variation of the two-dimensional compass model, where a directionally ordered state is realized at low temperatures [8–10]. These findings have been confirmed by high-temperature series expansions for the two-dimensional quantum model, but with the same method no signs of a finite-temperature phase transition could be found in the three-dimensional quantum model [11].

For this diploma thesis I have set out to analyze the three-dimensional classical compass model by the means of Monte Carlo simulations to search for a directional ordering transition and to understand its properties if it exists. In a preliminary study I have also re-examined the transition in the two-dimensional model. The compass model features high degeneracy, long autocorrelation times in simulations and strong finite-size effects. For these reasons a considerable methodical effort is required to obtain quantitative results. My main emphasis in this thesis is on the numerical methods that I have applied to this end. The thesis is organized in the following way:

In the first chapter a brief introduction into the formalism of equilibrium statistical physics, a short discussion of phase transitions and a presentation of the compass model and its symmetries are given.

The second chapter motivates and explains the methodology of Markov chain Monte Carlo computer simulations and discusses the effects of statistical correlations and how to account for them.

In the third chapter the concrete numerical methods used in this thesis are explained in detail. These include special screw-periodic boundary conditions, a local and a cluster update, parallel tempering and multiple histogram reweighting. Especially for the latter two points I present and compare various different approaches. In the context of multiple histogram reweighting, interesting effects in the structure of the statistical uncertainties of individual histogram bins have been found. These are described in appendix B.

In chapter four my results for the two-dimensional model are presented. They agree with those found in the literature.

Then chapter five gives the results found for the three-dimensional classical compass model. In contrast to the earlier publication on the quantum model, I have found a directional ordering phase transition, which appears to be of a first-order nature.

Finally in chapter six the results obtained in this thesis are summarized in a short form.

1 Statistical physics and the compass model

1.1 Fundamentals of equilibrium statistical physics

Macroscopic physical systems consisting of many particles cannot be described by specifying the microscopic states of all particles. That would require the solution of an intractably high number of coupled equations of motion. Also experimentally the complete microscopic information can never be acquired. But with the formalisms and methods of statistical physics the gross properties, described by thermodynamic observables like the internal energy, of such systems can often be described in a probabilistic manner. The expected behavior is derived statistically from the microscopic interactions of the components, which are encoded in the Hamiltonian $\mathcal{H}$. Here I can only give a concise presentation of some fundamental concepts of statistical physics. For a more detailed introduction see for instance the (German language) textbooks [12–14].

The following discussion will be limited to systems of fixed particle number in contact with a heat bath at constant temperature T that have relaxed into thermal equilibrium. Their macroscopic properties no longer change, although there are still microscopic fluctuations. Such a system can then be described in the canonical ensemble and the probability for it to be in a microstate $\boldsymbol{x}$ is given by the Boltzmann distribution

$$\mathcal{P}_{\mathrm{eq}}(\boldsymbol{x}) = \frac{1}{\mathcal{Z}} \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})}, \tag{1.1}$$

which only depends on the energy $E = \mathcal{H}(\boldsymbol{x})$. Here $\beta = {}^{1}/_{k_{\mathrm{B}}T}$ is the inverse temperature with and the normalizing constant $\mathcal{Z}$ is the partition function, which is given by

$$\mathcal{Z} = \int \mathrm{d}\boldsymbol{x} \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} = \int \mathrm{d}E \Omega(E) \mathrm{e}^{-\beta E}, \tag{1.2}$$

where in the first part the high-dimensional integral covers all possible microstates $\boldsymbol{x}$ and in the second part the energy density of states $\Omega(E)$ has been introduced to write $\mathcal{Z}$ in the form of a one-dimensional integral over the energy. The expectation values of an observable $\mathcal{O}$ is given by

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \mathrm{d}\boldsymbol{x} \mathcal{O}(\boldsymbol{x}) \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} = \frac{1}{\mathcal{Z}} \int \mathrm{d}E \langle \mathcal{O} \rangle_E \mathrm{e}^{-\beta E}, \tag{1.3}$$

where in the second part

$$\langle \mathcal{O} \rangle_E = \frac{1}{\Omega(E)} \int \mathrm{d}\boldsymbol{x} \mathcal{O}(\boldsymbol{x}) \delta(\mathcal{H}(\boldsymbol{x}) - E) \tag{1.4}$$

has been set.

From the partition function $\mathcal{Z}$ all thermodynamic state variables can be derived. The internal energy $U = \langle E \rangle$ is

$$\begin{aligned} \langle E \rangle &= \frac{1}{\mathcal{Z}} \int \mathrm{d}\boldsymbol{x} \mathcal{H}(\boldsymbol{x}) \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} = \frac{1}{\mathcal{Z}} \int \mathrm{d}\boldsymbol{x} \mathcal{H}(\boldsymbol{x}) \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} = -\frac{1}{\mathcal{Z}} \int \mathrm{d}\boldsymbol{x} \frac{\partial}{\partial \beta} \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} \\ &= -\frac{1}{\mathcal{Z}} \frac{\partial Z}{\partial \beta} = -\frac{\partial \ln \mathcal{Z}}{\partial \beta}. \end{aligned} \tag{1.5}$$

The heat capacity is given by the first derivative of U:

$$C = \frac{\partial U}{\partial T} = -k_{\mathrm{B}} \beta^2 \frac{\partial U}{\partial \beta} = k_{\mathrm{B}} \beta^2 \frac{\partial^2 \ln \mathcal{Z}}{\partial \beta^2}. \tag{1.6}$$

In analogy to Eq. (1.5) one can calculate $\langle E^2 \rangle = \frac{1}{\mathcal{Z}} \frac{\partial^2 Z}{\partial \beta^2}$ and express the heat capacity as

$$C = k_{\mathrm{B}} \beta^2 \left[\frac{1}{\mathcal{Z}} \frac{\partial^2 Z}{\partial \beta^2} - \left(\frac{1}{\mathcal{Z}} \frac{\partial Z}{\partial \beta} \right)^2 \right] = k_{\mathrm{B}} \beta^2 \left[\langle E^2 \rangle - \langle E \rangle^2 \right] = k_{\mathrm{B}} \beta^2 \langle [E - \langle E \rangle]^2 \rangle, \tag{1.7}$$

which relates it to the energy fluctuations. Moreover, C is related to the entropy S:

$$C = T \frac{\partial S}{\partial T} = -\beta \frac{\partial S}{\partial \beta}, \tag{1.8}$$

which by integration with respect to β and by setting $S(T = 0) = 0$ yields

$$S = -k_{\mathrm{B}} \beta \frac{\partial \ln \mathcal{Z}}{\partial \beta} + k_{\mathrm{B}} \ln \mathcal{Z}. \tag{1.9}$$

In combination with Eq. (1.5) an expression for the free energy F can be stated in relation to the partition function:

$$F = U - TS = -\frac{1}{\beta} \ln \mathcal{Z}. \tag{1.10}$$

In the canonical ensemble F is the thermodynamic potential that is minimized in thermal equilibrium.

1.2 Phase transitions and finite-size scaling

Phase transitions are characterized by singularities in a thermodynamic potential such as the free energy. If there is a finite discontinuity in a first derivative of the potential at an inverse temperature β_0, a thermal phase transition is classified as *first-order*. If the first derivative is continuous, but discontinuities or infinite divergences appear in second partial derivatives, the transition is called *second-order*, *continuous* or *critical*.

The partition function of finite systems is always analytic. Therefore, thermodynamic potentials derived from it do not show discontinuities. However, by the means of finite-size scaling theory properties of a phase transition occurring in the thermodynamic limit can be extrapolated from the behavior of finite systems.

For a more detailed account of the theory of phase transitions than I can give here see the standard textbooks on statistical physics or Stanley's introduction [15].

1.2.1 Order parameters

A characteristic feature of an order-disorder transition is the breaking of symmetries, which are present in the high-temperature phase, at low temperatures. Simultaneously, order increases and entropy is reduced.

To describe the ordering it is useful to introduce an order parameter ϕ, which can be taken as an observable that is zero in the disordered phase and which takes up a finite value in the ordered phase. If the Hamiltonian $\mathcal{H}$ does not already contain a term in ϕ, one can introduce a linear "perturbation" term to $\mathcal{H}$:

$$\mathcal{H}_h = \mathcal{H} - h\phi. \tag{1.11}$$

In this way ϕ is the conjugate variable coupling to a virtual field h of fixed value. The expectation value of ϕ can then be expressed as a first derivative of the free energy:

$$\langle\phi\rangle = \frac{1}{\mathcal{Z}}\int \mathrm{d}\boldsymbol{x}\phi(\boldsymbol{x})\mathrm{e}^{-\beta\mathcal{H}_h(\boldsymbol{x})} = \frac{1}{\beta\mathcal{Z}}\frac{\partial}{\partial h}\int \mathrm{d}\boldsymbol{x}\mathrm{e}^{-\beta\mathcal{H}_h(\boldsymbol{x})} = \frac{1}{\beta}\frac{\partial \ln \mathcal{Z}_h}{\partial h} = -\frac{\partial F_h}{\partial h}. \tag{1.12}$$

The second derivative can be calculated akin to the heat capacity in Eq. (1.7):

$$\frac{\partial^2 F_h}{\partial h^2} = -\frac{\partial\langle\phi\rangle}{\partial h} = -\beta[\langle\phi^2\rangle - \langle\phi\rangle^2]. \tag{1.13}$$

This gives rise to the definition of the *susceptibility* of ϕ to h, which is a measure for the fluctuations of ϕ:

$$\chi = \frac{\partial \langle \phi \rangle}{\partial h} = -\frac{\partial^2 F_h}{\partial h^2} = \beta[\langle \phi^2 \rangle - \langle \phi \rangle^2]. \tag{1.14}$$

After having taken the derivatives, h may be set to zero to make it disappear from the Hamiltonian.

1.2.2 First-order transitions

A first-order phase transition is indicated by a finite jump in the energy or the order parameter ϕ at the transition point β_0, which is caused by the coexistence of two phases. If β is varied across β_0, effects of hysteresis can be observed, which are brought about by the metastability of the phases. At first-order transitions the spatial correlation length ξ stays finite.

In the thermodynamic limit the heat capacity C and the order parameter susceptibility χ show a finite jump. This cannot be observed in finite systems, where C and χ show narrow peaks if the system is large enough. If one took the formal derivative of the discontinuous functions $E(\beta)$ or $\phi(\beta)$ of an infinite system, one would obtain δ-function singularities. The peaks in finite systems are the rounded residuals of those.

For an overview over the finite-size scaling theory of first-order transitions see the review [16] and let me here just state the central results for the scaling of the inverse transition temperature β_0. Pseudo-transition temperatures $\beta_0(N)$ for finite systems of size N can be defined by the locations of the maxima of χ and C or the minima of the Binder parameter

$$Q_2 = 1 - \frac{1}{3}\frac{\langle \phi^4 \rangle}{\langle \phi^2 \rangle^2}. \tag{1.15}$$

For all definitions of $\beta_0(N)$ the expected dominating scaling behavior is the following:

$$\beta_0(N) = \beta_0 + \frac{a}{N} + \cdots, \tag{1.16}$$

which allows the extrapolation of β_0. Here the parameter a depends on the studied quantity.

1.2.3 Continuous transitions

Continuous phase transitions feature a divergence of the spatial correlation length ξ at the critical temperature $T_\mathrm{c} = 1/k_\mathrm{B}\beta_\mathrm{c}$. Near T_c the power law

$$\xi \propto |t|^{-\nu} + \cdots \tag{1.17}$$

with the reduced temperature $t = \frac{T-T_\mathrm{c}}{T_\mathrm{c}}$ and the critical exponent ν describes the behavior of ξ. At T_c fluctuations become important at all length scales and the singular behaviors of thermodynamic quantities are also described by relations with critical exponents such as

$$\phi \propto |t|^{\beta} + \cdots, \qquad t < 0, \tag{1.18}$$

$$\chi \propto |t|^{-\gamma} + \cdots, \tag{1.19}$$

$$C - C_0 \propto |t|^{-\alpha} + \cdots. \tag{1.20}$$

Being measures for fluctuations, ξ and C diverge at T_c. The critical behavior defined by these exponents is the same for many systems. For systems with short-ranged interactions they are assumed to depend only on the spatial dimension of the system and the symmetry of the order parameter. In this way different physical systems can be grouped into universality classes.

In finite systems the correlation length ξ cannot diverge as it is limited by the linear system size L. According to finite-size scaling theory, the same power laws still hold, but L replaces ξ near T_c, so that

$$|t| \propto L^{-1/\nu} + \cdots. \tag{1.21}$$

To leading order one then has the scaling laws

$$\phi(L) \propto L^{-\beta/\nu} + \cdots, \tag{1.22}$$

$$\chi(L) \propto L^{\gamma/\nu} + \cdots, \tag{1.23}$$

$$C = C_0 + aL^{\alpha/\nu} + \cdots \tag{1.24}$$

and the locations of the finite maxima of χ and C scale according to

$$T_\mathrm{max}(L) = T_\mathrm{c} + aL^{-1/\nu} + \cdots. \tag{1.25}$$

1.3 The classical compass model

Lattice spin models with local interactions like the q-state Potts models and $O(n)$ models such as the Ising, xy and classical Heisenberg models have been very important in the development of the theory of phase transitions and of various analytical and numerical methods. The key difference of the compass models to these well-studied models is an inherent coupling of real space symmetry, realized by the point group of the lattice, to the symmetry of the interactions encoded in the Hamiltonian. In the $O(n)$ models the lattice symmetry only defines which sites are linked with each other and hence which pairs of spins interact. In the compass model, on the other hand, it also enters in the pair-wise interaction terms so that the lattice anisotropy is reflected in spin space.

In d spatial dimensions the compass model is defined on a simple hypercubic lattice of size $N = L^d$ with the Hamiltonian

$$\mathcal{H} = \sum_{i=1}^{N} \sum_{k=1}^{d} J_k s_i^k s_{i+\hat{\boldsymbol{k}}}^k . \tag{1.26}$$

Here s_i^k is the k-th component of a spin $\boldsymbol{s}_i$ at lattice site i. J_k is a coupling constant depending on the lattice direction k. The nearest neighbor of site i in the k-th direction is indicated by $i + \hat{\boldsymbol{k}}$. Two spins on sites neighboring in direction k only interact in their k-th components. In this sense the interaction is one-dimensional. In the classical compass model the constituent spins are represented by vectors on the unit hypersphere in d-dimensional space: $\boldsymbol{s}_i \in S^{d-1}$.

The discussion in this thesis will be limited to equal coupling constants in every direction: $J_k \equiv J$. The two-dimensional model on a square lattice of size $N = L \times L$ is in this case defined by the Hamiltonian

$$\mathcal{H}^{(2\mathrm{D})} = J \sum_{i=1}^{N} \left[s_i^x s_{i+\hat{\boldsymbol{x}}}^x + s_i^y s_{i+\hat{\boldsymbol{y}}}^y \right] . \tag{1.27}$$

Each spin $\boldsymbol{s}_i \in S^1$ can be parametrized by a polar angle $\varphi_i \in [0, 2\pi)$:

$$\boldsymbol{s}_i = \boldsymbol{s}(\varphi_i) = \begin{pmatrix} s_i^x \\ s_i^y \end{pmatrix} = \begin{pmatrix} \cos \varphi_i \\ \sin \varphi_i \end{pmatrix} . \tag{1.28}$$

In direct extension the three-dimensional model on the cubic lattice of size $N = L \times L \times L$ is defined by the Hamiltonian

$$\mathcal{H}^{(3\mathrm{D})} = J \sum_{i=1}^{N} \left[s_i^x s_{i+\hat{\boldsymbol{x}}}^x + s_i^y s_{i+\hat{\boldsymbol{y}}}^y + s_i^z s_{i+\hat{\boldsymbol{z}}}^z \right] , \tag{1.29}$$

where the spins $\boldsymbol{s}_i \in S^2$ can be parametrized by azimuthal and polar angles $\theta_i \in [0, \pi]$ and $\varphi_i \in [0, 2\pi)$:

$$\boldsymbol{s}_i = \boldsymbol{s}(\theta_i, \varphi_i) = \begin{pmatrix} s_i^x \\ s_i^y \\ s_i^z \end{pmatrix} = \begin{pmatrix} \cos\varphi_i \sin\theta_i \\ \sin\varphi_i \sin\theta_i \\ \cos\theta_i \end{pmatrix}. \tag{1.30}$$

In this work a negative coupling constant of $J = -1$ corresponding to ferromagnetic interactions has been chosen. As long as the linear lattice size L is taken even, the sign is not important on periodic lattices. The lattice is then bipartite: It can be divided into two disjoint sublattices such that within each sublattice no sites are linked with each other. The bipartition is realized by a checkerboard decomposition or its three-dimensional generalization. If a coupling sign $J > 0$ is chosen, one may then perform a simple canonical transformation by multiplying each spin in one sublattice by a factor of -1 and recover the same energy.

The classical compass models correspond to taking the limit of large spin S of the quantum mechanical compass models, where the spins would be represented by $S = 1/2$ operators

$$\boldsymbol{s}_i = \frac{\hbar}{2}(\sigma_x, \sigma_z) \qquad \text{for } d = 2, \tag{1.31}$$

$$\boldsymbol{s}_i = \frac{\hbar}{2}(\sigma_x, \sigma_y, \sigma_z) \qquad \text{for } d = 3 \tag{1.32}$$

with the Pauli matrices σ_k.

1.3.1 Symmetries and degeneracies

Due to the coupling of spins and lattice there is no global *continuous* rotational symmetry of the spins like in the $O(n)$ models. But the two- and three-dimensional classical compass model Hamiltonians remain invariant under a number of discrete transformations if open or periodic boundary conditions are assumed:

1. a) $\mathcal{H}^{(2D)}$ remains invariant under a reflection of all spins:

$$(s_i^x, s_i^y) \to (s_i^y, s_i^x) \qquad \text{for all sites } i. \tag{1.33}$$

   This is a $\mathbb{Z}_2$ symmetry.

   b) In three dimensions one may choose any plane of the $N = L^3$ cubic lattice and reflect all spins situated in that plane. Accordingly, $\mathcal{H}^{(3D)}$ remains invariant under any of the following transformations:

$$\begin{aligned} (s_i^y, s_i^z) &\to (s_i^z, s_i^x) && \text{for all sites } i \text{ with the same } x\text{-coordinate,} \\ (s_i^x, s_i^z) &\to (s_i^z, s_i^x) && \text{for all sites } i \text{ with the same } y\text{-coordinate,} \\ (s_i^x, s_i^y) &\to (s_i^y, s_i^x) && \text{for all sites } i \text{ with the same } z\text{-coordinate,} \end{aligned} \tag{1.34}$$

where in each case the remaining third component is left unchanged. This corresponds to the $(\mathbb{Z}_2)^{3L}$ symmetry group.

2. Along each chain parallel to the lattice axis k the k-th component of each spin may be flipped, while the other components are left unchanged. Depending on the dimension d, the following transformations are symmetries:

   a) $\mathcal{H}^{(2\mathrm{D})}$ is symmetric under any spin flip transformation of one of the following forms:

   $$\begin{aligned} s_i^x &\to -s_i^x && \text{for all sites } i \text{ with the same } y\text{-coordinate,} \\ s_i^y &\to -s_i^y && \text{for all sites } i \text{ with the same } x\text{-coordinate,} \end{aligned} \tag{1.35}$$

   which in combination gives a $(\mathbb{Z}_2)^{2L}$ symmetry.

   b) Due to the added third dimension $\mathcal{H}^{(3\mathrm{D})}$ is invariant under any of the following spin flip symmetry operations:

   $$\begin{aligned} s_i^x &\to -s_i^x && \text{for all sites } i \text{ with the same } y\text{- and } z\text{-coordinates,} \\ s_i^y &\to -s_i^y && \text{for all sites } i \text{ with the same } x\text{- and } z\text{-coordinates,} \\ s_i^z &\to -s_i^z && \text{for all sites } i \text{ with the same } x\text{- and } y\text{-coordinates.} \end{aligned} \tag{1.36}$$

   The symmetry group is $(\mathbb{Z}_2)^{3L^2}$.

   In the literature these symmetries are called "gauge-like", standing between the global symmetries of $O(n)$ models and local gauge symmetries [4, 17].

As a consequence of these discrete symmetries each state of the two-dimensional system is at least 2^{2L+1}-fold degenerate. In the three-dimensional compass model the symmetries lead to at least a $2^{3(L^2+L)}$-fold degeneracy of each state.

For the ground states of the classical compass models there is an additional continuous symmetry since any constant spin field $\boldsymbol{s}_i \equiv \boldsymbol{s}$ is a ground state. This can be seen by first noting that for any spin $\boldsymbol{s}_i$ the sum $\sum_{k=1}^{d} (s_i^k)^2 = 1$ is a constant. Hence one can construct a Hamiltonian $\mathcal{H}'$, which is equal to $\mathcal{H}$ of Eq. (1.26) up to an irrelevant additive constant:

$$\begin{aligned}\mathcal{H}' &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{d} J_k[s_i^k - s_{i+\hat{k}}^k]^2 \\ &= \sum_{i=1}^{N}\sum_{k=1}^{d}[J_k s_i^k s_{i+\hat{k}}^k - \frac{1}{2}J_k((s_i^k)^2 + (s_{i+\hat{k}}^k)^2)] = \mathcal{H} - J_k N \qquad (1.37)\end{aligned}$$

Clearly $\mathcal{H}'$ is minimized by any constant spin field if the J_k have negative signs.

1.3.2 Ordering

It has been shown that in the compass model the expectation value of any local quantity that is not invariant under the one-dimensional flipping symmetries (1.35) and (1.36) vanishes at any finite temperature [4]. For this reason one has $\langle \mathbf{s}_i \rangle = 0$ at any site i. This precludes global magnetic ordering: $\langle m \rangle = \langle |\frac{1}{N}\sum_i \mathbf{s}_i| \rangle \equiv 0$ at any temperature. However, one can construct quantities that are invariant under these symmetries, for instance $\langle s_i^x s_{i+\hat{x}}^x \rangle$, $\langle s_i^y s_{i+\hat{y}}^y \rangle$ and $\langle s_i^z s_{i+\hat{z}}^z \rangle$. Order parameters can be constructed from combinations of such expressions. These order parameters measure directional ordering characterized by long-ranged correlations in the direction of fluctuations in spin and lattice spaces, even though magnetic ordering is absent. Order is realized by linear spin alignment parallel to the lattice axes.

A possible choice for the order parameter for the two-dimensional compass model is

$$D = \frac{1}{N}\left| J\sum_{i=1}^{N}\left(s_i^x s_{i+\hat{x}}^x - s_i^y s_{i+\hat{y}}^y\right)\right| = \frac{1}{N}\left|E_x - E_y\right| \qquad (1.38)$$

where $E_k = J\sum_{i=1}^{N} s_i^k s_{i+\hat{k}}^k$ is the total bond energy along lattice direction k. An energy difference between bond directions directly shows a broken lattice rotation symmetry. Such a directional energy excess is measured by D. If one has $D > 0$, the system has long-ranged directional order; $D = 0$ corresponds to a disordered state. Previous Monte Carlo studies [8–10] have shown that there is a temperature-driven continuous phase transition from a high-temperature disordered phase to an directionally ordered phase at low temperatures, which apparently lies in the universality class of the two-dimensional Ising model. These findings agree with the results presented in this work in chapter 4.

An extension of D to the third dimension, that has also been employed in Monte Carlo studies of the related 120°-model on cubic lattices [18], is

$$D = \frac{1}{N}\sqrt{(E_z - E_y)^2 + (E_y - E_x)^2 + (E_z - E_x)^2}. \qquad (1.39)$$

This is a measure for the averaged directional energy excess. In a disordered state one has again $D = 0$; in a directionally ordered state D can have a maximum value of $\sqrt{2}|J|$, which is realized if all spins are perfectly aligned with the same lattice axis. See chapter 5 for the results of a study of directional ordering in the three-dimensional model.

In Fig. 4.6 on p. 88 an illustration of typical directionally ordered and disordered configurations of the two-dimensional compass model is given.

2 Fundamentals of Monte Carlo simulations

In the following the fundamental computer simulation methods used in this work shall be motivated and explained. This introductory presentation is mainly based on the textbooks by Newman and Barkema [19], Berg [20], Landau and Binder [21] and Krauth [22], as well as lecture notes by Janke [23], Katzgraber [24], Rummukainen [25] and Weinzierl [26].

2.1 Numerical integration schemes

Few statistical systems are simple enough for their partition functions to be computable analytically. Therefore one has to resort to numerical methods.

For systems with continuous degrees of freedom the partition function in Eq. (1.2) has to be expressed by a multi-dimensional integral, which can only be solved analytically in special cases. For example the phase space of the N-spin three-dimensional compass model has an dimension of $d = 2N$ and the determination of its partition function requires the evaluation of the integral

$$\mathcal{Z} = \int \frac{\mathrm{d}(\cos\theta_1)}{2} \int \frac{\mathrm{d}\varphi_1}{2\pi} \cdots \int \frac{\mathrm{d}(\cos\theta_N)}{2} \int \frac{\mathrm{d}\varphi_N}{2\pi} \mathrm{e}^{-\beta\mathcal{H}^{(3\mathrm{D})}(\{\theta_i,\varphi_i\})} \tag{2.1}$$

Such d-dimensional integrals could be solved by iteratively computing one-dimensional integrals of the form

$$I = \int\limits_a^b \mathrm{d}x f(x). \tag{2.2}$$

The most simple scheme is to partition the interval $[a, b]$ into M slices of equal width $\delta = (b - a)/M$ and to interpolate the function $f(x)$ to the k-th order in each slice, which delivers an approximation of the integral as a discrete sum. To first order the *midpoint rule* can be applied, where the area under f in the j-th slice is approximated by a rectangle of width δ and height $f\left((x_i + x_{i+1})/2\right)$. Then

$$I \approx \sum_{i=0}^{M-1} \delta \cdot f\left(\frac{x_i + x_{i+1}}{2}\right) \tag{2.3}$$

and for $M \to \infty$ the sum converges to the true integral I. It can be shown that the error due to the approximation of the function is in this case proportional to $\sim M^{-2}$. Higher order approximations have better convergence properties. For instance quadratic interpolation using Simpson's rule has an error proportional to $\sim M^{-4}$. An overview over more advanced integration schemes is given in Numerical Recipes [27].

With all these methods the error scales as $\sim M^{-n}$ for some power n. If they are applied to a d-dimensional integral, the partitioning has to be done independently in each space component. Therefore the error for the total integral scales as $\sim M^{-n/d}$. Evidently convergence becomes very slow for high dimensions d.

2.2 Simple sampling Monte Carlo

The integration of high-dimensional volumes can be made possible by so-called *Monte Carlo* methods. This term encompasses a variety of procedures that employ random variables to arrive at approximate solutions to mathematical problems.

The idea of *simple sampling* for the estimation of the integral of a function is to randomly choose the points of evaluation. While the original problem of integration is purely deterministic, this approach is inherently probabilistic and the results depend on the random numbers taken and are therefore subject to statistical uncertainty. The important difference to the approach outlined in the previous using a regularly spaced grid is that this statistical error does no longer depend on the spatial dimension, which shall be demonstrated in the following.

Consider the following d-dimensional integral of a function f

$$I = \int_{a^1}^{b^1} \mathrm{d}x_1 \cdots \int_{a^d}^{b^d} \mathrm{d}x_d f(x^1, \dots, x^d) = \int_V \mathrm{d}\boldsymbol{x} f(\boldsymbol{x}) \tag{2.4}$$

with points $\boldsymbol{x} = (x^1, \dots, x^d)$ from the hyper-rectangle $V = \{\, \boldsymbol{x} \mid a^1 \leq b^1, \dots, a^d \leq b^d \,\}$. For the Monte Carlo estimate of I first a sample of M randomly chosen points in V is taken, which are denoted by $\{\, \boldsymbol{x}_1, \dots, \boldsymbol{x}_M \,\}$. Each $\boldsymbol{x}_i$ is a realization of the d-dimensional random variable $X = (X^1, \dots, X^d)$. For simple sampling the X^j are taken independently and uniformly distributed from $[a^j, b^j]$. Their probability density function is

$$p_{X^j}(x) = \begin{cases} 1/(b^j - a^j), & \text{if } x \in [a^j, b^j], \\ 0, & \text{else.} \end{cases} \tag{2.5}$$

Since the X^j are independent, the probability density function of X is the joint probability density function of the X^j:

$$p_X(\boldsymbol{x}) = p_{X^1}(x^1) \cdots p_{X^d}(x^d). \tag{2.6}$$

The evaluation of the function $f(\boldsymbol{x_i})$ at the randomly sampled points $\boldsymbol{x_i}$ can then be understood as the realization of a function of the random variable X: $Y = f(X)$. The expectation value of Y is

$$\langle Y \rangle = \int\limits_V \mathrm{d}\boldsymbol{x} p_X(\boldsymbol{x}) f(\boldsymbol{x}) = \frac{1}{V} \int\limits_V \mathrm{d}\boldsymbol{x} f(\boldsymbol{x}) = I/V, \tag{2.7}$$

where V also stands for the measure of the region of integration. Therefore $\langle V \cdot Y \rangle = I$, which makes $V \cdot f(X)$ an unbiased estimator for I. Furthermore the variance of Y is

$$\sigma_Y^2 = \langle Y^2 \rangle - \langle Y \rangle^2 = \frac{1}{V} \int\limits_V \mathrm{d}\boldsymbol{x} \, [f(\boldsymbol{x})]^2 - \left[\frac{I}{V}\right]^2, \tag{2.8}$$

and hence one has for the variance of $V \cdot Y$:

$$\sigma_{V \cdot Y}^2 = V \int\limits_V \mathrm{d}\boldsymbol{x} \, [f(\boldsymbol{x})]^2 - I^2. \tag{2.9}$$

The simple sampling Monte Carlo estimator of I can then be defined as the sample mean of $V \cdot f(X)$ over M realizations $\boldsymbol{x}_i$ of X:

$$I_{\mathrm{MC}} = \frac{V}{M} \cdot \sum_{i=1}^{M} f(\boldsymbol{x}_i) \tag{2.10}$$

with an expectation value of

$$\langle I_{\mathrm{MC}} \rangle = \frac{V}{M} \cdot \langle \sum_{i=1}^{M} f(\boldsymbol{x}_i) \rangle = \frac{V}{M} \cdot M \cdot \langle Y \rangle = \frac{V}{M} \cdot M \cdot \frac{I}{V} = I. \tag{2.11}$$

Its variance is

$$\sigma_{I_{\mathrm{MC}}}^2 = \mathrm{Var}\left[\frac{V}{M} \cdot \sum_{i=1}^{M} f(\boldsymbol{x}_i)\right] = \frac{1}{M^2} \sum_{i=1}^{M} \sigma_{V \cdot Y}^2 = \frac{1}{M} \sigma_{V \cdot Y}^2. \tag{2.12}$$

I_{MC} is an unbiased estimator for I. To quantify its statistical uncertainty or "error" an unbiased estimate of $\sigma^2_{I_{\mathrm{MC}}}$ is needed. An unbiased estimator of σ^2_Y is the sample variance

$$s^2_M[Y] = \frac{1}{M-1}\left[\sum_{i=1}^{M} f(\boldsymbol{x}_i)^2 - M\left[\sum_{i=1}^{M}\frac{f(\boldsymbol{x}_i)}{M}\right]^2\right]. \tag{2.13}$$

Thus an unbiased estimate of $\sigma^2_{I_{\mathrm{MC}}}$ is

$$s^2_M[I_{\mathrm{MC}}] = \frac{V^2}{M}s^2_M[Y] = \frac{V^2}{M(M-1)}\left[\sum_{i=1}^{M} f(\boldsymbol{x}_i)^2 - \left[\sum_{i=1}^{M}\frac{f(\boldsymbol{x}_i)}{M}\right]^2\right]. \tag{2.14}$$

The square root of Eq. (2.14) can be taken as an estimate of the statistical uncertainty of I_{MC}. With the shorthands $\widehat{Y} = \sum_{i=1}^{M} f(\boldsymbol{x}_i)/M$ and $\widehat{Y^2} = \sum_{i=1}^{M} f(\boldsymbol{x}_i)^2/M$ it can be stated as

$$\delta I_{\mathrm{MC}} = V \cdot \sqrt{\frac{\widehat{Y^2} - \widehat{Y}^2}{M-1}}. \tag{2.15}$$

The important result is that this error really does not depend on the spatial dimension and scales merely as $\sim M^{-1/2}$. However, it must be stressed that with Monte Carlo integration it is only possible to acquire a probabilistic error bound: One can only give a probability that the estimate lies within a certain range of the true value.

While it is clear that Monte Carlo integration outperforms traditional integration schemes for high dimensions, the presented approach of simple sampling still suffers from very slow convergence of the estimates for the integral value and its uncertainty if the variance of the integrand is large. As the domain of integration is sampled uniformly, regions with little weight for the integral are sampled with the same probability as the important regions, where the function has high values.

2.3 Importance sampling

To overcome the shortcomings of simple sampling a method is desirable that samples the random points preferably in the region with the highest contribution to the result. In doing so it is necessary to ensure that the sampling of random numbers is still done in a manner which does not bias the result. A procedure complying with these requirements is known as *importance sampling*.

If one takes a probability density function $p(\boldsymbol{x})$ that still fulfills $p(\boldsymbol{x}) > 0$ if and only if $\boldsymbol{x} \in V$ and $\int_V \mathrm{d}\boldsymbol{x} p(\boldsymbol{x}) = 1$, but no longer has to be uniform, the integral of Eq. (2.4) can be written as

$$I = \int_V \mathrm{d}\boldsymbol{x} f(\boldsymbol{x}) = \int_V \mathrm{d}\boldsymbol{x} \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} p(\boldsymbol{x}) = \langle f/p \rangle_p. \tag{2.16}$$

So mathematically importance sampling amounts to a change of integration variables. A generalization of Eq. (2.10) then gives a new estimator

$$I_{\mathrm{IS}} = \frac{1}{M} \sum_{i=1}^{M} \frac{f(\boldsymbol{y}_i)}{p(\boldsymbol{y}_i)}, \tag{2.17}$$

where the $\boldsymbol{y}_i$ are now chosen p-distributed. Then the expectation value results as

$$\langle I_{\mathrm{IS}} \rangle_p = \frac{1}{M} \sum_{i=1}^{M} \left\langle \frac{f(\boldsymbol{y}_i)}{p(\boldsymbol{y}_i)} \right\rangle_p = \frac{1}{M} \sum_{i=1}^{M} \int_V \mathrm{d}\boldsymbol{x} \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} p(\boldsymbol{x}) = I. \tag{2.18}$$

and the variance is

$$\begin{aligned} \sigma_{I_{\mathrm{IS}}}^2 &= \frac{1}{M} \mathrm{Var}\left[\frac{f}{p}\right] = \frac{1}{M} \left[\int_V \mathrm{d}\boldsymbol{x} \left(\frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} \right)^2 p(\boldsymbol{x}) - \left(\int_V \mathrm{d}\boldsymbol{x} \, \frac{f(\boldsymbol{x})}{p(\boldsymbol{x})} p(\boldsymbol{x}) \right)^2 \right] \\ &= \frac{1}{M} \left[\int_V \mathrm{d}\boldsymbol{x} \, \frac{f(\boldsymbol{x})^2}{p(\boldsymbol{x})} - I^2 \right]. \end{aligned} \tag{2.19}$$

Here the effect of a good choice of $p(x)$ is clear: I_{IS} is minimized with $p(x) = f(x)/I$. This is of course a rather theoretical result, as knowing I means having already solved the problem at hand. Still, choosing p as close to f as it is practically possible leads to a drastically reduced statistical error.

2.4 Markov chain Monte Carlo

Importance sampling of thermodynamic observables

As outlined in Sec. 1.1 the expectation value of an observable $\mathcal{O}$ in a statistical system is given by

$$\langle \mathcal{O} \rangle = \frac{\int \mathrm{d}\boldsymbol{x} \, \mathcal{O}(\boldsymbol{x}) \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})}}{\int \mathrm{d}\boldsymbol{x} \, \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})}}. \tag{2.20}$$

If the phase space points $\boldsymbol{x}_i$ are sampled with a probability distribution function $\mathcal{P}(\boldsymbol{x})$, the important sampling estimator of Eq. (2.17) can be used in numerator and denominator to estimate $\langle \mathcal{O} \rangle$:

$$\widehat{\mathcal{O}} = \frac{\sum_{i=1}^{M} \mathcal{O}(\boldsymbol{x}_i) \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x}_i)} \mathcal{P}(\boldsymbol{x}_i)^{-1}}{\sum_{i=1}^{M} \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x}_i)} \mathcal{P}(\boldsymbol{x}_i)^{-1}}. \tag{2.21}$$

The ideal choice for $\mathcal{P}(\boldsymbol{x})$ is $\mathcal{P}_{\text{eq}}(\boldsymbol{x}) = \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{x})} / \mathcal{Z}$ since then Eq. (2.21) reduces to just the sample mean

$$\widehat{\mathcal{O}} = \frac{1}{M} \sum_{i=1}^{M} \mathcal{O}(\boldsymbol{x}_i) = \overline{\mathcal{O}}, \tag{2.22}$$

where the states $\boldsymbol{x}_i$ are now sampled according to the Boltzmann distribution $\mathcal{P}_{\text{eq}}$.

The remaining problem is that the distribution $\mathcal{P}_{\text{eq}}$ cannot be sampled directly. While there are methods available to generate pseudo-random numbers from certain classes of one-dimensional distributions (see Ref. [27]), no general-purpose method applicable to high dimensions is known. Also, again the normalization constant in $\mathcal{P}_{\text{eq}}$, which is the partition function $\mathcal{Z}$, is unknown.

2.4.1 Markov processes

To sample $\mathcal{P}_{\text{eq}}$ more involved schemes are necessary. The standard solutions rely on *Markov processes* to produce the states to be sampled. A Markov process is a stochastic process that generates a sequence of states $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ in such a way that once the system is in a state $\boldsymbol{x}_m$ the next state $\boldsymbol{x}_n$ is generated randomly in a fashion that only depends on $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$. Formally the *transition probability* has to satisfy

$$W(\boldsymbol{x}_m \to \boldsymbol{x}_n) = P(X = \boldsymbol{x}_n | X = \boldsymbol{x}_m). \tag{2.23}$$

Thus in contrast to the previously presented sampling methods the sampled states are no longer independent of each other. But a newly generated state depends only on the previous one, not on the whole history of this trajectory in state space.

Here and in the following a discrete and finite state space is assumed to keep the notation simple. While for continuous systems there are equivalent results, one should keep in mind that in the context of Monte Carlo simulations all continuous systems are eventually discretized for treatment in a computer, which can only deal with countable sets of states.

If the transition probabilities satisfy the following conditions:

$$(i)\ W(\boldsymbol{x}_m \to \boldsymbol{x}_n) \geq 0 \quad \text{for all } m, n, \tag{2.24a}$$

$$(ii)\ \sum_n W(\boldsymbol{x}_m \to \boldsymbol{x}_n) = 1 \quad \text{for all } m, \tag{2.24b}$$

the $W(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ form a stochastic matrix W. The sequence of states $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ generated by the Markov process is also called a *Markov chain*. It has to be set up in a way that regardless of the initial configuration $\boldsymbol{x}_0$ eventually the sequence will converge towards a succession of states that appear with probabilities according to the Boltzmann distribution. To achieve this two additional conditions have to be placed on the Markov process: *Ergodicity* and *detailed balance*.

Ergodicity

The condition of ergodicity is the requirement that from a state $\boldsymbol{x}_m$ any other state $\boldsymbol{x}_n$ can eventually be reached in the Markov chain. Some of the individual transition probabilities may well be set to zero, but ultimately there must be a trajectory of non-zero transition probabilities between any two states. If this were not the case, there would be states $\boldsymbol{x}_k$ that are never sampled, but in the Boltzmann distribution a non-zero probability is assigned to every state.

Detailed balance

If a pseudo-time is measured in steps along a Markov chain, a state $\boldsymbol{x}_m$ occurs at the k'th time step with probability $p_m(k)$. The evolution of these probabilities in the Markov process is described by a *master equation* for each state $\boldsymbol{x}_m$:

$$p_m(k+1) = p_m(k) + \sum_n p_n(k) W(\boldsymbol{x}_n \to \boldsymbol{x}_m) - \sum_n p_m(k) W(\boldsymbol{x}_m \to \boldsymbol{x}_n). \tag{2.25}$$

The first term on the right hand side describes the rate of transitions into $\boldsymbol{x}_m$, the second term describes the rate of transitions from $\boldsymbol{x}_m$ into the other states. The goal is that for $k \to \infty$ the probabilities $p_m(k)$ reach a stationary distribution corresponding to the equilibrium distribution of the system. Requiring $p_m(k) = p_m = \text{const}$ one has:

$$\sum_n p_m W(\boldsymbol{x}_m \to \boldsymbol{x}_n) = \sum_n p_n W(\boldsymbol{x}_n \to \boldsymbol{x}_m). \tag{2.26}$$

Eq. (2.26) is the condition of *balance* that has to be imposed on the transition probabilities of the Markov process. Any set of transition probabilities $W(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ that satisfies Eq. (2.26) leads to a stationary distribution p_m. However, as it turns out this condition is not strong enough to warrant that the probability distribution will tend to this distribution p_m from any initial state of the system.

Making use of condition (2.24b), Eq. (2.25) can be simplified to

$$p_m(k+1) = \sum_n p_n(k) W(\boldsymbol{x}_n \to \boldsymbol{x}_m). \tag{2.27}$$

Introducing the vector $\boldsymbol{p}(k)$ with the entries $p_n(k)$ and writing W for the stochastic matrix formed by the $W(\boldsymbol{x}_m \to \boldsymbol{x}_n)$, this can be written as

$$\boldsymbol{p}(k+1) = W \cdot \boldsymbol{p}(k). \tag{2.28}$$

If an equilibrium state $\boldsymbol{p}(\infty)$ is reached by the process as $k \to \infty$, it can be characterized as an eigenvector of W to the eigenvalue 1:

$$\boldsymbol{p}(\infty) = W \cdot \boldsymbol{p}(\infty). \tag{2.29}$$

Apart from this there also remains the possibility that the process reaches a *limit circle*, in which the probability distribution $\boldsymbol{p}$ rotates around a set of different values. Then $\boldsymbol{p}(\infty)$ would satisfy

$$\boldsymbol{p}(\infty) = W^n \cdot \boldsymbol{p}(\infty), \tag{2.30}$$

where n is the length of this limit circle and $\boldsymbol{p}(\infty)$ would be an eigenvector to an eigenvalue which is a complex root of one.

The occurrence of limit circles besides the desired equilibrium distribution p_m must be avoided to make sure that the states generated by the Markov process will have the right probability distribution. In order to achieve this a stronger condition than Eq. (2.26) is enforced on the transition probabilities:

$$p_m W(\boldsymbol{x}_m \to \boldsymbol{x}_n) = p_n W(\boldsymbol{x}_n \to \boldsymbol{x}_m). \tag{2.31}$$

This is the condition of *detailed balance*. By summing over n on both sides of Eq. (2.31), Eq. (2.26) is recovered. So transition probabilities that fulfill detailed balance always comply with the condition of balance. Condition (2.31) implies that the process has to be reversible. As it forces transitions from $\boldsymbol{x}_m$ to $\boldsymbol{x}_n$ to be equally probable as transitions from $\boldsymbol{x}_n$ to $\boldsymbol{x}_m$, limit circles are prevented: In a limit circle at a particular time step there would have to be more transitions into a certain state than out of it on average.

Once limit circles are excluded, it is straightforward to show that the system will always reach $\boldsymbol{p}(\infty)$ as $k \to \infty$, if one accepts that since W is a stochastic matrix, the largest absolute value of its eigenvalues is one. Since limit circles are excluded this eigenvalue has to be exactly one. Because of ergodicity there can only be one eigenvector of eigenvalue 1 satisfying Eq. (2.29) because else the system would subdivide into more than one mutually inaccessible subsets.

By iterating Eq. (2.28) from $k = 0$ one arrives at

$$\boldsymbol{p}(k) = W^k \cdot \boldsymbol{p}(0), \tag{2.32}$$

where the initial distribution $\boldsymbol{p}(0)$ can be expressed as a linear combination of the right eigenvectors $\boldsymbol{v}_i$ of W:

$$\boldsymbol{p}(0) = \sum_i \mu_i \boldsymbol{v}_i \tag{2.33}$$

with $\lambda_i \boldsymbol{v}_i = W \cdot \boldsymbol{v}_i$, so that

$$\boldsymbol{p}(k) = W^k \cdot \sum_i \mu_i \boldsymbol{v}_i = \sum_i \mu_i \lambda_i^k \boldsymbol{v}_i. \tag{2.34}$$

Now as $k \to \infty$, $\boldsymbol{p}(k)$ will be dominated by the largest eigenvalue λ_0 of W. The convergence is exponential.

As this eigenvalue is 1 and pertains to the unique stationary state $\boldsymbol{p}(\infty)$, it is clear that $\boldsymbol{p}(k)$ tends towards the equilibrium distribution of the system.

Choosing transition probabilities

By choosing transition probabilities that satisfy the condition of detailed balance (2.31) the Markov process can be made to generate states distributed according to any distribution p_m. In thermal equilibrium they should follow the Boltzmann distribution, which forces the transition probabilities to satisfy the following more specific condition of detailed balance:

$$\frac{W(\boldsymbol{x}_m \to \boldsymbol{x}_n)}{W(\boldsymbol{x}_n \to \boldsymbol{x}_m)} = \frac{p_n}{p_m} = \mathrm{e}^{-\beta(E_n - E_m)} \equiv \mathrm{e}^{-\beta \Delta E(\boldsymbol{x}_m \to \boldsymbol{x}_n)}, \tag{2.35}$$

where $E_m = \mathcal{H}(\boldsymbol{x}_m)$ and $E_n = \mathcal{H}(\boldsymbol{x}_n)$ label the energies in the states $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$. Of importance are only the *relative* energy differences ΔE between the states.

In the practical implementation of a Markov chain in a Monte Carlo simulation in each step a new state $\boldsymbol{x}_n$ has to be proposed based on an old state $\boldsymbol{x}_m$. Let $g(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ be the probability of this proposal and $A(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ be the probability of acceptance of the new state. Then the transition probability can be expressed as

$$W(\boldsymbol{x}_m \to \boldsymbol{x}_n) = g(\boldsymbol{x}_m \to \boldsymbol{x}_n)A(\boldsymbol{x}_m \to \boldsymbol{x}_n) \tag{2.36}$$

In principle any choice of $g(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ is valid as long as it does not break ergodicity and the acceptance probabilities $A(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ still satisfy detailed balance. However, there may be concerns of performance in an actual simulation that make certain choices more favorable than others.

Thermalization

If the Markov process has been set up correctly, one can start to generate a sequence of states, beginning with some initial configuration $\boldsymbol{x}_0$. Eventually after sufficiently many steps have been taken, equilibrium will be reached and one can start to take measurements of observables to be used in estimations according to Eq. (2.22). The period allowed for equilibration is also called the *thermalization phase*. Generally, the question of how long one has to wait until thermalization is completed is not trivial. It can be judged by monitoring the pseudo-time evolution of suitable observables, which should converge to their constant behavior in equilibrium. To stay on the safe side, this can be repeated for different initial conditions and different sets of pseudo-random numbers used in the simulation. To warrant the correctness of the results obtained in simulations, it is advisable to rather discard more samples than necessary for equilibration than to risk starting the actual measurements too early. A rule of thumb, that has been followed in this work, is to spend at least 10% of the simulation time in the period of thermalization.

2.4.2 The Metropolis algorithm

A viable choice for the transition probabilities is given by the famous *Metropolis* algorithm [28]. Here $A(\boldsymbol{x}_m \to \boldsymbol{x}_n) = \min\{1, \mathrm{e}^{-\beta \Delta E(\boldsymbol{x}_m \to \boldsymbol{x}_n)}\}$, i.e.

$$A(\boldsymbol{x}_m \to \boldsymbol{x}_n) = \begin{cases} \mathrm{e}^{-\beta \Delta E(\boldsymbol{x}_m \to \boldsymbol{x}_n)}, & \text{if } \Delta E > 0, \\ 1, & \text{else.} \end{cases} \tag{2.37}$$

If the update from $\boldsymbol{x}_m$ to the proposed new state $\boldsymbol{x}_n$ lowers the energy, it is always accepted. Updates that increase the energy are accepted with a probability depending on temperature. This can be understood as the influence of entropy. In thermal equilibrium the free energy $F = E - TS$ is minimized and not the internal energy E.

