4 Monte Carlo Methods in Classical Statistical Physics

Wolfhard Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences, Universität Leipzig, 04009 Leipzig, Germany

The purpose of this chapter is to give a brief introduction to Monte Carlo simulations of classical statistical physics systems and their statistical analysis. To set the general theoretical frame, first some properties of phase transitions and simple models describing them are briefly recalled, before the concept of importance sampling Monte Carlo methods is introduced. The basic idea is illustrated by a few standard local update algorithms (Metropolis, heat-bath, Glauber). Then methods for the statistical analysis of the thus generated data are discussed. Special attention is payed to the choice of estimators, autocorrelation times and statistical error analysis. This is necessary for a quantitative description of the phenomenon of critical slowing down at continuous phase transitions. For illustration purposes, only the two-dimensional Ising model will be needed. To overcome the slowing-down problem, non-local cluster algorithms have been developed which will be described next. Then the general tool of reweighting techniques will be explained which is extremely important for finite-size scaling studies. This will be demonstrated in some detail by the sample study presented in the next section, where also methods for estimating spatial correlation functions will be discussed. The reweighting idea is also important for a deeper understanding of so-called generalized ensemble methods which may be viewed as dynamical reweighting algorithms. After first discussing simulated and parallel tempering methods, finally also the alternative approach using multicanonical ensembles and the Wang-Landau recursion are briefly outlined.

4.1 Introduction

Classical statistical physics is a well understood subject which poses, however, many difficult problems when a concrete solution for interacting systems is sought. In almost all non-trivial applications, analytical methods can only provide approximate answers. Numerical computer simulations are, therefore, an important complementary method on our way to a deeper understanding of complex physical systems such as (spin) glasses and disordered magnets or of biologically motivated problems such as protein folding. Quantum statistical problems in condensed matter or the broad field of elementary particle physics and quantum gravity are other major applications which, after suitable mappings, also rely on classical simulation techniques.

In these lecture notes we shall confine ourselves to a survey of computer simu-
lations based on Markov chain Monte Carlo methods which realize the importance
sampling idea. Still, not all aspects can be discussed in these notes in detail, and for
further reading the reader is referred to recent textbooks [1, 2, 3, 4], where some
of the material is presented in more depth. For illustration purposes, here we shall
focus on the simplest spin models, the Ising and Potts models. From a theoretical
point of view, also spin systems are still of current interest since they provide the
possibility to compare completely different approaches such as field theory, series
expansions, and simulations. They are also the ideal testing ground for general con-
cepts such as universality, scaling or finite-size scaling, where even today some new
features can still be discovered. And last but not least, they have found a revival in
slightly disguised form in quantum gravity and and random network theory, where
they serve as idealized matter fields on Feynman diagrams or fluctuating graphs.

This chapter is organized as follows. In Sect. 4.2, first the definition of the stan-
dard Ising model is recalled and the most important observables (specific heat, mag-
netization, susceptibility, correlation functions, . . .) are briefly discussed. Next some
characteristic properties of phase transitions, their scaling properties, the definition
of critical exponents and finite-size scaling are briefly summarized. In Sect. 4.3,
the basic method underlying all importance sampling Monte Carlo simulations is
described and some properties of local update algorithms (Metropolis, heat-bath,
Glauber) are discussed. The following Sect. 4.4 is devoted to non-local cluster al-
gorithms which in some cases can dramatically speed up the simulations. A fairly
detailed account of the initial non-equilibrium period and ageing phenomena as well
as statistical error analysis in equilibrium is given in Sect. 4.5. Here temporal corre-
lation effects are discussed, which explain the problems with critical slowing down
at a continuous phase transition and exponentially large flipping times at a first-
order transition. In Sect. 4.6, we discuss reweighting techniques which are particu-
larly important for finite-size scaling studies. A worked out example of such a study
is presented in the following Sect. 4.7. Finally, more refined generalized ensemble
simulation methods are briefly outlined in Sect. 4.8, focusing on simulated and par-
allel tempering, the multicanonical ensemble and the Wang-Landau recursion. The
lecture notes close in Sect. 4.9 with a few concluding remarks.

4.2 Statistical Physics Primer

To set the scenery for the simulation methods discussed below, we need to briefly
recall a few basic concepts of statistical physics [5, 6, 7, 8]. In these lecture notes
we will only consider classical systems and mainly focus on the canonical ensemble
where the partition function is generically given as

$$\mathcal{Z} = \sum_{\text{states}} e^{-\beta\mathcal{H}} = e^{-\beta\mathcal{F}} , \qquad (4.1)$$

with the summation running over all possible states of the system. The state space
may be continuous or discrete. As usual $\beta \equiv 1/k_{\text{B}}T$ denotes the inverse temperature

fixed by an external heat bath, k_B is Boltzmann's constant, $\mathcal{H}$ is the Hamiltonian of the system, encoding the details of the interactions which may be short-, medium-, or long-ranged, and $\mathcal{F}$ is the free energy. Expectation values denoted by angular brackets $\langle \ldots \rangle$ then follow as

$$\langle \mathcal{O} \rangle = \sum_{\text{states}} \mathcal{O} e^{-\beta \mathcal{H}} / \mathcal{Z} , \tag{4.2}$$

where $\mathcal{O}$ stands symbolically for any observable, e.g., the energy $E \equiv \mathcal{H}$.

As we will see in the next section, the most elementary Monte Carlo simulation method (Metropolis algorithm) can, in principle, cope with all conceivable variants of this quite general formulation. Close to a phase transition, however, this basic algorithm tends to become very time consuming and for accurate quantitative results one needs to employ more refined methods. Most of them are much more specific and take advantage of certain properties of the model under study. One still quite broad class of systems are lattice models, where one assumes that the degrees of freedom live on the sites or/and links of a D-dimensional lattice. These are often taken to be hypercubic, but more complicated regular lattice types (e.g., triangular (T), body-centered cubic (BCC), face-centered cubic (FCC), etc.) and even random lattices do not cause problems in principle. The degrees of freedom may be continuous or discrete field variables such as a gauge field or the height variable of a crystal surface, continuous or discrete spins such as the three-dimensional unit vectors of the classical Heisenberg model or the ± 1 valued spins of the Ising model, or arrow configurations along the links of the lattice such as in Baxter's vertex models, to give only a few popular examples.

To be specific and to keep the discussion as simple as possible, most simulation methods will be illustrated with the minimalistic Ising model [9, 10] where

$$\mathcal{H} = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \tag{4.3}$$

with $\sigma_i = \pm 1$. Here J is a coupling constant which is positive for a ferromagnet ($J > 0$) and negative for an anti-ferromagnet ($J < 0$), h is an external magnetic field, and the symbol $\langle ij \rangle$ indicates that the lattice sum is restricted to run only over all nearest-neighbor pairs. In the examples discussed below, usually periodic boundary conditions are applied. And to ease the notation, we will always assume units in which $k_B = 1$ and $J = 1$.

Basic observables are the internal energy per site, $u = U/V$, with $U = -d \ln \mathcal{Z}/d\beta \equiv \langle \mathcal{H} \rangle$, and the specific heat

$$C = \frac{du}{dT} = \beta^2 \frac{\langle E^2 \rangle - \langle