In the Metropolis algorithm the transition probabilities $g(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ are chosen such that

$$\frac{g(\boldsymbol{x}_m \to \boldsymbol{x}_n)}{g(\boldsymbol{x}_n \to \boldsymbol{x}_m)} = 1. \tag{2.38}$$

It is then straightforward to see that the Metropolis transition probabilities satisfy the condition of detailed balance (2.35):

$$\begin{aligned}\frac{W(\boldsymbol{x}_m \to \boldsymbol{x}_n)}{W(\boldsymbol{x}_n \to \boldsymbol{x}_m)} &= \frac{g(\boldsymbol{x}_m \to \boldsymbol{x}_n)A(\boldsymbol{x}_m \to \boldsymbol{x}_n)}{g(\boldsymbol{x}_n \to \boldsymbol{x}_m)A(\boldsymbol{x}_n \to \boldsymbol{x}_m)} = \frac{A(\boldsymbol{x}_m \to \boldsymbol{x}_n)}{A(\boldsymbol{x}_n \to \boldsymbol{x}_m)} \\ &= \left\{\begin{matrix} \dfrac{1}{\mathrm{e}^{-\beta(-\Delta E)}}, & \text{if } \Delta E \le 0 \\ \dfrac{\mathrm{e}^{-\beta\Delta E}}{1}, & \text{if } \Delta E > 0 \end{matrix}\right\} = \mathrm{e}^{-\beta\Delta E}.\end{aligned} \tag{2.39}$$

To avoid large energy differences $\Delta E > 0$, which would lead to low acceptance probabilities, the Metropolis algorithm is usually implemented as a *local update* algorithm: In each Monte Carlo step only a single degree of freedom is altered. The $g(\boldsymbol{x}_m \to \boldsymbol{x}_n)$ are taken equal for all states $\boldsymbol{x}_n$ possible in this way and zero otherwise, automatically satisfying Eq. (2.38). Thus no huge changes of the system configuration occur and the energies before and after the update will be close. For a classical system of N spins $\boldsymbol{s}_i$ on a lattice one would choose a site i_0 in each step and propose only a change of the spin $\boldsymbol{s}_{i_0}$. A sequence of N such local updates is commonly called a Monte Carlo *sweep* and can be thought of as an update of the whole spin system.

Besides the Metropolis update many other valid choices of transition probabilities can be made, which may be more adapted to the problem at hand. For instance non-local *cluster* algorithms and methods for generalized ensemble sampling can lead to significantly improved performance. In chapter 3 several algorithms adapted to the simulation of the classical compass model will be presented.

2.5 Statistical uncertainties and autocorrelations

As mentioned above the configurations which have been sampled in a Markov chain are not independent of each other. Therefore observable measurements made from these configurations are correlated. The statistical uncertain of estimates computed from them are related to these correlations. An analysis of this connection for Monte Carlo simulations has been proposed for the first time by Müller-Krumbhaar and Binder [29]. A recent general review is given by Janke [30].

The expectation value $\langle \mathcal{O} \rangle$ of an observable $\mathcal{O}$ can be estimated by the sample mean $\overline{\mathcal{O}} = \frac{1}{M}\sum_{i=1}^{M} \mathcal{O}_i$ of the values $\mathcal{O}_i = \mathcal{O}(\boldsymbol{x}_i)$, which have been measured from the successively sampled system configurations x_i. $\overline{\mathcal{O}}$ is a random variable and for finite sample sizes M it fluctuates around the theoretically expected value $\langle \mathcal{O} \rangle$. To assess the quality of the results obtained from a simulation, it is necessary to quantify the statistical uncertainty caused by these fluctuations. An appropriate measure is the variance of $\overline{\mathcal{O}}$:

$$\sigma^2_{\overline{\mathcal{O}}} = \langle [\overline{\mathcal{O}} - \langle \overline{\mathcal{O}} \rangle]^2 \rangle = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2. \tag{2.40}$$

It can be estimated by conducting many simulations with identical setup, which only differ in the (pseudo-)random numbers that have been used. However, this approach is often prohibited by the computational cost. Instead $\sigma^2_{\overline{\mathcal{O}}}$ can also be computed from the series of measurements $\mathcal{O}_i$ taken in a single simulation:

$$\begin{aligned}
\sigma^2_{\overline{\mathcal{O}}} = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 &= \frac{1}{M^2} \left\langle \left[\sum_{i=1}^{M} \mathcal{O}_i \right]^2 \right\rangle - \frac{1}{M^2} \left[\sum_{i=1}^{M} \langle \mathcal{O}_i \rangle \right]^2 \\
&= \frac{1}{M^2} \left[\sum_{i=1}^{M} \langle \mathcal{O}_i^2 \rangle + \sum_{\substack{i,j=1 \\ i \neq j}}^{M} \langle \mathcal{O}_i \mathcal{O}_j \rangle \right] - \frac{1}{M^2} \left[\sum_{i=1}^{M} \langle \mathcal{O}_i \rangle^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{M} \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right] \\
&= \frac{1}{M^2} \sum_{i=1}^{M} \left[\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right] + \frac{1}{M^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{M} \left[\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right].
\end{aligned} \tag{2.41}$$

Since in the first sum $\langle \mathcal{O}_i \rangle = \langle \mathcal{O} \rangle$ and $\langle \mathcal{O}_i^2 \rangle = \langle \mathcal{O}^2 \rangle$ hold for all i and since the second sum is symmetric in i and j, Eq. (2.41) can be rewritten as follows:

$$\begin{aligned}
\sigma^2_{\overline{\mathcal{O}}} &= \frac{1}{M} \left[\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right] + \frac{2}{M^2} \sum_{i=1}^{M} \sum_{t=1}^{M-i} \left[\langle \mathcal{O}_i \mathcal{O}_{i+t} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_{i+t} \rangle \right] \\
&= \frac{1}{M} \sigma^2_{\mathcal{O}_i} + \frac{2}{M^2} \sum_{i=1}^{M} \sum_{t=1}^{M-i} \left[\langle \mathcal{O}_i \mathcal{O}_{i+t} \rangle - \langle \mathcal{O}_i \rangle^2 \right].
\end{aligned} \tag{2.42}$$

Here the first term has been identified with the variance of $\mathcal{O}_i$ and in the second term use of the identity $\langle \mathcal{O}_{i+t} \rangle = \langle \mathcal{O}_i \rangle$ has been made. By reordering the remaining two sums in the second term one obtains:

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{M} \sigma^2_{\mathcal{O}_i} + \frac{2}{M^2} \sum_{t=1}^{M-1} \sum_{i=1}^{M-t} \left[\langle \mathcal{O}_i \mathcal{O}_{i+t} \rangle - \langle \mathcal{O}_i \rangle^2 \right]. \tag{2.43}$$

The $\mathcal{O}_i$ are taken from an equilibrated Markov chain, which is the realization of a stationary stochastic process. Hence there is pseudo-time translational invariance and the $\langle \mathcal{O}_i \mathcal{O}_{i+t} \rangle$ under the sum are equal for all i:

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{M}\sigma^2_{\mathcal{O}_i} + \frac{2}{M^2}\sum_{t=1}^{M-1}(M-t)\left[\langle \mathcal{O}_1 \mathcal{O}_{1+t}\rangle - \langle \mathcal{O}_1\rangle^2\right]. \tag{2.44}$$

With the introduction of the *normalized autocorrelation function*

$$C_{\mathcal{O}}(t) = \frac{\langle \mathcal{O}_1 \mathcal{O}_{1+t}\rangle - \langle \mathcal{O}_1\rangle^2}{\langle \mathcal{O}_1^2\rangle - \langle \mathcal{O}_1\rangle^2} = \frac{\langle \mathcal{O}_1 \mathcal{O}_{1+t}\rangle - \langle \mathcal{O}_1\rangle^2}{\sigma^2_{\mathcal{O}_i}} \tag{2.45}$$

Eq. (2.44) can be expressed as

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{M}\sigma^2_{\mathcal{O}_i} + \frac{2}{M}\sum_{t=1}^{M-1}\left(\frac{M-t}{M}\right)\sigma^2_{\mathcal{O}_i}\cdot C_{\mathcal{O}}(t). \tag{2.46}$$

By additionally introducing the *integrated autocorrelation time*

$$\tau_{\mathcal{O},\text{int}} = \sum_{t=1}^{M}\left(1-\frac{t}{M}\right)\cdot C_{\mathcal{O}}(t) \tag{2.47}$$

Eq. (2.46) can be written as

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{\sigma^2_{\mathcal{O}_i}}{M}\left(1+2\tau_{\mathcal{O},\text{int}}\right) = \frac{\sigma^2_{\mathcal{O}_i}}{M/g_{\mathcal{O}}}. \tag{2.48}$$

The divisor $g_{\mathcal{O}} \equiv 1 + 2\tau_{\mathcal{O},\text{int}}$ can be understood as a statistical inefficiency:

For a series of uncorrelated measurements $\mathcal{O}_i$ one has $C_{\mathcal{O}}(t) \equiv 0$ and the variance of the mean is equal to the variance of the individual measurements scaled by the total number of measurements: $\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_i}/M$.

In general however, data from a Markov Chain Monte Carlo simulation has temporal correlations, which leads to $g_{\mathcal{O}} > 1$ and an amplification of the statistical uncertainty. The variance can be compared to the uncorrelated case by writing $M_{\text{eff}} = M/g_{\mathcal{O}}$ for the statistically effective size of the sample, which leads to $\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_i}/M_{\text{eff}}$. Only measurements $g_{\mathcal{O}}$ time steps apart are approximately uncorrelated.

By the central limit theorem the probability distribution of the mean value $\overline{\mathcal{O}}$ will assume a Gaussian distribution with a squared width of $\sigma^2_{\overline{\mathcal{O}}}$ in the asymptotic limit of large M. The statistical uncertainty of $\overline{\mathcal{O}}$ is usually given by the standard deviation or standard error $\delta_{\overline{\mathcal{O}}} = \sqrt{\sigma^2_{\overline{\mathcal{O}}}}$. If the assumption of a Gaussian distribution is valid, 68.2% of a set of equivalently performed simulations would give a mean value in the range $[\overline{\mathcal{O}} - \delta_{\overline{\mathcal{O}}}, \overline{\mathcal{O}} + \delta_{\overline{\mathcal{O}}}]$ as a result, the so-called one-sigma interval.

2.5.1 Critical and supercritical slowing down

For Markov processes the normalized autocorrelation function defined in Eq. (2.45) is expected to decay exponentially for large time separations t:

$$C_{\mathcal{O}}(t) = \frac{\langle \mathcal{O}_1 \mathcal{O}_{1+t} \rangle - \langle \mathcal{O}_1 \rangle^2}{\sigma^2_{\mathcal{O}_i}} \xrightarrow{t \to \infty} c \mathrm{e}^{-t/\tau_{\mathcal{O},\mathrm{exp}}}, \tag{2.49}$$

where c is some constant prefactor and the *exponential autocorrelation time* $\tau_{\mathcal{O},\mathrm{exp}}$ is defined. Typically at small values of t, additional modes contribute and the behavior of $C_{\mathcal{O}}(t)$ is not purely exponential. If the weight of those modes is not too high, $\tau_{\mathcal{O},\mathrm{exp}}$ is approximately equal to the integrated autocorrelation time $\tau_{\mathcal{O},\mathrm{int}}$ defined in Eq. (2.47).

Close to the critical point of a continuous phase transition, typically the autocorrelation time scales with the spatial correlation length ξ as

$$\tau_{\mathcal{O},\mathrm{exp}} \propto \min(L, \xi)^z \tag{2.50}$$

with a dynamical critical exponent $z \leq 0$. In finite systems ξ is capped by the linear system size L. Due to the divergence of $\xi \propto |T - T_c|^{-\nu}$ near the critical temperature in a temperature-driven transition, the autocorrelation time diverges as well for $T \to T_\mathrm{c}$:

$$\tau_{\mathcal{O},\mathrm{exp}} \propto |T - T_c|^{-\nu z}. \tag{2.51}$$

This effect is called *critical slowing down*.

The value of z depends strongly on the update algorithm used in a Monte Carlo simulation. For a spin system simulated with local dynamics one usually has $z \approx 2$. Non-local algorithms adapted to the model, that for instance make use of cluster updates, can have considerably lower values of z.

If a system which undergoes a first-order phase transition is simulated with Monte Carlo methods, a different phenomenon appears due to the coexistence of multiple phases at the finite size pseudo-transition point. In the course of the simulation the system can switch from one metastable pure phase to the other, but to do so it must cross a mixed-phase region in state space. The probability of configurations from that region is exponentially suppressed in relation to the weight of the pure phases by an additional Boltzmann factor $\mathrm{e}^{-2\sigma L^{d-1}}$, where σ is the interface tension between the coexisting phases and L^{d-1} is the projected area of their interface. The factor 2 is assumed with periodic boundary conditions, where the system topology enforces an even number of interfaces.

In a regular canonical simulation, the time spent between crossings of the suppressed region scales in inverse proportion to its relative probability. Therefore the autocorrelation time scales exponentially with the system size:

$$\tau_{\mathcal{O},\mathrm{exp}} \propto e^{2\sigma L^{d-1}}. \tag{2.52}$$

This effect is known as *supercritical slowing down*. In contrast to critical slowing down it is caused directly by the shape of the probability distribution and cannot be overcome by non-local updates. The impact of supercritical slowing down can however often be reduced by methods such as multicanonical sampling[76, 77], in which adapted weights are used to enhance the sampling of configurations with suppressed probabilities.

There is an immediate impact of high autocorrelation times on the effective sample sizes in the statistical uncertainties of observable expectation values. Therefore it is clear that special care in the choice and design of algorithms is necessary to obtain reliable results in the vicinity of phase transitions.

2.5.2 Numerical estimation of autocorrelation times

To estimate the autocorrelation function from empirical time series data, the expectation values in Eq. (2.44) can be replaced by mean values of the respective quantities:

$$\widehat{C}_{\mathcal{O}}(t) = \frac{\overline{\mathcal{O}_i \mathcal{O}_{i+t}} - \overline{\mathcal{O}}^2}{\overline{\mathcal{O}^2} - \overline{\mathcal{O}}^2}. \tag{2.53}$$

But for large time separations t the statistical uncertainty of this estimator becomes strong and as a result empirically determined autocorrelation functions are usually very noisy. This makes the determination of $\tau_{\mathcal{O},\mathrm{exp}}$ by a fit to the exponential function rather hard, as the range used must be limited both at low t to exclude the influence of additional modes before the exponential decay sets in and at high t to exclude noisy data.

Instead one usually resorts to estimating the integrated autocorrelation time $\tau_{\mathcal{O},\mathrm{int}}$. Since $C_{\mathcal{O}}(t)$ decays exponentially with t, the factor $(1 - t/M)$ in Eq. (2.47), which becomes important for high t, is usually left out. Also the sum is restricted by an upper bound $t_{\max}$ to exclude contributions where statistical noise would dominate:

$$\widehat{\tau}_{\mathcal{O},\mathrm{int}}(t_{\max}) = \sum_{t=1}^{t_{\max}} \widehat{C}_{\mathcal{O}}(t). \tag{2.54}$$

The upper limit $t_{\max}$ can be chosen by various approaches, for instance by self-consistently cutting off the summation once e.g. $t_{\max} \geq 6\widehat{\tau}_{\mathcal{O},\mathrm{int}}(t_{\max})$. A simpler alternative, that is independent of such a choice, is to stop accumulating to the sum once the autocorrelation function crosses zero. Except for a few special situations negative values of the correlation function can be considered unphysical and their occurrence can be understood as an indication of the dominance of statistical noise such that the remainder of the autocorrelation function will be indistinguishable from zero.

2.5.3 The jackknife resampling method

For quantities that have not been measured directly in the simulation, but are computed as functions of the observables or their moments, the previously outlined methods are not sufficient to estimate their statistical uncertainties. Examples are the heat capacity $C = k_{\mathrm{B}}\beta^2\langle(E - \langle E\rangle)^2\rangle$ or quantities obtained by a reweighting process like the one introduced in Sec. 3.6. A systematic way to estimate the uncertainty of arbitrary functions $f(\mathcal{O}^{(a)}, \langle\mathcal{O}^{(a)}\rangle, \dots, \mathcal{O}^{(b)}, \langle\mathcal{O}^{(b)}\rangle, \dots)$ is the *jackknife* method. It works by dividing the time series $\{\mathcal{O}_i^{(a)}, \mathcal{O}_i^{(b)}, \dots\}_{i=1}^{M}$ into N_{B} blocks of length $k \gg \tau$, which should be larger than the autocorrelation times of all observables.

Then for $j = 1, \dots, N_{\mathrm{B}}$ the quantity f is estimated from the data set that is obtained by leaving out the j'th block and including all the other blocks. Thus from the original time series a resampled data set of N_{B} jackknife block estimates $\widehat{f}^{(j)}$ is constructed. Each $\widehat{f}^{(j)}$ is computed from a sample of size $M - k$. Compared to a simple blocking or binning scheme, where estimates would be computed directly on the small blocks of length k, this approach reduces the biasing effects of small sample sizes.

The effect of temporal correlations is eliminated by the blocking procedure, but since the same data is re-used $N_{\mathrm{B}} - 1$ times, there are still trivial correlations between the $\widehat{f}^{(j)}$. If these correlations were not present, an estimate for the squared statistical uncertainty of $\widehat{f}$ would be given by the unbiased estimator for the variance of the mean of the $\widehat{f}^{(j)}$:

$$\widehat{\sigma_f^2}' = \frac{1}{N_{\mathrm{B}}(N_{\mathrm{B}} - 1)} \sum_{j=1}^{N_{\mathrm{B}}} (\widehat{f}^{(j)} - \overline{f^{(j)}}), \tag{2.55}$$

where $\overline{f^{(j)}} = \frac{1}{N_{\mathrm{B}}} \sum_{j=1}^{N_{\mathrm{B}}} \widehat{f}^{(j)}$. To account for the trivial correlations, it has to be scaled by a factor of $(N_{\mathrm{B}} - 1)^2$, so that the final jackknife estimate for the squared statistical uncertainty is:

$$\widehat{\delta^2 f} = \frac{N_{\mathrm{B}} - 1}{N_{\mathrm{B}}} \sum_{j=1}^{N_{\mathrm{B}}} (\widehat{f}^{(j)} - \overline{f^{(j)}}). \tag{2.56}$$

Since the jackknife method generally over-estimates the statistical uncertainty, it can be considered a relatively safe choice to control their magnitude. See Efron's monograph [31] for a detailed treatment of jackknife resampling.

3 Methods applied to the compass model

3.1 Observables

The basic quantities measured in the Monte Carlo simulations of the classical compass model in both two and three dimensions in this work are the total energy E and the directional ordering parameter D introduced in Sec. 1.3.2. Writing $E_x = J \sum_{i=1}^{N} s_i^x s_{i+\hat{x}}^x$, $E_y = J \sum_{i=1}^{N} s_i^y s_{i+\hat{y}}^y$ and $E_z = J \sum_{i=1}^{N} s_i^z s_{i+\hat{z}}^z$ for the total directional bond energies, the quantities to be measured are in two dimensions:

$$
\begin{aligned}
E^{(2\mathrm{D})} &= E_x + E_y, \\
D^{(2\mathrm{D})} &= \frac{1}{N} \left| E_x - E_y \right|
\end{aligned}
\tag{3.1}
$$

and in three dimensions

$$
\begin{aligned}
E^{(3\mathrm{D})} &= E_x + E_y + E_z, \\
D^{(3\mathrm{D})} &= \frac{1}{N} \sqrt{(E_z - E_y)^2 + (E_y - E_x)^2 + (E_z - E_x)^2}.
\end{aligned}
\tag{3.2}
$$

A number of derived quantities can be estimated from the measured time series of D and E to help with the analysis of the phase transitions. Among these are the specific heat capacity C/N, the susceptibility χ of D and the Binder parameter Q_2 for D:

$$
\begin{aligned}
C/N &= \frac{k_{\mathrm{B}} \beta^2}{N} \left[\langle E^2 \rangle - \langle E \rangle^2 \right], \\
\chi &= N \left[\langle D^2 \rangle - \langle D \rangle^2 \right], \\
Q_2 &= 1 - \frac{1}{3} \frac{\langle D^4 \rangle}{\langle D^2 \rangle^2}.
\end{aligned}
\tag{3.3}
$$

The factor of N in the definition of χ is included to give the susceptibility per spin as also D is an average over the lattice. Compared with the general definition (1.14) in Sec. 1.2.1 the leading factor of β is omitted. This does not change the thermodynamic limit extrapolated from the finite-size maxima of the susceptibility, but by experience can improve the scaling behavior [32].

3.2 Choice of boundary conditions

Most simulations of statistical models on finite lattices are carried out with periodic boundary conditions, where the topology of a torus is used. The assumption is that compared with open or fixed boundary conditions this choice minimizes finite size surface effects, which are irrelevant in the thermodynamic limit.

In previous studies of the two-dimensional classical compass model, however, periodic boundary conditions have not turned out to be the ideal choice. In the directionally ordered low-temperature phase the spins form essentially one-dimensional chains with decoupled rows and columns of spins on the square lattice. With periodic boundary conditions the spins tend to form closed aligned loops along the lattice. These excitations are particularly stable against thermal fluctuations. In their studies Mishra et al. have noticed such an effect spoiling the finite-size scaling with periodic boundary conditions [8] and suggested that the reason may lie in the existence of a one-dimensional magnetic correlation length $\xi_{1\mathrm{D}}$ which exceeds the linear system size L at low temperatures. Wenzel et al. have confirmed this claim [10].

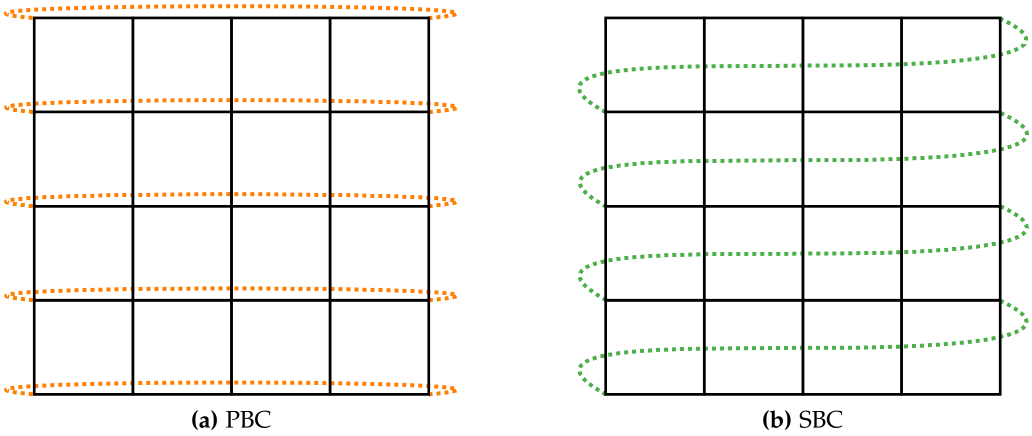
As a solution the authors of Ref. [8] have adopted special fluctuating or annealed boundary conditions. Here the sign of the coupling constants on the bonds along the lattice boundaries is allowed to fluctuate thermally. In this way, one-dimensional chains are effectively broken up. A disadvantage is that actually a different model is simulated, although one can assume that the influence of these dL^{d-1} fluctuating bonds becomes unimportant in the thermodynamic limit as $N = L^d \to \infty$. However, no good finite-size scaling theory is available for this type of boundary conditions.

As an alternative the authors of Ref. [10] have proposed screw-periodic boundary conditions, which are a deformation of the torus topology of regular periodic boundary conditions. In two dimensions they are defined by the following specification of nearest neighbors of a site $i = (x, y)$ in x- and y-directions:

$$\begin{aligned} (x,y) + \hat{\mathbf{x}} &= \begin{cases} (x+1, y), & \text{if } x < L-1, \\ (0, [y+S] \bmod L), & \text{if } x = L-1, \end{cases} \\ (x,y) + \hat{\mathbf{y}} &= \begin{cases} (x, y+1), & \text{if } y < L-1, \\ ([x+S] \bmod L, 0), & \text{if } y = L-1. \end{cases} \end{aligned} \tag{3.4}$$

The screw parameter S can be varied. If S is taken as one of the distinct divisors of L, the lattice can be subdivided in x- and y-direction into S groups of sites (called "loops"), which are linked as pairs of neighbors in that direction.

With $S = 0$ or $S = L$ regular periodic boundary conditions are recovered. With $S = 1$


(a) PBC

(b) SBC

Figure 3.1: Sketch of periodic (PBC) and screw-periodic boundary conditions (SBC) along the x-axis of a two-dimensional lattice. In the SBC picture the link between the lower right and the upper left corner is not shown. Equivalent boundary conditions are applied to the y-direction. Here a screw parameter of $S = 1$ is used.

all sites are put into only one loop in each direction. In Fig. 3.1 these two cases are illustrated. The power of screw-periodic boundary conditions lies in the fact that with a sufficiently low choice of S, the loop length exceeds the magnetic correlation length $\xi_{1\mathrm{D}}$ already for small L. Hence, excitations of linear alignment are broken up more easily than with regular periodic boundary conditions.

In chapter 5 it will become clear that also in the three-dimensional case regular periodic boundary conditions are not the ideal choice. The above definition (3.4) can easily be extended to have screw-periodic boundary conditions in three dimensions, for instance by setting:

$$
\begin{aligned}
(x,y,z)+\hat{\mathbf{x}} &= \begin{cases} (x+1,y,z), & \text{if } x < L-1, \\ (0,y,[z+S] \bmod L), & \text{if } x = L-1, \end{cases} \\
(x,y,z)+\hat{\mathbf{y}} &= \begin{cases} (x,y+1,z), & \text{if } y < L-1, \\ ([x+S] \bmod L,0,z), & \text{if } y = L-1, \end{cases} \\
(x,y,z)+\hat{\mathbf{z}} &= \begin{cases} (x,y,z+1), & \text{if } z < L-1, \\ (x,[y+S] \bmod L,0), & \text{if } z = L-1. \end{cases}
\end{aligned}
\tag{3.5}
$$

In this way, lattice planes are connected with each other. The choice of screw-periodic boundary conditions should not have an influence on the thermodynamic limit. But the high degeneracy of finite system configurations is greatly reduced by the connection of planes and chains, which lowers the number of available symmetries considerably from the enumeration in Sec. 1.3.1.

3.3 Metropolis update

Since interactions in the compass model Hamiltonian

$$\mathcal{H} = J \sum_{i=1}^{N} \sum_{k=1}^{d} s_i^k s_{i+\hat{\boldsymbol{k}}}^k \tag{3.6}$$

are restricted to nearest neighbor pairs of spins, the use of a local spin update with the Metropolis algorithm, that has already been introduced in general in Sec. 2.4.2, is an obvious choice. The energy difference between the spin configurations before and after the proposed update of a single spin $\boldsymbol{s}_j \rightarrow \boldsymbol{s}'_j$ can be evaluated locally:

$$\begin{aligned}
\Delta E &= \mathcal{H}(\{\, \boldsymbol{s}_1, \dots, \boldsymbol{s}'_{j-1}, \boldsymbol{s}_j, \boldsymbol{s}_{j+1}, \dots, \boldsymbol{s}_N \,\}) - \mathcal{H}(\{\, \boldsymbol{s}_1, \dots, \boldsymbol{s}_{j-1}, \boldsymbol{s}_j, \boldsymbol{s}_{j+1}, \dots, \boldsymbol{s}_N \,\}) \\
&= J \sum_{k=1}^{d} \left[s_j^{k\prime} s_{j+\hat{\boldsymbol{k}}}^k + s_j^{k\prime} s_{j-\hat{\boldsymbol{k}}}^k \right] - J \sum_{k=1}^{d} \left[s_j^k s_{j+\hat{\boldsymbol{k}}}^k + s_j^k s_{j-\hat{\boldsymbol{k}}}^k \right] \\
&= J \sum_{k=1}^{d} \left[s_j^{k\prime} (s_{j+\hat{\boldsymbol{k}}}^k + s_{j-\hat{\boldsymbol{k}}}^k) - s_j^k (s_{j+\hat{\boldsymbol{k}}}^k + s_{j-\hat{\boldsymbol{k}}}^k) \right] \\
&= J \sum_{k=1}^{d} (s_j^{k\prime} - s_j^k)(s_{j+\hat{\boldsymbol{k}}}^k + s_{j-\hat{\boldsymbol{k}}}^k).
\end{aligned} \tag{3.7}$$

One Monte Carlo step consists then of a sequential sweep of the lattice, i.e. updates for all spins $\boldsymbol{s}_i$ are proposed in lexicographical order for $i = 1, \dots, N$. The spins to be updated can be chosen by more elaborate schemes, e.g. by first visiting all odd sites and then all even sites or by picking them in random order. Such procedures might reduce correlations between the configurations before and after a sweep, but they generally also increase computational costs due to no longer aligned memory accesses or the need to draw a larger number of pseudo random numbers.

Since the spins are different objects depending on the dimensionality of the system, the actual update procedure must be adapted specifically. Except for the case of $d = 1$, where the compass model is reduced to a simple Ising chain [33, 34], the individual spins have continuous degrees of freedom. Some attention must be paid to the range in which these are altered, i.e. to the *step size* of the update algorithm. On the one hand, if it is chosen too wide, only a small fraction of the proposed updates will be accepted because often the energy will increase by a relatively large portion. If, on the other hand, it is chosen too narrow, a high fraction of the proposed updates will be accepted, but without substantial change of the spin configuration. Both extreme cases lead to high autocorrelation times and therefore to small statistically effective sample sizes. Instead it is advisable to adjust the step size in such a way that the ratio of accepted to rejected updates reaches an intermediary value of about 50%.

3.3.1 Metropolis in two dimensions

In the two-dimensional case the spins are represented by vectors on the unit circle that can be parametrized by a single angle $\varphi_i \in [0, 2\pi)$:

$$\boldsymbol{s}_i = \begin{pmatrix} s_i^x \\ s_i^y \end{pmatrix} = \begin{pmatrix} \cos\varphi_i \\ \sin\varphi_i \end{pmatrix}. \tag{3.8}$$

Given a spin $\boldsymbol{s}$, a new spin $\boldsymbol{s}'$ is proposed in one step of the Metropolis algorithm. It is chosen randomly, equally distributed on the arc with a central angle $\Delta\varphi$ around $\boldsymbol{s}$. See Fig. 3.2 for an illustration.

Given a uniformly distributed random number r from the interval $[0, 1)$ and the old angle φ, the new angle φ' can be chosen according to

$$\varphi' = \varphi + (r - 1/2) \times \Delta\varphi. \tag{3.9}$$

Detailed balance is satisfied because the angle range $\Delta\varphi$ is kept constant during the simulation.

A pseudo code rendition for one Metropolis sweep is listed in Algorithm 3.1. In the implementation the current values of s_i^x, s_i^y and φ_i are stored for all sites i. While actually only the angles or the Cartesian coordinates are necessary to fully characterize the spin configuration, the computational cost of the trigonometric functions can be reduced in this way. I have not optimized the implementation very intensively, as for the two-dimensional model high-precision results are already available [9, 10] and because the cluster update discussed in Sec. 3.4.2 provides great improvements.

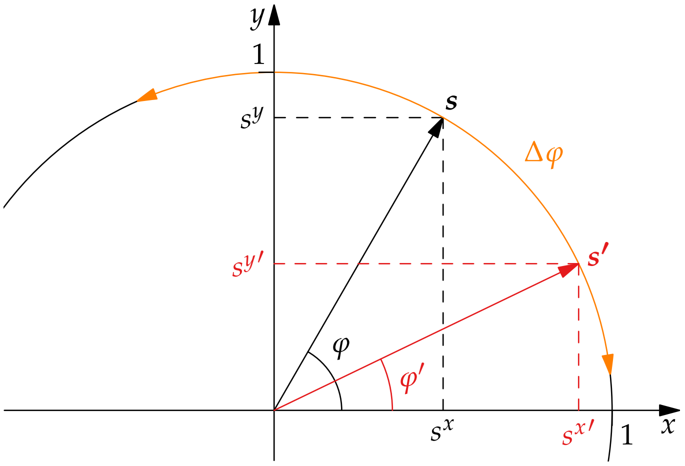


Figure 3.2: Illustration of the two-dimensional single spin Metropolis update

Algorithm 3.1 Metropolis sweep for the 2D compass model

```
for site $i = 1$ to $N$ do
    $r_1 \leftarrow$ random number from $[0, 1)$
    $\varphi' \leftarrow \varphi_i + (r_1 - 1/2) \cdot \Delta\varphi$
    $s^{x\prime} \leftarrow \cos\varphi'$
    $s^{y\prime} \leftarrow \sin\varphi'$
    $\Delta E \leftarrow J[(s^{x\prime} - s^x_i)(s^x_{i+\hat{x}} + s^x_{i-\hat{x}}) + (s^{y\prime} - s^y_i)(s^y_{i+\hat{y}} + s^y_{i-\hat{y}})]$
    if $\Delta E \leq 0$ then
        accept update
    else
        $r_2 \leftarrow$ random number from $[0, 1)$
        if $r_2 \leq \mathrm{e}^{-\beta \Delta E}$ then
            accept update
        else
            reject update
        end if
    end if
    if update accepted then
        $(s^x_i, s^y_i) \leftarrow (s^{x\prime}, s^{y\prime})$
        $\varphi_i \leftarrow \varphi'$
    end if
end for
```

3.3.2 Metropolis in three dimensions

In the three-dimensional case the spins are represented by vectors on the unit sphere that can be parametrized by two angles $\varphi_i \in [0, 2\pi)$ and $\theta_i \in [0, \pi]$:

$$\boldsymbol{s}_i = \boldsymbol{s}(\theta_i, \varphi_i) = \begin{pmatrix} s_i^x \\ s_i^y \\ s_i^z \end{pmatrix} = \begin{pmatrix} \cos\varphi_i \sin\theta_i \\ \sin\varphi_i \sin\theta_i \\ \cos\theta_i \end{pmatrix} . \tag{3.10}$$

Some attention must be paid to correctly select points randomly from a uniform distribution over the spherical surface. The requirements for the probability density of a point $\widetilde{\boldsymbol{s}}$ chosen on the surface are

$$p(\widetilde{\boldsymbol{s}})\mathrm{d}\Omega = p\mathrm{d}\Omega \tag{3.11}$$

with $p = \text{const}$ and

$$\iint \mathrm{d}\Omega p = 1, \tag{3.12}$$

which implies $p = 1/(4\pi)$. From the Jacobian of the spherical coordinates $(\delta\theta, \delta\varphi)$ defined according to Eq. (3.10) one finds for the solid angle element:

$$\mathrm{d}\Omega = \mathrm{d}(\delta\varphi)\sin(\delta\theta)\mathrm{d}(\delta\theta) = \mathrm{d}(\delta\varphi)\mathrm{d}(\cos\delta\theta) = \mathrm{d}(\delta\varphi)\mathrm{d}c \tag{3.13}$$

with $\delta\varphi \in [0, 2\pi)$ and $c \equiv \cos\delta\theta \in [-1, 1]$. If c and $\delta\varphi$ are random numbers from a uniform distribution, the corresponding point

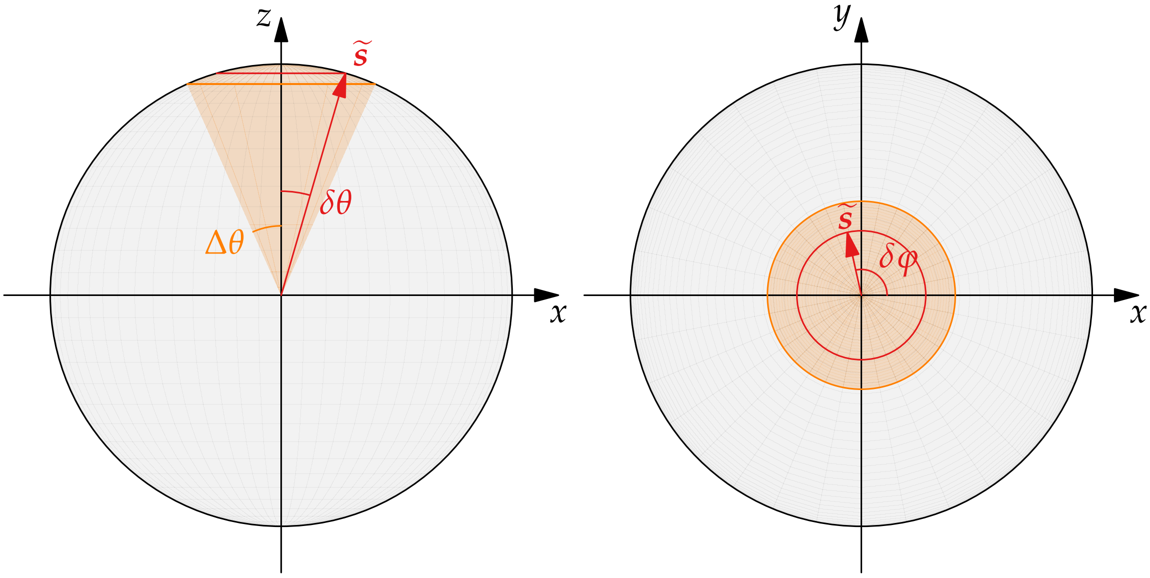
$$\widetilde{\boldsymbol{s}}(c, \delta\varphi) = \begin{pmatrix} \cos\delta\varphi\sqrt{1-c^2} \\ \sin\delta\varphi\sqrt{1-c^2} \\ c \end{pmatrix} \tag{3.14}$$

is a random variable with uniform distribution over the spherical surface.

By restricting c to an interval $[c_{\text{min}}, 1]$, a uniform distribution of vectors pointing onto the spherical cap with opening angle $\Delta\theta = \arccos(c_{\text{min}})$ is obtained. The construction of such a vector $\widetilde{\boldsymbol{s}}$ from two angles $\delta\theta$ and $\delta\varphi$ on a spherical cap centered around the z-axis is illustrated in Fig. 3.3.

In one step of the Metropolis algorithm a new spin vector $\boldsymbol{s}'$ is proposed from a spherical cap around the original spin

$$\boldsymbol{s} = \boldsymbol{s}(\theta, \varphi) = (s^x, s^y, s^z)^\intercal . \tag{3.15}$$



(a) Front view **(b)** Top view

Figure 3.3: Choice of a spin from a spherical cap around the z-axis for the three-dimensional Metropolis single spin update

For an illustration see Fig. 3.4. Since the surface area described by this cap is always held constant, detailed balance is not broken. To find $\boldsymbol{s}'$, $\widetilde{\boldsymbol{s}}$ is rotated into a coordinate system whose z-axis coincides with $\boldsymbol{s}$:

$$\boldsymbol{s}' = R^{-1} \cdot \widetilde{\boldsymbol{s}}, \tag{3.16}$$

where R is a rotational matrix that satisfies $R \cdot \boldsymbol{s} = \hat{\boldsymbol{z}}$. One such matrix is given by $R = R_{\hat{\boldsymbol{n}}}(\theta)$, which realizes a right-handed rotation around the normal vector

$$\hat{\boldsymbol{n}} = \frac{\boldsymbol{s} \times \hat{\boldsymbol{z}}}{|\boldsymbol{s} \times \hat{\boldsymbol{z}}|} = \frac{1}{\sqrt{(s^x)^2 + (s^y)^2}} \begin{pmatrix} s^y \\ -s^x \\ 0 \end{pmatrix} \tag{3.17}$$

by the angle $\theta = \arccos(s^z)$. Using the most natural basis, the application of $R = R_{\hat{\boldsymbol{n}}}(\theta)$ on a vector $\boldsymbol{x}$ results in

$$R_{\hat{\boldsymbol{n}}}(\theta) \cdot \boldsymbol{x} = \hat{\boldsymbol{n}}(\hat{\boldsymbol{n}} \cdot \boldsymbol{x}) + \cos\theta[(\hat{\boldsymbol{n}} \times \boldsymbol{x}) \times \hat{\boldsymbol{n}}] + \sin\theta[\hat{\boldsymbol{n}} \times \boldsymbol{x}]. \tag{3.18}$$

The inverse of R is given by $R^{-1} = R_{\hat{\boldsymbol{n}}}(-\theta)$. Hence one finds

$$\boldsymbol{s}' = R_{\hat{\boldsymbol{n}}}(-\theta) \cdot \widetilde{\boldsymbol{s}} = \hat{\boldsymbol{n}}(\hat{\boldsymbol{n}} \cdot \widetilde{\boldsymbol{s}}) + \cos(\theta)[(\hat{\boldsymbol{n}} \times \widetilde{\boldsymbol{s}}) \times \hat{\boldsymbol{n}}] - \sin(\theta)[\hat{\boldsymbol{n}} \times \widetilde{\boldsymbol{s}}]. \tag{3.19}$$

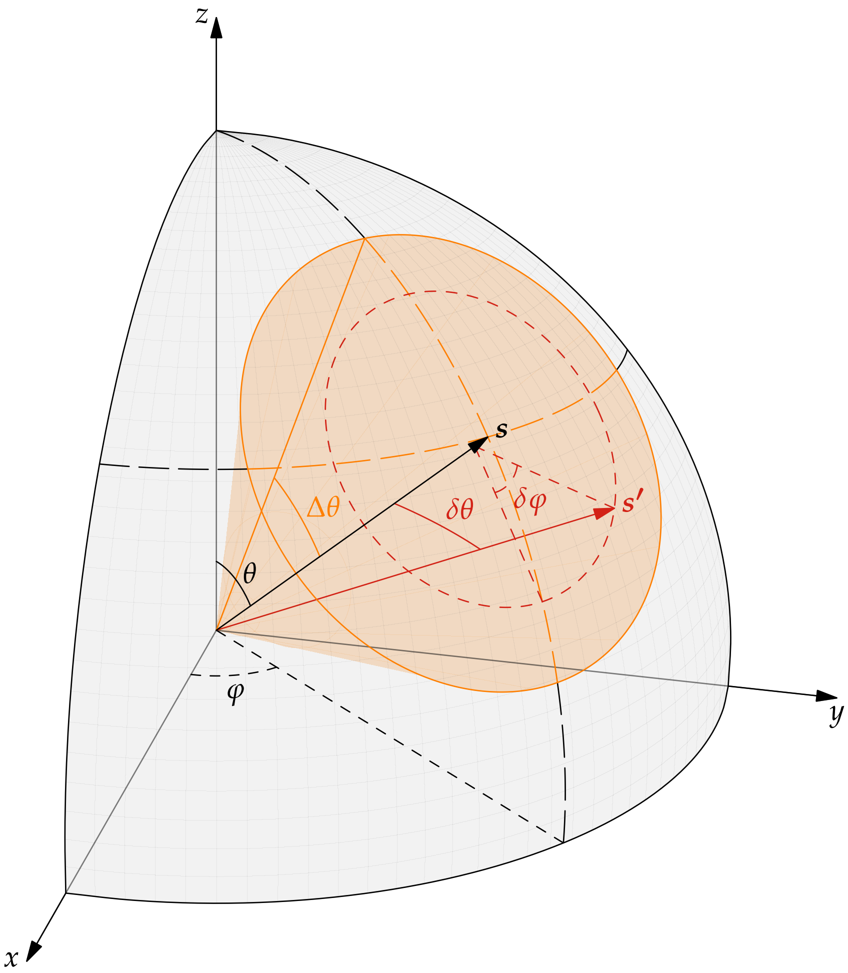


Figure 3.4: Illustration of the three-dimensional single spin Metropolis update. Shown is one octant of the unit sphere, the original spin $\boldsymbol{s}$, a spherical cap with opening angle $\Delta\theta$ from which the new spin is chosen and one possibility for a resulting spin $\boldsymbol{s}'$.

By inserting $\cos\theta = s^z$, $\sin\theta = \sqrt{1-(s^z)^2}$, as well as

$$\hat{\boldsymbol{n}} \times \widetilde{\boldsymbol{s}} = \frac{1}{\sqrt{(s^x)^2+(s^y)^2}} \begin{pmatrix} -s^x \cdot c \\ -s^y \cdot c \\ [s^x \cdot \cos(\delta\varphi) + s^y \cdot \sin(\delta\varphi)]\sqrt{1-(s^z)^2} \end{pmatrix}, \tag{3.20}$$

$$(\hat{\boldsymbol{n}} \times \widetilde{\boldsymbol{s}}) \times \hat{\boldsymbol{n}} = \frac{\sqrt{1-c^2} s^x \cdot \cos(\delta\varphi) + s^y \cdot \sin(\delta\varphi)}{(s^x)^2+(s^y)^2} \begin{pmatrix} s^x \\ s^y \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} \quad \text{and} \tag{3.21}$$

$$\hat{\boldsymbol{n}}(\hat{\boldsymbol{n}} \cdot \widetilde{\boldsymbol{s}}) = \frac{\sqrt{1-c^2}}{(s^x)^2+(s^y)^2} (s^y \cos(\delta\varphi) - s^x \sin(\delta\varphi)) \begin{pmatrix} s^y \\ -s^x \\ 0 \end{pmatrix} \tag{3.22}$$

into Eq. (3.19), a final expression for the updated spin $\mathbf{s}' = (s^{x\prime}, s^{y\prime}, s^{z\prime})^\intercal$ is retrieved:

$$\mathbf{s}' = \begin{pmatrix} \dfrac{\sqrt{1-c^2}}{(s^x)^2+(s^y)^2}\left([(s^x)^2 s^z + (s^y)^2]\cos\delta\varphi + (s^z-1)s^x s^y\sin\delta\varphi\right) + s^x c \\ \dfrac{\sqrt{1-c^2}}{(s^x)^2+(s^y)^2}\left([s^z-1]s^x s^y\cos\delta\varphi + [(s^x)^2+(s^y)^2 s^z]\sin\delta\varphi\right) + s^y c \\ -\sqrt{1-c^2}\,(s^x\cos\delta\varphi + s^y\sin\delta\varphi) + s^z c \end{pmatrix} \tag{3.23}$$

In Algorithm 3.2 the procedure for one Metropolis sweep of the three-dimensional compass model is given in pseudo code.

Since the procedure is rather elaborate, it is a good idea to verify that it produces correct results. The same update algorithm can also be applied to the classical Heisenberg model, which is also a lattice model with spins taken as vectors on the unit sphere S^2. In the simple case of a one dimensional chain of spins with free boundary conditions it can be solved analytically. See appendix A for a quantitative comparison of Monte Carlo results with the exact solution.

Eq. (3.23) has a singularity if the initial spin is aligned with the z-axis, $\mathbf{s} = \pm\hat{\mathbf{z}}$. But due to the way pseudo-random numbers are chosen, this situation never occurs in the simulations unless one initializes the system with such a pathological configuration.

Algorithm 3.2 Metropolis sweep for the 3D compass model

```
for site $i = 0$ to $N$ do
  $r_1 \leftarrow$ random number from $[0,1)$
  $r_2 \leftarrow$ random number from $[0,1)$
  $c \leftarrow c_{\min} + r_1 \cdot (1 - c_{\min})$
  $\delta\varphi \leftarrow 2\pi \cdot r_2$
  Calculate $(s^{x\prime}, s^{y\prime}, s^{z\prime})$ from $(s_i^x, s_i^y, s_i^z)$ and $(c, \delta\varphi)$ according to Eq. (3.23)
  $\Delta E \leftarrow J[(s^{x\prime} - s_i^x)(s_{i+\hat{x}}^x + s_{i-\hat{x}}^x) + (s^{y\prime} - s_i^y)(s_{i+\hat{y}}^y + s_{i-\hat{y}}^y) + (s^{z\prime} - s_i^z)(s_{i+\hat{z}}^z + s_{i-\hat{z}}^z)]$
  if $\Delta E \leq 0$ then
    accept update
  else
    $r_3 \leftarrow$ random number from $[0,1)$
    if $r_3 \leq \mathrm{e}^{-\beta\Delta E}$ then
      accept update
    else
      reject update
    end if
  end if
  if update accepted then
    $(s_i^x, s_i^y, s_i^z) \leftarrow (s^{x\prime}, s^{y\prime}, s^{z\prime})$
  end if
end for
```

3.3.3 Adjustment of step sizes

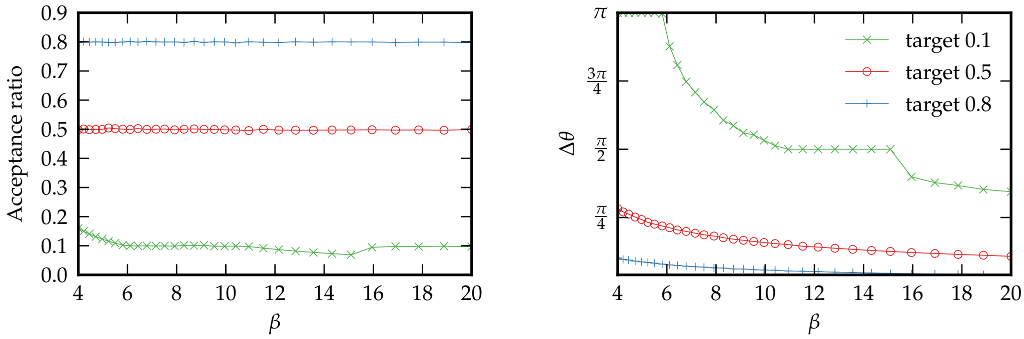
Both implementations of the Metropolis algorithm that have been discussed above feature an adjustable parameter $\alpha = \Delta\varphi$ or $\alpha = -c_{\min}$ that can be tuned to achieve a desired ratio of accepted spin updates. A higher value of α leads to lower acceptance ratios.

In some Monte Carlo studies analogous parameters which set a maximum step size were modified dynamically in the course of the simulations. Of course, such an approach breaks detailed balance and the stochastic process underlying the Monte Carlo simulation is no longer Markovian. Indeed Miller et al. have shown in Ref. [35] that a dynamic adjustment of the maximum step size in conjunction with the Metropolis algorithm can lead to systematic errors. While Swendsen has argued in a recent publication [36] that by waiting sufficiently long between adjustments the bias introduced by this procedure can be made negligible compared to the statistical error, I have refrained from using such a scheme.

Instead I have restricted such an adjustment to the first half of the thermalization phase to find the value of α which is optimal for the simulation temperature. Every 100 sweeps the parameter α is adjusted iteratively in such a way that the acceptance ratio measured over the last 100 sweeps converges towards some target ratio t. The simple scheme is outlined in Algorithm 3.3. Empirically the method works very well to realize an acceptance ratio of $\approx 50\%$ for the ranges of temperatures used with the compass model in this work.

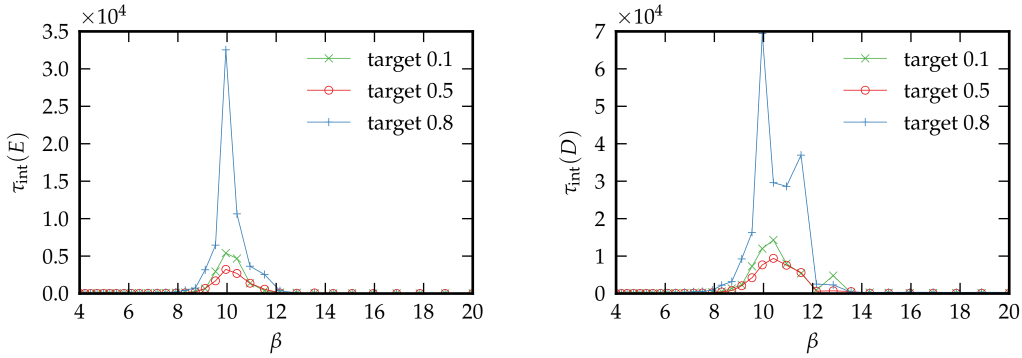
Algorithm 3.3 Adjustment of the maximum step size for a target Metropolis acceptance ratio t

```
$(\alpha_{\min}, \alpha_{\max}) \leftarrow \begin{cases} (0, 2\pi), & \text{2D} \\ (-1, 1), & \text{3D} \end{cases}$
$\alpha \leftarrow \alpha_{\min}$
for the first half of thermalization do
    Perform 100 Metropolis using the parameter $\alpha$ and measure the acceptance ratio $r$
    if $r < t$ and $\alpha > \alpha_{\min}$ then
        $\alpha_{\max} \leftarrow \alpha$
        $\alpha \leftarrow \alpha - (\alpha - \alpha_{\min})/2$
    else if $r > t$ and $\alpha < \alpha_{\max}$ then
        $\alpha_{\min} \leftarrow \alpha$
        $\alpha \leftarrow \alpha + (\alpha_{\max} - \alpha)/2$
    end if
end for
for the remaining half of thermalization do
    Update the system using the final parameter $\alpha$ in the Metropolis routine
end for
```



(a) Measured acceptance ratios

(b) Opening angle $\Delta\theta$



(c) Autocorrelation times of the energy E

(d) Autocorrelation times of the order parameter D

Figure 3.5: Comparison of target acceptance ratios 0.1, 0.5, 0.8 in canonical simulations with purely local updates of the three-dimensional compass model with screw-periodic boundary conditions at $L = 8$.

Efficiency

I have compared preset target acceptance ratios of 0.1, 0.5 and 0.8 for the three-dimensional compass model. The method of determining the maximum opening angle $\Delta\theta$ for the updated spins during thermalization works well for the higher values, see Figs. 3.5(a) and 3.5(b). The acceptance ratio of 0.1 could not be sustained at the high temperatures (low β), where spins were already chosen from the whole sphere. The deviations at medium temperatures can probably be explained by a faulty handling of the zero crossing of $\cos(\Delta\theta)$ in the iterative procedure. For this qualitative comparison this does not play a role, however, and neither in any of the production simulation runs. A target acceptance ratio of 0.5 is a reasonable choice as it leads to lower autocorrelation times than the wider or narrower settings as one can see in Figs. 3.5(c) and 3.5(d). Further improvements are made using the non-local update described in the next section.

3.4 Cluster update

3.4.1 Cluster algorithms for spin models

Cluster update algorithms have originally been devised as a remedy for the problem of critical slowing down at second-order phase transitions, which is caused by the magnitude of spatial correlations at the transition point, see Sec. 2.5.1. In spin models this leads to large connected droplets of spins with equal orientation. Instead of trying to flip individual spins with a local update algorithm, the knowledge of these structures is exploited in non-local cluster updates. Their central idea is to flip entire clusters of equally oriented spins at the same time.

The first successful algorithm of this class was the multiple-cluster update by Swendsen and Wang [37] for discrete spin q-state Potts models defined by the Hamiltonian

$$\mathcal{H}^{\text{Potts}} = J \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j} \tag{3.24}$$

with $\sigma_i = 1, \ldots, q$ and where the sum goes over nearest neighbor pairs of lattice sites. In one step of the Swendsen-Wang algorithm the whole lattice is first decomposed into stochastic clusters, then a random new common spin state is assigned to each cluster.

The second important cluster algorithm is the single-cluster update by Wolff [38], which has been formulated from the beginning for spin models with $O(n)$-symmetry defined by the Hamiltonian

$$\mathcal{H}^{O(n)} = J \sum_{\langle i,j \rangle} \boldsymbol{s}_i \cdot \boldsymbol{s}_j, \tag{3.25}$$

where the spins $\boldsymbol{s}_i$ are n-dimensional vectors on the hyper-sphere S^{n-1}. For $n = 1$ it corresponds to the Ising-model, for $n = 2$ to the xy-model and for $n = 3$ to the classical Heisenberg-model.

In one step of the Wolff-cluster update a random lattice site is picked as a seed from which a stochastic cluster is constructed. Then all spins in the cluster are reflected about a common randomly chosen plane.

Cluster updates always need to be tailored specifically to the model under examination.

3.4.2 A Wolff-like single linear cluster update

In Ref. [10] Wenzel, Janke and Läuchli have proposed a one-dimensional adaption of the Wolff-cluster update from to the two-dimensional classical compass model. It exploits the special coupling of lattice and spin symmetries in the compass model, which leads to the formation of one-dimensional spin chains in the ordered phase. The generalization of this algorithm to arbitrary dimensions d is straightforward.

The Hamiltonian of a model of spins with nearest neighbor interactions on a periodic hyper-cubic lattice can be written in the form

$$\mathcal{H} = \sum_{\langle i,j \rangle} \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j), \tag{3.26}$$

where $\mathcal{H}_{ij}$ is the bond energy of a pair of nearest neighbor spins. For an $O(n)$-model it is $\mathcal{H}_{ij} = J\boldsymbol{s}_i \cdot \boldsymbol{s}_j$ and for the d-dimensional compass model it is $\mathcal{H}_{ij} = J\sum_{k=1}^{d}(s_i^k s_j^k)\delta_{j,i+\hat{\boldsymbol{k}}}$.

Wolff defines a spin flip operation by the reflection of a spin $\boldsymbol{s}_i$ along the hyperplane orthogonal to a vector $\boldsymbol{r} \in S^{n-1}$:

$$R(\boldsymbol{r})\boldsymbol{s}_i = \boldsymbol{s}_i - 2(\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r}, \tag{3.27}$$

which is idempotent:

$$\begin{aligned} R(\boldsymbol{r})^2\boldsymbol{s}_i &= R(\boldsymbol{r})R(\boldsymbol{r})\boldsymbol{s}_i = R(\boldsymbol{r})\boldsymbol{s}_i - 2(R(\boldsymbol{r})\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r} \\ &= \boldsymbol{s}_i - 2(\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r} - 2(\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r} + 4(\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r} = \boldsymbol{s}_i. \end{aligned} \tag{3.28}$$

Clusters will be constructed from the spins on pairs of neighboring sites $\langle i,j \rangle$, also called *bonds*, and flipped using $R(\boldsymbol{r})$. Detailed balance is achieved in the Wolff algorithm if the energies of included bonds are invariant under the reflection $R(\boldsymbol{r})$:

$$\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j) = \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) \tag{3.29}$$

and if one has

$$\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j) = \mathcal{H}_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j). \tag{3.30}$$

A proof will be given further below. In an $O(n)$-model, these conditions are of course satisfied for any choice of bonds $\langle i,j \rangle$ and vectors $\boldsymbol{r}$, as one can verify immediately:

$$\begin{aligned} R(\boldsymbol{r})\boldsymbol{s}_i \cdot R(\boldsymbol{r})\boldsymbol{s}_j &= [\boldsymbol{s}_i - 2(\boldsymbol{s}_i \cdot \boldsymbol{r})\boldsymbol{r}] \cdot [\boldsymbol{s}_j - 2(\boldsymbol{s}_j \cdot \boldsymbol{r})\boldsymbol{r}] \\ &= \boldsymbol{s}_i \cdot \boldsymbol{s}_j - 4(\boldsymbol{s}_i \cdot \boldsymbol{r})(\boldsymbol{s}_j \cdot \boldsymbol{r}) + 4(\boldsymbol{s}_i \cdot \boldsymbol{r})(\boldsymbol{s}_j \cdot \boldsymbol{r}) = \boldsymbol{s}_i \cdot \boldsymbol{s}_j, \end{aligned} \tag{3.31}$$

and

$$(R(\boldsymbol{r})\boldsymbol{s}_i) \cdot \boldsymbol{s}_j = \boldsymbol{s}_i \cdot \boldsymbol{s}_j - 2(\boldsymbol{s}_i \cdot \boldsymbol{r})(\boldsymbol{s}_j \cdot \boldsymbol{r}) = \boldsymbol{s}_i \cdot R(\boldsymbol{r})\boldsymbol{s}_j. \tag{3.32}$$

Clearly, the compass model bond energy is not generally conserved under this reflection. However, the reflection operations $R(\hat{\boldsymbol{k}}), k = 1, \dots, d$, with respect to the lattice planes orthogonal to $\hat{\boldsymbol{k}}$ each leave $\mathcal{H}_{ij}$ invariant for a subset of bonds $\langle i, j \rangle = \left\langle i, i + \hat{\boldsymbol{k}} \right\rangle$ (see the earlier discussion in Sec. 1.3.1). In three dimensions these are the reflections with

$$R(\hat{\boldsymbol{x}})(s_i^x, s_i^y, s_i^z) = (-s_i^x, s_i^y, s_i^z), \tag{3.33}$$
$$R(\hat{\boldsymbol{y}})(s_i^x, s_i^y, s_i^z) = (s_i^x, -s_i^y, s_i^z), \tag{3.34}$$
$$R(\hat{\boldsymbol{z}})(s_i^x, s_i^y, s_i^z) = (s_i^x, s_i^y, -s_i^z), \tag{3.35}$$

which are symmetries on one-dimensional subsets of the lattice:

$$\mathcal{H}_{i,i+\hat{\boldsymbol{x}}}(R(\hat{\boldsymbol{x}})\boldsymbol{s}_i, R(\hat{\boldsymbol{x}})\boldsymbol{s}_{i+\hat{\boldsymbol{x}}}) = \mathcal{H}_{i,i+\hat{\boldsymbol{x}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{x}}}), \tag{3.36}$$
$$\mathcal{H}_{i,i+\hat{\boldsymbol{y}}}(R(\hat{\boldsymbol{y}})\boldsymbol{s}_i, R(\hat{\boldsymbol{y}})\boldsymbol{s}_{i+\hat{\boldsymbol{y}}}) = \mathcal{H}_{i,i+\hat{\boldsymbol{y}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{y}}}), \tag{3.37}$$
$$\mathcal{H}_{i,i+\hat{\boldsymbol{z}}}(R(\hat{\boldsymbol{z}})\boldsymbol{s}_i, R(\hat{\boldsymbol{z}})\boldsymbol{s}_{i+\hat{\boldsymbol{z}}}) = \mathcal{H}_{i,i+\hat{\boldsymbol{z}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{z}}}). \tag{3.38}$$

Moreover the $R(\hat{\boldsymbol{k}})$ also satisfy Eq. (3.30):

$$\mathcal{H}_{i,i+\hat{\boldsymbol{k}}}(\boldsymbol{s}_i, R(\hat{\boldsymbol{k}})\boldsymbol{s}_{i+\hat{\boldsymbol{k}}}) = -J s_i^k s_{i+\hat{\boldsymbol{k}}}^k = \mathcal{H}_{i,i+\hat{\boldsymbol{k}}}(\boldsymbol{s}_i, R(\hat{\boldsymbol{k}})\boldsymbol{s}_{i+\hat{\boldsymbol{k}}}). \tag{3.39}$$

Therefore the Wolff algorithm can be used with the compass model if it is restricted to strictly linear clusters constructed along the lattice axes.

One update step of the algorithm consists of the following sequence:

1. Choose a random reflection plane defined by a normal $\boldsymbol{r}$ taken from the set of vectors suitable for the model.

   Choose a random lattice site i as the seed of a cluster C.

2. Flip the spin at site i: $\boldsymbol{s}_i \to R(\boldsymbol{r})\boldsymbol{s}_i$, and mark the site as visited.

3. Visit all neighboring unmarked sites j of i for which $\mathcal{H}_{ij}$ is invariant under the reflection $R(\boldsymbol{r})$.

   Then with probability

$$P_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) = 1 - \mathrm{e}^{\min\left\{0, -\beta\left[\mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j)\right]\right\}} \tag{3.40}$$

   add site j to C: Flip $\boldsymbol{s}_j \to R(\boldsymbol{r})\boldsymbol{s}_j$, and mark the site as visited. In Eq. (3.40) $\boldsymbol{s}_i$ already has the flipped orientation.

4. Iterate step 3 with i taking the place of every newly adjoined site until no further sites are added to the cluster C.

For an $O(n)$-model as defined in Eq. (3.25) $P_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)$ in step 3 can be evaluated to

$$P_{ij}^{O(n)}(\boldsymbol{s}_i, \boldsymbol{s}_j) = 1 - \mathrm{e}^{\min\left\{0, -\beta \boldsymbol{s}_i [1 - R(\boldsymbol{r})] \boldsymbol{s}_j\right\}} = 1 - \mathrm{e}^{\min\left\{0, -2\beta J (\boldsymbol{r} \cdot \boldsymbol{s}_i)(\boldsymbol{r} \cdot \boldsymbol{s}_j)\right\}}. \tag{3.41}$$

In this way the cluster is constructed only from spins that are mostly aligned: Only those spins for which the dot products with r have the same sign before the flip are considered for addition. Moreover, spins that are nearly aligned are added with higher probability than others.

For the compass model the addition probabilities P_{ij} for bonds $\left\langle i, i + \hat{\boldsymbol{k}} \right\rangle$ are given by

$$P_{i,i+\hat{\boldsymbol{k}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{k}}}) = 1 - \mathrm{e}^{\min\left\{0, -2\beta J s_i^k s_{i+\hat{\boldsymbol{k}}}^k\right\}}. \tag{3.42}$$

As an example consider the three-dimensional case. With the Wolff algorithm it is possible to flip one-dimensional clusters parallel to the x-, y- or z-axis of the lattice, where the cluster-growth probabilities are given by

$$P_{i,i+\hat{\boldsymbol{x}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{x}}}) = 1 - \mathrm{e}^{\min\left\{0, -2\beta J s_i^x s_{i+\hat{\boldsymbol{x}}}^x\right\}}, \tag{3.43}$$

$$P_{i,i+\hat{\boldsymbol{y}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{y}}}) = 1 - \mathrm{e}^{\min\left\{0, -2\beta J s_i^y s_{i+\hat{\boldsymbol{y}}}^y\right\}}, \tag{3.44}$$

$$P_{i,i+\hat{\boldsymbol{z}}}(\boldsymbol{s}_i, \boldsymbol{s}_{i+\hat{\boldsymbol{z}}}) = 1 - \mathrm{e}^{\min\left\{0, -2\beta J s_i^z s_{i+\hat{\boldsymbol{z}}}^z\right\}}. \tag{3.45}$$

An implementation of the single cluster update for the compass model can be done according to Algorithm 3.4 on p. 45, which allows for periodic boundary conditions. The implementation is simpler than for higher-dimensional clusters in other models because no stack of visited sites needs to be handled.

See Fig. 3.6 on p. 46 for an illustration of the single linear cluster update with the two-dimensional compass model. Since screw-periodic boundary conditions are applied, this cluster is longer than the linear lattice size.

Detailed balance

To see that the single cluster algorithm satisfies detailed balance according to Eq. (2.35) in Sec. 2.4.1, consider the ratio of transition probabilities between two spin configurations $\{ \boldsymbol{s}_i \}$ and $\{ \boldsymbol{s}_i' \}$ that differ by a reflection operation $R(\boldsymbol{r})$ applied to a cluster C of size $|C|$. Also let ∂C denote the surface of C, which consists of all pairs of neighboring lattice sites $\langle i, j \rangle$ with $i \in C$ and $j \notin C$. The transition probabilities can be expressed as products

Algorithm 3.4 Single cluster update for the d-dimensional compass model

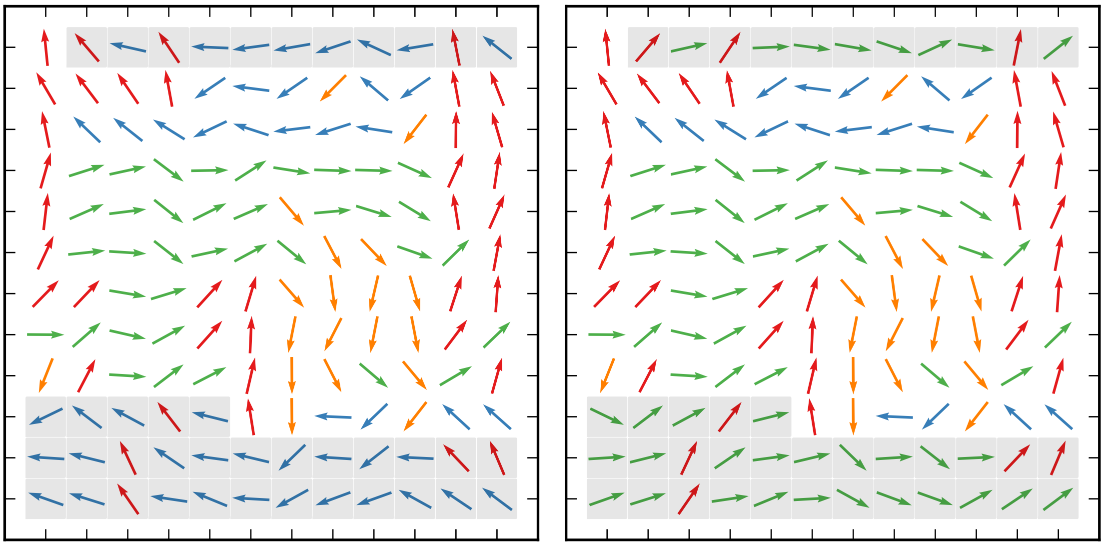
```
$\hat{\boldsymbol{k}} \leftarrow$ random direction from $\{ \hat{\boldsymbol{e}}_1, \dots, \hat{\boldsymbol{e}}_d \}$
seed site: $i_0 \leftarrow$ random site from $\{ 1, \dots, N \}$
boundary site: $i_\mathrm{f} \leftarrow i_0$
$i \leftarrow i_0$
repeat
  flip: $s_i^k \leftarrow -s_i^k$
  next site $j \leftarrow i + \hat{\boldsymbol{k}}$
  $r \leftarrow$ random number from $[0, 1)$
  if $r < 1 - \mathrm{e}^{-2J\beta s_i^k s_j^k}$ then
    $i \leftarrow j$
    $i_\mathrm{f} \leftarrow i$
  end if
until $i \neq j$ {no site added in this iteration} or $j = i_0$ {percolating cluster}
if $j \neq i_0$ then
  $i \leftarrow i_0$
  repeat
    next site $j \leftarrow i - \hat{\boldsymbol{k}}$
    if $j \neq i_\mathrm{f}$ then
      $r \leftarrow$ random number from $[0, 1)$
      if $r < 1 - \mathrm{e}^{-2J\beta s_i^k s_j^k}$ then
        $i \leftarrow j$
        flip: $s_i^k \leftarrow -s_i^k$
      end if
    end if
  until $i \neq j$ {no site added in this iteration} or $j = i_\mathrm{f}$ {percolating cluster}
end if
```

of the probabilities of choices that lead to the construction of C in both directions:

$$W\left(\{ \boldsymbol{s}_i \} \xrightarrow[\text{on } C]{R(\boldsymbol{r})} \{ \boldsymbol{s}'_i \}\right) = \underbrace{P(\boldsymbol{r})}_{\substack{\text{flip}\\ \text{direction}}} \cdot \underbrace{\frac{|C|}{N}}_{\substack{\text{cluster}\\ \text{seed}}} \cdot \underbrace{\prod_{\langle i,j\rangle \in C} P_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j)}_{\substack{\text{addition of}\\ \text{inner sites}}} \cdot \underbrace{\prod_{\langle i,j\rangle \in \partial C} (1 - P_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j))}_{\substack{\text{sites outside the cluster,}\\ \text{not added}}} \tag{3.46}$$

and similarly

$$W\left(\{ \boldsymbol{s}'_i \} \xrightarrow[\text{on } C]{R(\boldsymbol{r})} \{ \boldsymbol{s}_i \}\right) = P(\boldsymbol{r}) \cdot \frac{|C|}{N} \cdot \prod_{\langle i,j\rangle \in C} P_{ij}(R(\boldsymbol{r})\boldsymbol{s}'_i, \boldsymbol{s}'_j) \cdot \prod_{\langle i,j\rangle \in \partial C} (1 - P_{ij}(R(\boldsymbol{r})\boldsymbol{s}'_i, \boldsymbol{s}_j)). \tag{3.47}$$



(a) Before update

(b) After update

Figure 3.6: Snapshots from a Monte Carlo simulation of the two-dimensional compass model with $L = 12$ and screw-periodic boundary conditions at $\beta = 6$ right before and after the update of a single cluster in x-direction of size 40. Sites included in the cluster are marked with gray boxes. Color is used to ease the discrimination of spin orientations.

The probability $P(\boldsymbol{r})$ of choosing the normal $\boldsymbol{r}$ of the plane of reflection is given as discrete here, which is valid for the compass model. For models like an $O(n)$-model with $n \geq 2$ it would have to be formulated in terms of a probability density, but the important point is that the probabilities are equal in both directions. The probabilities of adding new sites to the cluster P_{ij} in Eq. (3.47) satisfy

$$P_{ij}(R(\boldsymbol{r})\boldsymbol{s}'_i, \boldsymbol{s}'_j) = P_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j) \overset{!}{=} P_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j). \tag{3.48}$$

The last equality holds because by the definition in Eq. (3.40) one has for the site addition probabilities

$$\begin{aligned} P_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j) &= 1 - \mathrm{e}^{\min\left\{0, -\beta\left[\mathcal{H}_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)\right]\right\}} \\ &\overset{!}{=} 1 - \mathrm{e}^{\min\left\{0, -\beta\left[\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j) - \mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j)\right]\right\}} \\ &= P_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j), \end{aligned} \tag{3.49}$$

where the symmetries (3.29) and (3.30) have been used. Hence the transition probabilities in Eqs. (3.46) and (3.47) only differ by the factors related to the cluster surface and their ratio is:

$$\frac{W\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}{W(\{\,\boldsymbol{s}'_i\,\} \to \{\,\boldsymbol{s}_i\,\})} = \frac{\prod_{\langle i,j\rangle \in \partial C}(1 - P_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j))}{\prod_{\langle i,j\rangle \in \partial C}(1 - P_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j))} \tag{3.50}$$

All P_{ij} appearing here are smaller than one. Else $\langle i, j\rangle$ would be part of C. Thus:

$$\frac{W\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}{W(\{\,\boldsymbol{s}'_i\,\} \to \{\,\boldsymbol{s}_i\,\})} = \prod_{\langle i,j\rangle \in \partial C} \frac{1 - \mathrm{e}^{-\beta[\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j) \,-\, \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)]}}{1 - \mathrm{e}^{-\beta[\mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) \,-\, \underbrace{\mathcal{H}_{ij}(\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j)}_{=\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j)}]}} \tag{3.51}$$

Each factor is of the form $\frac{1-\alpha}{1-1/\alpha} = -\alpha$ with $\alpha = \mathrm{e}^{-\beta[\dots]}$, which allows to rewrite the product in the following way:

$$\frac{W\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}{W(\{\,\boldsymbol{s}'_i\,\} \to \{\,\boldsymbol{s}_i\,\})} = \prod_{\langle i,j\rangle \in \partial C} (-1) \cdot \mathrm{e}^{-\beta[\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)]} \tag{3.52}$$

The factors of -1 vanish because on the periodic, hyper-cubic lattices considered here the number of surface bonds of a cluster is always even. With the help of an added zero it becomes clear that the expression is completely independent of the choice of the cluster C:

$$\begin{aligned}
\frac{W\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}{W(\{\,\boldsymbol{s}'_i\,\} \to \{\,\boldsymbol{s}_i\,\})} &= \exp\Big[-\beta \sum_{\langle i,j\rangle \in \partial C} [\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, \boldsymbol{s}_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)] \\
&\qquad\quad - \beta \sum_{\langle i,j\rangle \in C} \underbrace{[\mathcal{H}_{ij}(R(\boldsymbol{r})\boldsymbol{s}_i, R(\boldsymbol{r})\boldsymbol{s}_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)]}_{=0}\Big] \\
&= \exp\Big[-\beta \sum_{\langle i,j\rangle} [\mathcal{H}_{ij}(\boldsymbol{s}'_i, \boldsymbol{s}'_j) - \mathcal{H}_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j)]\Big]
\end{aligned} \tag{3.53}$$

So the transition probabilities satisfy detailed balance:

$$\frac{W\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}{W(\{\,\boldsymbol{s}'_i\,\} \to \{\,\boldsymbol{s}_i\,\})} = \mathrm{e}^{-\beta \Delta E\,\left(\{\,\boldsymbol{s}_i\,\} \to \{\,\boldsymbol{s}'_i\,\}\right)}. \tag{3.54}$$

Ergodicity

For $O(n)$-models the single cluster update is ergodic: There is always a reflection $R(\boldsymbol{r})$ connecting any two spin orientations. And since clusters of size 1 are constructed with positive probability, there is always the possibility to reach any configuration of spins from any given configuration in a series of updates.

However, due to the restricted choice of possible reflection vectors $\boldsymbol{r}$ in the compass model, the Wolff update on its own is not ergodic here. The solution proposed by

Wenzel et al. [10] is to mix cluster updates with local Metropolis updates. For the two-dimensional compass model they have combined on average L linear cluster updates in both x- and y-direction with $N = L^2$ single spin updates. For most of the studies of the three-dimensional compass model in this work, one update step has consisted of $3L$ cluster updates in randomly chosen directions followed by $N = L^3$ local updates.

Choosing a higher number of cluster updates per step like $3L^2$ does not provide benefits. Due to the limited set of possible cluster flips after a certain threshold too many cluster flips in a row will recreate very similar configurations again and again.

Such a combination of local with non-local updates into a cluster hybrid algorithm has already been discussed earlier for discrete spin systems. Plascak et al. have found the combination of single-spin Metropolis and Wolff-cluster updates to both increase the efficiency and to reduce numerical errors caused by correlated pseudo-random numbers with respect to the uncombined algorithms [39]. The Wolff algorithm has been found particularly sensitive to the quality of the random number generator in use [40].

3.5 Parallel tempering

The previously described algorithms always treat the simulated system in an canonical ensemble, where it is kept in thermal equilibrium with a heat bath and hence at constant temperature. Further improvements can be made by relaxing this restriction and considering different ensembles. The methods presented in the following have been invented in this spirit.

3.5.1 Tempering methods

The idea of tempering methods is to treat the temperature at which a simulation is carried out as a dynamic variable. In this way short correlation times at high temperatures can be exploited to overcome free energy barriers to visit regions of phase space at lower temperatures that would otherwise only be reached with a significant amount of additional simulation time.

Simulated tempering

With simulated tempering [41, 42] the inverse temperature β takes on different values from a set $\{\, \beta_k \,\}_{k=1}^{K}$ and instead of the canonical ensemble at a single temperature an "expanded ensemble" with a joint partition function

$$\mathcal{Z}_{\mathrm{ST}} = \sum_{k=1}^{K} \mathrm{e}^{\alpha_k} \mathcal{Z}(\beta_k) = \sum_{k=1}^{K} \mathrm{e}^{\alpha_k} \int \mathrm{d}\boldsymbol{x}\, \mathrm{e}^{-\beta_k \mathcal{H}(\boldsymbol{x})} \tag{3.55}$$

is sampled. Temperatures are exchanged during the simulation of a single system according to the Metropolis criterion, where the transition probabilities assume the form

$$W(\boldsymbol{x}, \beta_k \to \beta_l) = \min \left\{ 1, \mathrm{e}^{-(\beta_l - \beta_k)\mathcal{H}(\boldsymbol{x})} + \alpha_l - \alpha_k \right\}. \tag{3.56}$$

The system configuration $\boldsymbol{x}$ is altered separately with regular update algorithms. A major disadvantage of simulated tempering is that the parameters α_k are not known in advance and can only be adjusted in an iterative fashion.

Parallel tempering

The method of parallel tempering, which is also known under the name of *replica exchange Monte Carlo*, has been proposed independently by multiple authors: Swendsen and Wang [43], Geyer [44] and Hukushima and Nemoto [45].

In contrast to simulated tempering, an "extended ensemble" characterized by a product of partition functions

$$\mathcal{Z}_{\mathrm{PT}} = \prod_{k=1}^{K} \mathcal{Z}(\beta_k) = \prod_{k=1}^{K} \int_{\{\, \boldsymbol{x} \,\}_k} \mathrm{d}\boldsymbol{x} \; \mathrm{e}^{-\beta_k \mathcal{H}(\boldsymbol{x})} \tag{3.57}$$

is sampled. Here, K different *replicas* of the system are simulated simultaneously. Taken on its own, each replica is kept in a canonical ensemble associated with a distinct inverse temperature β_k. A state of the extended ensemble is given by the set $\{\, \boldsymbol{x}_k, \beta_k \,\}_{k=1}^{K}$ and in equilibrium its probability distribution is given by the product of the canonical distributions of the individual replicas:

$$\mathcal{P}_{\mathrm{eq}}(\{\, \boldsymbol{x}_k, \beta_k \,\}_{k=1}^{K}) = \prod_{k=1}^{K} \mathcal{P}_{\mathrm{eq}}(\boldsymbol{x}_k, \beta_k) = \frac{1}{\mathcal{Z}_{\mathrm{PT}}} \prod_{k=1}^{K} \mathrm{e}^{-\beta_k \mathcal{H}(\boldsymbol{x}_k)} \tag{3.58}$$

In a parallel tempering simulation a Markov process in the extended ensemble is realized by exchanging system configurations among the replicas. In one step, configurations are swapped between two replicas k and l with a transition probability $W(\boldsymbol{x}_k, \beta_k | \boldsymbol{x}_l, \beta_l)$. According to Sec. 2.4 a valid choice of transition probabilities is given if they satisfy detailed balance:

$$\begin{aligned} &\mathcal{P}_{\mathrm{eq}}(\{\, \dots; \boldsymbol{x}_k, \beta_k; \dots; \boldsymbol{x}_l, \beta_l; \dots \,\}) W(\boldsymbol{x}_k, \beta_k | \boldsymbol{x}_l, \beta_l) \\ &\qquad = \mathcal{P}_{\mathrm{eq}}(\{\, \dots; \boldsymbol{x}_l, \beta_k; \dots; \boldsymbol{x}_k, \beta_l; \dots \,\}) W(\boldsymbol{x}_l, \beta_k | \boldsymbol{x}_k, \beta_l) \end{aligned} \tag{3.59}$$

By Eq. (3.58) this fixes their ratio:

$$\frac{W(\boldsymbol{x}_k, \beta_k | \boldsymbol{x}_l, \beta_l)}{W(\boldsymbol{x}_l, \beta_k | \boldsymbol{x}_k, \beta_l)} = \frac{\mathrm{e}^{-\beta_k \mathcal{H}(\boldsymbol{x}_l)} \mathrm{e}^{-\beta_l \mathcal{H}(\boldsymbol{x}_k)}}{\mathrm{e}^{-\beta_k \mathcal{H}(\boldsymbol{x}_k)} \mathrm{e}^{-\beta_l \mathcal{H}(\boldsymbol{x}_l)}} = \mathrm{e}^{[\beta_l - \beta_k][\mathcal{H}(\boldsymbol{x}_l) - \mathcal{H}(\boldsymbol{x}_k)]}. \tag{3.60}$$

Usually the transition probabilities are realized by an adaption of the Metropolis algorithm from Sec. 2.4.2 to the extended ensemble. Writing $\Delta = \Delta\beta\Delta E = [\beta_l - \beta_k][\mathcal{H}(\boldsymbol{x}_l) - \mathcal{H}(\boldsymbol{x}_k)]$, this leads to

$$W(\boldsymbol{x}_k, \beta_k | \boldsymbol{x}_l, \beta_l) = \min\left\{ 1, \mathrm{e}^{\Delta} \right\} = \begin{cases} \mathrm{e}^{\Delta}, & \text{if } \Delta \leq 0, \\ 1, & \text{if } \Delta > 0. \end{cases} \tag{3.61}$$

In contrast to simulated tempering there are no additional weighting factors that would have to be determined.

The inverse temperatures of replica exchange partners should be chosen relatively close to each other so that the canonical energy distributions at both temperatures overlap. Else the magnitude of ΔE would be large, but enter with the sign opposing $\Delta\beta$. Hence acceptance probabilities would be small and there would be little to no efficiency gain. For this reason typically only exchanges between pairs of nearest-neighbor inverse temperature points β_k are proposed.

The parallel tempering routine can be combined with any canonical algorithm used to update the replicas. After a certain number of canonical updates replica exchanges are proposed and accepted with a probability according to Eq. (3.61).

In this way, an estimate for the expectation value of some observable $\mathcal{O}$ at one of the inverse temperatures β_k used in the simulation can be determined as the usual sample mean:

$$\widehat{\mathcal{O}}(\beta_k) = \frac{1}{M} \sum_{t=1}^{M} \mathcal{O}(\boldsymbol{x}_k(t)) \tag{3.62}$$

See Algorithm 3.5 for a pseudo-code rendition of the parallel tempering procedure. The actual implementation differs from that outline in so far as not the system conformations $\boldsymbol{x}_k$ are exchanged between the replicas, but the inverse temperatures β_k. This is fully equivalent, but reduces the computational overhead at the price of slightly obstructed notation due to additional "bookkeeping" of replica and temperature indexes. With this scheme, one needs to keep track of the temperature time series $\{ \beta_l(t) \}_{t=1}^{M}$ of each replica l, $l = 1, \ldots, K$, to estimate expectation values:

$$\widehat{\mathcal{O}}(\beta_k) = \frac{1}{M} \sum_{t=1}^{M} \sum_{l=1}^{K} \mathcal{O}(\boldsymbol{x}_l(t)) \delta_{\beta_l(t), \beta_k}. \tag{3.63}$$

Algorithm 3.5 Parallel tempering update step

```
for $k = 1$ to $K$ do
    for a certain number of Monte Carlo sweeps do
        update replica $k$ canonically at $\beta_k$
    end for
end for
for $k = 1$ to $K - 1$ do
    $\Delta \leftarrow [\beta_k - \beta_{k+1}][\mathcal{H}(\boldsymbol{x}_k) - \mathcal{H}(\boldsymbol{x}_{k+1})]$
    if $\Delta \geq 0$ then
        accept exchange
    else
        $r \leftarrow$ random number from $[0, 1)$
        if $r \leq \mathrm{e}^{\Delta}$ then
            accept exchange
        end if
    end if
    if exchange accepted then
        swap configurations $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}$
    end if
end for
```

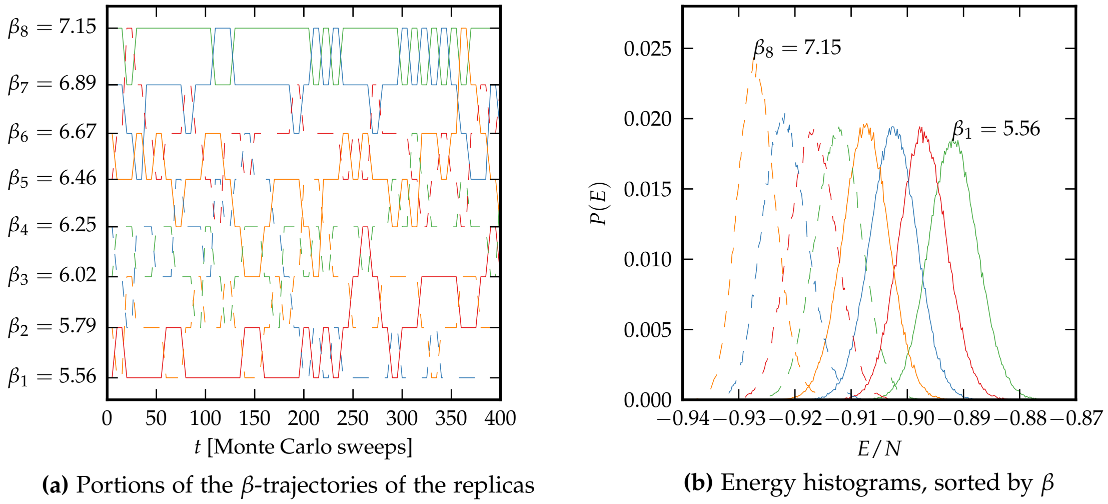
In Fig. 3.7 example replica-trajectories in temperature space are given as an illustration along with the measured energy distributions at the different simulation temperatures. The energy histograms at neighboring temperatures show significant overlap.

Of course the simulation of multiple replicas costs CPU time. However, for many systems the improvement of statistics due to reduced correlations outweighs these considerations. Even if that is not fully the case, one is often interested in results over a range of temperatures, which are given directly with parallel tempering. Moreover, the algorithm can be parallelized quite easily with little need for communication, so if sufficient resources are available, the impact on real "wall clock" time is limited.

3.5.2 Selection of temperature intervals

It has already been mentioned above that the energy distributions at the inverse temperatures $\{ \beta_k \}$ that are taken as replica exchange partners need to have sufficient overlap because else exchanges would only be accepted very rarely. An order of magnitude estimate of the logarithm of the probability e^{Δ} to exchange β_k and $\beta_{k+1} = \beta_k + \delta\beta$ is [45]:

$$\begin{aligned}\Delta &= \delta\beta \cdot [\mathcal{H}(\boldsymbol{x}_{k+1}) - \mathcal{H}(\boldsymbol{x}_k)] \\ &\sim (\delta\beta)^2 \cdot \frac{\langle E\rangle(\beta_k + \delta\beta) - \langle E\rangle(\beta_k)}{\delta\beta} \sim (\delta\beta)^2 \cdot \frac{\mathrm{d}E(\beta_k)}{\mathrm{d}\beta} = -\frac{(\delta\beta)^2}{k_\mathrm{B}\beta_k^2} \cdot C(\beta_k). \qquad (3.64)\end{aligned}$$



(a) Portions of the β-trajectories of the replicas

(b) Energy histograms, sorted by β

Figure 3.7: Example β-trajectories and energy histograms from a short parallel tempering simulation of the two-dimensional compass model with screw-periodic boundary conditions at $L = 36$ with 8 replicas. Temperature swaps were proposed every 5 sweeps. The energy histograms overlap, which is necessary for the exchanges to be accepted with non-negligible probabilities.

Eq. (3.64) gives raise to two considerations:

1. The energy E and the heat capacity C are extensive quantities that scale with the system size N. To maintain acceptance probabilities on a constant level the spacing of the inverse temperatures $\delta\beta$ should be chosen on the order of $1/\sqrt{N}$.

2. The heat capacity C diverges at a phase transition. This leads to low acceptance probabilities if the effect is not compensated by a closer spacing $\delta\beta$ in the vicinity of the pseudo-transition temperature of the finite system. A more rigorous relation between heat capacity and acceptance probabilities has been given by Predescu et al. [46].

The second effect has been very noticeable in simulations of the three-dimensional compass model, see Fig. 3.8 for an example.

3.5.2.1 Constant acceptance probabilities

To compensate for the effect of the heat capacity many schemes have been proposed in the literature that aim for constant exchange acceptance probabilities. Of note is the work by Predescu et al. [46, 47], Kofke and Kone [48–50] and Rathore et al. [51].

By having these local transition probabilities in temperature space uniform, one hopes to achieve a random walk of the replicas over the whole temperature range instead of them being restricted to disjoint segments.

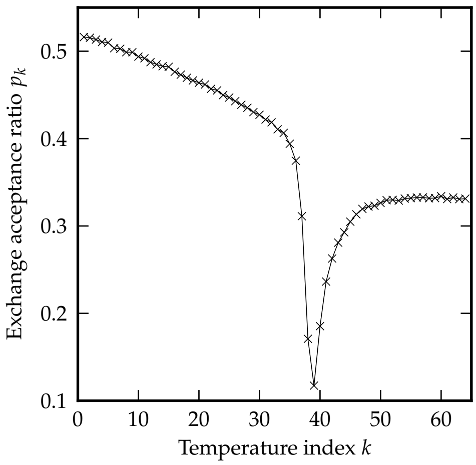


Figure 3.8: Measured exchange acceptance ratios of neighboring temperature pairs in a parallel tempering simulation of the three-dimensional compass model with screw-periodic boundary conditions at $L = 28$. The model was simulated at 64 inverse temperatures between $\beta_1 = 8$ and $\beta_{64} = 12.5$ chosen with constant separation $\delta\beta$. The minimum of the acceptance ratio is at $\beta_{39} = 10.29$, close to the transition point.

Following Ref. [48], the previously alluded relation between energy distribution overlap and exchange acceptance probability can be quantified: Let the expectation value of that probability between two inverse temperatures β_k and β_l be denoted by $\langle W_{kl} \rangle \equiv \langle W(\boldsymbol{x}_k, \beta_k | \boldsymbol{x}_l, \beta_l) \rangle_{\boldsymbol{x}_k, \boldsymbol{x}_l}$ and write $P_k(E) = \mathcal{P}_{\text{eq}}(E, \beta_k)$. Without loss of generality assume $\beta_k > \beta_l$. Then

$$\begin{aligned}\langle W_{kl} \rangle &= \int_{-\infty}^{\infty} \mathrm{d}E_k P_k(E_k) \int_{-\infty}^{\infty} \mathrm{d}E_l P_l(E_l) \min\left\{ 1, \mathrm{e}^{[\beta_k - \beta_l][E_k - E_l]} \right\} \\ &= \int_{-\infty}^{\infty} \mathrm{d}E_k \int_{-\infty}^{E_k} \mathrm{d}E_l P_k(E_k) P_l(E_l) \mathrm{e}^{[\beta_k - \beta_l][E_k - E_l]} + \int_{-\infty}^{\infty} \mathrm{d}E_k \int_{E_k}^{\infty} \mathrm{d}E_l P_k(E_k) P_l(E_l)\end{aligned}$$

and with the insertion of $\mathcal{P}_{\text{eq}}(E, \beta) = \Omega(E) \mathrm{e}^{-\beta E} / \mathcal{Z}(\beta)$, where $\Omega(E)$ is the energy density of states, into the first term and the application of Fubini's theorem on the second term one obtains

$$\begin{aligned}\langle W_{kl} \rangle &= \int_{-\infty}^{\infty} \mathrm{d}E_k \frac{\Omega(E_k) \mathrm{e}^{-\beta_l E_k}}{\mathcal{Z}(\beta_l)} \int_{-\infty}^{E_k} \mathrm{d}E_l \frac{\Omega(E_l) \mathrm{e}^{-\beta_k E_l}}{\mathcal{Z}(\beta_k)} + \int_{-\infty}^{\infty} \mathrm{d}E_k \int_{-\infty}^{E_k} \mathrm{d}E_l P_k(E_k) P_l(E_l) \\ &= 2 \int_{-\infty}^{\infty} \mathrm{d}E_k P_k(E_k) \int_{-\infty}^{E_k} \mathrm{d}E_l P_l(E_l). \qquad (3.65)\end{aligned}$$

The double integral in Eq. (3.65) is a measure for the overlap of the two distributions. Its value ranges from 0 if they are completely disjoint to 1 if they are identical.

Consequently, in order to have constant exchange acceptance probabilities it suffices to find a range of inverse temperatures $\{ \beta_k \}$ that leads to canonical energy distributions $\mathcal{P}_{\text{eq}}(E, \beta_k)$ with constant overlap between neighbors. I have applied an empirical scheme to find such a set $\{ \beta_k \}$:

1. A short preliminary parallel tempering simulation is run with an ad hoc distribution of temperatures $\{ \beta_k^0 \}$ spanning the range of interest. This yields rough estimates of $\mathcal{P}_{\text{eq}}(E, \beta_k^0)$ in the form of energy histograms.
2. These histograms are subjected to the multiple histogram analysis method outlined in Sec. 3.6 to approximately estimate the discretized energy density of states $\Omega(E)$.
3. From that estimate the energy distribution at arbitrary inverse temperatures contained in the original range can be reweighted directly:

$$\widehat{\mathcal{P}_{\text{eq}}}(E, \beta) = \frac{\widehat{\Omega}(E) e^{-\beta E}}{\widehat{\mathcal{Z}}(\beta)}, \tag{3.66}$$

   where $\widehat{\mathcal{Z}}(\beta)$ is determined as a normalization constant.
4. Iteratively the β_k are adjusted until the histogram overlap between neighboring temperature points is approximately constant over the whole range. Here the overlap is quantized as the common area under the distribution functions. The method used for this iteration is sketched in Algorithm 3.6.

This method works very well to ensure approximately constant exchange probabilities although already the preliminary simulations can be quite time consuming. Nevertheless, in simulations of the three-dimensional compass model on relatively large lattices, the goal of achieving an unbiased random walk in temperature space could not be achieved in this way. See Fig. 3.9 on p. 56 for an example trajectory in temperature space of one replica of the $N = 32^3$ compass model. There is a clear separation into two regions below and above the transition point with few crossings in between. Evidently the influence of the phase transition has not been eliminated by compensating for the effect of a diverging heat capacity on the exchange probabilities.

3.5.2.2 Optimized flow

In a different approach for the selection of temperatures in parallel tempering one sets out to minimize the total mean first passage time or average round-trip time τ_{rt}, which is the number of proposed exchange updates a replica needs to travel from the lowest

Algorithm 3.6 Reweight to approximately constant energy overlap

```
Require: original inverse temperatures $\{ \beta_k^0 \}_{k=1}^K$, ordered $\beta_k^0 < \beta_{k+1}^0$
Ensure: new inverse temperatures $\{ \beta_k \}_{k=1}^K$ with $\beta_1 = \beta_1^0$, $\beta_K = \beta_K^0$, $\beta_k < \beta_{k+1}$ and
    overlap($\widehat{\mathcal{P}_{\text{eq}}}(E, \beta_k), \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_{k+1})$) approximately constant
$t \leftarrow$ average next neighbor overlap of $\left\{ \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_k^0) \right\}$
$\beta_1 \leftarrow \beta_1^0$, $\beta_K \leftarrow \beta_K^0$
repeat
    {Find inverse temperatures closing in from above and below:}
    $l \leftarrow 1, h \leftarrow K$
    while $l < h$ do
        $\beta_{l+1} \leftarrow$ try to find between $\beta_l$ and $\beta_h$ such that overlap($\widehat{\mathcal{P}_{\text{eq}}}(E, \beta_l), \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_h)$)
        is $t$ within 5% by continuously bisecting the $\beta$-interval and reweighting trial
        histograms; if impossible take the last value from the bisection
        $\beta_{h-1} \leftarrow$ try to find in the same way between $\beta_l$ and $\beta_h$ such that
        overlap($\widehat{\mathcal{P}_{\text{eq}}}(E, \beta_{h-1}), \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_h)$) $\approx t$
        $l \leftarrow l + 1, h \leftarrow h - 1$
    end while
    {Handle central temperature interval:}
    $\beta_{l+1} \leftarrow$ try to find in the same way between $\beta_l$ and $\beta_{l+2}$ such that
    overlap($\widehat{\mathcal{P}_{\text{eq}}}(E, \beta_l), \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_{l+1})$) $\approx t$
    {Adjust $t$ if some overlap is much smaller:}
    for $k \leftarrow 1$ to $K - 1$ do
        $o \leftarrow$ overlap($\widehat{\mathcal{P}_{\text{eq}}}(E, \beta_k), \widehat{\mathcal{P}_{\text{eq}}}(E, \beta_{k+1})$)
        if $\frac{o-t}{t} < -5\%$ then
            $t \leftarrow t + \frac{o-t}{K-1}$
        end if
    end for
until no adjustment of $t$ was necessary
```

inverse temperature β_1 to the highest β_K and back. Such an optimization would prevent situations like the one in Fig. 3.9. Adapting work by Trebst et al. [52], Katzgraber et al. have shown [53] under some assumptions on the temperature dependence of the local diffusion coefficient that the set of inverse temperatures $\{ \beta_k \}$ optimized in the sense of minimizing $\overline{\tau_{\text{rt}}}$ can be found by studying the fraction of replicas moving up in inverse temperature space [1]

$$f_{\text{up}}(\beta_k) = \frac{n_{\text{up}}(\beta_k)}{n_{\text{up}}(\beta_k) + n_{\text{down}}(\beta_k)}. \tag{3.67}$$

---

[1] Actually the discussion in Ref. [53] is formulated in terms of temperatures T_k, but the results are equivalent.

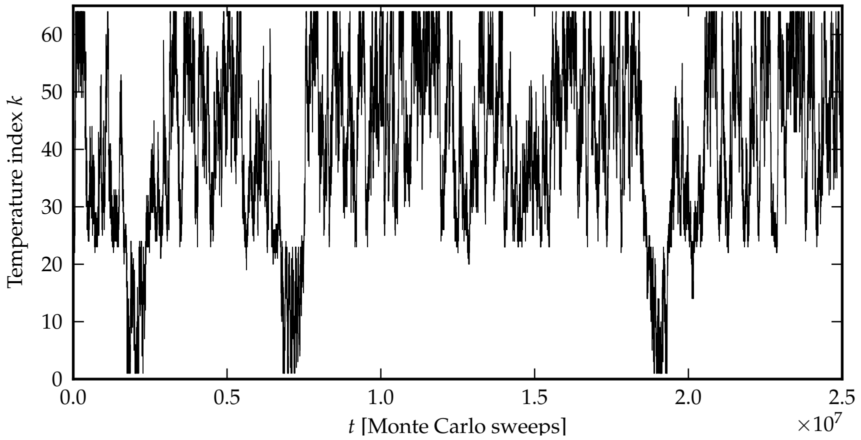


Figure 3.9: Temperature index time series of one replica from a parallel tempering simulation of the three-dimensional compass model at $L = 32$ with β ranging from 9 to 12.5. Exchange acceptance probabilities are approximately constant at 0.42 ± 0.03 for all pairs of temperatures, but there still is a clear bottleneck close to the location of the phase transition. The trajectory over the whole simulation of $2.5 \cdot 10^7$ sweeps is shown heavily scaled down. Still all crossings into the region of $k \lesssim 20$ are visible.

Here $n_{\mathrm{up}}(\beta_k)$ counts the number of times a replica has visited β_k in the course of the simulation after having last visited β_1 and not β_K. Accordingly $n_{\mathrm{down}}(\beta_k)$ counts the number of replicas currently moving down from β_K when they visit β_k. The boundary conditions are of course set to $f_{\mathrm{up}}(\beta_1) = 1$ and $f_{\mathrm{up}}(\beta_K) = 0$. Now the optimal fraction is found to be linear in the inverse temperature index:

$$f_{\mathrm{up}}^{\mathrm{opt}}(\beta_k) = 1 - \frac{k-1}{K-1}. \tag{3.68}$$

The iteration scheme known as feedback-optimized parallel tempering has been proposed to find temperature ranges optimized in this sense. It has been applied successfully to the ferromagnetic and the fully frustrated Ising model [53] and polymer models [54]. A concise formulation of the algorithm is found in the paper by Nadler and Hansmann [55]. Some improvements are given by Hamze et al. [56]. However, for convergence the scheme requires to collect sufficient round-trip data, which can make the preliminary determination of the temperature range very time consuming.

An effect of this procedure of optimization is that exchange acceptance probabilities are in general no longer uniform for all temperatures. Instead, temperature points are concentrated near bottle-necks such as phase transitions, which leads to higher exchange rates in those regions.

A different, but similar approach of optimization has been presented by Bittner et al. [57]. Here the range of simulation temperatures is held fixed with constant exchange acceptance probabilities of 50%. The quantity that is adapted instead is the number of Monte Carlo sweeps a replica spends at a temperature before the next exchange is attempted. By scaling this time with the autocorrelation time measured in a preliminary canonical simulation at that temperature, optimal flow akin to the feedback-optimized parallel tempering could be achieved in simulations of the ferromagnetic Ising model. Otherwise, if waiting times before swaps are chosen too low, critical slowing down at the second-order phase transition causes the replicas to be highly correlated in time: Coming from the high-temperature phase they are not given the chance to assume an energy that would allow them to be exchanged with a low-temperature replica before they are already swapped back to a higher temperature.

A practical problem of this approach is that it cannot be parallelized as well as regular parallel tempering, where normally replicas are distributed over parallel computer nodes and updated concurrently until exchanges are coordinated from a master node. If some replicas are given significantly more time than others before the exchange, a lot of CPU time will be wasted on idle waiting.

3.5.2.3 Constant entropy difference

Kofke has provided the idea to relate replica exchange acceptance probabilities with the difference of the entropy S between the exchange partners [48]. Building on this, Sabo et al. have suggested [58] to choose the set of inverse temperatures $\{ \beta_k \}_{k=1}^{K}$ with the requirement that the difference of entropy between adjacent temperatures is constant. The authors do not give a rigorous motivation for this choice, but it has some intuitive appeal in the sense that each successive temperature point contributes the same amount of additional "information".

If volume is constant, the entropy difference between given inverse temperatures $\beta_1 = 1/k_B T_1$ and $\beta_K = 1/k_B T_K$ can be computed from the heat capacity C_V:

$$\Delta S = \int_{T_1}^{T_K} \mathrm{d}T \frac{C_V(T)}{T} = -\int_{\beta_1}^{\beta_K} \mathrm{d}\beta \frac{C_V(\beta)}{\beta}, \tag{3.69}$$

which for most physical systems is negative as long as $\beta_K > \beta_1$. Then the intermediate inverse temperatures can be found by sequentially solving

$$\int_{\beta_k}^{\beta_{k+1}} \mathrm{d}\beta \frac{C_V(\beta)}{\beta} = \frac{-\Delta S}{K-1} \tag{3.70}$$

for β_{k+1}. In general the heat capacity is not known beforehand. One could estimate it for the system size of interest in a preliminary simulation run or simply use results from a smaller system if available. In simulations of the 2D Ising model the authors of Ref. [58] have obtained results with high accuracy close to that achieved with the feed-back optimization algorithm of Ref. [53] even when an approximate model of the heat capacity was used.

As this approach promises to improve the replica exchange process in simulations of the compass model without requiring exhaustive preliminary iterations, I have decided to implement it. The following scheme is used to find a set $\{ \beta_k \}$ with approximately constant entropy difference between given inverse temperatures β_1 and β_K:

1. Use data from an earlier parallel tempering simulation at a smaller lattice size which included at least β_1 and β_K and apply the multiple histogram analysis from Sec. 3.6. This provides the ability to estimate $C_V(\beta)$.

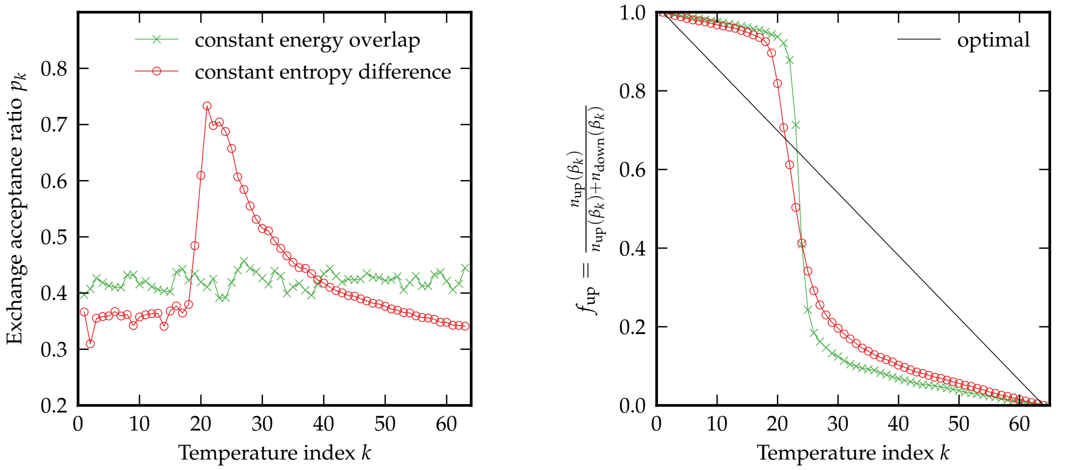
2. Estimate ΔS by Eq. (3.69) using an adaptive numerical integration routine from the GNU scientific library [59].

3. For $k = 1, \ldots, K-2$ solve Eq. (3.70) numerically using an algorithm which combines bisection and a Newton-Raphson solver similar to `rtsafe` in chapter 9.4 of Numerical Recipes [27].

I have found it to be quick and to deliver stable results.

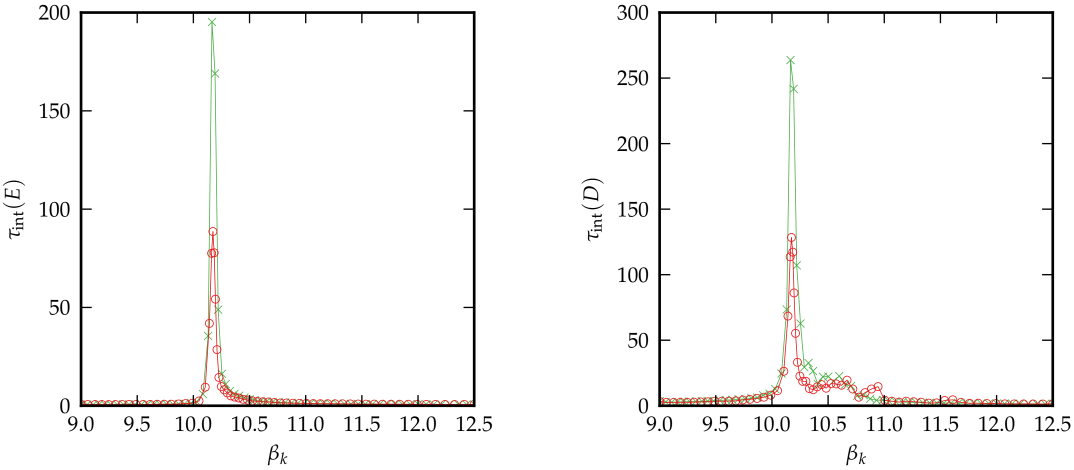
To estimate the quality of the generated inverse temperature range $\{ \beta_k \}$, I have repeated the parallel tempering simulation of the three-dimensional compass model with screw-periodic boundary conditions at $L = 32$, where diffusion in temperature space was strongly inhibited with the constant acceptance ratio ensemble of temperatures. All parameters were kept identical except for the distribution of the 64 temperature points in the β-range from 9 to 12.5, which was now chosen for constant entropy difference. Both simulations were run for $2.5 \cdot 10^7$ sweeps after $2.5 \cdot 10^6$ sweeps of thermalization. Measurements were made every 50 sweeps, yielding $5 \cdot 10^5$ samples at each temperature. Replica exchanges were attempted every 100 sweeps.

The exchange acceptance ratios in Fig. 3.10(a) are no longer flat, but higher in the vicinity of the phase transition. By Fig. 3.10(b) the diffusion in temperature space is improved, but sill far from the optimal behavior. Finally, the squared statistical accuracy of the measurements in the vicinity of the transition temperature is improved approximately by a factor of two as one can see in the integrated autocorrelation times shown in Figs. 3.10(c) and 3.10(d). They have been estimated by summing up the autocorrelation function of the time series ordered by temperature and should not be understood as precise measurements. Moreover it should be noted that they are given in units of 50 Monte Carlo sweeps.



(a) Exchange acceptance ratios

(b) Fraction f_{up} characterizing flow



(c) Autocorrelation times of the energy E

(d) Autocorrelation times of the order parameter D

Figure 3.10: Comparison of parallel tempering simulations with constant energy distribution overlap against constant entropy differences between adjacent temperatures

3.6 Multiple histogram reweighting

The multiple histogram method was originally introduced by Ferrenberg and Swendsen in 1989 [60] as a procedure to join together several sets of simulation data generated at different temperatures in an optimized way. It has been made particularly popular by Kumar et al. in 1992 in a slightly extended version under the name Weighted Histogram Analysis Method, frequently abbreviated WHAM [61].

The multiple histogram method allows one to estimate expectation values of an arbitrary observable $\mathcal{O}$ over a wide range of inverse temperatures β. A simpler reweighting technique that uses simulation data obtained at only a single temperature to extrapolate at nearby temperatures had already been proven to be effective earlier by Ferrenberg and Swendsen [62]. But thanks to the combination of data acquired independently at various temperatures the multiple histogram method has the potential to provide results with lower statistical uncertainties.

When parallel tempering is used in Monte Carlo simulations, generally a wide range of temperatures is sampled. To estimate expectation values of observables in the canonical ensemble at specific temperatures a method that discards as little of the simulation data as possible is preferable. The multiple histogram reweighting method as described in the following almost suggests itself for such a scheme of analysis.

3.6.1 Analysis of independent simulations

The following presentation of the multiple histogram method mostly follows the textbook by Newman and Barkema [19] and the derivation given by Chodera et al. [63].

The fundamental idea is to combine the data from K independent simulations that sample from the canonical ensembles at inverse temperatures $\beta_1, \beta_2, \dots, \beta_K$ to estimate the internal energy density of states $\Omega(E)$, which is independent of temperature, for all energies reached in the simulations. $\Omega(E)$ can then be used to estimate expectation values of any observable at arbitrary inverse temperatures β, that do not need to match any of the β_k at which the simulations were originally carried out:

$$\langle \mathcal{O} \rangle_\beta = \frac{\int \mathrm{d}E \Omega(E) \mathrm{e}^{-\beta E} \mathcal{O}(E)}{\int \mathrm{d}E \Omega(E) \mathrm{e}^{-\beta E}}, \tag{3.71}$$

where $\mathcal{O}(E)$ is the average of $\mathcal{O}$ over all states with internal energy E:

$$\mathcal{O}(E) = \frac{\int \mathrm{d}\boldsymbol{x} \delta(E(\boldsymbol{x}) - E) \mathcal{O}(E)}{\int \mathrm{d}\boldsymbol{x} \delta(E(\boldsymbol{x}) - E)}. \tag{3.72}$$

3.6.1.1 Estimating the density of states from single simulations

For each $k = 1, \dots, K$ first individual estimates $\widehat{\Omega}_k(E)$ of the density of states using only the information from a single simulation at β_k are made. The $\widehat{\Omega}_k(E)$ will then be combined in an optimized way for the complete estimate $\widehat{\Omega}(E)$.

In the following the probability density function $p(E|\beta)$ of the energy at inverse temperature β will be estimated based on histograms. While this will introduce some systematic discretization error in the case of a continuous system such as the compass model, this approach is simple and efficient to implement. With this in mind, $p(E|\beta)$ will be considered only at I different energy levels E_i, which are chosen equally spaced ΔE apart and span the whole energy range sampled in the simulation. Writing $p_i(\beta) \equiv p(E_i|\beta)$ and $\Omega_i \equiv \Omega(E_i)$, one has:

$$p_i(\beta) = \Omega_i \mathrm{e}^{-\beta E_i} / \mathcal{Z}(\beta). \tag{3.73}$$

The normalizing constant is the partition function $\mathcal{Z}(\beta) \approx \Delta E \sum_{i=1}^{I} \Omega_i \mathrm{e}^{-\beta E_i}$. Let $\{E_{km}\}_{m=1}^{M_k}$ denote the time series of length M_k containing energy measurements from the k'th simulation. Then define $\psi_i(E)$ as the characteristic function for the energy bin of width ΔE centered about E_i:

$$\psi_i(E) = \begin{cases} 1, & \text{if } E \in [E_i - \Delta E/2, E_i + \Delta E/2), \\ 0, & \text{else.} \end{cases} \tag{3.74}$$

With the shorthand $\psi_{ikm} \equiv \psi_i(E_{km})$ the histogram which counts samples in energy bin i from simulation k can be expressed as $H_{ik} = \sum_{m=1}^{M_k} \psi_{ikm}$. Then $p_i(\beta_k)$ can be estimated as

$$\widehat{p}_i(\beta_k) = \frac{1}{\Delta E} \cdot \frac{H_{ik}}{M_k}, \tag{3.75}$$

which can be biased by discretization, if $\langle \psi_{ikm} \rangle = \Delta E \cdot p_i(\beta_k)$ does not hold. Combining this with Eq. (3.73) an estimate of the density of states from simulation k can be obtained:

$$\widehat{\Omega}_{ik} = \frac{1}{\Delta E} \cdot \frac{H_{ik}}{M_k} \cdot \frac{\mathcal{Z}(\beta)}{\mathrm{e}^{-\beta E_i}}. \tag{3.76}$$

3.6.1.2 Optimizing the estimate of the density of states

To construct an optimal estimator for the density of states a weighted combination of the independent estimates $\widehat{\Omega}_{ik}$ from the K individual simulations can be formed:

$$\widehat{\Omega}_i = \sum_{k=1}^{K} r_{ik} \cdot \widehat{\Omega}_{ik} \tag{3.77}$$

Here relative weights r_{ik}, which are independent for each energy bin i, have been introduced. They are subject to the constraint

$$\sum_{k=1}^{K} r_{ik} = 1. \tag{3.78}$$

As long as the $\widehat{\Omega}_{ik}$ are not biased, which could be a consequence of the discretization applied earlier, $\widehat{\Omega}_i$ is an unbiased estimator for any choice of r_{ik}. A choice of r_{ik} deemed "optimal" is the one that minimizes the variance

$$\delta^2\widehat{\Omega}_i = \left\langle \left(\widehat{\Omega}_i - \langle\widehat{\Omega}_i\rangle\right)^2 \right\rangle = \langle\widehat{\Omega}_i^2\rangle - \langle\widehat{\Omega}_i\rangle^2. \tag{3.79}$$

Since the $\widehat{\Omega}_{ik}$ are independent for each i, the variance results in

$$\delta^2\widehat{\Omega}_i = \sum_{k=1}^{K} {r_{ik}}^2 \cdot \delta^2\widehat{\Omega}_{ik}. \tag{3.80}$$

To minimize $\delta^2\widehat{\Omega}_i$ under the constraint (3.78) one can introduce a Lagrange multiplier λ and solve

$$\frac{\partial}{\partial r_{ik}} \left[\delta^2\widehat{\Omega}_i - \lambda \left(\sum_{k'=1}^{K} r_{ik'} - 1 \right) \right] = 0, \tag{3.81}$$

which yields $r_{ik} = \lambda / (2 \cdot \delta^2\widehat{\Omega}_{ik})$ and another application of Eq. (3.78) gives $\lambda = 1/\left[\sum_{k=1}^{K} [2 \cdot \delta^2\widehat{\Omega}_{ik}]^{-1}\right]$. Thus one has

$$r_{ik} = \frac{[\delta^2\widehat{\Omega}_{ik}]^{-1}}{\sum_{k'=1}^{K} [\delta^2\widehat{\Omega}_{ik'}]^{-1}} \tag{3.82}$$

and the optimal estimator for the density of states results in

$$\widehat{\Omega}_i = \frac{\sum_{k=1}^{K} [\delta^2\widehat{\Omega}_{ik}]^{-1} \cdot \widehat{\Omega}_{ik}}{\sum_{k=1}^{K} [\delta^2\widehat{\Omega}_{ik}]^{-1}} \tag{3.83}$$

with corresponding variance

$$\delta^2\widehat{\Omega}_i = \left[\sum_{k=1}^{K}[\delta^2\widehat{\Omega}_{ik}]^{-1}\right]^{-1}. \tag{3.84}$$

3.6.1.3 Estimation of histogram uncertainties

To evalutate the expressions (3.83) and (3.84) the uncertainties in the estimates of the density of states from the individual simulations $\widehat{\Omega}_{ik}$ need to be estimated. In the expression (3.76) β, E_i, ΔE and M_k are parameters known exactly. While the partition function $\mathcal{Z}(\beta)$ is unknown beforehand, it has a well-defined value as a constant of normalization. Here it does not enter as a quantity estimated with uncertainty. The squared uncertainty $\delta^2\widehat{\Omega}_{ik}$ thus stems solely from the uncertainty of the measured histogram count and can now be stated in dependence of $\mathcal{Z}(\beta)$:

$$\delta^2\widehat{\Omega}_{ik} = \frac{\delta^2 H_{ik}\cdot\mathcal{Z}(\beta)}{\Delta E\cdot M_k\cdot \mathrm{e}^{-\beta E_i}}. \tag{3.85}$$

For the evaluation of $\delta^2 H_{ik}$ it is fruitful to express H_{ik} as a scaled time average of the characteristic function $\psi_i(E)$, defined in Eq. (3.74), evaluated on the correlated energy measurements sampled in the simulation:

$$H_{ik} = M_k\cdot\frac{1}{M_k}\sum_{m=1}^{M_k}\psi_{ikm} = M_k\cdot\widehat{\psi_{ik}}, \qquad \widehat{\psi_{ik}} = \frac{1}{M_k}\sum_{m=1}^{M_k}\psi_{ikm}. \tag{3.86}$$

Then, considering temporal autocorrelations of ψ_{ik} with the statistical inefficiency factor $g_{ik} = 1 + 2\tau_{\mathrm{int}}(\psi_{ik})$, as introduced in Sec. 2.5, one finds:

$$\begin{aligned}\delta^2 H_{ik} &= N_k{}^2\cdot\frac{\sigma^2_{\psi_{ik}}}{M_k/g_{ik}} = g_{ik}\,M_k\left[\langle\psi_{ik}{}^2\rangle - \langle\psi_{ik}\rangle^2\right]\\ &= g_{ik}\,M_k\,\langle\psi_{ik}\rangle\left[1-\langle\psi_{ik}\rangle\right] = g_{ik}\,\langle H_{ik}\rangle\left[1-\frac{\langle H_{ik}\rangle}{M_k}\right],\end{aligned} \tag{3.87}$$

where $\psi_i(E)^2 = \psi_i(E)$ as ψ_i only takes values of 1 and 0. Often it is assumed that there are sufficiently many energy bins so that each individual bin is sparsely populated. Then $\langle H_{ik}\rangle/M_k \ll 1$ and one can simplify Eq. (3.87) to

$$\delta^2 H_{ik} \approx g_{ik}\,\langle H_{ik}\rangle. \tag{3.88}$$

However, as energy fluctuations in the canonical ensemble typically are described by a narrow probability distribution around the expectation value, this claim should be tested on a case by case basis. Taking the limit $\langle H_{ik} \rangle / M_k \to 0$ corresponds to assuming a Poisson distribution for the histogram entries.

The statistical inefficiency g_{ik} reflects the number of measurements that is necessary to sample the energy bin without correlations. In general it depends on bin index, bin width and temperature. Kumar et al. left out all three of these dependencies in their presentation [61], but depending on the simulation method correlation times can be expected to be shorter at higher temperatures and especially in the vicinity of phase transitions, which cause (super-) critical slowing down. Ferrenberg and Swendsen allowed for a temperature dependence of the statistical inefficiency factor [60], but proposed to calculate it from the the autocorrelation time of the whole energy time series, neglecting any variation for different bins, although depending on the energy landscape the simulation can be expected to spend longer times in certain regions of phase space than in others. Furthermore the relation between the statistical uncertainty of the time average of some observable such as the energy and the uncertainties of the individual histogram counts is not trivial.

The full dependency on temperature and bin index is acknowledged in later work by Gallicchio et al. [64]. A clear derivation is contained in the paper by Chodera et al. [63].

Up to some to discretization error $\langle H_{ik} \rangle$ can be evaluated to

$$\begin{aligned} \langle H_{ik} \rangle &\approx M_k \Delta E p_i(\beta_k) = M_k \Delta E \Omega_i \, \mathrm{e}^{-\beta_k E_i} / \mathcal{Z}(\beta) \\ &= M_k \Delta E \Omega_m \, \mathrm{e}^{-\beta_k E_i + f_k}, \end{aligned} \tag{3.89}$$

where the dimensionless free energies $f_k = -\ln \mathcal{Z}(\beta_k)$ have been introduced. The best estimate of $\langle H_{ik} \rangle$ is obtained using the best estimate of Ω_i:

$$\widehat{H}_{ik} = M_k \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k}. \tag{3.90}$$

By inserting this into Eq. (3.87) one obtains an estimate of the squared uncertainty $\delta^2 H_{ik}$:

$$\delta^2 \widehat{H}_{ik} = g_{ik} M_k \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k} (1 - \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k}) \tag{3.91}$$

and subsequently by the means of Eq. (3.85) an estimate of $\delta^2 \widehat{\Omega}_{ik}$:

$$\delta^2 \widehat{\Omega}_{ik} = \frac{g_{ik} \widehat{\Omega}_i (1 - \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k})}{M_k \Delta E \, \mathrm{e}^{-\beta_k E_i + f_k}}. \tag{3.92}$$

3.6.1.4 Equations for estimators of the density of states

Finally using Eqs. (3.76) and (3.92) the optimized estimator for the density of states from Eq. (3.83) can be evaluated to:

$$\widehat{\Omega}_i = \frac{\sum_{k=1}^{K} H_{ik} \cdot [g_{ik}(1 - \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k})]^{-1}}{\sum_{k=1}^{K} M_k \Delta E \, \mathrm{e}^{-\beta_k E_i + f_k} [g_{ik}(1 - \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta_k E_i + f_k})]^{-1}}. \tag{3.93}$$

The nonlinear dependency on $\widehat{\Omega}_i$ vanishes if the preceding calculations are carried out with the approximation (3.88):

$$\widehat{\Omega}_i = \frac{\sum_{k=1}^{K} H_{ik} \cdot {g_{ik}}^{-1}}{\sum_{k=1}^{K} M_k \Delta E \, \mathrm{e}^{-\beta_k E_i + f_k} \cdot {g_{ik}}^{-1}}. \tag{3.94}$$

In both cases there remains a dependency on f_k, which can be evaluated from an estimate of $\widehat{\Omega}_i$ to:

$$f_k = -\ln \sum_{i=1}^{I} \widehat{\Omega}_i \Delta E \, \mathrm{e}^{-\beta_k E_i}. \tag{3.95}$$

By iteration of Eqs. (3.93) and (3.95) or Eqs. (3.94) and (3.95) the f_k can be determined in a self-consistent manner for all of the original inverse temperatures β_k. The $\widehat{\Omega}_i$ and f_k are expected to converge towards a fixed point eventually, although there is neither a guarantee for convergence nor an estimate for the speed of convergence. In practice convergence is usually quite fast from arbitrary starting values, but an educated "first guess" can be beneficial.

Instead of minimizing the variance of the estimator $\widehat{\Omega}_i$, Bartels and Karplus [65] have used a maximum likelihood approach to derive the same set of equations (3.94) and (3.95), but neglected differences of autocorrelation times. In their language statistical inefficiencies would be less easy to take into account.

Under the assumption that the approximation (3.88) is valid, one has

$$\delta^2 \widehat{\Omega}_{ik} = \frac{\widehat{\Omega}_i}{{g_{ik}}^{-1} M_k \Delta E \, \mathrm{e}^{-\beta_k E_i + f_k}} \tag{3.96}$$

and by Eq. (3.84) there is also an estimate for the statistical uncertainty of the estimate of the density of states:

$$\delta^2 \widehat{\Omega}_i = \frac{\widehat{\Omega}_i}{\sum_{k=1}^{K} {g_{ik}}^{-1} M_k \Delta E \, \mathrm{e}^{-\beta_k E_i + f_k}}. \tag{3.97}$$

Combining this with Eq. (3.94), the squared relative uncertainty of the estimated density of states can be calculated:

$$\frac{\delta^2 \widehat{\Omega}_i}{\widehat{\Omega}_i^2} = \left[\sum_{k=1}^{K} H_{ik} \cdot {g_{ik}}^{-1} \right]^{-1} \tag{3.98}$$

In the case of totally uncorrelated data, i.e. if all $g_{ik} = 1$, it is equal to the inverse of the total number of samples in energy bin i.

3.6.1.5 Estimating observables

Once an estimate of the density of states has been obtained, it can be used to estimate expectation values of observables at arbitrary temperatures β that lie within the range of original simulation temperatures $\{ \beta_k \}$ or slightly outside of it.

It is straightforward to estimate the expectation values of arbitrary functions of the energy $\langle f(E) \rangle_\beta$. Since

$$\langle f(E) \rangle_\beta = \frac{\int \mathrm{d}E \Omega(E) \mathrm{e}^{-\beta E} f(E)}{\int \mathrm{d}E \Omega(E) \mathrm{e}^{-\beta E}} \tag{3.99}$$

an estimator based on the estimated density of states is

$$\begin{aligned} \widehat{f}(E, \beta) &= \frac{1}{\widehat{\mathcal{Z}}(\beta)} \sum_{i=1}^{I} \Delta E \widehat{\Omega}_i \mathrm{e}^{-\beta E_i} f(E_i) \quad \text{with} \\ \widehat{\mathcal{Z}}(\beta) &= \sum_{i=1}^{I} \Delta E \widehat{\Omega}_i \mathrm{e}^{-\beta E_i}. \end{aligned} \tag{3.100}$$

For instance an estimator for the heat capacity $C_V = k_\mathrm{B} \beta^2 (\langle E^2 \rangle - \langle E \rangle^2)$ is

$$\widehat{C}_V(\beta) = k_\mathrm{B} \beta^2 \left(\frac{1}{\widehat{\mathcal{Z}}(\beta)} \sum_{i=1}^{I} \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta E_i} \, [E_i]^2 - \left[\frac{1}{\widehat{\mathcal{Z}}(\beta)} \sum_{i=1}^{I} \Delta E \widehat{\Omega}_i \, \mathrm{e}^{-\beta E_i} E_i \right]^2 \right) \tag{3.101}$$

and the probability density function of the energy $p(E'|\beta) = \langle \delta(E - E') \rangle_\beta$ can be estimated for discretized energy points E_i via

$$\widehat{p}_i(\beta) = \widehat{\Omega}_i \mathrm{e}^{-\beta E_i} / \widehat{\mathcal{Z}}(\beta). \tag{3.102}$$

Moreover expectation values of functions $f(\mathcal{O})$ of an observable $\mathcal{O}$ that is not a function of the energy can also be estimated easily. One could extend Eq. (3.100) using a two-dimensional histogram of E and $\mathcal{O}$, but this can become quite cumbersome in the implementation and potentially introduces additional systematic errors due to the discretization of continuous observables. A better alternative is to formulate an estimator

$\widehat{\mathcal{O}}(\beta)$ for the expectation value $\langle\mathcal{O}\rangle_\beta$ that makes direct use of the time series $\{\mathcal{O}_{km}\}_{m=1}^{M_k}$ of sampled observable measurements from the K simulations. Based on Eqs. (3.71) and (3.102) one can write for discretized energies

$$\widehat{\mathcal{O}}(\beta) = \frac{1}{\widehat{\mathcal{Z}}_\beta}\sum_{i=1}^{I}\widehat{\Omega}_i \Delta E \mathrm{e}^{-\beta E_i}\widehat{\mathcal{O}}_i, \tag{3.103}$$

where $\widehat{\mathcal{O}}_i$ is an estimator for $\mathcal{O}_i$, the mean of $\mathcal{O}$ over all system configurations with an energy that lies within energy bin i:

$$\mathcal{O}_i = \frac{\int \mathrm{d}\boldsymbol{x}\mathcal{O}(\boldsymbol{x})\psi_i(E(\boldsymbol{x}))}{\int \mathrm{d}\boldsymbol{x}\psi_i(E(\boldsymbol{x}))}, \tag{3.104}$$

which can be estimated by considering the data from all K simulations:

$$\widehat{\mathcal{O}}_i = \frac{1}{H_i}\sum_{k=1}^{K}\sum_{m=1}^{M_k}\psi_{ikm}\mathcal{O}_{km}, \tag{3.105}$$

where $H_i = \sum_{k=1}^{K} H_{ik}$ is the total number of sampled configurations from all simulations that are consistent with energy bin i. Put together one has

$$\begin{aligned}\widehat{\mathcal{O}}(\beta) &= \frac{\sum_{i=1}^{I}\Delta E\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}\widehat{\mathcal{O}}_i}{\sum_{i=1}^{I}\Delta E\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}} = \frac{\sum_{i=1}^{I}\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}\frac{1}{H_i}\sum_{k=1}^{K}\sum_{m=1}^{M_k}\psi_{ikm}\mathcal{O}_{km}}{\sum_{i=1}^{I}\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}\underbrace{\frac{1}{H_i}\sum_{k=1}^{K}\sum_{m=1}^{M_k}\psi_{ikm}}_{=1}} \\ &= \frac{\sum_{k=1}^{K}\sum_{m=1}^{M_k}\mathcal{O}_{km}w_{km}(\beta)}{\sum_{k=1}^{K}\sum_{m=1}^{M_k}w_{km}(\beta)}\end{aligned} \tag{3.106}$$

with weights that are defined by

$$w_{km}(\beta) = \sum_{i=1}^{I}\frac{\psi_{ikm}\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}}{H_i}, \tag{3.107}$$

where only one term in the sum is not zero. These weights need only be determined up to a factor of normalization and can be computed once to be used in the estimation of expectation values of any observable function at an inverse temperature β. It may be advantageous to use Eq. (3.106) instead of Eq. (3.100) to estimate functions of the energy to reduce discretization error.

3.6.2 Non-iterative estimation of the density of states

An alternative to the iterative solution of Eqs. (3.94) and (3.95) has been proposed by Fenwick [66]. Instead of optimizing a linear combination of individual simulation estimates of the density of states $\widehat{\Omega}_{ik} = H_{ik} / [\Delta E M_k \mathrm{e}^{-\beta_k E_i + f_k}]$, this approach deals with the logarithm of the ratio between estimates for neighboring energy bins:

$$\widehat{s}_{ik} = \ln \frac{\widehat{\Omega}_{i+1,k}}{\widehat{\Omega}_{i,k}} = \ln \left[\frac{H_{i+1,k}}{H_{i,k}} \cdot \frac{\mathrm{e}^{-\beta E_i}}{\mathrm{e}^{-\beta E_{i+1}}} \right] = \ln H_{i+1,k} - \ln H_{i,k} + \beta \Delta E. \tag{3.108}$$

$k_\mathrm{B} \cdot \widehat{s}_{ik}$ is an estimate for the change of the microcanonical entropy between energies E_i and E_{i+1}. Then an optimized estimate $\widehat{s}_i$ is computed from the individual simulation estimates in an analogical manner to Eq. (3.83), where $\widehat{s}_{ik}$ is only taken into account for $H_{i+1,k} > 0$ and $H_{ik} > 0$:

$$\widehat{s}_i = \ln \frac{\widehat{\Omega}_{i+1}}{\widehat{\Omega}_i} = \frac{\sum_{k=1}^{K} \widehat{s}_{ik} [\delta^2 \widehat{s}_{ik}]^{-1}}{\sum_{k=1}^{K} [\delta^2 \widehat{s}_{ik}]^{-1}}. \tag{3.109}$$

The error weights $\delta^2 \widehat{s}_{ik}$ are computed in Ref. [66] by error propagation in a first order approximation:

$$\delta^2 \widehat{s}_{ik} \approx \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{ik}} \right]^2 \delta^2 H_{ik} + \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{i+1,k}} \right]^2 \delta^2 H_{i+1,k} \tag{3.110}$$

Subsequently Poisson distributed histogram counts are assumed, which corresponds to the approximation (3.88), and autocorrelations are neglected, so that $\delta^2 H_{ik} = \langle H \rangle_{ik}$. Then the expectation value of the histogram count is estimated by the histogram actually measured in the simulation: $\delta^2 \widehat{H}_{ik} = H_{ik}$, which leads to the following expression to be used as error weights in Eq. (3.108):

$$\delta^2 \widehat{s}_{ik} = \frac{1}{H_{ik}} + \frac{1}{H_{i+1,k}} = \frac{H_{ik} + H_{i+1,k}}{H_{ik}\, H_{i+1,k}}. \tag{3.111}$$

Then Eq. (3.109) can be evaluated directly. Since the density of state is only defined up to a factor of proportionality, an estimate of its logarithm can be constructed for all energy bins i, starting from some arbitrarily chosen $\ln \widehat{\Omega}_1$:

$$\ln \widehat{\Omega}_{i+1} = \ln \widehat{\Omega}_i + \widehat{s}_i \tag{3.112}$$

The implementation of this method poses fewer numerical difficulties then the traditional approach of Sec. 3.6.1 and the density of states can be computed in a single pass.

One should note that the estimation of the uncertainty in Eq. (3.110) is not really correct because it assumes the histogram counts H_{ik} and $H_{i+1,k}$ to be uncorrelated. But since every configuration sampled in a simulation can only fall into a single energy bin, the energy histograms are actually purely anticorrelated. Making use of the characteristic function defined in Eq. (3.74), their covariance can be calculated directly:

$$\begin{aligned} \mathrm{Cov}\left[H_{ik}, H_{i+1,k}\right] &= \langle H_{ik} H_{i+1,k} \rangle - \langle H_{ik} \rangle \langle H_{i+1,k} \rangle \\ &= \sum_{m_1=1}^{M_k} \sum_{m_2=1}^{M_k} \underbrace{\langle \psi_{ikm_1} \psi_{i+1,k,m_2} \rangle}_{=0} - \langle H_{ik} \rangle \langle H_{i+1,k} \rangle \\ &= -\langle H_{ik} \rangle \langle H_{i+1,k} \rangle \\ &= -{M_k}^2 p_i(\beta_k) p_{i+1}(\beta_k). \end{aligned} \tag{3.113}$$

A version of Eq. (3.110) that takes these crosscorrelations into account is then:

$$\begin{aligned} \delta^2 \widehat{s}_{ik} \approx \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{ik}}\right]^2 \delta^2 H_{ik} + \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{i+1,k}}\right]^2 \delta^2 H_{i+1,k} + \\ + 2 \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{ik}}\right] \left[\frac{\partial \widehat{s}_{ik}}{\partial H_{i+1,k}}\right] \mathrm{Cov}\left[H_{ik}, H_{i+1,k}\right]. \end{aligned} \tag{3.114}$$

Using the estimate $\langle H_i \rangle \approx H_i$ this evaluates to

$$\begin{aligned} \delta^2 \widehat{s}_{ik} &\approx \frac{H_{ik} + H_{i+1,k}}{H_{i,k}\, H_{i+1,k}} + 2 \left[\frac{1}{H_{ik}}\right] \left[\frac{-1}{H_{i+1,k}}\right] \left[-H_{ik} H_{i+1,k}\right] \\ &= \frac{H_{ik} + H_{i+1,k}}{H_{ik}\, H_{i+1,k}} + 2, \end{aligned} \tag{3.115}$$

but the directly measured H_{ik} by itself is not the best estimate for its expectation value $\langle H_{ik} \rangle$ that can be made from the simulation data. No information from the simulations conducted at temperatures $\beta \neq \beta_k$ is taken into account, in contrast to the estimate from Eq. (3.90) used in the traditional approach. In the light of this it can be assumed that the estimator in Eq. (3.109) with the error weights from Eq. (3.111) is not necessarily the optimal choice.

Since the estimated $\widehat{s}_i$ is constructed from ratios of anticorrelated quantities based on energy bins i and $i+1$, it generally seems plausible that the statistical errors of the individual quantities are amplified in the combination. The method outlined in Sec. 3.6.1 is less susceptible to such effects because $\widehat{\Omega}_i$ is estimated primarily from histogram counts in the i'th bin only. The other energy bins only enter in the normalizing constant $\mathcal{Z}(\beta_k) = \mathrm{e}^{-f_k}$.

3.6.3 Application to parallel tempering simulations

As mentioned further above, the multiple histogram reweighting method as outlined in Sec. 3.6.1 is very useful in the context of parallel tempering simulations. However, it should be noted that the previously described method has been contrived for the analysis of *independent* simulations at different temperatures. In parallel tempering on the other hand, separate configurations of the studied system are simulated in a number of replicas that are no longer fully independent. Correlations between replicas are introduced by the swapping of temperatures during the replica exchange update. When a proposed exchange move is accepted, the configuration that is currently identified with a certain temperature is changed considerably because it has been exchanged with that of a different replica.

When the multiple histogram method is applied, these correlations are usually ignored. The conventional approach is to take the sampled configurations from all the replicas and to reorder them into pseudo-trajectories at constant temperatures. Chodera et al. argue [63] that by treating these temperature-grouped trajectories as if they had been sampled in independent canonical simulations, autocorrelation times are underestimated because the parallel tempering update is treated just like a regular single temperature Monte Carlo update. Accordingly this neglecting of crosscorrelations would lead to wrongly estimated statistical inefficiencies g_{ik} in Eqs. (3.93) and (3.94) and a no longer optimal estimator $\widehat{\Omega}_i$. To circumvent this problem they propose a scheme in which density of states estimators from individual *replicas* instead of temperatures are optimized. In this scheme the replica time series are treated as if they came from independent simulated tempering simulations. But the single replica trajectories are still not fully statistically independent, as they are coupled by the exchange of temperatures. The authors of Ref. [63] claim this to be a minor effect, which is dominated by the proper capture of the correlations between successively sampled configurations in each replica.

By still using the index k to label replicas, but introducing also the new index l to label inverse temperatures $\{ \beta_1, \dots, \beta_l, \dots, \beta_K \}$ because during the simulation each replica now visits different temperature values, a generalization of Eq. (3.73) leads to the following estimator for the probability to sample an energy E_i in replica k:

$$\widehat{p}_{ik} = \sum_{l=1}^{K} \frac{M_{kl}}{M_k} \cdot \frac{\widehat{\Omega}_{ik} \mathrm{e}^{-\beta_l E_i}}{\mathcal{Z}(\beta_l)} = \widehat{\Omega}_{ik} \sum_{l=1}^{K} \frac{M_{kl}}{M_k} \cdot \mathrm{e}^{-\beta_l E_i + f_l}, \tag{3.116}$$

where M_{kl}/M_k is the fraction of configurations sampled at β_l in replica k during the whole simulation and the dimensionless free energy parameters $f_l = -\ln \mathcal{Z}(\beta_l)$ remain functions solely of temperature. This probability no longer has a clear physical meaning as it is a joint quantity for all simulation temperatures. Nevertheless, it can be estimated by the histogram H_{ik} in which the occurrences of energy bin i during the simulation of replica k at all temperatures β_l is counted. If the energy bins have been chosen fine enough for discretization errors to be negligible, one can write the equivalent of Eq. (3.75):

$$\widehat{p}_{ik} \approx \frac{1}{\Delta E} \cdot \frac{H_{ik}}{M_k} \tag{3.117}$$

and by rearranging Eq. (3.116) an estimator for the density of states at energy bin i in replica k is obtained:

$$\widehat{\Omega}_{ik} = \frac{H_{ik}}{\sum_{l=1}^{K} M_{kl} \Delta E \mathrm{e}^{-\beta_l E_i + f_l}}. \tag{3.118}$$

Under the assumption that the single histogram bins are sparsely populated, which is reasonable if the replica visits many temperatures and hence many energy levels during the simulation, the arguments of Sec. 3.6.1.3 lead to the following estimate of the statistical uncertainty of the density of states:

$$\delta^2 \widehat{\Omega}_{ik} = \frac{g_{ik} \widehat{\Omega}_i}{\sum_{l=1}^{K} M_{kl} \Delta E \mathrm{e}^{-\beta_l E_i + f_l}}, \tag{3.119}$$

where g_{ik} is now the statistical inefficiency of energy bin i computed from the time series of replica k. Then the optimized estimator of the density of states given in Eq. (3.83), in which the estimates from all K replicas are combined, can be evaluated to

$$\widehat{\Omega}_i = \frac{\sum_{k=1}^{K} H_{ik} \cdot {g_{ik}}^{-1}}{\sum_{k=1}^{K} {g_{ik}}^{-1} \sum_{l=1}^{K} M_{kl} \Delta E \mathrm{e}^{-\beta_l E_i + f_l}}. \tag{3.120}$$

It should be noted that in this estimate the information of the relative probability of the energy bins at the various temperatures is no longer used directly. Histograms are only computed over the non-physical replica trajectories and then weighted globally by the proportion of simulation time spent at the individual temperatures.

In the same way as in Sec. 3.6.1.5, expectation values of observables $\mathcal{O}$ at arbitrary temperatures β can then be estimated as weighted averages over the *replica* time series. Eqs. (3.106) and (3.107) hold again:

$$\widehat{\mathcal{O}}(\beta) = \frac{\sum\limits_{k=1}^{K} \sum\limits_{m=1}^{M_k} \mathcal{O}_{km} w_{km}(\beta)}{\sum\limits_{k=1}^{K} \sum\limits_{m=1}^{M_k} w_{km}(\beta)}, \tag{3.121}$$

$$w_{km}(\beta) = \sum_{i=1}^{I} \frac{\psi_{ikm}\widehat{\Omega}_i \mathrm{e}^{-\beta E_i}}{H_i}, \tag{3.122}$$

where the index k now distinguishes only the replica. The information about the current temperature which an observable measurement $\mathcal{O}_{kn}$ has been made at is not used any more. The sum over i in Eq. (3.122) reduces to a single term. $H_i = \sum_{k=1}^{K} H_{ik}$ is the total number of times energy bin i has been sampled in all replicas.

In general it is not clear whether the more careful consideration of inter-replica cross-correlations in the way described in this section delivers more correct results than the more naive approach, where trajectories are pre-sorted by temperatures, but where also no physical temperature information is discarded.

3.6.4 Optimizing quantities and estimating statistical uncertainties

In Sec. 3.6.1.5 a method for time series reweighting based on the multiple histogram analysis has been presented. It allows the evaluation of arbitrary functions of observables at any inverse temperature β. It can easily be combined with Brent's algorithm for numerical minimization without derivatives [67]. The algorithm combines bracketing by the golden section with parabolic interpolation to find the minimum of a function within numeric machine precision. This way the number of points, where the function has to be evaluated, is kept low. This is beneficial because the reweighting procedure can be computationally expensive.

The maxima of a susceptibility or heat capacity can be found by applying it on $-\chi(\beta)$ and $-C(\beta)$. Likewise the algorithm can be used to find the inverse temperature β_{eqH} of phase coexistence by minimizing the peak height difference of bimodal histograms of an observable.

Statistical uncertainties of expectation values and these optimized quantities can be estimated by carrying out the whole multiple histogram analysis on jackknife blocked subsets of the data as described in Sec. 2.5.3.

3.6.5 Choices for the compass model

To see if the various methods for the estimation of the density of states that have been presented above deliver results of different quality, I have made extensive comparisons with the two-dimensional Ising model, for which exact results are readily available. The results are presented in appendix B.

Whether or not autocorrelation times were taken into account for statistical inefficiency factors made little difference there. Also the careful distinction of different autocorrelation times for different bins of the energy histograms did not lead to any improvements, rather this might actually open the door for more numerical errors. The little role different inefficiency factors play can be understood by realizing that they are only effective as relative weights between energy histograms with non-negligible overlap. But in realistic models only histograms from closely separated temperatures overlap and at these temperatures also autocorrelation times are very similar.

Furthermore, with parallel tempering the inclusion of autocorrelation times in the estimations lead to systematically wrong results. The method presented in Sec. 3.6.3, which has been devised to properly deal with correlations stemming from replica exchanges, actually delivered even stronger deviations. For this work I have therefore decided to always set the inefficiency factor to unity: $g_{ik} = 1$ and to sort the replica trajectories by temperature before commencing the multiple histogram analysis. But in general the question how to correctly handle correlated statistics generated in parallel tempering simulations remains open.

The non-iterative method of Sec. 3.6.2 has only been used as a quick way to get good starting values for the iterations, on its own it has shown to be rather susceptible to statistical errors. But compared to a naive choice of $\Omega_{i,0} \equiv 1$ several thousands of iteration steps could be saved in some cases with this choice of $\Omega_{i,0}$.

Generally the whole energy range sampled in a parallel tempering simulation has been discretized into 1000 bins. Expectation values have then been calculated with the scheme of Sec. 3.6.1.5, which avoids additional discretization errors.

3.7 Technical aspects

To obtain the numerical results that will be presented in the following two chapters I have implemented the algorithms described in this chapter in the form of several elaborate C++ programs. The g++-compiler from the GNU Compiler Collection version 4.4.5 was used. The programs are accompanied by many Python and shell scripts to assist with the management of the simulation runs, data analysis and preparation of graphics.

3.7.1 Implementation of the simulations

To run the Monte Carlo simulations of the different models studied in this work with the various possible combinations of update algorithms I have written an object-oriented framework that allows all parameters to be set at run-time in the form of configuration files. The time series of measurements are written to disk and stored for later analysis by separate programs.

Most of the parallel tempering simulations have been run on the "Grawp" Linux computer cluster at the ITP, which consists of 30 nodes, each equipped with two dual core AMD Opteron 2218 processors. The simulations were run in parallel with typically one replica assigned to one processor. In my implementation, temperature exchanges are coordinated by a master process, which after a set number of Monte Carlo sweeps gathers energy information from all replicas, determines the new distribution of temperatures and redistributes it among the replicas. For communication between the processes the distributed-memory parallelization library Open MPI [68] is used. Essentially only calls to the two routines `MPI_Gather` and `MPI_Scatter` are necessary. The computational overhead for communication is negligible compared with the time spent on canonical updates.

Random number generation

The correctness of the results obtained in Monte Carlo simulations depends crucially on the statistical quality of the mathematical recursion used to provide the pseudo-random numbers. Since generally a large portion of the execution time is spent on the generation of these numbers, also the speed of the random number generator is very important. For an introduction see for instance the reviews by Janke [69] or Katzgraber [70].

A fast generator of high-quality pseudo-random numbers, that is very well tested, is the Mersenne Twister [71], which employs a generalized feedback shift register. In its variation `mt19937` it has a very long period length of $2^{19937} - 1 \approx 10^{6001}$. I have used a highly optimized implementation, which is adapted to the processor architecture of the computer cluster and specialized in the generation of floating point random numbers: the double precision SIMD-oriented Fast Mersenne Twister (`dSFMT`, [72, 73]).

To avoid correlations between the different processes of the parallel tempering simulations each should use an independent stream of pseudo-random numbers. To achieve this each process uses a different initial seed for its generator. These seeds are generated randomly by the master process in the beginning and then distributed. Due to the very long period length of the Mersenne Twister it is very unlikely that the various streams overlap.

3.7.2 Implementation of the multiple histogram analysis

I have implemented most of the data analysis in one C++ program, which performs the multiple histogram analysis at its core. To reduce the necessary run time during the iterative estimation of the density of states and the time series reweighting procedure to compute estimates of expectation values, shared-memory parallelization with OpenMP [74] has been used to evaluate multiple sums concurrently.

Logarithmic computations

Special care must be taken for all calculations involving partition functions $\mathcal{Z}(\beta_k)$ or densities of states $\Omega(E_i)$. Since they grow exponentially with the total system energy, for large systems their range can quickly exceed that of floating point numbers representable on the computer. Standard double precision numbers have a range from about 10^{-308} to 10^{+308}. But for instance in the $N = 128^2$ Ising model, the order of magnitude of $\Omega(E)$ ranges from 10^0 to about 10^{4930}, which cannot be stored directly in a standard floating point number without overflows.

For this reason all calculations involving these quantities are not performed directly on them, but on their logarithms, which have a much smaller dynamic range. Sums like those in Eqs. (3.94), (3.95) or (3.106) are always evaluated in terms of $\ln \Omega(E_i)$ and $f_k = -\ln \mathcal{Z}(\beta_k)$.

To calculate the logarithm of the sum of two numbers x_1 and x_2, which are only known in terms of their logarithms $l_1 = \ln x_1$ and $l_2 = \ln x_2$, the following rule can be used:

$$\begin{aligned} \ln(x_1 + x_2) &= \ln \mathrm{e}^{l_1} + \mathrm{e}^{l_2}) = \ln(\mathrm{e}^{l_1}[1 + \mathrm{e}^{l_2 - l_1}]) \\ &= l_1 + \ln[1 + \mathrm{e}^{l_2 - l_1}], \end{aligned} \tag{3.123}$$

where the exponential can be safely evaluated without risk of overflow if $l_2 <= l_1$. Otherwise the sum has to be evaluated with l_2 and l_1 exchanged:

$$\ln(x_1 + x_2) = l_2 + \ln[1 + \mathrm{e}^{l_1 - l_2}]. \tag{3.124}$$

Usually $\alpha \equiv \mathrm{e}^{-|l_1 - l_2|}$ is very small. To evaluate $\ln[1 + \alpha]$ accurately in this case, the specialized standard C library function `log1p` can be used.

Iterative solution of equations

To estimate the density of states and the free energy parameters f_k, Eqs. (3.94) and (3.95) are solved by direct iteration. As a criterion of convergence the quantity

$$\Delta^2 = \sum_k \left[\frac{\mathcal{Z}_k^{(i)} - \mathcal{Z}_k^{(i-1)}}{Z_k^{(i)}} \right]^2 = \sum_k \left[1 - \mathrm{e}^{f_k^{(i)} - f_k^{(i-1)}} \right]^2 \tag{3.125}$$

is computed at each iteration (i). It measures the relative change of the estimated partition functions since the last iterative step. Once Δ^2 falls below 10^{-14}, the solution is deemed to have converged.

A good choice of starting values $f_k^{(0)}$ can save many iterations. To find them quickly the non-iterative method of Sec. 3.6.2 is very suitable. The convergence may be sped up further by procedures like the Newton-Raphson method or specialized algorithms like the one developed by Bereau and Swendsen [75], but for this work the direct iteration appeared to be fast enough in all cases.

Estimation of statistical inefficiencies

To calculate the bin statistical inefficiencies $g_{ik} = 1 + 2\tau_{\text{int},ik}$ used for the weights, the correlation function $C_{ik}(t)$ is directly summed up:

$$\widehat{\tau}_{ik,\text{int}} = \sum_{t=1} \widehat{C}_{ik}(t). \tag{3.126}$$

As suggested by Chodera et al. [63] the sum is stopped at the gap time t where $\widehat{C}_{ik}(t) < 0$ for the first time. The evaluation of the sums can be done in a single pass over the time series for all energy bins i.

4 Results for the two-dimensional compass model

In the following I report the results of a study of the directional ordering transition in the two-dimensional classical compass model using periodic and screw-periodic boundary conditions. As in the higher-precision studies by Wenzel et al. [9, 10] I find the finite-size scaling of the results obtained on regular periodic lattices to be problematic. In contrast, the results obtained with screw-periodic boundary conditions scale consistently with the expectations for a continuous phase transition without higher-order corrections. They comply with the results published in Ref. [10], where fluctuating boundary conditions were used. Since a large part of my computer code for simulation and data evaluation is shared with the three-dimensional case, this is also an indication for the trustworthiness of the results presented in the next chapter. All numeric results quoted here and in the following are given in units with $k_\mathrm{B} = 1$ and $|J| = 1$.

4.1 The directional ordering transition

The two-dimensional compass model has been simulated on square lattices with linear sizes $L \in \{16, 20, 24, 36, 48, 60, 72, 84, 96, 112, 128, 160, 208, 256\}$ in each case both with regular and screw-periodic boundary conditions. Parallel tempering has always been used with a number of temperature points ranging from 16 to 24. They have been chosen linearly spaced for $L \leq 36$ and according to the constant entropy difference scheme for $L \geq 48$, where specific heat capacity data from smaller lattices has been used. Replica exchanges have been proposed every 20 sweeps. The simulations at the various lattice sizes have each been run for a number of sweeps ranging from 10^6 to $5 \cdot 10^6$. Here the combination of $2L$ cluster updates and $N = L^2$ single-spin Metropolis update steps is counted as one sweep.

The behavior of the order parameter

$$D = \frac{1}{N} \left| J \sum_{i=1}^{N} \left(s_i^x s_{i+\hat{x}}^x - s_i^y s_{i+\hat{y}}^y \right) \right|, \tag{4.1}$$

whose estimates are shown in Fig. 4.2 on p. 83, and of its susceptibility

$$\chi = N(\langle D^2 \rangle - \langle D \rangle^2), \tag{4.2}$$

shown in Fig. 4.3 on p. 84, clearly indicate the existence of a temperature-driven phase transition with directional ordering $D > 0$ realized in the low temperature phase (higher β). It is immediately obvious that finite-size effects are stronger with regular periodic boundary conditions: With screw-periodic boundary conditions the locations of the maxima of the susceptibility $\beta_{\max}(L)$ converge much faster to their infinite system size limit. Furthermore, the curves of the Binder parameter related to D

$$Q_2 = 1 - \frac{1}{3} \frac{\langle D^4 \rangle}{\langle D^2 \rangle^2} \tag{4.3}$$

in Fig. 4.4 on p. 85 cross very close to each other with screw-periodic boundary conditions, while with regular periodic boundary conditions there is a very strong size-dependence, which does not agree with the conventional expectations for the scaling at a continuous phase transition.

To find the maxima of the susceptibility, I have applied Brent's algorithm on the results of the multiple histogram analysis of the time series data. The resulting transition point locations $\beta_{\max}(L)$ and susceptibility values $\chi_{\max}(L)$ are listed in Table 4.1. The uncertainties given in parentheses have been estimated by carrying out the whole procedure on 20 jackknife-resampled blocks. With the available data the locations $\beta_{\max}(L)$ of the increasingly sharper peaks could be determined quite precisely, whereas the relative statistical uncertainty of the values $\chi_{\max}(L)$ is somewhat higher.

From finite-size scaling theory the pseudo-transition temperatures $T_{\max}(L) = {}^{1}/_{k_\mathrm{B}\beta_{\max}(L)}$ are expected to fulfill the relation

$$T_{\max}(L) = T_\mathrm{c} \left(1 - aL^{-1/\nu} + \cdots \right) \tag{4.4}$$

with the critical temperature T_c and the critical exponent of the correlation length ν. Further corrections to the dominating power law behavior are hidden in the ellipsis in Eq. (4.4).

To extrapolate T_c and ν I have performed least squares fits of the data from Table 4.1 to the relation (4.4). While it is preferable to include many data points into the fit, additional scaling corrections, that have not been considered here, are expected to become important especially with the smallest lattices. For this reason various different ranges have been tried for the fit, where the lattice sizes below a varying cut-off $L_{\min}$ were excluded. The results are collected in Table 4.2.

Table 4.1: Locations and values of the maxima of the susceptibility χ in the two-dimensional compass model with regular periodic (PBC) and screw-periodic (SBC) boundary conditions

L	$\beta_{\max}(L)$		$\chi_{\max}(L)$	
	PBC	SBC	PBC	SBC
16	4.892(4)	6.194(5)	6.73(1)	8.32(3)
20	5.003(3)	6.309(5)	9.51(3)	11.82(5)
24	5.129(3)	6.387(5)	12.68(5)	15.94(8)
36	5.495(3)	6.543(6)	25.8(2)	31.8(3)
48	5.773(3)	6.612(4)	45.4(3)	52.7(5)
60	5.982(1)	6.655(3)	71.8(4)	77.0(4)
72	6.141(1)	6.684(2)	106.9(8)	104.5(7)
84	6.264(2)	6.705(3)	146(1)	136.3(9)
96	6.361(1)	6.726(3)	199(2)	172(2)
112	6.462(1)	6.739(3)	265(2)	225(2)
128	6.541(1)	6.751(2)	338(2)	281(2)
160	6.649(1)	6.768(3)	495(7)	421(6)
208	6.741(1)	6.784(3)	740(15)	651(9)
256	6.784(2)	6.799(2)	1034(19)	950(18)

Observing the quantity χ^2_{dof}, it is apparent that with periodic boundary conditions the conformity of the data with the scaling ansatz (4.4) is better for high cut-offs $L_{\min}$. Still, with this simple approach the discrepancy is high in all cases and it is clear that further correction terms or simulations on larger lattices would be necessary to obtain reliable results.

With screw-periodic boundary conditions on the other hand, the fit which includes all data points available starting from $L_{\min} = 16$ already has a χ^2_{dof}-value of order unity and the estimated uncertainties of the parameters T_c and ν are the smallest of all the ranges. The final results taken from this fit

$$\begin{aligned} T_c &= 0.14617 \pm 0.00006, \\ \nu &= 1.00 \pm 0.01 \end{aligned} \tag{4.5}$$

agree with those obtained using fluctuating boundary conditions in Ref.[10]. Plotted over $1/L$, $T_{\max}$ scales linearly in Fig. 4.1(a) on p. 82 with screw-periodic boundary conditions, whereas the asymptotic regime of Eq. (4.4) is not reached on the studied lattice sizes by the periodic boundary conditions data, which show multiple bends. From the figure it is conceivable that the curve for periodic boundary conditions may meet the one for screw-periodic boundary conditions for $L > 256$.

Standard finite-size scaling theory predicts a power law behavior of the maximum

Table 4.2: Results of non-linear least squares fits to estimate T_c and ν for the two-dimensional compass model from the data in Table 4.1 according to Eq. (4.4) with regular periodic (PBC) and screw-periodic (SBC) boundary conditions. Here n is the number of included data points ranging from the smallest considered lattice size L_{min} up to the largest $L_{\text{max}} = 256$. $\chi^2_{\text{dof}} = \chi^2/n-3$ is a measure to help with the estimation of the validity of the fit. The table only includes parameter sets where the fitting procedure converged. The results taken finally are marked bold.

		PBC			SBC		
n	L_{min}	T_c	ν	χ^2_{dof}	T_c	ν	χ^2_{dof}
14	**16**				**0.14617(6)**	**1.00(1)**	**0.75**
13	20				0.14619(7)	0.99(2)	0.75
12	24	0.136(2)	1.28(9)	281.35	0.14620(8)	0.99(2)	0.83
11	36	0.140(1)	1.06(6)	109.88	0.14600(9)	1.08(4)	0.43
10	48	0.1414(7)	0.95(4)	47.03	0.1460(1)	1.08(6)	0.49
9	60	0.1420(6)	0.90(4)	31.81	0.1460(2)	1.08(8)	0.57
8	72	0.1427(5)	0.84(4)	20.16	0.1460(2)	1.0(1)	0.68
7	84	0.1434(5)	0.77(4)	12.10	0.1459(4)	1.1(2)	0.82
6	96	0.1438(5)	0.73(5)	10.69	0.1444(8)	2.0(5)	0.14
5	112	0.1445(4)	0.64(5)	5.14	0.144(1)	2(1)	0.20
4	128	0.1449(5)	0.58(6)	5.27			

values of the susceptibility:

$$\chi_{\text{max}}(L) = bL^{\gamma/\nu} + \cdots , \tag{4.6}$$

with the critical exponents γ of the susceptibility and ν of the correlation length. If the power law in Eq. (4.6) is the dominating behavior, $\chi_{\text{max}}(L)$ should scale monotonically with L. In Fig. 4.1(b) on p. 82 one can see that this is not the case with regular periodic boundary conditions, except for maybe the largest lattice sizes. In the plot χ_{max} has been scaled by a factor of $1/L^2$ for clarity, which changes the exponent of the expected power law, but does not influence the condition of monotonicity.

By taking the logarithm on both sides of Eq. (4.6) with the ellipsis dropped, a linear relationship is obtained:

$$\ln \chi_{\text{max}}(L) = \tilde{b} + \frac{\gamma}{\nu} \ln L. \tag{4.7}$$

Then γ/ν can be estimated by linear fits of the data from Table 4.1 to Eq. (4.7). The results of such fits are collected in Table 4.3 on p. 82 on p. 82. As expected, the results at hand with regular periodic boundaries do not match Eq. (4.7) well, which is reflected in high values of χ^2_{dof}.

The screw-periodic data is in better agreement with the scaling relation: χ^2_{dof} is of order unity for $L_{\text{min}} \geq 20$. Taking the data range $20 \leq L \leq 256$, the ratio of exponents results in

$$\gamma/\nu = 1.712 \pm 0.003. \tag{4.8}$$

From this estimate the critical exponent γ can be computed by multiplication with $\nu = 1.00(1)$. The uncertainty of γ quoted in the following is a conservative estimate, calculated as the maximum uncertainty arising from the individual uncertainties ϵ_ν from Eq. (4.5) and $\epsilon_x \equiv \epsilon_{\gamma/\nu}$ from Eq. (4.8):

$$\epsilon_\gamma = \max\left\{\, (\nu+\epsilon_\nu)(x+\epsilon_x) - x\cdot\nu,\; x\cdot\nu - (\nu-\epsilon_\nu)(x-\epsilon_x)\,\right\} \tag{4.9}$$

This results in the estimate

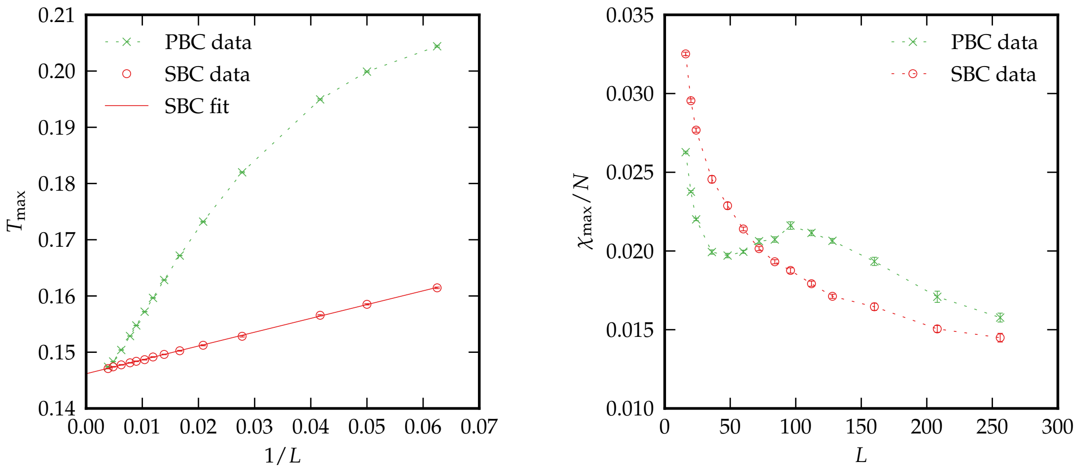
$$\gamma = 1.71 \pm 0.02. \tag{4.10}$$

The result obtained with fluctuating boundary conditions and higher computational expenses in Ref. [10] is $\gamma = 1.75(1)$. The results found here with screw-periodic boundary conditions agree with this value within error bars if a higher cut-off L_{min} of at least 72 is chosen. Whether this slight discrepancy can be explained by somewhat less favorable scaling properties of the screw-periodic boundary conditions or by insufficient accuracy in the determination of the $\chi_{\text{max}}(L)$ in this work, could only be cleared up by a more extended examination.

Nevertheless, the findings with screw-periodic boundary conditions agree well with the claim that the directional-ordering transition in the two-dimensional compass lies in the universality class of the Ising model in two dimensions, where $\nu = 1$ and $\gamma = 1.75$ are known exactly.

Table 4.3: Results of linear least squares fits to estimate γ/ν for the two-dimensional compass model from the data in Table 4.1 according to Eq. (4.7) with regular periodic (PBC) and screw-periodic (SBC) boundary conditions. From the SBC results for γ/ν, γ is estimated with $\nu = 1.00(1)$ (from Table 4.2). Here n is the number of included data points ranging from the smallest considered lattice size L_{min} up to the largest $L_{\mathrm{max}} = 256$. $\chi^2_{\mathrm{dof}} = \chi^2/n{-}2$ is a measure to help with the estimation of the validity of the fit. Results referred to in the text are marked bold.

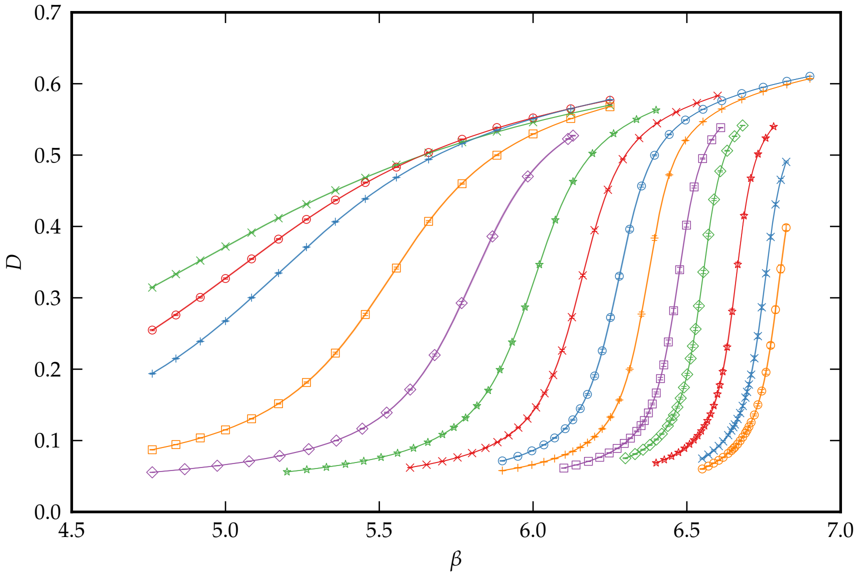
		PBC		SBC		
n	L_{min}	γ/ν	χ^2_{dof}	γ/ν	χ^2_{dof}	γ
14	16	1.86(2)	161.10	1.70(1)	6.26	1.70(2)
13	**20**	1.91(2)	77.07	**1.712(3)**	**1.14**	**1.71(2)**
12	24	1.95(2)	49.03	1.718(3)	0.61	1.72(2)
11	36	1.99(3)	42.26	1.721(5)	0.64	1.72(2)
10	48	1.98(4)	45.98	1.72(1)	0.72	1.72(2)
9	60	2.0(1)	46.59	1.72(1)	0.70	1.72(2)
8	**72**	1.9(1)	30.22	**1.73(1)**	**0.50**	**1.73(2)**
7	84	1.8(1)	21.72	1.73(1)	0.60	1.73(3)
6	96	1.69(4)	4.72	1.73(1)	0.74	1.73(3)
5	112	1.65(3)	2.72	1.74(2)	0.74	1.74(3)
4	128	1.62(2)	1.00	1.75(2)	0.79	1.75(4)
3	160	1.56(2)	0.16	1.72(5)	0.93	1.7(1)



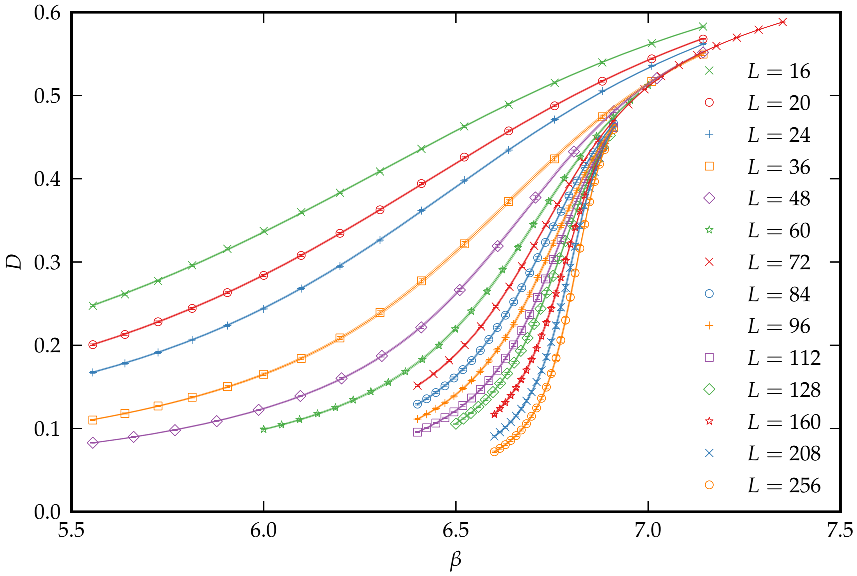
(a)

(b)

Figure 4.1: Finite-size scaling in the two-dimensional compass model with regular periodic (PBC) and screw-periodic (SBC) boundary conditions. (a) Pseudo-transition points $T_{\mathrm{max}}(L)$ and extrapolation of the critical temperature T_c from the SBC data. (b) The scaled susceptibility-maxima $\chi_{\mathrm{max}}(L)/L^2$ show a non-monotonic behavior with PBC, whereas no anomaly is apparent with SBC.

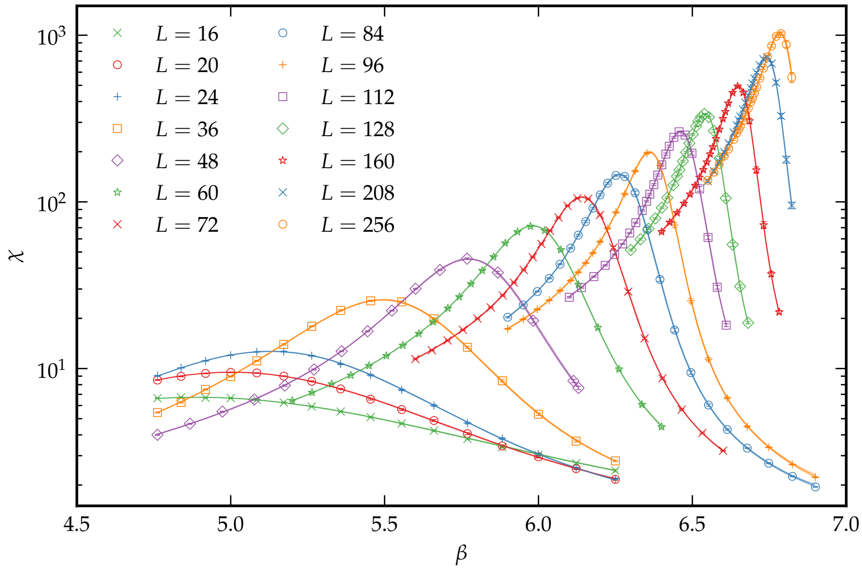


(a) PBC

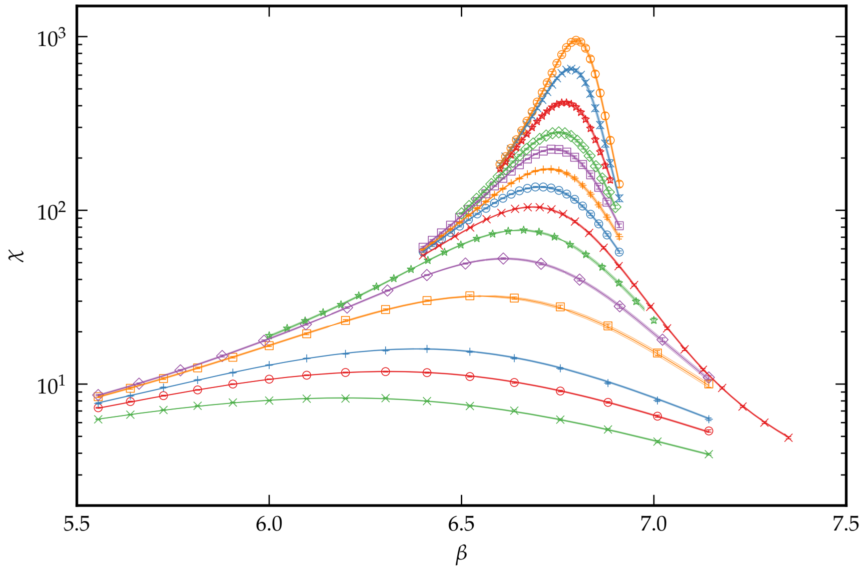


(b) SBC

Figure 4.2: Estimates of the order parameter D in the two-dimensional compass model for various lattice sizes L and inverse temperatures β with (a) regular periodic and (b) screw-periodic boundaries, $S = 1$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty. Colors and symbols correspond to the same lattice sizes in both plots.

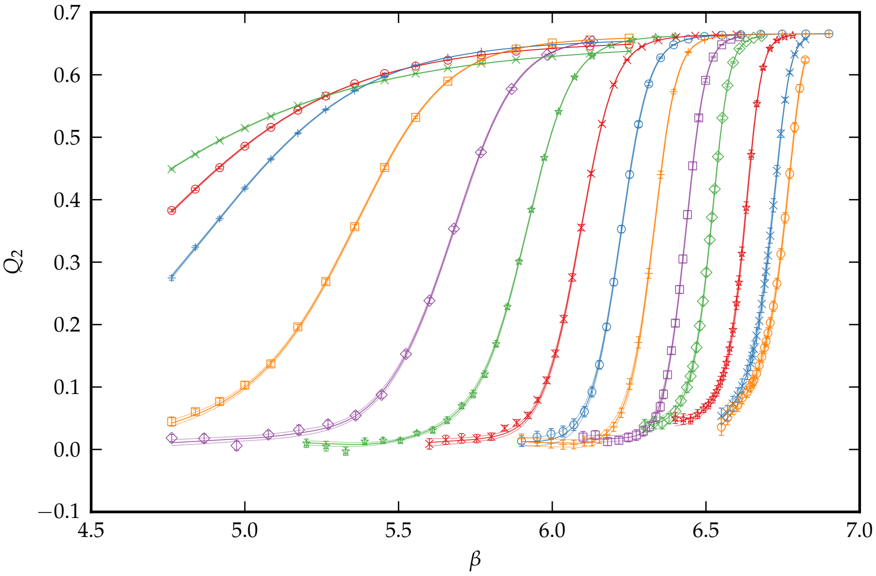


(a) PBC

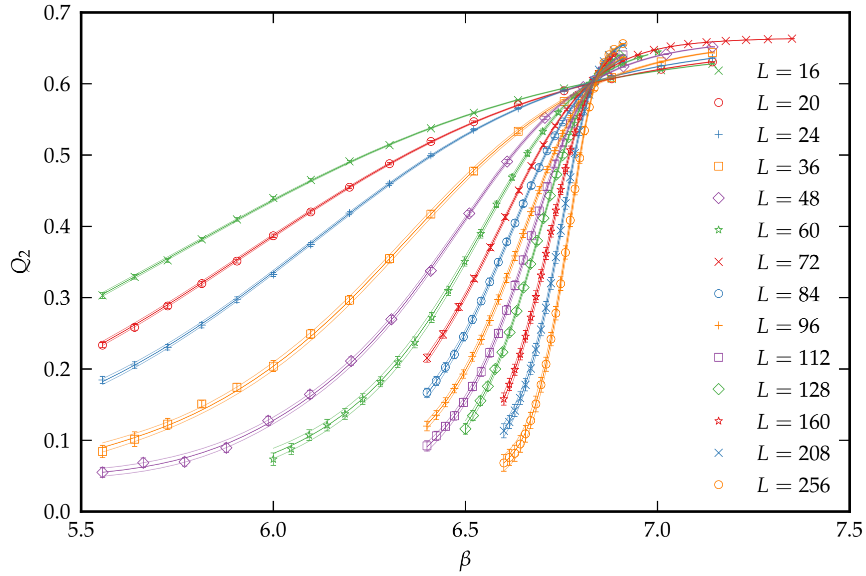


(b) SBC

Figure 4.3: Estimates of the order parameter susceptibility χ in the two-dimensional compass model for various lattice sizes L and inverse temperatures β with (a) regular periodic and (b) screw-periodic boundaries, $S = 1$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty. Colors and symbols correspond to the same lattice sizes in both plots.



(a) PBC



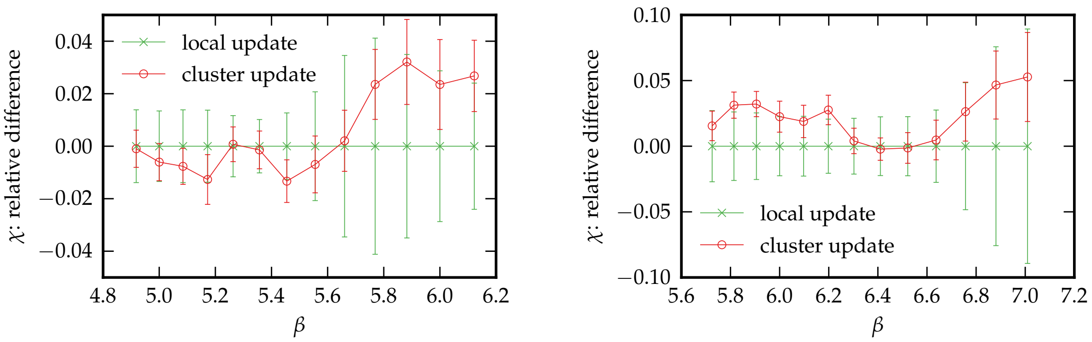
(b) SBC

Figure 4.4: Estimates of the Binder parameter Q_2 in the two-dimensional compass model for various lattice sizes L and inverse temperatures β with (a) regular periodic and (b) screw-periodic boundaries, $S = 1$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty. Colors and symbols correspond to the same lattice sizes in both plots.

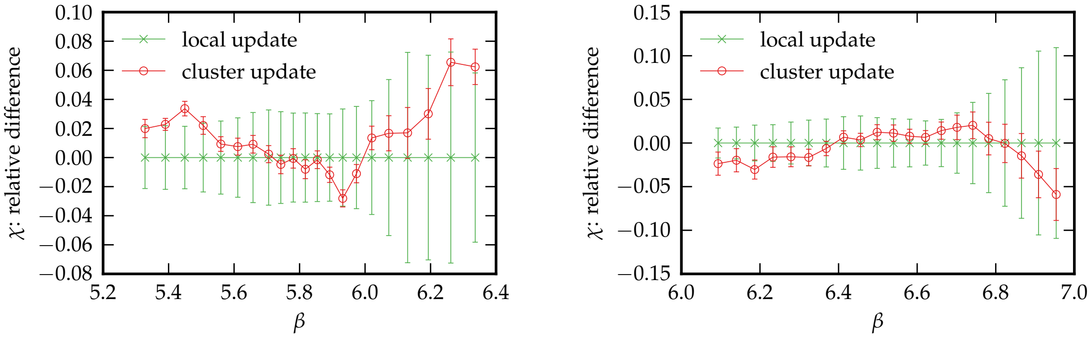
4.2 Local and cluster updates

To check that the cluster update presented in Sec. 3.4.2 produces correct results, I have also run parallel tempering simulations with purely local Metropolis updates of the two-dimensional model with both types of boundary conditions considered here on lattices of sizes $L = 36$ and $L = 60$. The resulting estimates of the susceptibility $\chi(\beta)$ agree well as one can see in Fig. 4.5.

In each case the same number of samples has been used with both update algorithms, however less than for the results in Sec. 4.1 as with the purely local update more initial measurements had to be discarded for thermalization. While the results obtained with the combined cluster and local update presented here are more precise, more spins are modified in one Monte Carlo sweep with that algorithm than with the purely local variant. This effect would need to be compensated for a quantitative comparison of the efficiency of the algorithms.



(a) PBC. $L = 36$ **(b)** SBC. $L = 36$



(c) PBC, $L = 60$ **(d)** SBC, $L = 60$

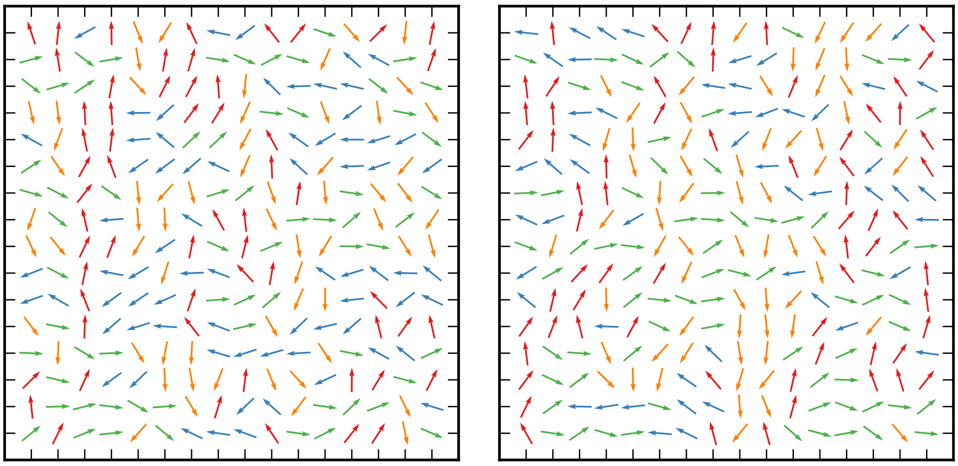
Figure 4.5: Comparison of results for the susceptibility χ from simulations with purely local and combined cluster and local updates in the two-dimensional compass model with regular periodic (PBC) and screw-periodic (SBC) boundary conditions on two different lattice sizes. The data points correspond to the relative differences of the measured values to those measured with purely local updates. (Therefore they all lie on the zero line for local updates). The error bars show the relative statistical uncertainties of the measurements within 1σ, estimated with the jackknife method.

4.3 Typical spin configurations

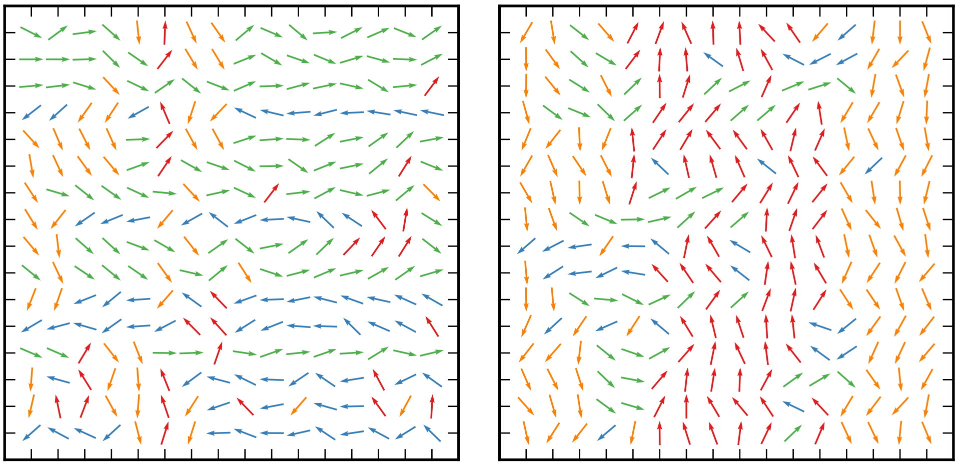
In Fig. 4.6 an illustration of typical spin configurations of the two-dimensional $L = 16$ compass model with regular periodic (PBC) and screw-periodic (SBC) boundary conditions with $S = 1$ is given. The configurations shown there are snapshots from canonical simulations in the disordered phase at $\beta_{\text{low}} = 1.0$, the ordered phase at $\beta_{\text{high}} = 10.0$ and at the pseudo-transition point $\beta_{\max}(L)$ taken from Table 4.1. In preliminary simulations the expectation value $D(\beta)$ was estimated and only spin configurations conforming to that value of D were considered to find the "typical" configurations.

In the low temperature snapshots at $\beta = 10.0$ directional ordering is apparent. The majority of the spins is predominantly aligned with one of the lattice axes. In the case of regular periodic boundary conditions spins on separate columns of the lattice are aligned independently of each other. For this reason one part is pointing mostly upwards and the other part is pointing mostly downwards. With screw-periodic boundary conditions on the other hand, all spins are linked in one chain along the lattice. Hence the aligned spins do not point in opposing directions.

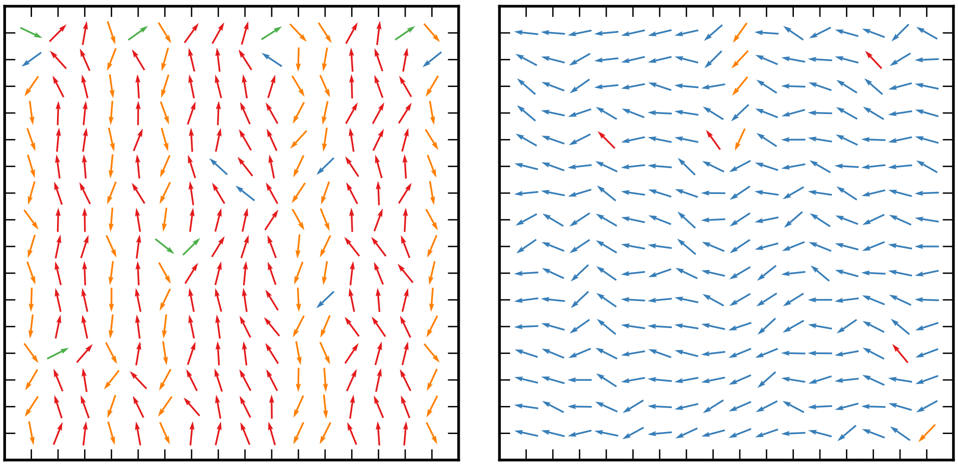
In the intermediate temperature snapshots at $\beta_{\max}(L)$ linear alignment of spins is already visible to a higher degree than in the high temperature snapshots at β_{low}. However, no global order is realized yet. At the same time some groups are aligned more in x-direction and others more in y-direction.



(a) PBC, $\beta_{\text{low}} = 1.0$, $D = 0.0433$ **(b)** SBC, $\beta_{\text{low}} = 1.0$, $D = 0.0433$



(c) PBC, $\beta_{\text{max}} = 4.892$, $D = 0.346$ **(d)** SBC, $\beta_{\text{max}} = 6.194$, $D = 0.382$



(e) PBC, $\beta_{\text{high}} = 10.0$, $D = 0.718$ **(f)** SBC, $\beta_{\text{high}} = 10.0$, $D = 0.723$

Figure 4.6: Typical spin configurations in the two-dimensional compass model. Top row: Disordered configurations. Middle row: Transition states. Bottom row: Directionally ordered configurations.

5 Results for the three-dimensional compass model

In this chapter I present the simulation results for the three-dimensional classical compass model. Parallel tempering has been used for all the principal simulations. The Metropolis update has been employed together with the linear cluster update. Here one Monte Carlo sweep is counted as the combination of $3L$ cluster updates with $N = L^3$ local Metropolis steps. For the first section of this chapter regular periodic boundary conditions have been applied. The introduction of screw-periodic boundary conditions in the second section has lead to substantial improvements.

As an order parameter for directional ordering of the bond energies the quantity

$$D = \frac{1}{N}\sqrt{(E_z - E_y)^2 + (E_y - E_x)^2 + (E_z - E_x)^2}, \tag{5.1}$$

has been measured. All numeric results quoted here and in the following are given in units with $k_\mathrm{B} = 1$ and $|J| = 1$.

5.1 Results with regular periodic boundary conditions

I have studied the three-dimensional model with regular periodic boundary conditions on lattices of linear sizes from $L = 8$ to $L = 36$. The parameters of the parallel tempering simulations are summarized in Table 5.1 on p. 90. For $L \leq 20$ the inverse temperature points $\{\, \beta_k \,\}_{k=1}^{K}$ used in the measurements have been determined by first running short preliminary simulations on lattices of the same size with arbitrarily chosen temperatures and then determining a new set with constant energy distribution overlap from that data. For $L \geq 24$ the β_k have been chosen with the constant entropy difference scheme, where heat capacity data from smaller lattices has been used. This has not been an ideal choice in this case as the heat capacity peak shifts considerably between lattice sizes. However, for the purposes of this section the quality of the results is sufficient.

In Fig. 5.1(a) on p. 91 the temperature dependence of D as estimated from the simulations is shown. A thermal phase transition from a disordered high-temperature (low values of β) phase with $D = 0$ to an ordered phase with $D > 0$ at lower temperatures

Table 5.1: Parameters used in parallel tempering simulations of the three-dimensional compass model with regular periodic boundary conditions. Listed are the linear lattice sizes L, the corresponding lattice volumes $N = L^3$, the lowest and highest included inverse temperatures β, the total number of replicas and temperature points K, the number of Monte Carlo sweeps used for thermalization and measurements and finally the interval in sweeps between proposed replica exchanges.

L	N	β-range	K	Sweeps for Thermalization	Sweeps for Measurement	Sweeps between Exchanges
8	512	4 — 20	32	$3 \cdot 10^6$	$3 \cdot 10^7$	100
12	1728	4 — 20	32	10^6	$7 \cdot 10^6$	10
16	4096	4 — 20	64	10^6	10^7	100
20	8000	8 — 12.5	64	$1.5 \cdot 10^6$	$1.5 \cdot 10^7$	150
24	13824	8 — 10.5	48	$2 \cdot 10^6$	$2 \cdot 10^7$	100
28	21952	8 — 10.5	48	$2.5 \cdot 10^6$	$2.5 \cdot 10^7$	100
32	32768	8.3 — 9.5	48	$2.5 \cdot 10^6$	$2.5 \cdot 10^7$	100
36	46656	8.7 — 9.3	48	$3 \cdot 10^6$	$3 \cdot 10^7$	100

can clearly be made out. But there are two peculiarities to take note of:

1. The temperature location of the transition shifts considerably with increasing lattice sizes. This can also be seen in the positions of the maxima of the susceptibility

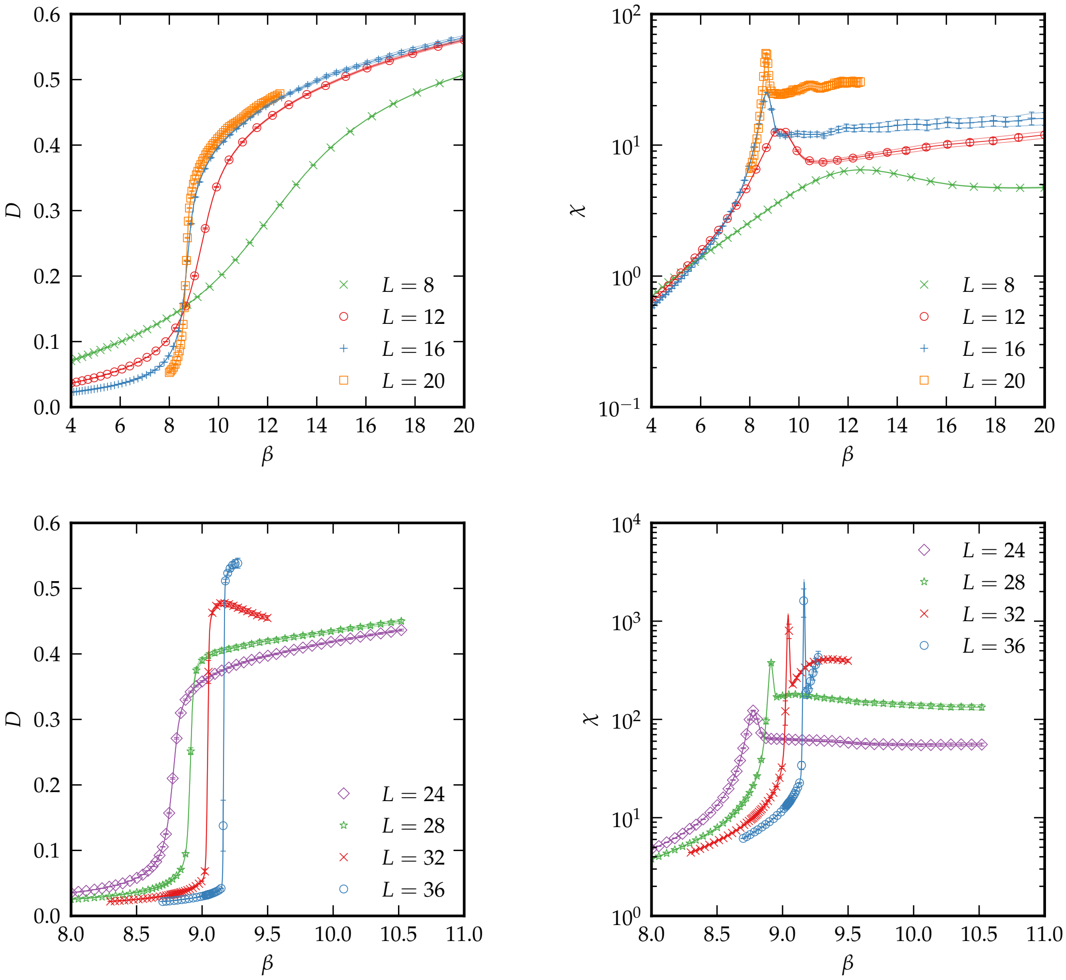
$$\chi = N(\langle D^2 \rangle - \langle D \rangle^2). \tag{5.2}$$

   The estimated $\chi(\beta)$ curves are shown in Fig. 5.1(b) on p. 91 The locations $\beta_{\max}(L)$ and values $\chi_{\max}(L)$ of the maxima are listed in Table 5.2 on p. 92. For lattice sizes $L < 20$ the values of $\beta_{\max}(L)$ seem to converge from higher values down to $\beta \approx 8.7$, but then start to rise again on the larger lattices.

2. The estimated values of χ approach zero for $\beta \to 0$ in the high-temperature phase, but they remain finite in the low-temperature phase.

The first point is reminiscent of the strong finite-size effects encountered in simulations of the two-dimensional model with periodic boundary conditions.

The second point can be understood by realizing that χ effectively measures the width of the probability distribution $P(D)$ of the order parameter. $P(D)$ can be evaluated at the low temperatures, where χ shows the unexpected behavior by examining the histograms of D measured in the simulations. An example for $L = 16$ and $\beta = 20$ is shown in Fig. 5.2(a) on p. 93. The distribution of D is characterized by a structure with multiple peaks. The configurations in the different groups of D-values all lie in the same range of energies, which is illustrated by the example in Fig. 5.2(b) on p. 93. Even if the width of these individual D-peaks shrinks with increasing β, the value of the susceptibility χ estimated from the simulation data remains on mostly the same level as long as this



(a) Order parameter D

(b) Susceptibility χ

Figure 5.1: $D(\beta)$ and $\chi(\beta)$ estimated from measurements obtained in simulations of the three-dimensional compass model with regular periodic boundary conditions for different lattice sizes. Top row: Linear sizes $L = 8, \ldots, 20$. Bottom row: Linear sizes $L = 24, \ldots, 36$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.

degeneracy is present. In the high-temperature phase, where χ goes to zero with β, the distribution has only one peak of steadily shrinking width.

The multiple peak structures could only be reproduced in parallel tempering simulations, where the replica exchange process frequently feeds new decorrelated configurations into the low temperature phase. In a canonical simulation after thermalization configurations corresponding to only one of these D-peaks are realized at least within a reasonable number of Monte Carlo sweeps. Simulations with different random initial configurations lead to different peaks of D.

Table 5.2: Locations and values of the maxima of the susceptibility χ in the three-dimensional compass model with regular periodic boundary conditions found by multiple histogram reweighting

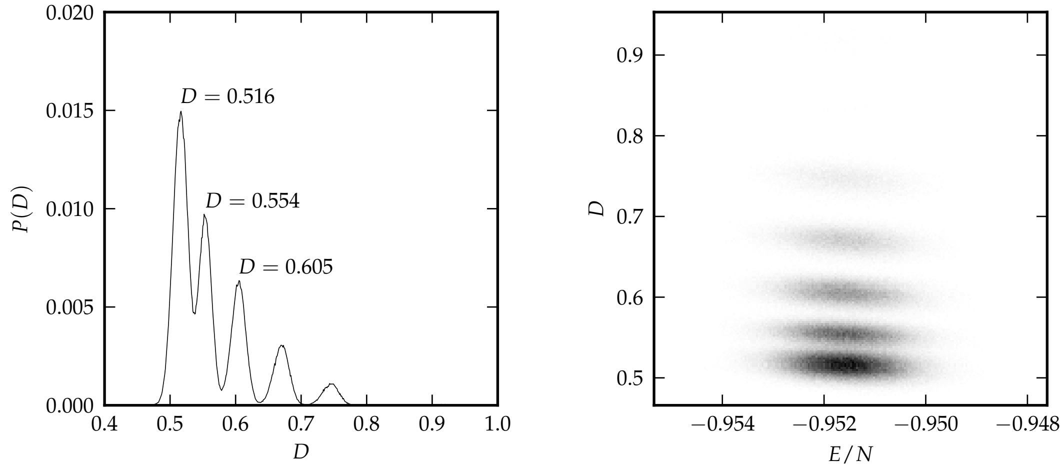
L	$\beta_{\max}(L)$	$\chi_{\max}(L)$
8	12.55(2)	6.48(2)
12	9.24(1)	13.2(1)
16	8.696(3)	25.2(2)
20	8.671(1)	51.0(3)
24	8.778(1)	123(1)
28	8.911(1)	380(8)
32	9.042(1)	1159(16)
36	9.165(1)	2509(19)

Some insight on the qualities that distinguish these configurations from each other can be gained by directly looking at typical snapshots taken in these Monte Carlo simulations. In the illustration of a part of the spins as three-dimensional arrows in Fig. 5.3(a) on p. 94 linear alignment in different directions can be recognized. However, it is very difficult to get a clear picture of the whole system in this way. In this example 4096 arrows would have to be shown. A better visualization of the same snapshot is presented in Fig. 5.3(b) on p. 94.

There the averaged z-component of all spins at sites with the same x- and y-, but varying z-coordinates is shown color-coded on the xy-plane. In this way one can see the degree to which the spins are aligned with the z-direction orthogonal to the plane. Correspondingly, projections onto the y-direction are shown on the xz-plane and projections onto the x-direction on the yz-plane. Evidently, in this low-temperature snapshot the spins are predominantly aligned in linear chains. These chains lie in parallel planes. In these planes the chains are oriented along one of the coplanar lattice axes, some in positive and some in negative direction. In different planes chains parallel to either possible axis are formed.

A comparison with the snapshots in Fig. 5.4 on p. 94, where the order parameter D takes on values corresponding to two of the other peaks in Fig. 5.2(a), then clarifies why multiple phases characterized by D seem to coexist. While in each case all spins are aligned linearly in parallel planes, they can have one of two orthogonal directions in the different planes. In the case of a high value of D one direction is taken more often than the other one. With lower values of D the ratio is more balanced.

As far as the linear directional ordering of the spins is concerned, the different possible values of D in the low temperature phase should be considered equivalent. One could define an alternative order parameter that is immune against this effect in the following



(a) Histogram of the order parameter D

(b) 2D histogram of D and the energy E

Figure 5.2: Distributions measured in the three-dimensional compass model with periodic boundary conditions with $L = 16$ in the low temperature phase at $\beta = 20$. In (b) dark points correspond to high densities.

way:

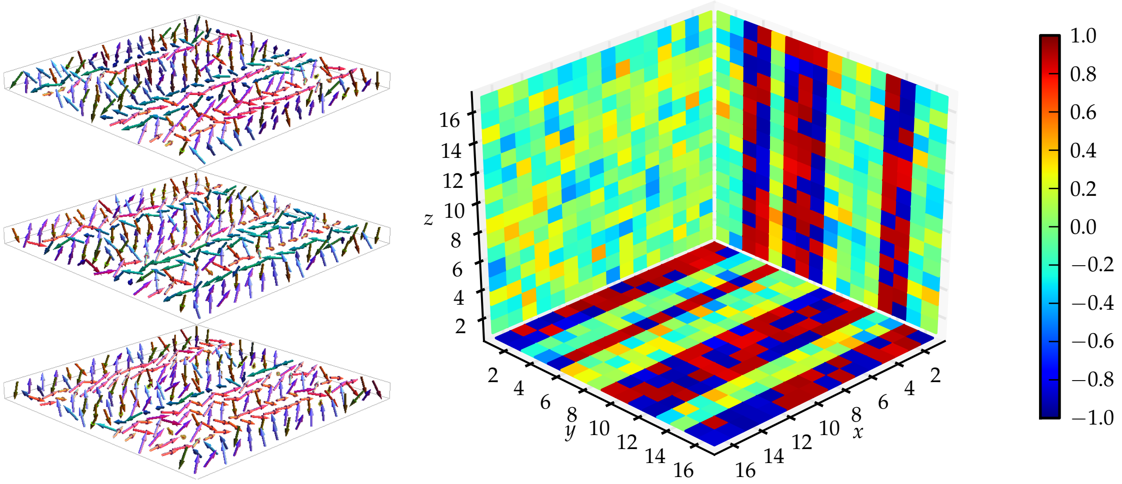
$$D' = \frac{1}{N}\sqrt{(D_z - D_y)^2 + (D_y - D_x)^2 + (D_z - D_x)^2} \tag{5.3}$$

with

$$\begin{aligned} D_z &= \sum_{j=1}^{L} \max\left\{\left|E_y^{(z=j)}\right|, \left|E_x^{(z=j)}\right|\right\}, \\ D_y &= \sum_{j=1}^{L} \max\left\{\left|E_x^{(y=j)}\right|, \left|E_z^{(y=j)}\right|\right\}, \\ D_x &= \sum_{j=1}^{L} \max\left\{\left|E_y^{(x=j)}\right|, \left|E_z^{(x=j)}\right|\right\}, \end{aligned} \tag{5.4}$$

where $E_y^{(z=j)}$ is the total energy of bonds in y-direction restricted to the lattice plane defined by $z = j$ and so on. However, as it turns out, this is not necessary with screw-periodic boundary conditions, where the problems of the order parameter are effectively avoided and where finite size effects are also weaker. Details are presented in the next section.

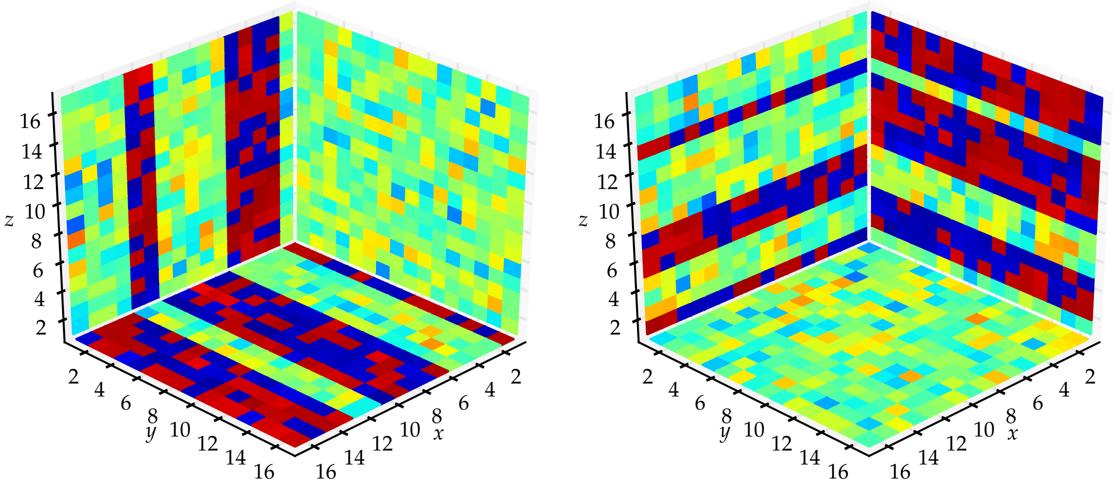
In the larger systems studied here there are also signs of phase coexistence at temperatures close to the transition. For instance the histogram of D in Fig. 5.5 on p. 95, which has been measured in the $L = 28$ system at the inverse temperature $\beta_{\max}$, where the susceptibility χ is maximized, apparently features a bimodal distribution, which could be seen as the signature of a first order transition. But again there are also contributions at larger values of D, which complicate the analysis.



(a) 3 layers of spins taken from the complete system of 16 layers

(b) Averaged spin projections of the complete system

Figure 5.3: Example spin configuration with $D = 0.516$ of the three-dimensional system with $L = 16$ and regular periodic boundary conditions at $\beta = 20$. In 7 planes there is alignment parallel to the x-axis, in 9 planes parallel to the z-axis.



(a) $D = 0.554$

(b) $D = 0.605$

Figure 5.4: Averaged spin projections of two example configurations of the three-dimensional system with $L = 16$ and regular periodic boundary conditions at $\beta = 20$. The same color coding as in Fig. 5.3(b) has been used. (a) 6 planes aligned parallel to the y-axis, 10 planes parallel to the z-axis. (b) 5 planes aligned parallel to the y-axis, 11 planes parallel to the z-axis.

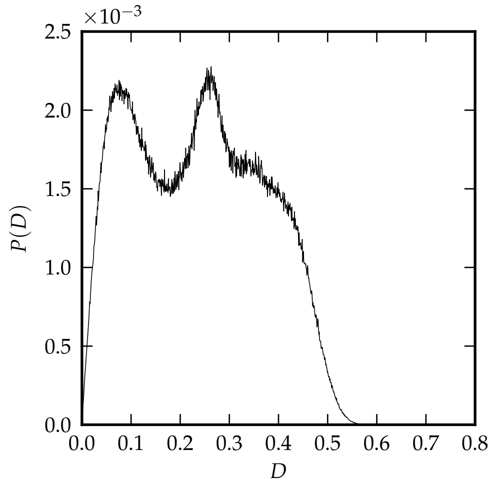


Figure 5.5: Histogram of the order parameter D at $\beta_{\max}(L = 28)$ in the three-dimensional compass model with $L = 28$ and regular periodic boundary conditions

5.2 Results with screw-periodic boundary conditions

In this section I present the results of Monte Carlo simulations of the three-dimensional compass model with screw-periodic boundary conditions with screw parameter $S = 1$ on lattices of linear sizes from $L = 8$ to $L = 48$. In all parallel tempering simulations replica exchanges were proposed after 100 sweeps of canonical updates. Detailed information on the parameters of the simulations is given in Table 5.3 on p. 96. For $L \leq 28$ the inverse temperatures $\{ \beta_k \}_{k=1}^{K}$ used in the measurements have been optimized for constant energy distribution overlap. For $L \geq 32$ the β_k have been chosen with the constant entropy difference scheme employing heat capacity data from smaller lattices. This has worked rather well in this case as the shape and location of the specific heat peak is subject to rather low finite-size influences.

As seen with regular periodic boundary conditions before, the estimated $D(\beta)$ curves shown in Fig. 5.9 on p. 106 indicate a thermal phase transition from a disordered high-temperature phase into a directionally ordered phase with $D > 0$. Here the temperature location of the transition does not move as much between the different lattice sizes as with the previously studied boundary conditions. This can also be seen in the locations of the peaks of the susceptibility χ in Fig. 5.10 on p. 107. For $L < 32$ there are some irregularities in the shapes of the susceptibility peaks, which disappear on larger lattices. The Binder parameter Q_2 shows distinct minima in Fig. 5.13 on p. 110. It tends to a constant value of about $1/3$ at high temperatures and to approximately $2/3$ at low temperatures.

The curves of the normalized energy $E(\beta)/N$ in Fig. 5.11 on p. 108 also show bends in the same temperature region. The corresponding peaks of the specific heat capac-

Table 5.3: Parameters used in parallel tempering simulations of the three-dimensional compass model with screw-periodic boundary conditions. Listed are the linear lattice sizes L, the corresponding lattice volumes $N = L^3$, the lowest and highest inverse temperatures β, the total number of replicas and temperature points K and the number of Monte Carlo sweeps used for thermalization and measurement. For comparison also the highest autocorrelation times of the energy time series ordered by temperature (measured close to the transition temperature) are given for each lattice size.

L	N	β-range	K	Sweeps for Thermalization	Sweeps for Measurement	$\max \tau_{\mathrm{int}}(E)$
8	512	4 — 20	32	10^6	10^7	135
12	1728	4 — 20	32	10^6	10^7	979
16	4096	4 — 20	64	10^6	10^7	554
20	8000	8 — 12.5	64	2×10^6	2×10^7	898
24	13824	8 — 12.5	64	2×10^6	2×10^7	1886
28	21952	8 — 12.5	64	2.5×10^6	2.5×10^7	5144
32	32768	9 — 12.5	64	2.5×10^6	2.5×10^7	4315
36	46656	9 — 11.5	48	2.5×10^6	2.5×10^7	7723
40	64000	9.5 — 11.5	64	3×10^6	3×10^7	8230
44	85184	9.5 — 11.5	64	3.5×10^6	3.5×10^7	19934
48	110592	9.5 — 11.5	64	3.8×10^6	3.8×10^7	36971

ity $C(\beta)/N$ in Fig. 5.12 on p. 109 have more homogeneous shapes than those of the susceptibility χ.

The maxima of χ and C/N and the minima of Q_2 have been determined by multiple histogram analysis and are listed in Table 5.4.

At low temperatures the distribution $P(D)$ of the order parameter does not show the peculiarities observed with regular periodic boundary conditions. See Fig. 5.16 on p. 112 for an example comparable to that of the last section. The histogram is strictly unimodal. An example spin configuration corresponding to the peak value of D can be seen in Fig. 5.17 on p. 113. Again linear alignment of the spins is clearly visible.

However, resulting from the connection of neighboring lines and planes with the screw-periodic boundary conditions, now all spins in one plane always point in the same direction and also spins in different planes are oriented in parallel. This is an effect of the reduced degeneracy with these boundary conditions because there are are less reflection symmetries available for the lattice planes than outlined in Sec. 1.3.1. Hence, there is no problem in using the simple definition of D as an order parameter.

Close to the transition temperature there are signs for phase coexistence realized in a double peak distribution of D. By applying the multiple histogram analysis the inverse temperatures $\beta^D_{\mathrm{eqH}}(L)$, where the histogram of D has two peaks of equal height, can be determined quite precisely. Those histograms are shown in Figs. 5.14(a)–(c) on p. 111. The double peak structure is already present in the smallest system studied here with

Table 5.4: Locations and values of the extrema of the susceptibility χ, the specific heat capacity C/N and the Binder parameter Q_2 in the three-dimensional compass model with screw-periodic boundary conditions

L	$\beta^{\chi}_{\max}$	$\chi_{\max}$	$\beta^{C}_{\max}$	$C_{\max}/N$	$\beta^{Q_2}_{\min}$	$Q_{2,\min}$
8	9.902(4)	20.10(4)	9.834(5)	1.904(3)	8.97(1)	0.230(2)
12	10.26(1)	51(1)	10.21(1)	2.83(3)	9.72(4)	0.297(4)
16	10.42(1)	75(1)	10.246(3)	2.99(2)	9.76(2)	0.293(3)
20	10.205(2)	111(1)	10.208(1)	3.53(2)	9.98(1)	0.272(3)
24	10.192(1)	190(2)	10.199(1)	4.44(3)	10.059(3)	0.244(3)
28	10.180(1)	310(3)	10.188(1)	5.6(1)	10.104(2)	0.17(1)
32	10.177(1)	457(4)	10.183(1)	6.99(5)	10.123(1)	0.11(1)
36	10.176(1)	662(5)	10.180(1)	8.8(1)	10.139(1)	0.05(1)
40	10.173(1)	916(8)	10.176(1)	10.8(1)	10.147(1)	−0.05(1)
44	10.1724(2)	1237(10)	10.1744(3)	13.3(1)	10.1521(4)	−0.17(1)
48	10.1728(4)	1688(26)	10.1742(4)	16.7(2)	10.157(1)	−0.35(2)

$L = 8$. But from $L = 16$ to $L = 28$ the central relative suppression of the probability distributions successively goes down. Also up to $L = 24$ the two peaks move successively closer together.

Then, starting from $L = 32$, the behavior changes again. The dip between the two peaks grows steadily with L and also their separation no longer shrinks. Moreover, from $L = 36$ on there are also double peak structures in the energy E. The distributions $P(E)$ measured at the corresponding inverse temperatures $\beta^{E}_{\mathrm{eqH}}(L)$ are shown in Fig. 5.15 on p. 112.

Additionally, in Fig. 5.14(d) on p. 111 the distributions $P(D)$ at inverse temperatures $\beta_{\mathrm{eqW}}(L)$ are shown for lattice sizes $L = 32, \ldots, 48$. Here the two peaks have equal weight as proposed by Janke et al. [16]. They have been found by first defining $D_{\mathrm{cut}}(L)$ as the value of D where at $\beta_{\mathrm{eqH}}(L)$ the distribution has its central minimum and then searching for the inverse temperature $\beta_{\mathrm{eqW}}(L)$ where $\int_0^{D_{\mathrm{cut}}(L)} \mathrm{d}D P(D) = \int_{D_{\mathrm{cut}}(L)}^{D_{\max}} \mathrm{d}D P(D)$ holds. Again this is easily done by a combination of multiple histogram reweighting with Brent's algorithm for minimization.

$\beta^{D}_{\mathrm{eqH}}(L)$, $\beta^{E}_{\mathrm{eqH}}(L)$ and $\beta^{D}_{\mathrm{eqW}}(L)$ are listed in Table 5.5 on p. 98 together with the suppression of the probability density at the central dip of the respective histograms as indicated by $P_{\max}/P_{\min}$. Two example configurations of the $L = 32$ system at $\beta^{D}_{\mathrm{eqH}}(L = 32) = 10.172$, which were taken at the peak-values of D corresponding to the ordered and disordered phases, are shown in Fig. 5.18 on p. 113.

The coexistence of two phases and the minima of the Binder parameter are indicators for a first-order phase transition. At such a transition the maxima of susceptibility and

Table 5.5: Inverse temperatures where the histograms of the order parameter D or the energy E have two peaks of equal height or, in the case of D, of equal weight, together with the ratio of the estimated probabilities at the highest peak and at the lowest point in the dip between the peaks

L	β^D_{eqH}	$P_{\max}/P_{\min}$	β^D_{eqW}	$P_{\max}/P_{\min}$	β^E_{eqH}	$P_{\max}/P_{\min}$
8	9.906(4)	1.40(4)				
12	10.29(1)	1.7(1)				
16	10.39(1)	1.5(1)				
20	10.26(1)	1.20(3)				
24	10.199(3)	1.14(2)				
28	10.176(1)	1.0(2)				
32	10.172(1)	1.17(4)	10.172(1)	1.17(4)		
36	10.173(1)	1.48(5)	10.173(1)	1.48(5)	10.177(1)	1.0(1)
40	10.1719(5)	1.8(1)	10.1719(5)	1.8(1)	10.175(1)	1.23(2)
44	10.1716(2)	2.1(1)	10.1716(2)	2.1(1)	10.1740(2)	1.47(3)
48	10.1727(4)	3.1(2)	10.1727(4)	3.1(2)	10.1744(4)	1.9(1)

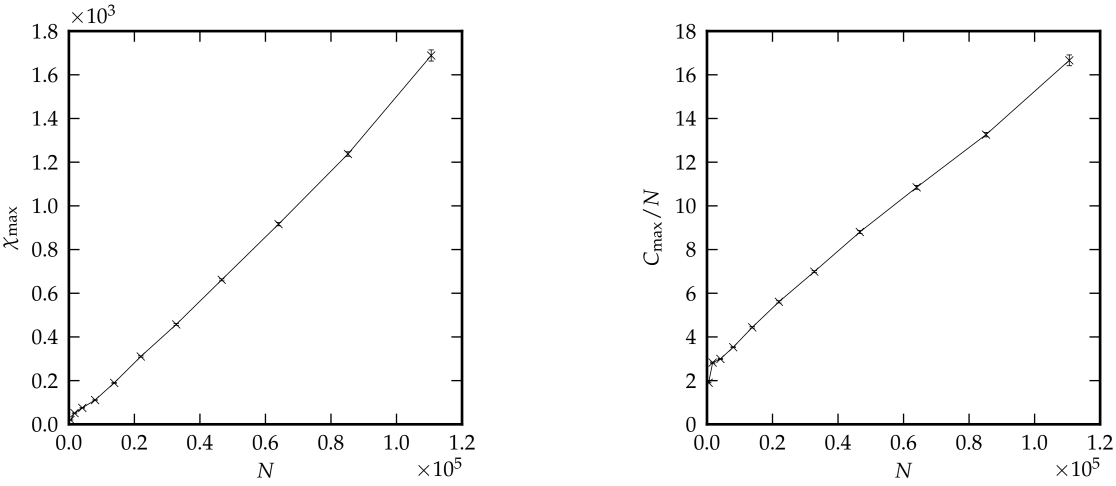
specific heat capacity are expected to scale with the system size $N = L^3$ as

$$\begin{aligned} \chi_{\max}(L) &= a_1 + b_1 \cdot N + \cdots, \\ C_{\max}(L)/N &= a_2 + b_2 \cdot N + \cdots. \end{aligned} \tag{5.5}$$

Plotted over the system size these maximum values indeed show an approximately linear growth in Fig. 5.6 if the attention is restricted to the larger lattices. The quality of attempted linear fits to Eqs. (5.5) without further scaling corrections is summarized in Table 5.6. The agreement is better for the specific heat capacity than for the susceptibility.

Table 5.6: The maximum values of the susceptibility ($\mathcal{O} = \chi$) and specific heat capacity ($\mathcal{O} = C/N$) from Table 5.4 have been fitted to the linear relation $\mathcal{O}_{\max} = a + b \cdot N$, where $N = L^3$. Listed are the resulting values of $\chi^2_{\text{dof}} = \chi^2/n-2$ for different ranges of included lattice sizes.

$L_{\max}$	$L_{\min}$	n	$\chi^2_{\text{dof}}[\mathcal{O} = \chi]$	$\chi^2_{\text{dof}}[\mathcal{O} = C/N]$
48	8	11	77.91	347.75
	12	10	79.26	11.61
	16	9	42.19	11.44
	20	8	5.01	11.93
	24	7	3.79	4.18
	28	6	4.46	1.60
	32	5	2.72	1.56
	36	4	3.25	1.16
	40	3	3.43	2.09



(a) Susceptibility χ

(b) Specific heat capacity C/N

Figure 5.6: Maximum values taken from Table 5.4 plotted over the lattice volume $N = L^3$

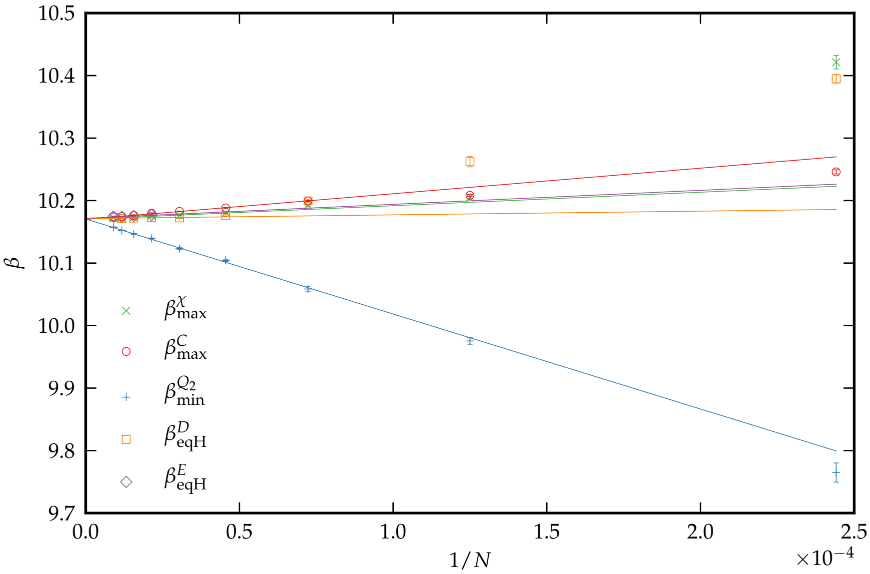


Figure 5.7: Finite-size scaling of inverse pseudo-transition temperatures of the three-dimensional compass model with screw-periodic boundary conditions for $L \geq 16$ together with the best fits from Tables 5.6 and 5.7, which allow to extrapolate the infinite-volume transition point β_0.

5.2.1 Transition point

With $\beta^{C}_{\max}(L)$, $\beta^{\chi}_{\max}(L)$, $\beta^{Q_2}_{\min}(L)$, $\beta^{D}_{\mathrm{eqH}}(L)$ and $\beta^{E}_{\mathrm{eqH}}(L)$ there are various possible definitions of a lattice-size dependent inverse pseudo-transition temperature. For a first-order transition they are expected to have a displacement from the true infinite-volume transition point β_0 that to leading order scales proportionally to the reciprocal system size $1/N$:

$$\begin{aligned}
\beta^{C}_{\max}(L) &= \beta_0 + \frac{c_1}{N} + \cdots, \\
\beta^{\chi}_{\max}(L) &= \beta_0 + \frac{c_2}{N} + \cdots, \\
\beta^{Q_2}_{\min}(L) &= \beta_0 + \frac{c_3}{N} + \cdots, \\
\beta^{D}_{\mathrm{eqH}}(L) &= \beta_0 + \frac{c_4}{N} + \cdots, \\
\beta^{E}_{\mathrm{eqH}}(L) &= \beta_0 + \frac{c_5}{N} + \cdots.
\end{aligned} \tag{5.6}$$

The results of linear fits to Eqs. (5.6) are given in Tables 5.7 and 5.8. For a visualization of the scaling see Fig. 5.7. The fits are better for $\beta^{C}_{\max}(L)$, $\beta^{\chi}_{\max}(L)$ and $\beta^{Q_2}_{\min}(L)$ than for $\beta^{D}_{\mathrm{eqH}}(L)$ and $\beta^{E}_{\mathrm{eqH}}(L)$, but they all agree well. The best estimate for the infinite-volume transition point can be made from the specific heat capacity data from $L \geq 24$:

$$\beta_0 = 10.1700(3), \qquad \chi^2_{\mathrm{dof}} = 1.12. \tag{5.7}$$

This corresponds to the transition temperature

$$T_0 = 0.098328(3). \tag{5.8}$$

While it is possible to consider additional terms with higher powers of $(1/N)$ or exponential corrections in the scaling laws (5.6), this also leads to a higher number of free parameters and in this case does not improve the quality of the fits.

Table 5.7: Results of least-squares fits of the inverse temperatures taken from Table 5.4, where the maxima of χ and C and the minima of Q_2 are located, to estimate the infinite volume transition point β_0 by a relation of the form $\beta(L) = \beta_0 + c/N$. Various ranges of lattice sizes have been tried. The best fits are marked bold.

$L_{\max}$	$L_{\min}$	n	$\beta_0^{\chi,\max}$	χ^2_{dof}	$\beta_0^{C,\max}$	χ^2_{dof}	$\beta_0^{Q_2,\min}$	χ^2_{dof}
48	8	11	10.176(3)	268.62	10.180(4)	445.25	10.157(4)	148.63
	12	10	10.170(2)	51.07	10.173(2)	62.05	10.170(2)	14.69
	16	9	10.169(2)	42.43	10.171(1)	6.85	10.171(1)	1.97
	20	**8**	10.1693(5)	3.52	10.171(1)	7.15	**10.171(1)**	**1.57**
	24	**7**	10.169(1)	4.18	**10.1700(3)**	**1.12**	10.170(1)	1.57
	28	**6**	**10.1702(5)**	**1.60**	10.1699(5)	1.38	10.170(1)	1.59
	32	5	10.170(1)	2.10	10.170(1)	1.64	10.170(1)	1.79
	36	4	10.170(1)	3.11	10.169(1)	2.34	10.170(1)	2.02
	40	3	10.172(2)	2.66	10.171(2)	2.16	10.171(2)	1.80

Table 5.8: Results of least-squares fits of the inverse temperatures taken from Table 5.5, where equal peak height and equal peak weight histograms are located, to estimate the infinite volume transition point β_0 by a relation of the form $\beta(L) = \beta_0 + c/N$. Various ranges of lattice sizes have been tried. The best fits are marked bold.

$L_{\max}$	$L_{\min}$	n	$\beta_0^{\text{eqH},D}$	χ^2_{dof}	$\beta_0^{\text{eqH},E}$	χ^2_{dof}
48	8	11	10.174(3)	259.64		
	12	10	10.168(2)	86.60		
	16	9	10.165(3)	78.28		
	20	8	10.169(2)	22.40		
	24	7	10.170(1)	12.28		
	28	**6**	**10.171(1)**	**2.85**		
	32	5	10.172(1)	2.64		
	36	**4**	10.172(1)	3.95	**10.172(1)**	**2.84**
	40	3	10.173(2)	4.51	10.173(2)	3.22

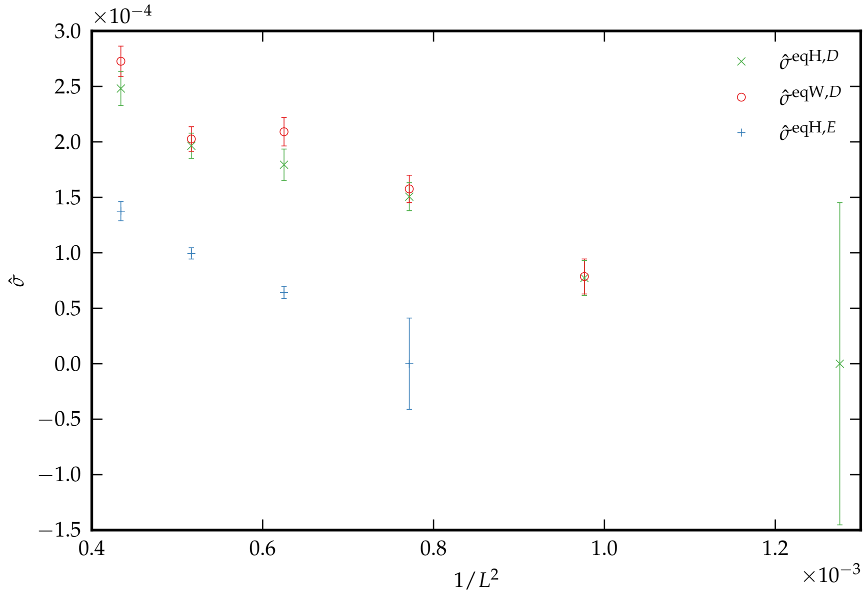


Figure 5.8: Reduced interface tensions calculated from $P(D)$ histograms at $\beta^D_{\text{eqH}}(L)$ and $\beta^D_{\text{eqW}}(L)$ and from $P(E)$ histograms at $\beta^E_{\text{eqH}}(L)$ plotted over $1/L^2$

5.2.2 Interface tension

If a system exhibiting a first-order phase transition is considered on finite periodic lattices of size L^3, the suppression of the minimum between the two peaks of the probability distribution of the energy or the order parameter is expected to grow exponentially with L^2:

$$P_{\text{max}}/P_{\text{min}} \propto \mathrm{e}^{2\beta\sigma L^2}. \tag{5.9}$$

Configurations corresponding to P_{min} are in a mixture of the ordered and the disordered phases with two interfaces that contribute an excess free energy of $2\sigma L^2$. The free energy density σ is called interface tension. It is a common practice to extract the reduced interface tension $\hat{\sigma} = \beta\sigma$ from

$$\hat{\sigma}(L) = \frac{1}{2L^2} \ln \left[\frac{P_{\text{max}}(L)}{P_{\text{min}}(L)} \right], \tag{5.10}$$

where $P_{\text{max}}(L)/P_{\text{min}}(L)$ is determined from the double-peaked distributions, and then to determine the infinite volume limit $\hat{\sigma}_0$ of $\hat{\sigma}(L)$.

For the three-dimensional compass model I have calculated $\hat{\sigma}(L)$ at $\beta^D_{\text{eqH}}(L)$ and $\beta^D_{\text{eqW}}(L)$ for $P(D)$ and at $\beta^E_{\text{eqH}}(L)$ for $P(E)$. The results are listed in Table 5.9 and are plotted over $1/L^2$ in Fig. 5.8 for $L \geq 28$. For these lattice sizes the reduced interface tension grows with L. If it vanished in the limit of large systems, this would be an argument against the first-order nature of the transition. The results of the attempts to calculate $\hat{\sigma}_0$ by fitting to the relation

$$\hat{\sigma}(L) = \hat{\sigma}_0 + \frac{c}{L^2} \tag{5.11}$$

are summarized in Table 5.10 on p. 104. The fits are not of high accuracy, but yield an estimate of approximately

$$\hat{\sigma}_0 \approx 3 \cdot 10^{-4} \tag{5.12}$$

However, the asymptotic regime, where Eq. (5.11) is valid, may not yet have been reached. In general it is also not clear if the same interface tension would be obtained for both observables D and E.

The usage of the histograms of equal peak weight was mainly motivated by the desire to reduce the effect of unconsidered scaling corrections, but in this case they do not increase the quality of the fits. Similarly, carrying out this analysis directly at the estimated true inverse transition temperature $\beta_0 = 10.17$ for every lattice size does not yield better results. Except for the largest lattices the histograms are still highly asymmetric there.

Table 5.9: Reduced interface tensions calculated from the data in Table 5.5 for $P(D)$ histograms at $\beta^D_{\mathrm{eqH}}(L)$ and $\beta^D_{\mathrm{eqW}}(L)$ and for $P(E)$ histograms at $\beta^E_{\mathrm{eqH}}(L)$

L	$\hat{\sigma}^{\mathrm{eqH},D}(L)$	$\hat{\sigma}^{\mathrm{eqW},D}(L)$	$\hat{\sigma}^{\mathrm{eqH},E}(L)$
8	0.0026(2)		
12	0.0018(2)		
16	0.0008(1)		
20	0.00023(3)		
24	0.00011(2)		
28	0.0000(1)		
32	0.00008(2)	0.00008(2)	
36	0.00015(1)	0.00016(1)	0.00000(4)
40	0.00018(1)	0.00021(1)	0.00006(1)
44	0.00020(1)	0.00020(1)	0.00010(1)
48	0.00025(2)	0.00027(1)	0.00014(1)

Table 5.10: Results of least-squares fits of the data from Table 5.9 to estimate the infinite volume limit of the interface tension $\hat{\sigma}_0$ by the relation $\hat{\sigma}(L) = \hat{\sigma}_0 + c/L^2$ from different ranges of included lattice sizes between L_{min} and L_{max}. The quantity $\chi^2_{\mathrm{dof}} = \chi^2/n-2$ is given in each case to help with the estimation of the validity of the fit.

L_{max}	L_{min}	n	$\hat{\sigma}_0^{\mathrm{eqH},D}$	χ^2_{dof}	$\hat{\sigma}_0^{\mathrm{eqW},D}$	χ^2_{dof}	$\hat{\sigma}_0^{\mathrm{eqH},E}$	χ^2_{dof}
48	**28**	**6**	**$3.1(3) \cdot 10^{-4}$**	**6.32**				
	32	**5**	$3.2(4) \cdot 10^{-4}$	7.05	**$3.4(5) \cdot 10^{-4}$**	**10.52**		
	36	**4**	$4(1) \cdot 10^{-4}$	6.16	$4(1) \cdot 10^{-4}$	9.55	**$2.8(3) \cdot 10^{-4}$**	**4.18**
	40	3	$5(1) \cdot 10^{-4}$	6.48	$5(2) \cdot 10^{-4}$	16.27	$2.9(4) \cdot 10^{-4}$	6.28
44	28	5	$2.9(2) \cdot 10^{-4}$	2.19				
	32	4	$3.0(2) \cdot 10^{-4}$	1.86	$3.2(2) \cdot 10^{-4}$	2.09		
	36	3	$3.4(2) \cdot 10^{-4}$	0.70	$3.7(1) \cdot 10^{-4}$	0.29	$2.5(1) \cdot 10^{-4}$	0.61

5.2.3 Evaluation of the algorithms

The simulation of the $L = 48$ system, the largest considered here, ran for 10 days on the parallel computing cluster, where each replica was assigned to a separate node. Despite the high computational effort, which amounts to a total of more than 15000 CPU hours, one finds that only about 500 independent configurations have been sampled at the temperature with the strongest temporal correlations if one considers the integrated autocorrelation time of the energy time series sorted by temperature (see Table 5.3 on p. 96). Of course the situation is better at the other temperatures and the combination of statistics by multiple histogram analysis further improves matters. Moreover, compared with simple canonical simulations using only the Metropolis algorithm the cluster update and parallel tempering already reduce correlations considerably. But different choices of algorithms might lead to further improvements.

Parallel tempering usually is not the best choice for the analysis of first-order phase transitions. When high autocorrelation times at the transition temperature are mainly caused by the exponentially growing long times the system needs to flip between the two coexisting phases, parallel tempering is only of limited help. It accelerates the simulation in both phases, but does not assist much with the crossover between them. Since the different phases at the temperature of coexistence correspond to different peaks in the energy distribution, in this case at least for the large systems, replica exchanges between the phases are rarely accepted because of little overlap. This is manifested in the bottleneck of replica diffusion discussed in Sec. 3.5.2.

It might be worthwhile to check whether multicanonical sampling [76, 77] can be a viable alternative for the relatively weak first-order transition in the three-dimensional compass model. With the multicanonical approach, one first determines adapted weighting factors for the probability density in an initial recursion. Ideally, those then allow to sample all regions of phase space with the same probability, which would speed up the traversal of the mixed-phase region in the large lattices. On the other hand, the recursion can be very time consuming and the inclusion of the cluster update would be nontrivial. Another related method of interest, which provides many of the advantages of multicanonical sampling to parallel tempering simulations without requiring the costly initial adaption of weighting factors, could be the algorithm using multiple Gaussian modified ensembles presented in Ref. [78].

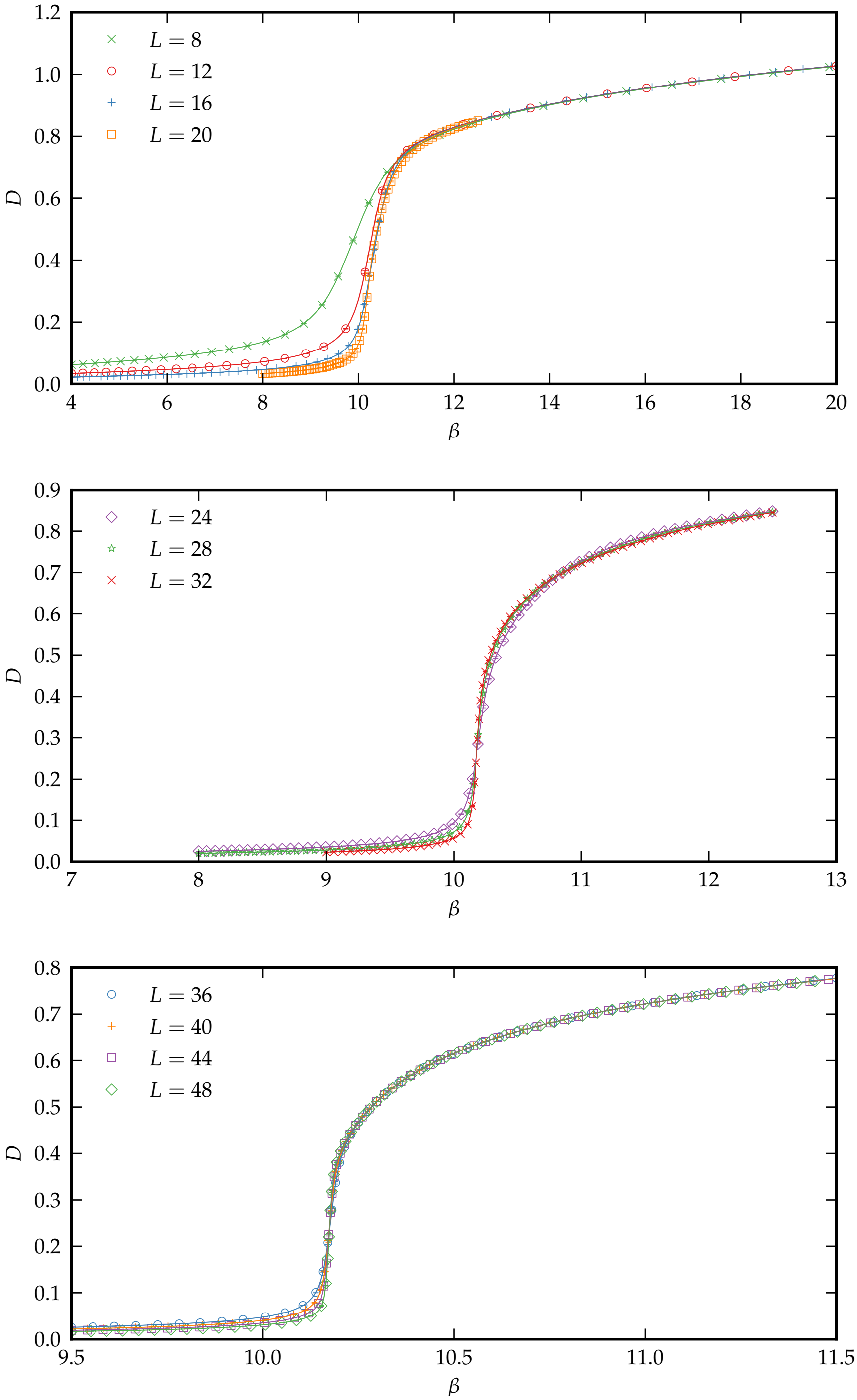


Figure 5.9: Estimates of the order parameter D for various lattice sizes and inverse temperatures β obtained in simulations of the three-dimensional compass model with screw-periodic boundary conditions. Top row: Linear sizes $L = 8, \dots, 20$. Middle row: Linear sizes $L = 24, \dots, 32$. Bottom row: Linear sizes $L = 36, \dots, 48$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.

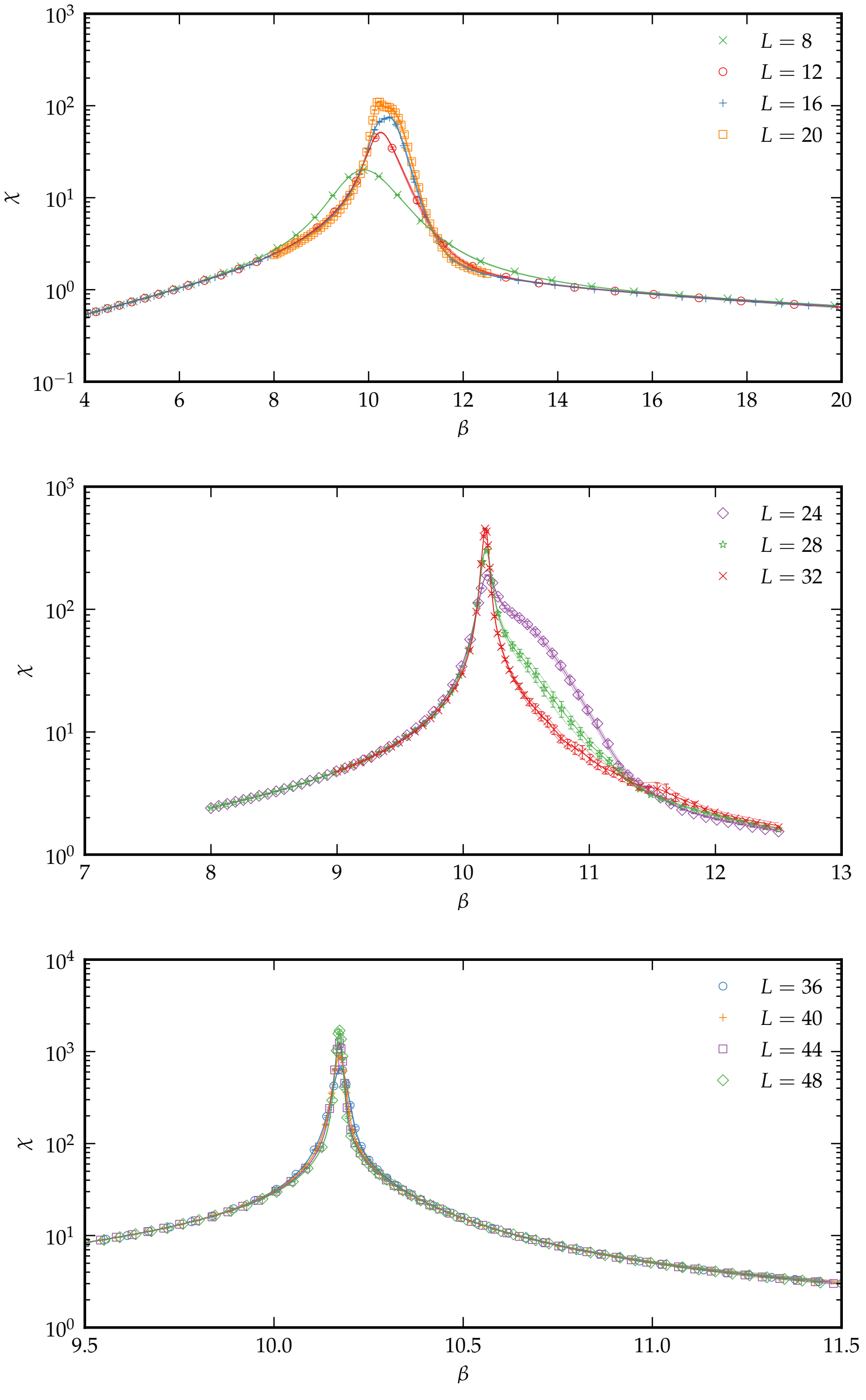


Figure 5.10: Estimates of the order parameter susceptibility χ for various lattice sizes and inverse temperatures β obtained in simulations of the three-dimensional compass model with screw-periodic boundary conditions. Top row: Linear sizes $L = 8, \ldots, 20$. Middle row: Linear sizes $L = 24, \ldots, 32$. Bottom row: Linear sizes $L = 36, \ldots, 48$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.

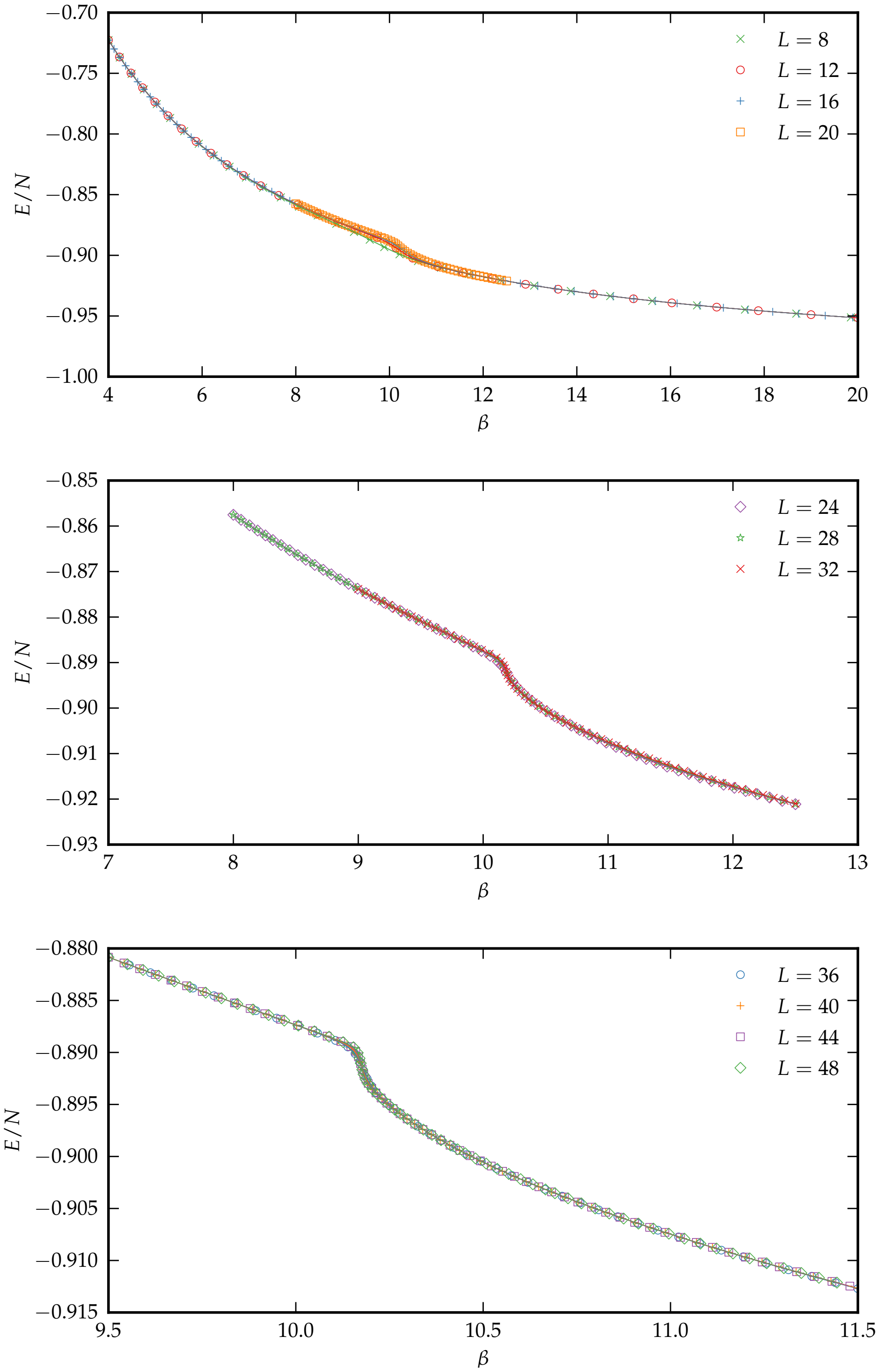


Figure 5.11: Estimates of the energy per site E/N for various lattice sizes and inverse temperatures β obtained in simulations of the three-dimensional compass model with screw-periodic boundary conditions. Top row: Linear sizes $L = 8, \ldots, 20$. Middle row: Linear sizes $L = 24, \ldots, 32$. Bottom row: Linear sizes $L = 36, \ldots, 48$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.

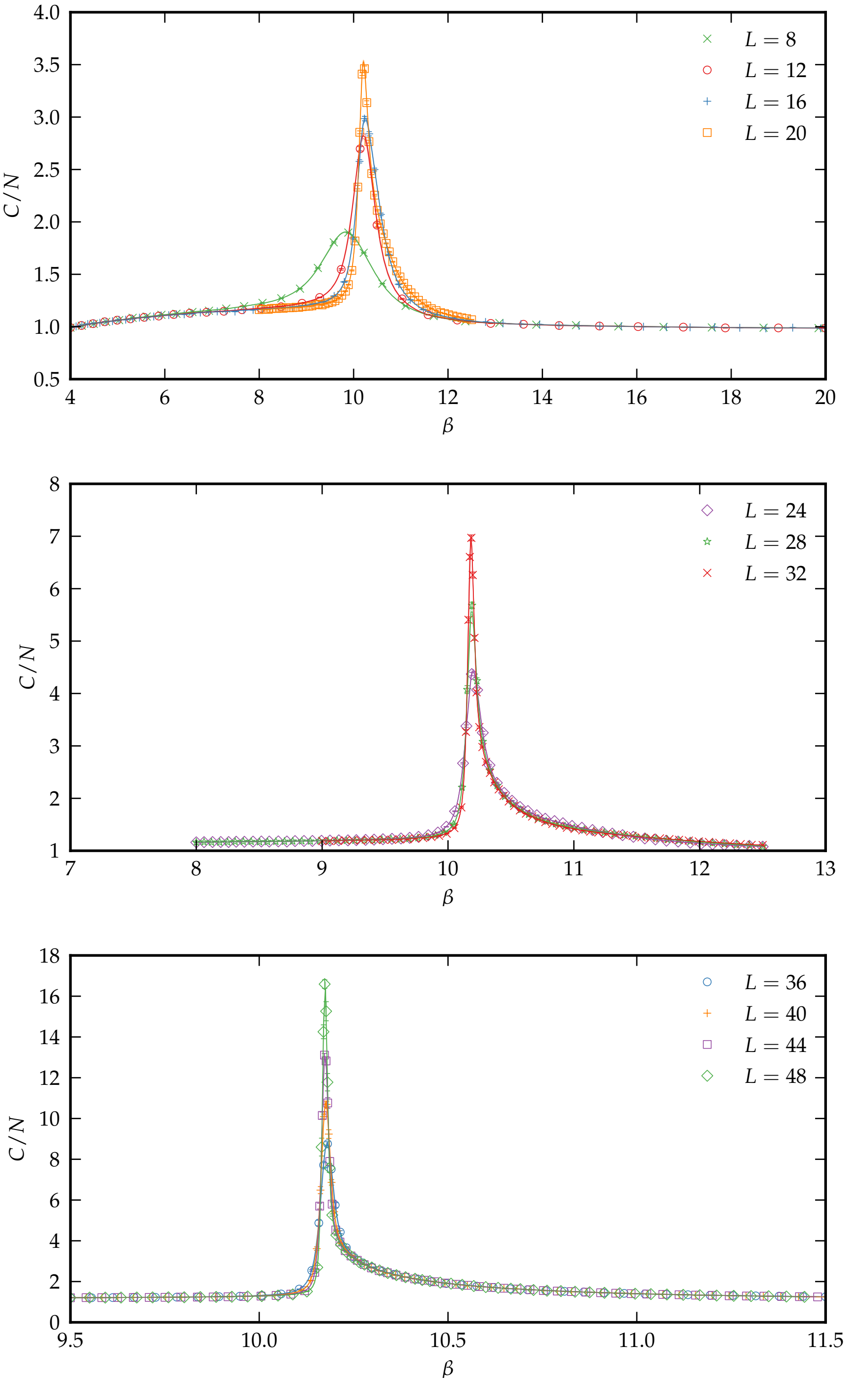


Figure 5.12: Estimates of the specific heat capacity C/N for various lattice sizes and inverse temperatures β obtained in simulations of the three-dimensional compass model with screw-periodic boundary conditions. Top row: Linear sizes $L = 8, \dots, 20$. Middle row: Linear sizes $L = 24, \dots, 32$. Bottom row: Linear sizes $L = 36, \dots, 48$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.

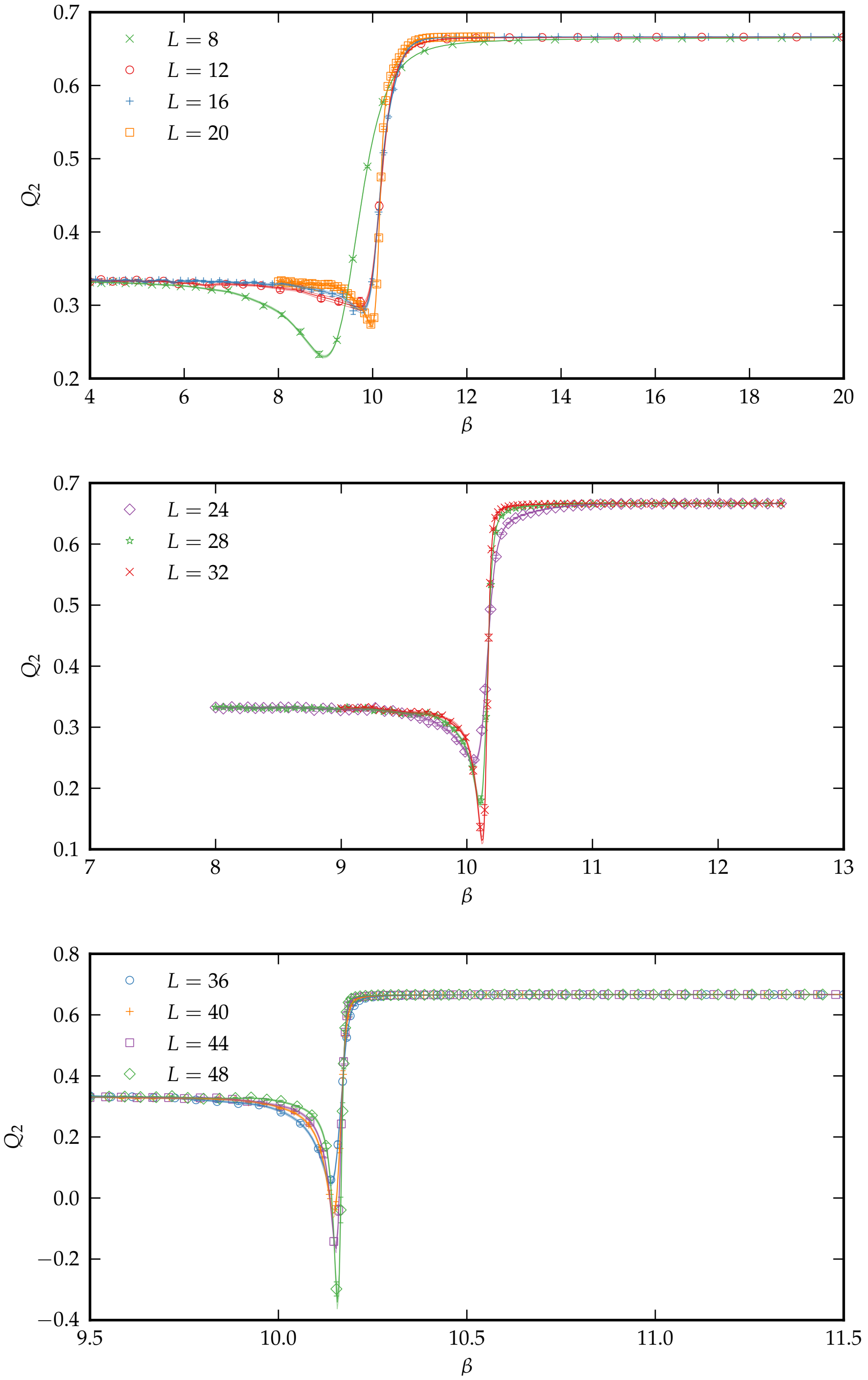
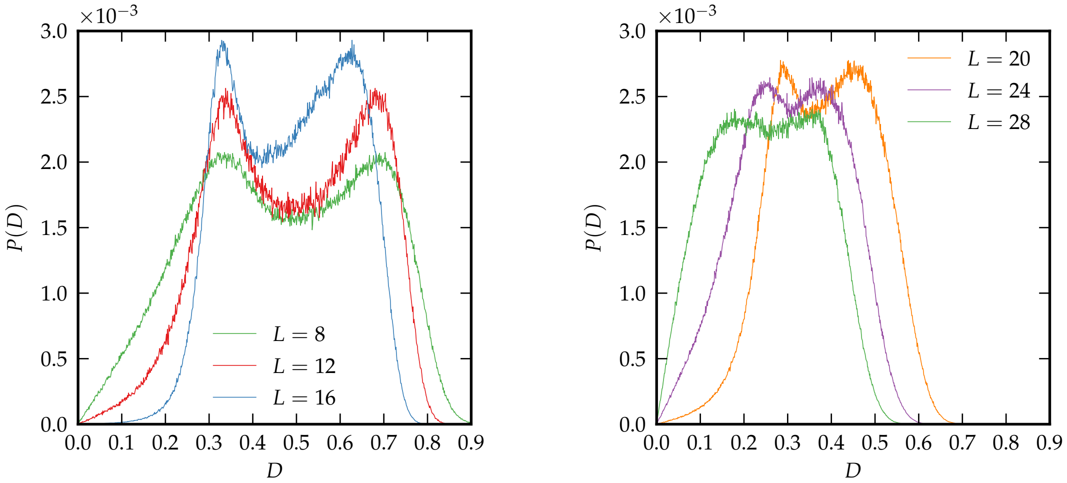
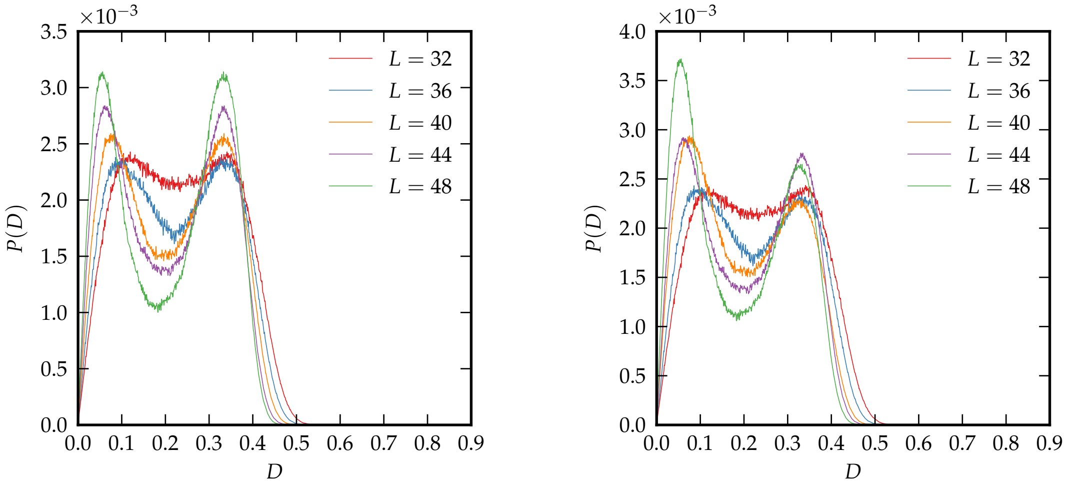


Figure 5.13: Estimates of the Binder parameter Q_2 for the order parameter D for various lattice sizes and inverse temperatures β obtained in simulations of the three-dimensional compass model with screw-periodic boundary conditions. Top row: Linear sizes $L = 8, \ldots, 20$. Middle row: Linear sizes $L = 24, \ldots, 32$. Bottom row: Linear sizes $L = 36, \ldots, 48$. Markers with error bars are estimates from single-temperature time series. Continuous lines are from the multiple histogram analysis with faint surrounding lines indicating the 1σ-margin of statistical uncertainty.



(a) Equal height

(b) Equal height



(c) Equal height

(d) Equal weight

Figure 5.14: Histograms of the order parameter D in the three-dimensional compass model with screw-periodic boundary conditions for various lattice sizes. Different factors of normalization have been used on different lattice sizes. (a)-(c): Equal peak height. (d): Equal peak weight

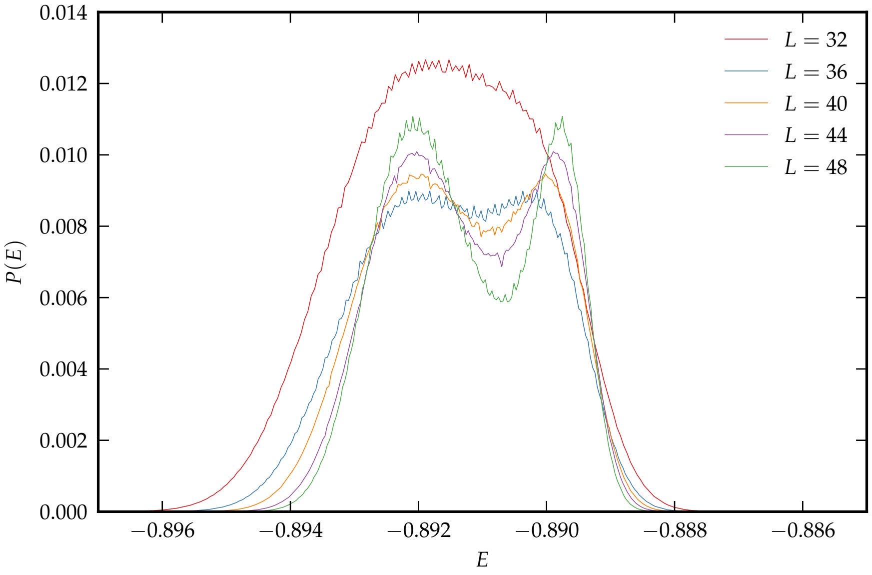
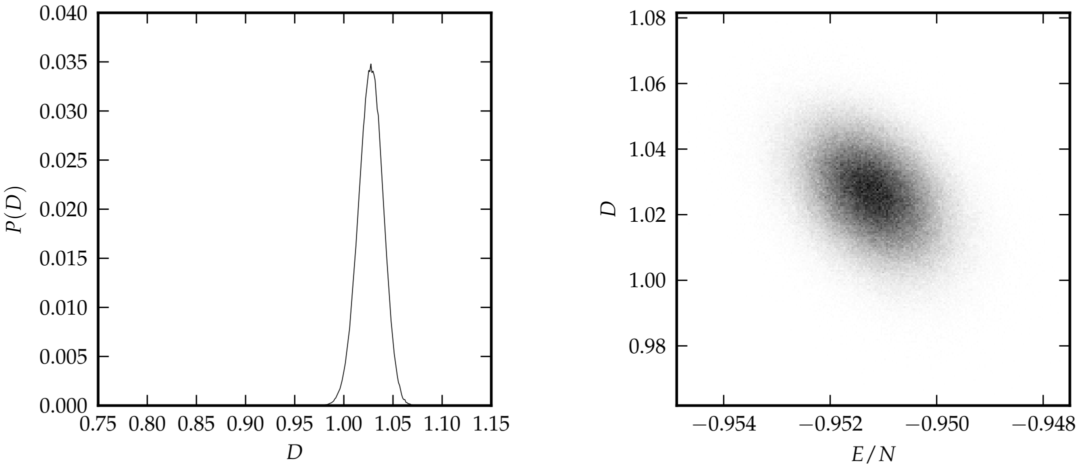


Figure 5.15: Equal peak height histograms of the energy E in the three-dimensional compass model with screw-periodic boundary conditions for various lattice sizes. For $L = 32$ and smaller lattices no double-peak distribution can be found at any temperature. The $L = 32$ histogram in the plot is shown only for comparison and is taken at a temperature close to that of the others. Different factors of normalization have been used on different lattice sizes.



(a) Histogram of the order parameter D

(b) 2D histogram of D and the energy E

Figure 5.16: Distributions measured in the three-dimensional compass model with screw-periodic boundary conditions with $L = 16$ in the low temperature phase at $\beta = 20$. In (b) dark points correspond to high densities.

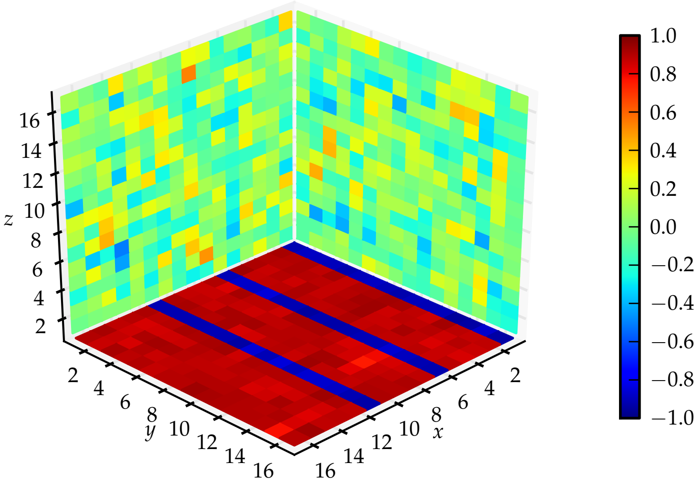
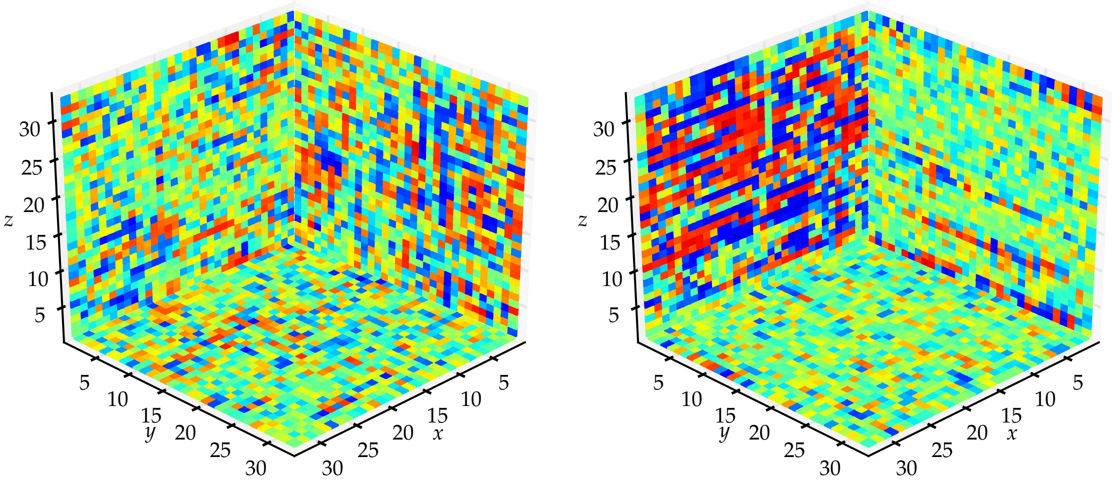


Figure 5.17: Averaged spin projections of an example configuration with $D = 1.028$ of the three-dimensional system with $L = 16$ and screw-periodic boundary conditions at $\beta = 20$.



(a) $D = 0.12$

(b) $D = 0.34$

Figure 5.18: Averaged spin projections of two example configurations of the three-dimensional compass model with screw-periodic boundary and conditions and $L = 32$ at $\beta = 10.172$ close to the phase transition. (a) is taken from the disordered phase, (b) from the ordered phase. The same color coding as in Fig. 5.17 has been used.

6 Conclusion

My findings for the two-dimensional classical compass model are in good agreement with the well-accepted results that have been published earlier. There is a temperature-driven continuous phase transition between a disordered high-temperature phase and a phase characterized by long-range directional ordering at low temperatures. The critical temperature in the thermodynamic limit has been estimated as

$$T_c = 0.14617 \pm 0.00006$$

in units with $k_B = 1$ and $|J| = 1$. The critical exponents of the spatial correlation length and the order parameter susceptibility have been found to be

$$\nu = 1.00 \pm 0.01 \qquad \text{and}$$
$$\gamma = 1.71 \pm 0.02.$$

The transition may belong to the two-dimensional Ising universality class.

The main result of this thesis is that there is also a directional ordering transition in the three-dimensional classical compass model, which is first-order. I have extrapolated a value of

$$T_0 = 0.098328 \pm 0.000003$$

for the infinite volume transition temperature.

In contrast to conventional magnetic-like ordering the essentially one-dimensional directional ordering of the compass model is less stable in higher dimensions. This is reflected by a lower value of the transition temperature in three than in two dimensions. Generally the directional ordering is broken up faster by thermal fluctuations. The Ising model for instance has significantly higher critical temperatures of $T_c \approx 2.27$ in two dimensions and $T_c \approx 4.51$ in three dimensions.

In both the two- and the three-dimensional compass model finite-size effects are severe. Only with the help of unusual screw-periodic boundary conditions reliable results could be obtained. Still it has been necessary to go to $L > 32$ (corresponding to $N > 32768$ spins) to see clear first-order scaling behavior in the three-dimensional model. This may be explained by a length scale set by a spatial correlation length, for which, however, a suitable observable would have to be found.

As a byproduct of this thesis the multiple histogram reweighting method has been studied quite intensively. While the technique has been well-known for more than twenty years, there has been an upsurge of popularity more recently, which can be explained by the frequent application of parallel tempering in many fields. Nevertheless, it is still not fully clear how to best deal with correlated data in the estimation of the density of states. Unfortunately the best conclusion I can give in this regard is that it seems to hardly matter how one accounts for autocorrelation times.

When work on this thesis had already been well advanced, an independent study by Wenzel and Läuchli was published [79], which also mentions a first-order transition in the three-dimensional compass model. The authors quote a value of $T_0 \approx 0.098$ for the transition temperature, which matches the estimate given above. The absence of a signature of a phase transition in the three-dimensional quantum model in the work by Oitmaa and Hamer [11] may be explained by a limitation of their methods, which can only detect well continuous transitions, or by an intrinsic difference of the classical and the quantum model in this case. It would be an intriguing future project to pursue further investigations of the quantum compass model in three dimensions. Monte Carlo methods may be of help to shed light on this open question and to gain better understanding of orbital-only models. It may furthermore be interesting to see how an asymmetric version of the three dimensional compass model reacts to a change of the ratio of the directional coupling constants or which influence the addition of an external field can have.

Appendix A

Check for correctness of the three-dimensional Metropolis update

A model with the same constituents $\boldsymbol{s}_i \in S^2$ as the three-dimensional compass model is the classical Heisenberg model. In the case of a one-dimensional chain of N spins with free boundary conditions it can be solved exactly[1]. This suggests the adaption of the Metropolis algorithm from Sec. 3.3.2 to this model to verify the correctness of its results with the analytic solution.

The Hamiltonian of the one-dimensional Heisenberg chain with free boundary conditions is given by

$$\mathcal{H} = -J \sum_{i=1}^{N-1} \boldsymbol{s}_i \cdot \boldsymbol{s}_{i+1} = -J \sum_{i=1}^{N-1} \cos(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}) = -J \sum_{i=1}^{N-1} \cos(\theta_i). \tag{A.1}$$

The energy only depends on the $N-1$ angles θ_i between neighboring spins. Conveniently, the partition function factorizes with this choice of boundary conditions:

$$\begin{aligned} \mathcal{Z}_{N-1} &= \int \mathrm{d}\boldsymbol{s}_1 \cdots \int \mathrm{d}\boldsymbol{s}_N \mathrm{e}^{\frac{J}{k_\mathrm{B}T} \sum_{i=1}^{N-1} \boldsymbol{s}_i \boldsymbol{s}_{i+1}} = \int \mathrm{d}\Omega_1 \cdots \int \mathrm{d}\Omega_N \mathrm{e}^{\frac{J}{k_\mathrm{B}T} \sum_{i=1}^{N-1} \cos(\theta_i)} \\ &= \left[\int \mathrm{d}\Omega_1 \mathrm{e}^{\frac{J}{k_\mathrm{B}T} \cos(\theta_1)} \right]^{N-1} = \mathcal{Z}_1^{N-1}. \end{aligned} \tag{A.2}$$

The remaining integral can be solved directly:

$$\mathcal{Z}_1 = \int \mathrm{d}\Omega \mathrm{e}^{\frac{J}{k_\mathrm{B}T} \cos(\theta)} = 2\pi \int_{-1}^{1} \mathrm{d}(\cos\theta) \mathrm{e}^{\frac{J}{k_\mathrm{B}T} \cos(\theta)} = \frac{4\pi k_\mathrm{B} T}{J} \sinh\left(\frac{J}{k_\mathrm{B}T}\right). \tag{A.3}$$

Making use of thermodynamic relations, the free energy, entropy, internal energy and heat capacity of the two-spin system, which has only one degree of freedom, follow from

---

[1] An early discussion of the model is given by Fisher [80].

Eq. (A.3):

$$F_1 = -k_\mathrm{B} T \ln \mathcal{Z}_1 = -k_\mathrm{B} T \ln \left[\frac{4\pi k_\mathrm{B} T}{J} \sinh\left(\frac{J}{k_\mathrm{B} T} \right) \right] \tag{A.4}$$

$$\begin{aligned} S_1 &= -\frac{\partial F_1}{\partial T} = k_\mathrm{B} \ln \mathcal{Z}_1 + \frac{k_\mathrm{B} T}{\mathcal{Z}_1} \frac{\partial \mathcal{Z}_1}{\partial T} \\ &= k_\mathrm{B} \left[1 + \frac{J}{k_\mathrm{B} T} \coth\left(\frac{J}{k_\mathrm{B} T} \right) + \ln \left[\frac{4\pi k_\mathrm{B} T}{J} \sinh\left(\frac{J}{k_\mathrm{B} T} \right) \right] \right] \end{aligned} \tag{A.5}$$

$$E_1 = F_1 + T S_1 = k_\mathrm{B} T - J \coth\left(\frac{J}{k_\mathrm{B} T} \right) \tag{A.6}$$

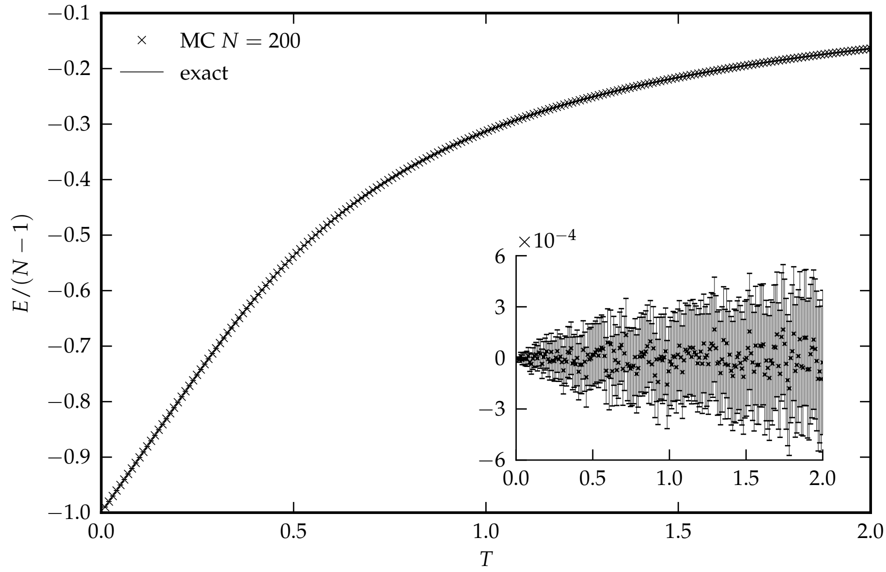
$$C_1 = \frac{\partial E_1}{\partial T} = k_\mathrm{B} - \frac{J^2}{k_\mathrm{B} T^2} \sinh^{-2}\left(\frac{J}{k_\mathrm{B} T} \right) \tag{A.7}$$

Since $\mathcal{Z}_{N-1} = \mathcal{Z}_1^{N-1}$, one has $F_1 = {}^{F_{N-1}}\!/_{N-1}$ and consequently $E_1 = {}^{E_{N-1}}\!/_{N-1}$ and $C_1 = {}^{C_{N-1}}\!/_{N-1}$. There are no finite-size effects.

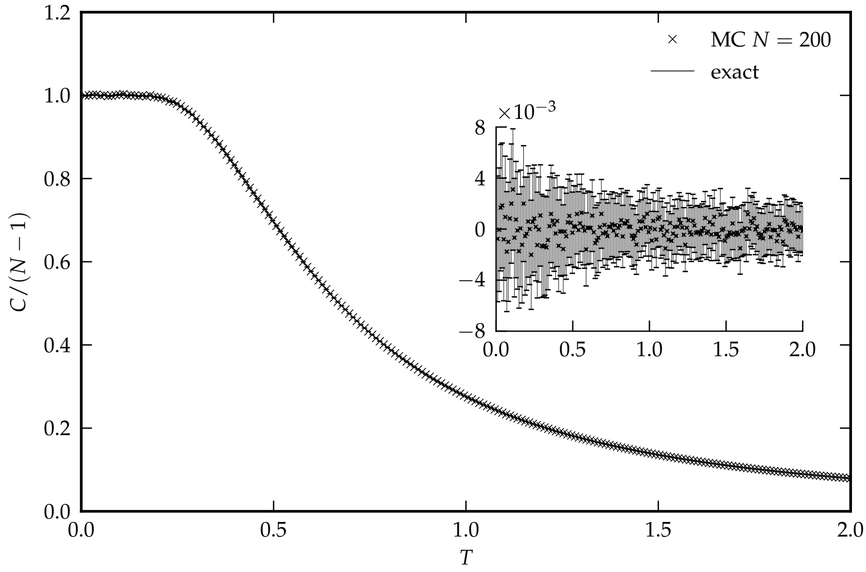
I have implemented a Monte Carlo simulation of the classical Heisenberg model using the update from Algorithm 3.2 on p. 38, but with random selection of the site i and the energy difference term adapted to the new model:

$$\begin{aligned} \Delta E &= -J \left[(\boldsymbol{s}_{i-1} \cdot \boldsymbol{s}'_i + \boldsymbol{s}'_i \cdot \boldsymbol{s}_{i+1}) - (\boldsymbol{s}_{i-1} \cdot \boldsymbol{s}_i + \boldsymbol{s}_i \cdot \boldsymbol{s}_{i+1}) \right] \\ &= -J \left[(\boldsymbol{s}_{i-1} + \boldsymbol{s}_{i+1}) \cdot (\boldsymbol{s}'_i - \boldsymbol{s}_i). \right] \end{aligned} \tag{A.8}$$

The $N = 200$ system has been simulated in the canonical ensemble at different temperatures ranging from $T_\mathrm{min} = 0.01$ to $T_\mathrm{max} = 2$ for 10^7 sweeps each following a thermalization phase of 10^5 sweeps. I have attempted to set an acceptance ratio of 50% using the method presented in Sec. 3.3.3, but this was not possible for the higher temperature regions, which lead to acceptance ratios of up to 68%. The measurements of the energy and heat capacity shown in Fig. A.1 match the theory well with small deviations distributed uniformly around the expected values.



(a) Internal energy



(b) Heat capacity

Figure A.1: Monte Carlo results of energy and heat capacity for the $N = 200$ classical Heisenberg spin chain compared with the analytic solution. The insets show the relative deviations from the exact results with statistical uncertainties estimated by the jackknife method.

Appendix B

Numerical comparisons for the multiple histogram method

In Sec. 3.6.1.3 an analysis of energy histogram uncertainties with bin-dependent autocorrelations taken into account for the multiple histogram reweighting method has been presented. Since such a procedure has not yet found widespread use, an examination of the nature of these uncertainties shall be given in the following. Moreover the results of this analysis with

(a) statistical inefficiencies distinguished by energy bin and temperature: $\delta^2 H_{ik} = g_{ik} \langle H_{ik} \rangle$,

(b) only temperature-dependent inefficiency factors: $\delta^2 H_{ik} = g_k \langle H_{ik} \rangle$, or

(c) no consideration of varied inefficiencies at all: $\delta^2 H_{ik} = \langle H_{ik} \rangle$

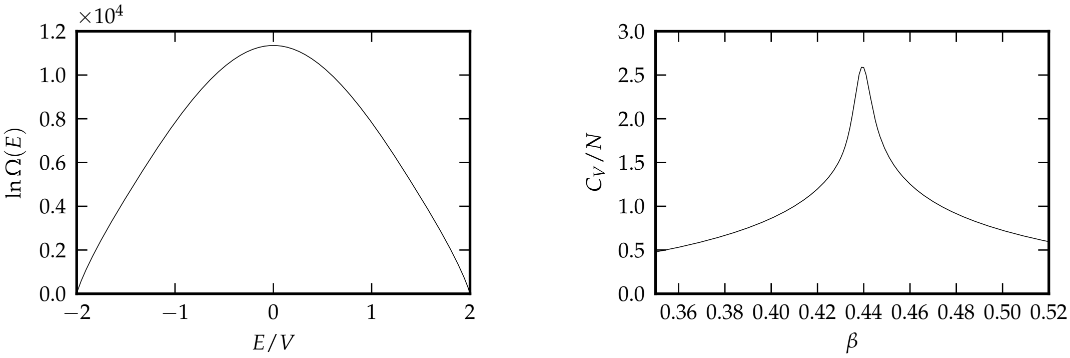
will be compared.

As a reference system the two-dimensional Ising model [33, 34] without external magnetic field, defined by the Hamiltonian

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j, \qquad \text{with} \qquad \sigma_i = \pm 1, \tag{B.1}$$

is a good choice. It has not only been solved analytically on infinite lattices by Onsager [81], but also on finite square lattices by Kaufman [82]. Beale used the latter solution for an algorithm to compute the exact partition function [83]. An implementation in Mathematica is readily available [84], by the means of which the density of states for lattices up to a size of 128×128 can be obtained, see Fig. B.1(a). From the density of states exact expectation values of thermodynamic quantities such as the specific heat capacity are easily calculated at any temperature. To reduce notational overhead, in the following units with $k_\mathrm{B} = 1$ and $J = 1$ will be assumed.

On the infinite lattice the two-dimensional Ising model has a continuous phase transition at the critical point $\beta_c = 1/T_c = 1/2 \cdot \ln(1 + \sqrt{2}) \approx 0.44$. There the specific heat



(a) Logarithm of the density of states

(b) Specific heat capacity

Figure B.1: Density of states and specific heat capacity for the 2D Ising model on a $N = 128 \times 128$ square lattice, computed from the Beale solution

diverges. Due to finite-size effects of the $N = 128^2$ system the discontinuity is rounded into the peak seen in Fig. B.1(b). Close to this second-order phase transition Monte Carlo simulations suffer from critical slowing down: Autocorrelation times are much longer than at higher or lower temperatures and the statistically effective number of samples is reduced.

B.1 Different estimates for histogram uncertainties

I have carried out canonical Monte Carlo simulations at different temperatures below, close to and above the critical point. The measured data has then been processed by the multiple histogram reweighting routines to estimate the energy density of states. In the simulations a simple Metropolis update with local spin flips has been used. While this is not an ideal choice for high accuracy near the transition temperature, where a cluster update as devised by Swendsen and Wang [37] or Wolff [38] would be vastly superior, it assures that the autocorrelation times of the energy trajectories measured in the different simulations cover a wide range because critical slowing down is not attenuated at all. Arguably, this should highlight eventual differences between the results of the various weighting schemes under examination, which can be discerned by the way autocorrelation times are handled. The results of this analysis for systems of sizes $N = 32^2$ and $N = 128^2$ are discussed in the following.

In the Ising model all accessible energy levels are discrete. On the square lattice they can be enumerated by $E_i = \{ -2N, -2N+8, -2N+12, \ldots, 2N-12, 2N-8, 2N \}$. They are generally spaced $\Delta E = 4$ apart, but the energies $-2N+4$ and $2N-4$ are not possible. Because of this no additional binning is necessary and there is no concern of discretization error.

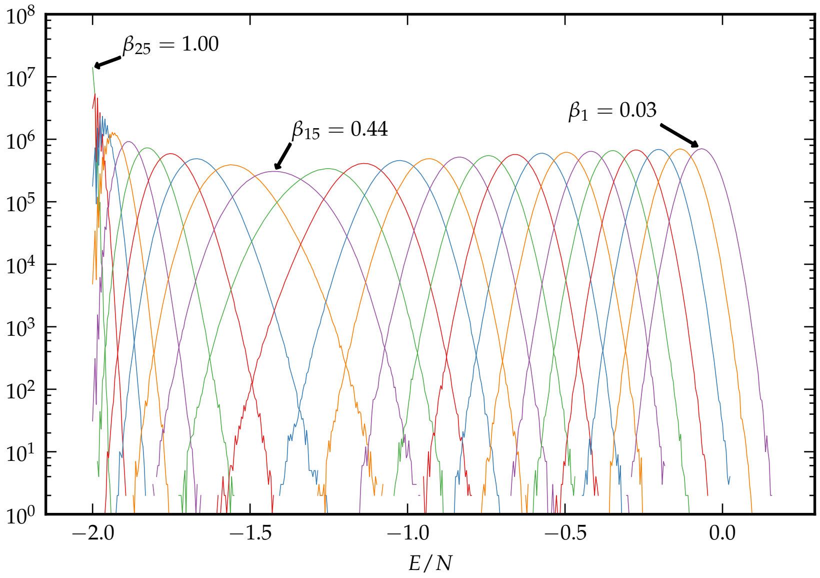


Figure B.2: Energy histograms measured in 25 canonical Monte Carlo simulations of the $N = 32^2$ 2D Ising model at different inverse temperatures ranging from 0.03 to 1.00. Counts are displayed on a logarithmic scale. Color is only used to ease the visual discrimination of overlapping histograms.

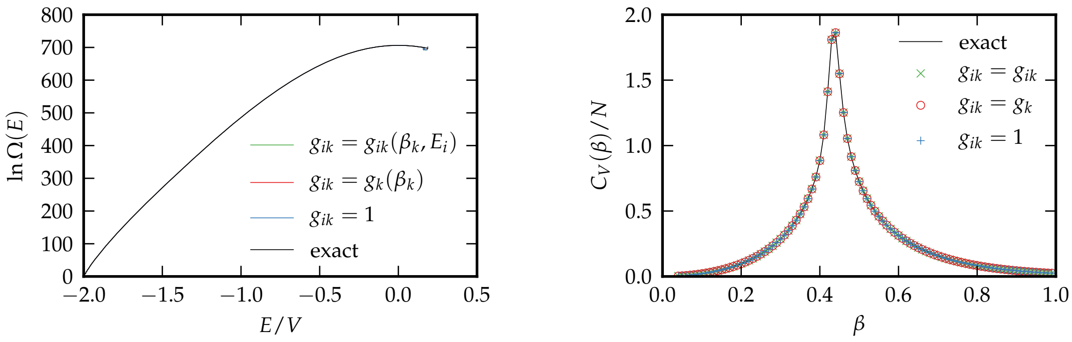
B.1.1 Results for $N = 32^2$

The Ising model on the $N = 32^2$ square lattice has been simulated at 25 different inverse temperatures $\beta_k = 0.03, \ldots, 1.00$ for $2 \cdot 10^7$ Monte Carlo sweeps each. The temperatures have been chosen such that energy histograms at neighboring temperatures have an overlap of $45\% \pm 5\%$. The measured histograms are shown in Fig. B.2.

The range of temperatures is wide enough to have sampled the ferromagnetic phase below the critical temperature T_c starting from the ground state $E_0 = -2N$ as well as the paramagnetic phase above T_c up to energies slightly above zero. For $J > 0$ the expectation value of the energy $\langle E \rangle$ is negative at all temperatures.

Using the collected simulation data, the density of states has been estimated over the sampled range of energies using the multiple histogram analysis method with three different levels of sophistication in the consideration of autocorrelation effects. The same data has been used repeatedly to compare the results of the analysis with the three schemes. Statistical errors were estimated by performing the whole analysis separately on 100 jackknife blocks and calculating the corrected variance of the block results.

The density of states $\Omega(E)$ is only defined up to a factor of proportionality. But since in the Ising model there are exactly two configurations that realize the ground state energy E_0, $\sigma_i \equiv +1$ or $\sigma_i \equiv -1$ for all spins σ_i, one usually sets $\Omega(E_0) = 2$. Because the ground state has been reached in the simulations, the estimated densities of states can be scaled appropriately.



(a) Logarithm of the density of states

(b) Specific heat capacity

Figure B.3: Shown are exact solutions and estimations from the multiple histogram analysis for the $N = 32^2$ Ising model. The error bars of the heat capacity estimates are too small to be visible.

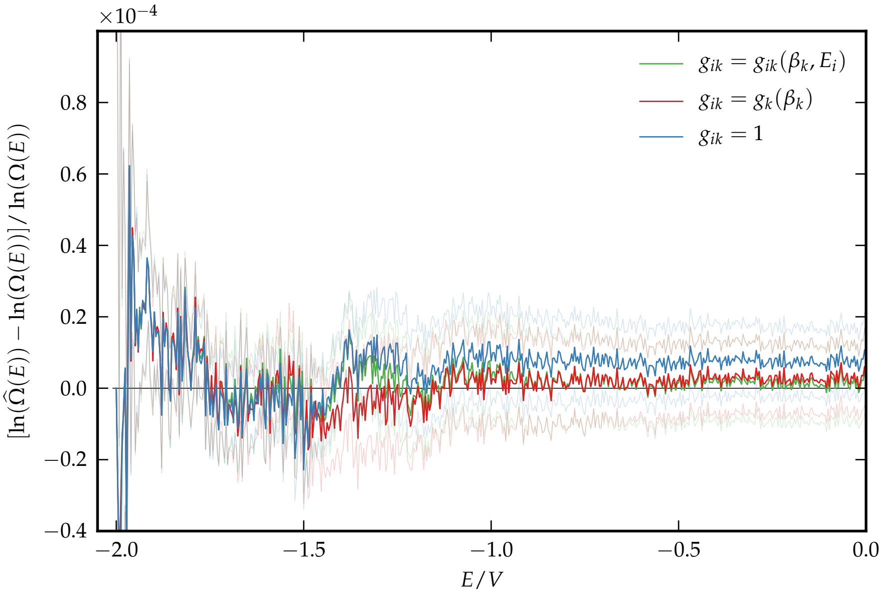


Figure B.4: Relative deviation of the logarithm of the estimated density of states from the exact solution for the $N = 32^2$ Ising model. Faint lines indicate the statistical uncertainty within 1σ.

Regardless of the way autocorrelation times were considered, the estimated density of state matches the exact solution very well. Perceivable deviations only occur at the highest energies, which were hardly visited in the simulations. Additionally the specific heat capacity was calculated at various temperatures different from the original simulation temperatures. It matches the exact solution with high precision in each case. The results are shown in Fig. B.3.

The high resolution of the agreement with the exact solution can be seen in Fig. B.4, where the relative deviation of the estimated logarithm of the density of states $\ln\widehat{\Omega}(E)$ from the exact value $\ln\Omega(E)$ is plotted over the energy. Within statistical error all estimates match and a clear best candidate cannot be made out.

B.1.2 Results for $N = 128^2$

I have repeated the same procedure as in Sec. B.1.1 on the $N = 128^2$ square lattice with 40 different inverse temperatures $\beta_k = 0.36, \ldots, 0.83$. Each simulation has been run for 10^7 Monte Carlo sweeps. Energy histograms at neighboring temperatures have an average overlap of 35%, see Fig. B.5.

While the range of temperatures is now no longer wide enough to have sampled the ground state, the ferromagnetic and paramagnetic phases are still both included within the energy range. For this reason the estimated densities of states have not been normalized by the value at the lowest energy, but by that at the median energy sampled in the simulations, which could still be related to the exact result from Beale's solution.

Again the multiple histogram estimates match the analytical results for the density of states and the heat capacity very well as one can see in Fig. B.6 on p. 126. Also a close look at the relative deviations of the estimated logarithm of the density of states $\ln \widehat{\Omega}(E)$ from the exact value $\ln \Omega(E)$ as shown in Fig. B.7 on p. 126 reveals only slight variations of the quality of the results. Somewhat surprisingly, the conceptually most sophisticated approach of taking autocorrelations into account ($g_{ik} = g_{ik}(\beta_k, E_i)$) moves further from the zero-line than the other two estimates over a wide range of energies. Most likely, these deviations can be attributed to an insufficient accuracy in the numerical estimation of the integrated autocorrelation times for the bin-characteristic functions at low temperatures. Nevertheless, all three estimates match within the statistical uncertainty and these relative differences become visible only on a small scale.

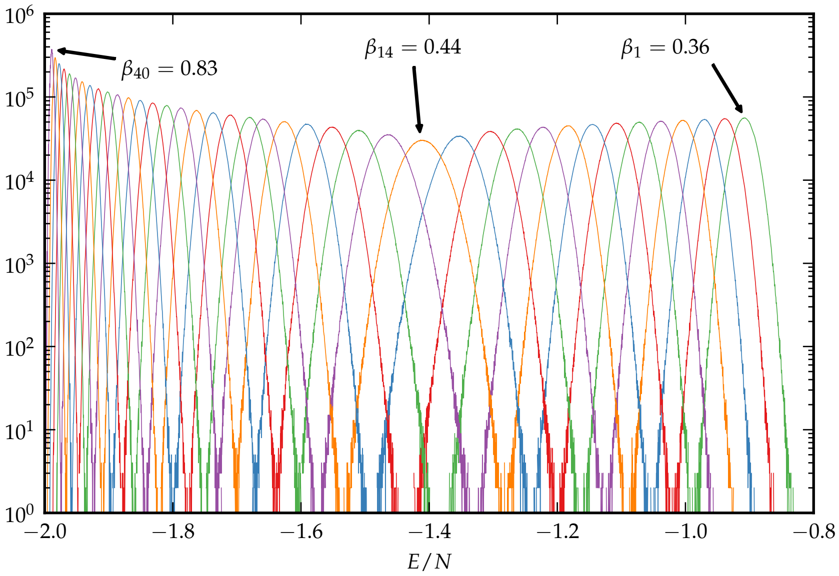
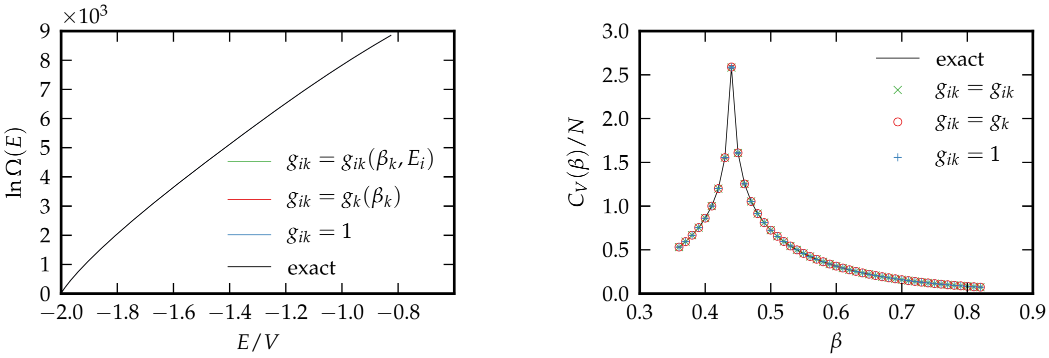


Figure B.5: Energy histograms measured in 40 canonical Monte Carlo simulations of the $N = 128^2$ 2D Ising model at different inverse temperatures ranging from 0.36 to 0.83. Counts are displayed on a logarithmic scale. Color is only used to ease the visual discrimination of overlapping histograms.



(a) Logarithm of the density of states

(b) Specific heat capacity

Figure B.6: Shown are exact solutions and estimations from the multiple histogram analysis for the $N = 128^2$ Ising model. On the left hand plot deviations from the analytically computed curve are smaller than the line width. On the right hand plot error bars are almost invisible.

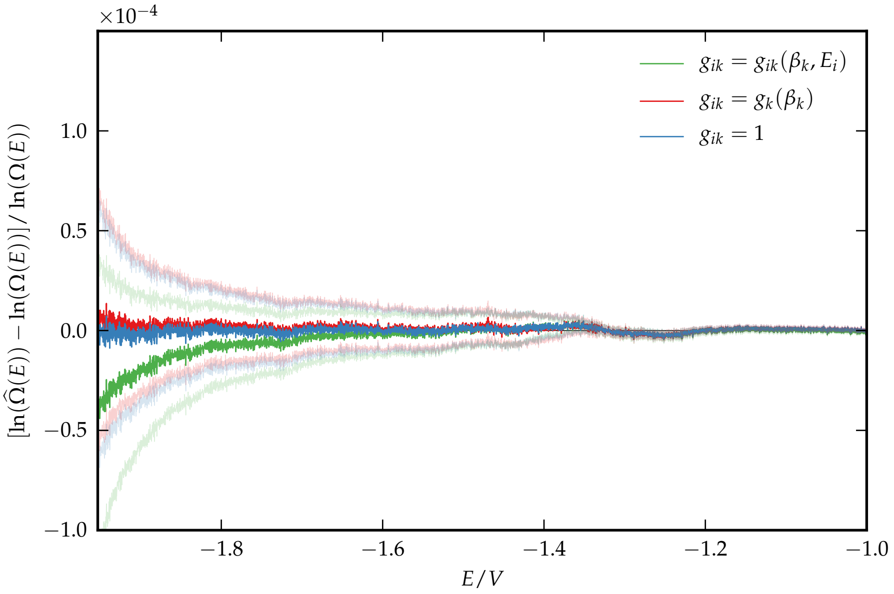


Figure B.7: Relative deviation of the logarithm of the estimated density of states from the exact solution for the $N = 128^2$ Ising model. Faint lines indicate the statistical uncertainty within 1σ.

B.1.3 Statistical uncertainties of histogram entries

The evaluation of the statistical uncertainties of the individual histogram entries has been done according to Eq. (3.87): They can be calculated directly from the expectation values of the histogram occupations $\langle H_{ik} \rangle$ and the statistical inefficiency factors $g_{ik} = 1 + 2\tau_{\mathrm{int},ik}$. To estimate the latter, the integrated autocorrelation times of the bin characteristic functions ψ_{ikm} defined in Eq. (3.74) have been computed by direct summation of their autocorrelation functions. Since each function ψ_{ikm} only has values of 0 and 1, one might think that the behavior of its autocorrelation function could be more problematic than for instance that of the energy E itself, which usually has a much wider range. Fortunately, this can be checked. The squared statistical uncertainties $\delta^2 H_{ik}$ can also be estimated by jackknife resampling the whole time series, building the histograms separately for each block and calculating their corrected variance $\delta^2 \widehat{H}_{ik}^{\mathrm{JK}}$. Then by Eq. (3.87) an estimate for g_{ik} is:

$$\widehat{g}_{ik}^{\mathrm{JK}} = \frac{\delta^2 \widehat{H}_{ik}^{\mathrm{JK}}}{H_{ik}\left(1 - H_{ik}/M_k\right)}, \tag{B.2}$$

for which only the data of a time series of length M_k from a single simulation is used.

In Fig. B.8 on p. 130 two energy histograms are shown that have been measured in the simulations of Sec. B.1.1 at inverse temperatures slightly below and slightly above the pseudo-critical point of the $N = 32^2$ Ising model, which is shifted downwards from $\beta_c \approx 0.44$ due to finite size effects. In the figure they are accompanied by plots of the statistical inefficiencies g_{ik} of each histogram bin from two estimates: $\widehat{g}_{ik}$ by summation of the autocorrelation function and $\widehat{g}_{ik}^{\mathrm{JK}}$ by Eq. (B.2) using 100 jackknife blocks. Considering that especially the latter estimate is rather noisy, the two curves match quite well, which supports the applicability of the integrated autocorrelation times of the bin characteristic functions to determine the weights used in the multiple histogram analysis.

The plots of the statistical inefficiencies in Fig. B.8 on p. 130 show that both the samples of the most probable energies in the peaks of the histograms and those of the least probable energies in the tails are almost uncorrelated, while those of the intermediate energies are more correlated. This leads to a characteristic double peak structure with a central dip. The same can also been observed at temperatures further away from the phase transition point, although the structure is less pronounced since autocorrelation times are generally shorter. See Fig. B.12 on p. 132 for a demonstration with the data for the $N = 128^2$ Ising model from Sec B.1.2. I have also observed this double peak structure in the uncertainties of the energy histograms of the compass model. It can also be seen in a simple model of bivariate Gaussian random variables [32]. Apparently this is a non-model-specific phenomenon.

A qualitative understanding may be found in the observation that if the system is first in a highly probable state at the peak of the energy histogram, it will likely remain there or, since it can only leave the peak for less probable states (be it with higher or lower energy), return to it after a short time. Accordingly, the autocorrelation function

$$C_{ik}(t) = \frac{\langle \psi_{ikm} \psi_{ik,m+t} \rangle - \langle \psi_{ikm} \rangle^2}{\langle {\psi_{ikm}}^2 \rangle - \langle \psi_{ikm} \rangle^2} \tag{B.3}$$

will show a sharp decay for these energy bins i.

The situation is less symmetric for configurations that have energies in the sloped sides of the histogram. The system can assume a less probable state, from which it should return rather quickly, but it is more probable for it to move in the direction of higher probability, from which it will return after a longer wait. In addition to the sharp decay, the autocorrelation function $C_{ik}(t)$ will also be positive for longer gap times t. This leads to larger integrated autocorrelation times.

Finally, the energy levels in the flat tails are very improbable and will only be revisited very rarely and are hence almost uncorrelated.

For illustrative purposes some example histogram bin autocorrelation functions, that were evaluated in the course of the estimation of $\widehat{g}_{ik}$, are shown in Fig. B.9 on p. 130.

The histograms shown in Fig. B.8 are skewed and not Gaussian because they were measured in the vicinity of the pseudo-critical temperature, where spin-spin correlations become important and $p(E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j)$ is no longer determined by the central limit theorem. Consequently, also the distribution of the $\widehat{g}_{ik}$ is not symmetric. The $\widehat{g}_{ik}$-peak lies on the "broader" side of the histogram because for intermediate energies the range of higher-probability energies is more extensive there.

If bin statistical inefficiencies are used to calculate weights for the multiple histogram analysis, their central suppression implies that the majority of sampled configurations, which have energies in the histogram peak region, are counted almost fully. Autocorrelation times have less influence than if only one inefficiency factor g_k is used for all histogram entries measured at an inverse temperature β_k.

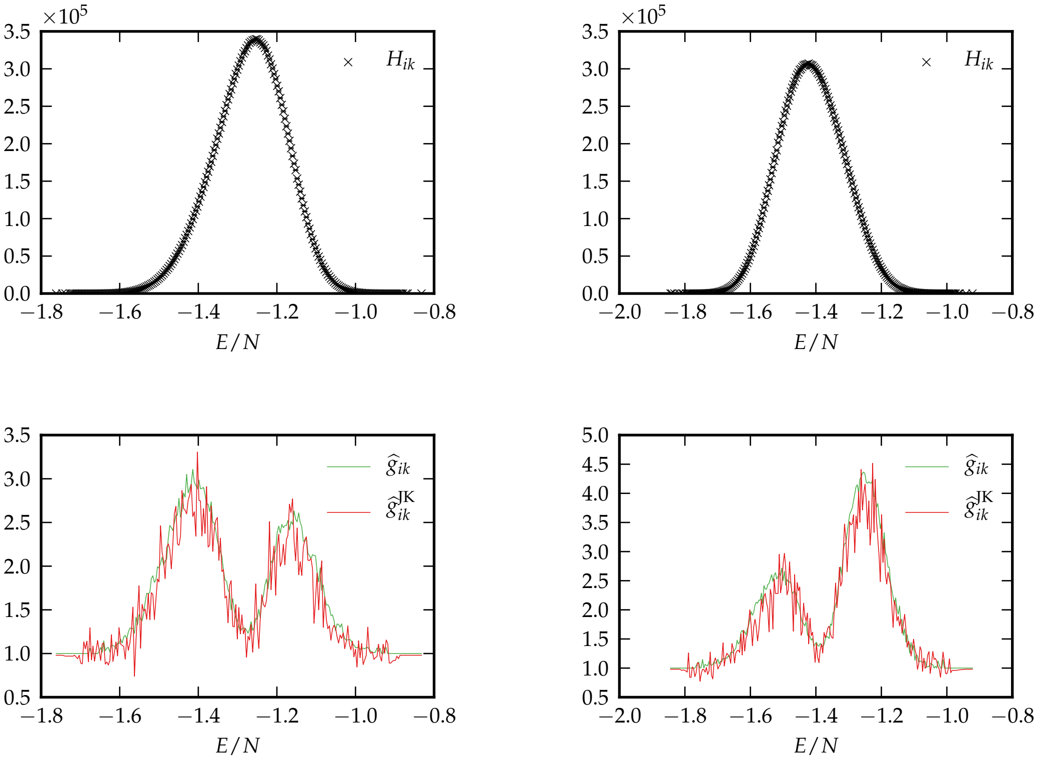
However, if histograms from low temperature simulations are considered, the situation is different. Their main constituent is the ground state energy and the statistical inefficiency of that histogram entry is highest too, see Fig. B.10 on p. 131 for an example. This can be understood from the high asymmetry of the distribution which is cut off at its peak.

To qualitatively compare the effect of statistical inefficiencies $g_{ik}(\beta_k, E_i)$ estimated per bin with inefficiencies estimated for the whole energy time series $g_k(\beta_k) = 1 + 2\tau_{\text{int},E}(\beta_k)$, one can introduce averaged statistical inefficiencies

$$\overline{g}_k = \overline{g}_k(\beta_k) = \frac{\sum_i g_{ik} H_{ik}}{\sum_i H_{ik}} = \frac{\sum_i g_{ik} H_{ik}}{N_k}, \tag{B.4}$$

where the individual bin inefficiencies are weighted by the occupation of each bin. $g_k(\beta_k)$ and $\overline{g}_k(\beta_k)$ are compared in Fig. B.11 on p. 131 for the $N = 32^2$ and $N = 128^2$ Ising model. If g_{ik} or g_k are used in the multiple histogram analysis, in both cases significantly less weight is given to the samples from the simulations close to the critical point. However, the ratio between the highest and lowest inefficiency-factors is larger when autocorrelation times are considered per temperature only. Moreover in the $N = 32^2$ case, where the ground state is reached in the low temperatures, the averaged statistical inefficiencies are also high there.

It should be noted that only the ratio of the statistical inefficiencies is significant and that their absolute values do not play a role in the multiple histogram analysis. Furthermore their ratio is only important for energy bins where histogram overlap is large.



(a) $\beta \approx 0.4237$ **(b)** $\beta \approx 0.4386$

Figure B.8: Top row: Energy histograms of the $N = 32^2$ Ising model measured at two different temperatures, error bars from the jackknife analysis. Bottom row: Corresponding statistical inefficiencies for each histogram bin

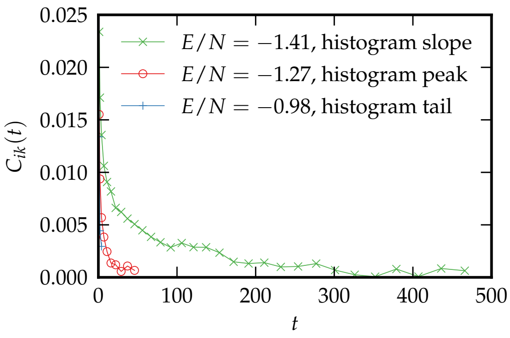


Figure B.9: Numerically evaluated autocorrelation functions that went into the evaluation of $\widehat{g}_{mk}$ in Fig. B.8(a) ($N = 32^2$, $\beta \approx 0.4237$), for selected energy bins

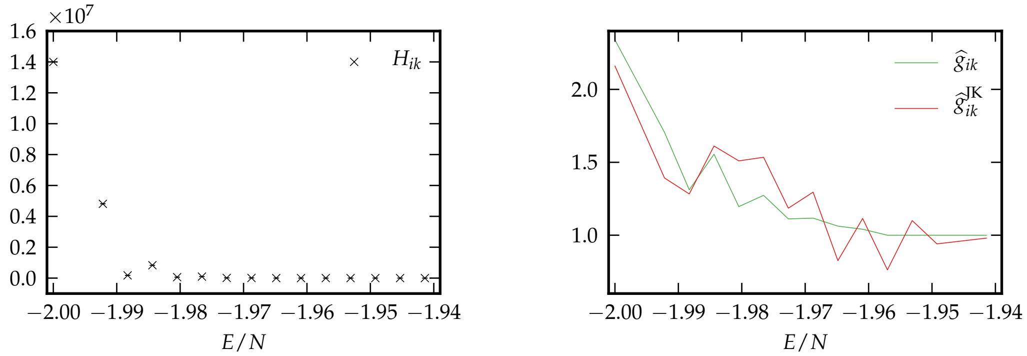
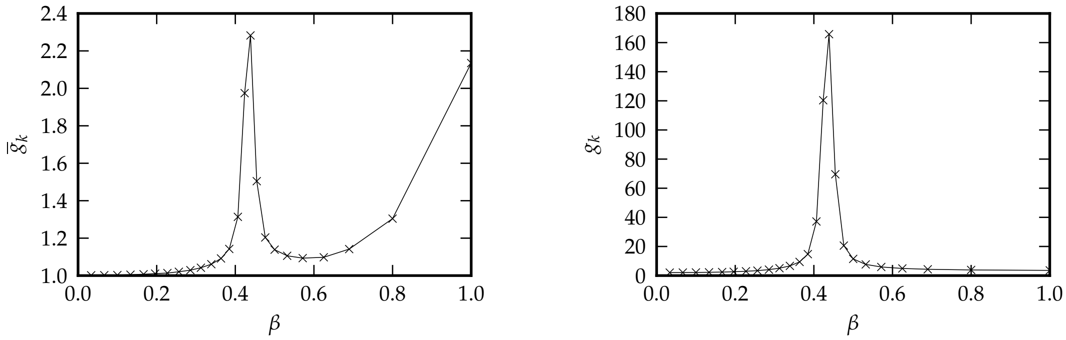
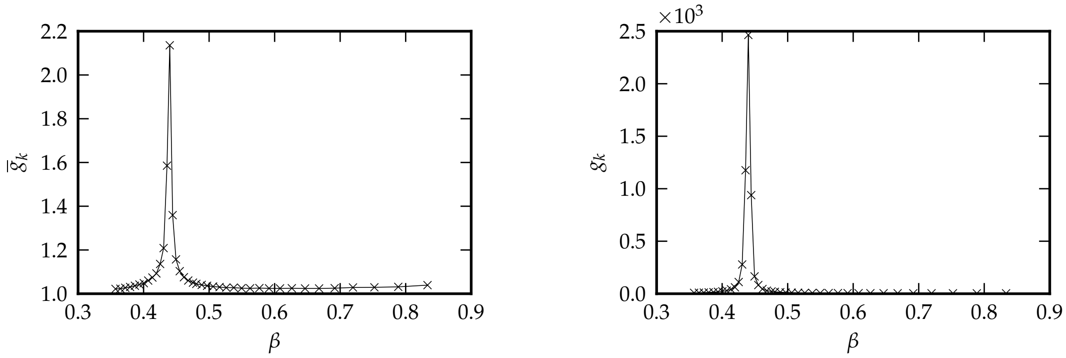


Figure B.10: Energy histogram of the $N = 32^2$ Ising model measured at $\beta = 1.0$ and statistical inefficiencies for each histogram bin.



(a) $N = 32^2$, energy bin statistical inefficiencies, averaged

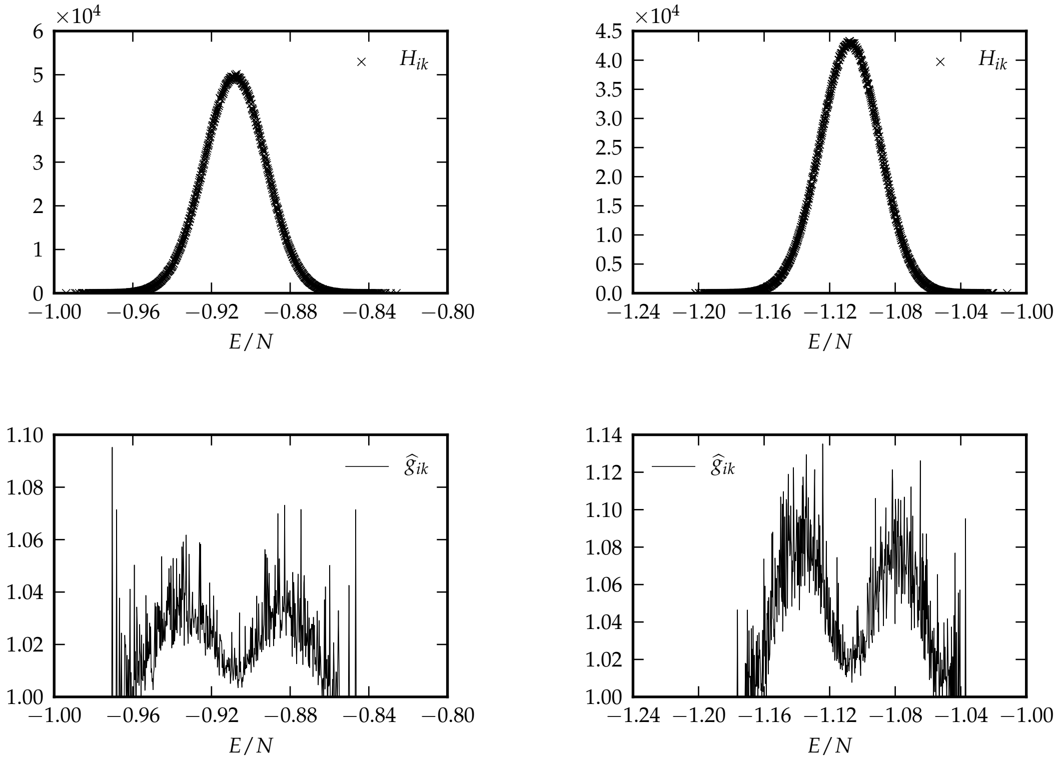
(b) $N = 32^2$, statistical inefficiencies of the energy



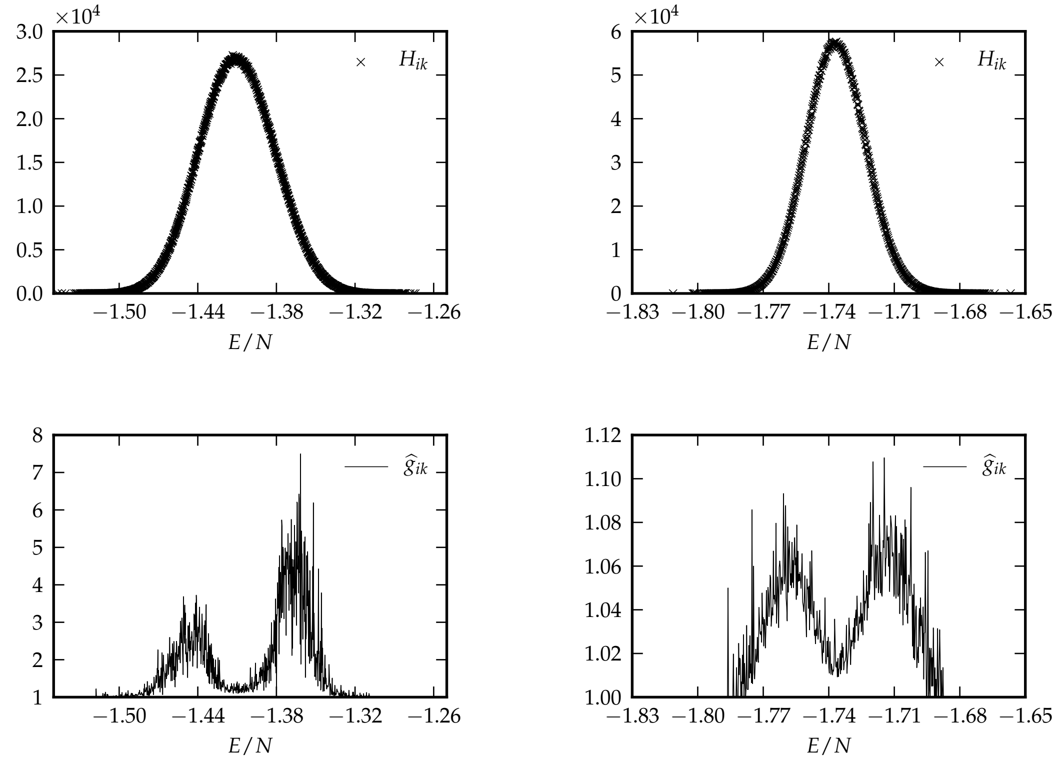
(c) $N = 128^2$, energy bin statistical inefficiencies, averaged

(d) $N = 128^2$, statistical inefficiencies of the energy

Figure B.11: Comparison of different measures for statistical inefficiencies in the 2D Ising model. Top row: $N = 32^2$. Bottom row: $N = 128^2$.



(a) $\beta \approx 0.3571$ **(b)** $\beta \approx 0.4003$



(c) $\beta \approx 0.4400$ **(d)** $\beta \approx 0.4969$

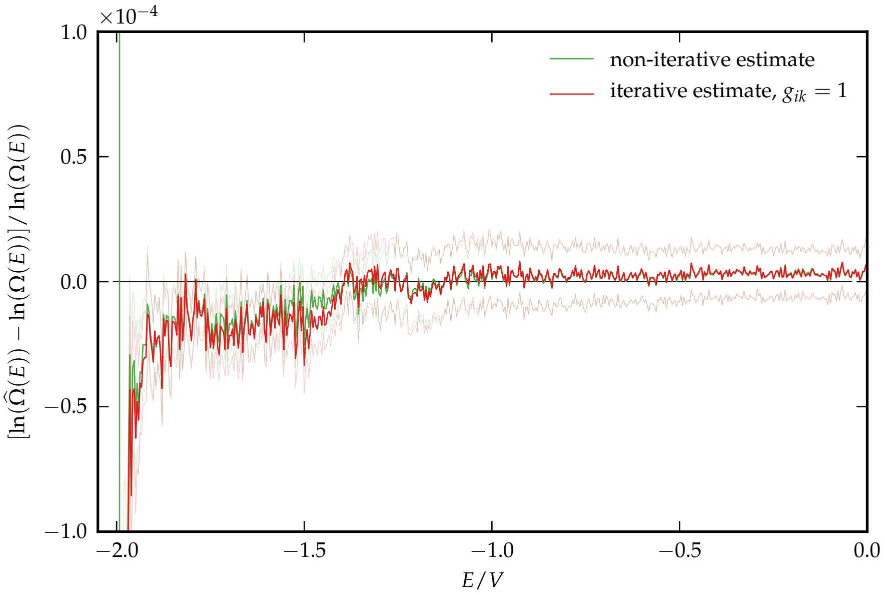
Figure B.12: Top rows: Energy histograms of the $N = 128^2$ Ising model measured at different temperatures. Bottom rows: Corresponding statistical inefficiencies for each histogram bin. Note the differing scales.

B.2 Non-iterative and iterative estimation of the density of states

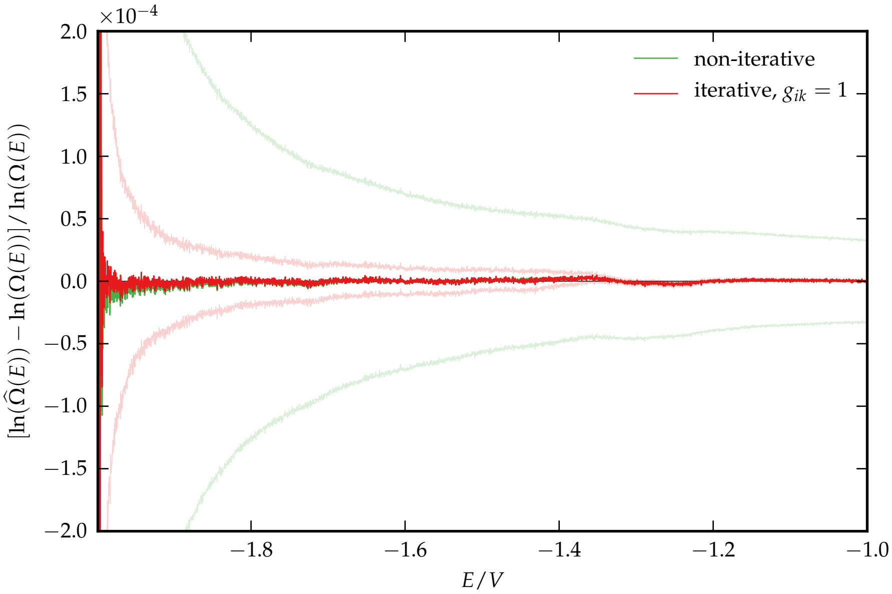
Fenwick's estimator for the density of states described in Sec. 3.6.2 does not require the iterative solution of a set of equations. To see whether it represents a viable alternative to the traditional approach outlined in Sec. 3.6.1, I have applied it also to the simulation data used in Sec. B.1. Since in my implementation the non-iterative method does not distinguish varying statistical inefficiencies, I have compared it only to the iterative estimator with all inefficiency factors set to unit: $g_{ik} \equiv 1$.

Here only plots of the relative deviation of the logarithm of the estimated density of states $\ln \widehat{\Omega}(E_i)$ from the exact solution $\ln \Omega(E_i)$ are shown, which highlight the differences of the results of the two schemes of analysis. Since the estimate for the density of states at the lowest energy computed by the non-iterative method is of rather low quality, the estimated densities of states for both $N = 32^2$ and $N = 128^2$ have been normalized by the central entry.

The results for $N = 32^2$ shown in Fig. B.13(a) on p. 134 are of comparable quality for both methods. For $N = 128^2$, however, the results obtained with the non-iterative method have a much higher statistical uncertainty over a wide range of energies, see Fig. B.13(b). Apparently the non-iterative approach is more vulnerable to lower quality statistics at the lower energies. These deficiencies are propagated to the higher energies by construction of the estimate $\widehat{\Omega}_i$. In contrast to this, the iterative method only takes into account data from the single energy bin E_i to estimate $\widehat{\Omega}_i$.



(a) $N = 32^2$



(b) $N = 128^2$

Figure B.13: Shown is the relative deviation of the logarithm of the estimated density of states from the exact solution for the Ising model with $N = 32^2$ and $N = 128^2$ for both the iterative and the non-iterative method. Faint lines indicate the statistical uncertainty within 1σ.

B.3 Parallel tempering

To get an impression of the applicability of the multiple histogram reweighting method for data generated in parallel tempering simulations, I have also simulated the Ising model on the $N = 16^2$ and $N = 32^2$ square lattices with replica exchange updates in addition to canonical Metropolis updates. 25 replicas have been simulated for $2 \cdot 10^7$ Monte Carlo sweeps each and exchanges between next neighbor temperatures have been proposed every 10 sweeps. A narrower range of inverse temperatures than in Sec. B.1.1 has been chosen to facilitate exchanges of temperatures, with β_l taking values between 0.3 and 0.5. The same range has been used for both lattice sizes.

Histograms of the energies measured in the simulations after they have been sorted by temperatures are shown in Fig. B.14 on p. 136; the broadening of the distributions close to the critical point is clearly visible. In contrast to this the energy histograms shown in Fig. B.15 on p. 136, which are taken directly from the replica time series without prior sorting by temperatures, are narrower. Apparently the replicas have not had the chance to significantly change their system configurations before the temperature exchanges. If each replica had explored large portions of configuration space as it would happen in a true random walk through the temperatures, the replica histograms would actually span wider ranges of energies than the temperature-sorted histograms.

The multiple histogram reweighting method has been applied on the measured data in different ways:

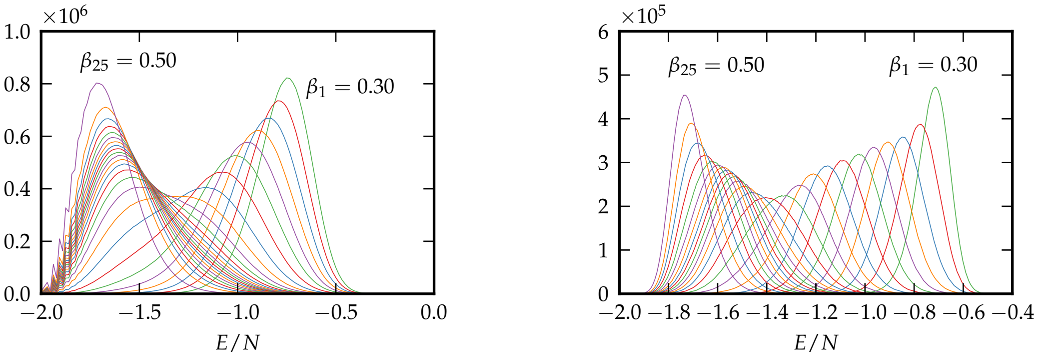
(i) with time series of energy measurements pre-ordered by temperature, which have then been treated as if obtained from independent canonical simulations with the method of Sec. 3.6.1 with autocorrelation times estimated for each energy bin,

(ii) with pre-ordered time series, but ignoring statistical inefficiencies, and

(ii) by applying the method outlined in Sec. 3.6.3 directly on the replica time series, again with autocorrelation times estimated for each energy bin.

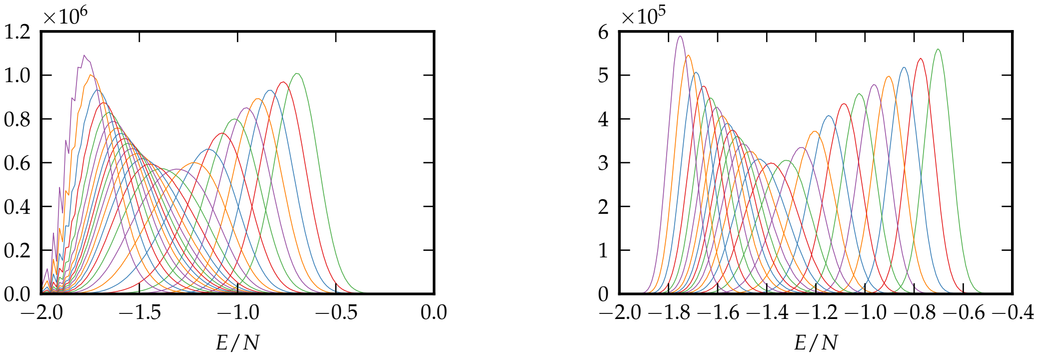
Fig. B.16 on p. 137 shows the relative deviation of the estimated logarithm of the density of states $\ln \widehat{\Omega}(E_i)$ from the exact result $\ln \Omega(E_i)$ for cases (i) to (iii) and the two lattice sizes. $\ln \Omega(E_i)$ has always been normalized at the central energy bin. In this example the results of approach (iii), that was specially designed to respect the correlations in parallel tempering time series, deviate from the exact result more strongly than the results of approach (i). But also there the results are worse than those obtained with the most naive approach (ii). Apparently neither approach (i) nor (iii) have correctly captured the optimal statistical weighting of the histograms.

A general decision on the preferable method cannot be made directly from these findings. It would be necessary to carry out a comparative examination with simulations set up in such a way that the single replicas explore broader portions of configuration space.



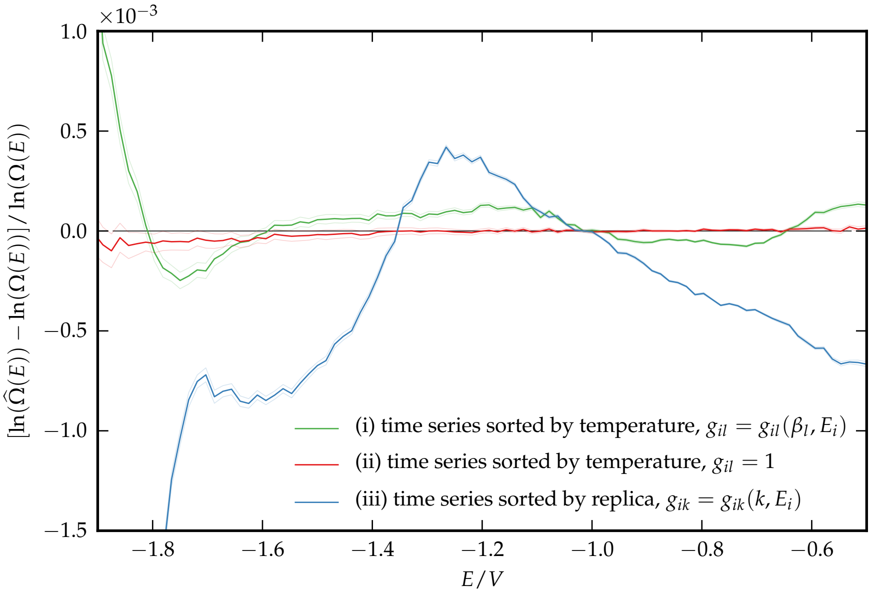
(a) $N = 16^2$ **(b)** $N = 32^2$

Figure B.14: Energy histograms of the 2D Ising model from parallel tempering simulations. The data from all replicas has been collected and sorted by temperatures

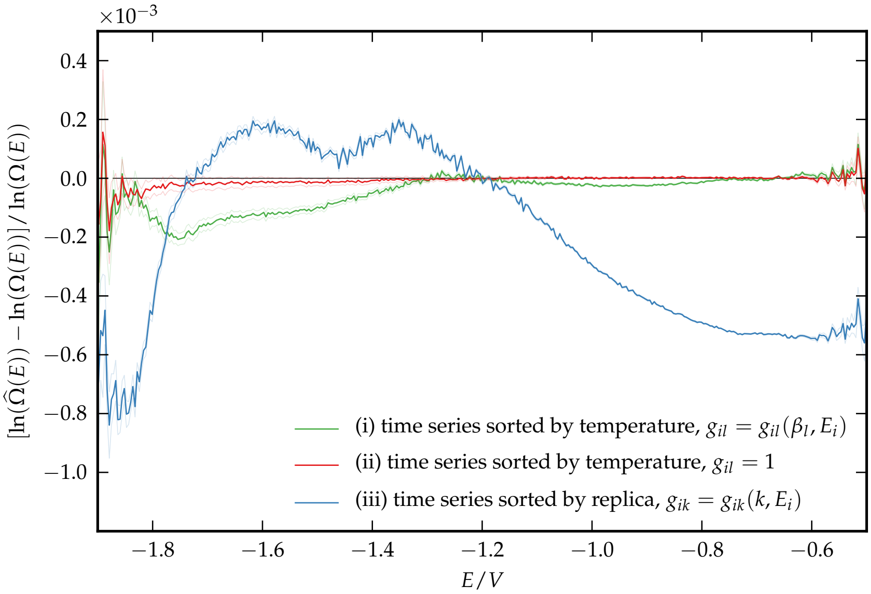


(a) $N = 16^2$ **(b)** $N = 32^2$

Figure B.15: Energy histograms of individual replicas of parallel tempering simulations of the 2D Ising model, not sorted by temperatures



(a) $N = 16^2$



(b) $N = 32^2$

Figure B.16: Shown is the relative deviation of the logarithm of the estimated density of states from the exact solution for the Ising model on the $N = 16^2$ and $N = 32^2$ square lattices with data obtained in parallel tempering simulations. Faint lines indicate the statistical uncertainty within 1σ.

References

[1] D. I. Khomskii and M. V. Mostovoy, "Orbital ordering and frustrations," J. Phys. A: Math. Gen. **36**, 9197 (2003).

[2] J. van den Brink, "Orbital-only models: Ordering and excitations," New J. Phys. **6**, 201 (2004).

[3] K. I. Kugel and D. I. Khomskii, "The Jahn-Teller effect and magnetism: Transition metal compounds," Sov. Phys. Usp. **25**, 231 (1982).

[4] C. Batista and Z. Nussinov, "Generalized Elitzur's theorem and dimensional reductions," Phys. Rev. B **72**, 045137 (2005).

[5] B. Douçot, M. Feigel'man, L. Ioffe and A. S. Ioselevich, "Protected qubits and Chern-Simons theories in Josephson junction arrays," Phys. Rev. B **71**, 024505 (2005).

[6] P. Milman, W. Maineult, S. Guibal, L. Guidoni, B. Douçot, L. Ioffe and T. Coudreau, "Topologically decoherence-protected qubits with trapped ions," Phys. Rev. Lett. **99**, 020503 (2007).

[7] S. Gladchenko, D. Olaya, E. Dupont-Ferrier, B. Doucot, L. B. Ioffe and M. E. Gershenson, "Superconducting nanocircuits for topologically protected qubits," Nat. Phys. **5**, 48 (2009).

[8] A. Mishra, M. Ma, F.-C. Zhang, S. Guertler, L.-H. Tang and S. Wan, "Directional ordering of fluctuations in a two-dimensional compass model," Phys. Rev. Lett. **93**, 207201 (2004).

[9] S. Wenzel and W. Janke, "Monte Carlo simulations of the directional-ordering transition in the two-dimensional classical and quantum compass model," Phys. Rev. B **78**, 064402 (2008).

[10] S. Wenzel, W. Janke and A. M. Läuchli, "Re-examining the directional-ordering transition in the compass model with screw-periodic boundary conditions," Phys. Rev. E **81**, 066702 (2010).

[11] J. Oitmaa and C. Hamer, "Quantum compass model on the square and simple-cubic lattices," Phys. Rev. B **83**, 094437 (2011).

[12] F. Schwabl, *Statistische Mechanik* (Springer-Verlag, Berlin, 2006).

[13] H. Haug, *Statistische Physik* (Springer-Verlag, Berlin, 2006).

[14] T. Fließbach, *Statistische Physik* (Spektrum Akademischer Verlag, Heidelberg, 2010).

[15] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971).

[16] W. Janke, “First-order phase transitions,” in *Computer Simulations of Surfaces and Interfaces, NATO Science Series II: Mathematics, Physics and Chemistry*, vol. 114, edited by B. Dünweg, D. P. Landau and A. I. Milchev (Kluwer Academic Publishers, Dordrecht, 2003), pp. 111–135.

[17] Z. Nussinov and E. Fradkin, “Discrete sliding symmetries, dualities, and self-dualities of quantum orbital compass models and p+ip superconducting arrays,” Phys. Rev. B **71**, 195120 (2005).

[18] S. Wenzel and A. M. Läuchli, “Monte Carlo study of the critical properties of the three-dimensional 120-degree model,” J. Stat. Mech. **2011**, P09010 (2011).

[19] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods In Statistical Physics* (Clarendon Press, Oxford, 1999).

[20] B. A. Berg, *Markov chain Monte Carlo simulations and their statistical analysis* (World Scientific, Singapore, 2004).

[21] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2009).

[22] W. Krauth, *Statistical Mechanics: Algorithms And Computations* (Oxford University Press, Oxford, 2006).

[23] W. Janke, “Monte Carlo methods in classical statistical physics,” in *Computational Many-Particle Physics, Lect. Notes Phys.*, vol. 739, edited by H. Fehske, R. Schneider and A. Weiße (Springer-Verlag, Berlin, 2008), pp. 79–140.

[24] H. G. Katzgraber, “Introduction to Monte Carlo methods,” (2009), `arXiv:cond-mat.stat-mech/0905.1629`.

[25] K. Rummukainen, “Simulation methods in physics,” (2003), URL `http://www.helsinki.fi/~rummukai/simu/`.

[26] S. Weinzierl, “Introduction to Monte Carlo methods,” (2000), `arXiv:hep-ph/0006269`.

[27] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, New York, 2007).

[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, “Equation of state calculations by fast computing machines,” J. Chem. Phys. **21**, 1087 (1953).

[29] H. Müller-Krumbhaar and K. Binder, “Dynamic properties of the Monte Carlo method in statistical mechanics,” J. Stat. Phys. **8**, 1 (1973).

[30] W. Janke, "Statistical analysis of simulations: Data correlations and error estimation," in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, NIC Series*, vol. 10, edited by J. Grotendorst, D. Marx and A. Muramatsu (John von Neunmann Institute for Computing, Jülich, 2002), pp. 423–445.

[31] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, 1982).

[32] W. Janke, personal communication.

[33] W. Lenz, "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern," Phys. Z. **21**, 613 (1920).

[34] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," Z. Phys. **31**, 253 (1925).

[35] M. A. Miller, L. M. Amon and W. P. Reinhardt, "Should one adjust the maximum step size in a Metropolis Monte Carlo simulation?" Chem. Phys. Lett. **331**, 278 (2000).

[36] R. H. Swendsen, "How the maximum step size in Monte Carlo simulations should be adjusted," Phys. Procedia **15**, 81 (2011).

[37] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," Phys. Rev. Lett. **58**, 86 (1987).

[38] U. Wolff, "Collective Monte Carlo updating for spin systems," Phys. Rev. Lett. **62**, 361 (1989).

[39] J. Plascak, A. Ferrenberg and D. P. Landau, "Cluster hybrid Monte Carlo simulation algorithms," Phys. Rev. E **65**, 1 (2002).

[40] A. Ferrenberg, D. Landau and Y. Wong, "Monte Carlo simulations: Hidden errors from "good" random number generators," Phys. Rev. Lett. **69**, 3382 (1992).

[41] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov and P. N. Vorontsov-Velyaminov, "New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles," J. Chem. Phys. **96**, 1776 (1992).

[42] E. Marinari and G. Parisi, "Simulated tempering: A new Monte Carlo scheme," Europhys. Lett. **19**, 451 (1992).

[43] R. Swendsen and J. Wang, "Replica Monte Carlo simulation of spin glasses," Phys. Rev. Lett. **57**, 2607 (1986).

[44] C. J. Geyer, "Markov chain Monte Carlo maximum likelihood," in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, 1991), pp. 156–163, URL `http://users.stat.umn.edu/geyer/f05/8931/c.pdf`.

[45] K. Hukushima and K. Nemoto, "Exchange Monte Carlo method and application to spin glass simulations," J. Phys. Soc. Jpn. **65**, 1604 (1996).

[46] C. Predescu, M. Predescu and C. V. Ciobanu, “The incomplete beta function law for parallel tempering sampling of classical canonical systems,” J. Chem. Phys. **120**, 4119 (2004).

[47] C. Predescu, M. Predescu and C. V. Ciobanu, “On the efficiency of exchange in parallel tempering monte carlo simulations,” J. Phys. Chem. B **109**, 4189 (2005).

[48] D. A. Kofke, “On the acceptance probability of replica-exchange Monte Carlo trials,” J. Chem. Phys. **117**, 6911 (2002).

[49] D. A. Kofke, “Erratum: "On the acceptance probability of replica-exchange Monte Carlo trials" [J. Chem. Phys. 117, 6911 (2002)],” J. Chem. Phys. **120**, 10852 (2004).

[50] A. Kone and D. A. Kofke, “Selection of temperature intervals for parallel-tempering simulations,” J. Chem. Phys. **122**, 206101 (2005).

[51] N. Rathore, M. Chopra and J. J. de Pablo, “Optimal allocation of replicas in parallel tempering simulations,” J. Chem. Phys. **122**, 024111 (2005).

[52] S. Trebst, D. Huse and M. Troyer, “Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations,” Phys. Rev. E **70**, 046701 (2004).

[53] H. G. Katzgraber, S. Trebst, D. A. Huse and M. Troyer, “Feedback-optimized parallel tempering Monte Carlo,” J. Stat. Mech. **2006**, P03018 (2006).

[54] S. Trebst, M. Troyer and U. H. E. Hansmann, “Optimized parallel tempering simulations of proteins,” J. Chem. Phys. **124**, 174903 (2006).

[55] W. Nadler and U. H. E. Hansmann, “Generalized ensemble and tempering simulations: a unified view,” Phys. Rev. E **75**, 026109 (2007).

[56] F. Hamze, N. Dickson and K. Karimi, “Robust parameter selection for parallel tempering,” Int. J. Mod. Phys. C **21**, 603 (2010).

[57] E. Bittner, A. Nußbaumer and W. Janke, “Make life simple: Unleash the full power of the parallel tempering algorithm,” Phys. Rev. Lett. **101**, 130603 (2008).

[58] D. Sabo, M. Meuwly, D. L. Freeman and J. D. Doll, “A constant entropy increase model for the selection of parallel tempering ensembles,” J. Chem. Phys. **128**, 174109 (2008).

[59] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth and F. Rossi, *GNU Scientific Library: Reference Manual* (Network Theory Ltd., 2003), URL `http://www.gnu.org/software/gsl/`.

[60] A. M. Ferrenberg and R. H. Swendsen, “Optimized Monte Carlo data analysis,” Phys. Rev. Lett. **63**, 1195 (1989).

[61] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method,” J. Comput. Chem. **13**, 1011 (1992).

[62] A. Ferrenberg and R. Swendsen, "New Monte Carlo technique for studying phase transitions," Phys. Rev. Lett. **61**, 2635 (1988).

[63] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok and K. A. Dill, "Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations," J. Chem. Theory Comput. **3**, 26 (2007).

[64] E. Gallicchio, M. Andrec, A. K. Felts and R. M. Levy, "Temperature weighted histogram analysis method, replica exchange and transition paths," J. Phys. Chem. B **109**, 6722 (2005).

[65] C. Bartels and M. Karplus, "Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations," J. Comput. Chem. **18**, 1450 (1997).

[66] M. K. Fenwick, "A direct multiple histogram reweighting method for optimal computation of the density of states," J. Chem. Phys. **129**, 125106 (2008).

[67] R. P. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, Englewood Cliffs, 1973), chap. 5.

[68] E. Gabriel, G. E. Fagg, G. Bosilca, T. Angskun, J. J. Dongarra, J. M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R. H. Castain, D. J. Daniel, R. L. Graham and T. S. Woodall, "Open MPI: Goals, concept, and design of a next generation MPI implementation," in *Proceedings, 11th European PVM/MPI Users' Group Meeting* (Budapest, Hungary, 2004), pp. 97–104, URL `http://www.open-mpi.org`.

[69] W. Janke, "Pseudo random numbers: generation and quality checks," in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, *NIC Series*, vol. 10, edited by J. Grotendorst, D. Marx and A. Muramatsu (John von Neunmann Institute for Computing, Jülich, 2002), pp. 447–458.

[70] H. G. Katzgraber, "Random numbers in scientific computing: An introduction," (2010), `arXiv:physics.comp-ph/1005.4117v1`.

[71] M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator," ACM Trans. Model. Comput. Simul. **8**, 3 (1998).

[72] M. Saito and M. Matsumoto, "SIMD-oriented fast Mersenne Twister: A 128-bit pseudorandom number generator," in *Monte Carlo and Quasi-Monte Carlo Methods 2006*, edited by A. Keller, S. Heinrich and H. Niederreiter (Springer-Verlag, Berlin, 2008), pp. 607–622.

[73] M. Saito and M. Matsumoto, "A uniform real random number generator obeying the IEEE 754 format using an affine transition," in *Abstracts of 8th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC '08)* (Montreal, 2008), p. 151, URL `http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/SFMT/`.

[74] L. Dagum and R. Menon, "OpenMP: An industry-standard API for shared-memory programming," IEEE Comput. Sci. Eng. **5**, 46 (1998).

[75] T. Bereau and R. H. Swendsen, "Optimized convergence for multiple histogram analysis," J. Comput. Phys. **228**, 6119 (2009).

[76] B. A. Berg and T. Neuhaus, "Multicanonical algorithms for first order phase transitions," Phys. Lett. B **267**, 249 (1991).

[77] B. A. Berg and T. Neuhaus, "Multicanonical ensemble: A new approach to simulate first-order phase transitions," Phys. Rev. Lett. **68**, 9 (1992).

[78] T. Neuhaus and J. Hager, "Free-energy calculations with multiple Gaussian modified ensembles," Phys. Rev. E **74**, 036702 (2006).

[79] S. Wenzel and A. M. Läuchli, "Unveiling the nature of three-dimensional orbital ordering transitions: The case of e_g and t_{2g} models on the cubic lattice," Phys. Rev. Lett. **106**, 197201 (2011).

[80] M. E. Fisher, "Magnetism in one-dimensional systems—the Heisenberg model for infinite spin," Am. J. Phys. **32**, 343 (1964).

[81] L. Onsager, "Crystal statistics. I. A two-dimensional model with an order-disorder transition," Phys. Rev. **65**, 117 (1944).

[82] B. Kaufman, "Crystal statistics. II. Partition function evaluated by spinor analysis," Phys. Rev. **76**, 1232 (1949).

[83] P. D. Beale, "Exact distribution of energies in the two-dimensional Ising model." Phys. Rev. Lett. **76**, 78 (1996).

[84] P. D. Beale, "Mathematica program for the exact calculation of the partition function of the two-dimensionalIsing model," URL `http://spot.colorado.edu/~beale/MathematicaFiles/`.

Danksagung

Zunächst möchte ich mich bei Herrn Prof. Wolfhard Janke für die Betreuung während meiner Diplomarbeit und die Möglichkeit, in einer Abteilung mit ausgezeichneter Arbeitsatmosphäre zu arbeiten, bedanken. Er erlaubte mir eine sehr selbstständige Arbeitsweise und fand dabei die Zeit für viele Anregungen und Ratschläge bei Problemen.

Für viele fruchtbare Diskussionen zu physikalischen und technischen Fragestellungen danke ich meinen Kommilitonen und Kollegen am ITP, insbesondere Hannes Nagel, Johannes Zierenberg, Marco Müller, Martin Marenz, Martin Treffkorn, Momchil Ivanov, Monika Möddel, Norma Kühn und Rüdiger Kürsten.

Hannes Nagel und Martin Marenz, die sich während meiner Bearbeitungszeit um die Computersysteme am Institut gekümmert haben, gebührt mein ausdrücklicher Dank. Bei Problemen und Sonderwünschen konnten sie schnell und unkompliziert helfen.

Meinen Eltern, die mir das Studium ermöglicht haben, stets hinter mir standen und mir bei zahlreichen Gelegenheiten moralische Unterstützung gaben, danke ich besonders.

Schließlich geht mein großer Dank an Ellen Hahne für ihren andauernden Rückhalt und viel Verständnis für lange Arbeitsabende in der letzten Zeit.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die Diplomarbeit selbständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche kenntlich gemacht. Ferner erkläre ich, dass die Arbeit noch nicht in einem anderen Studiengang als Prüfungsleistung verwendet wurde.

Max Henner Gerlach

Ich bin einverstanden, dass die Arbeit nach positiver Begutachtung in der Universitätsbibliothek zur Verfügung steht.

Max Henner Gerlach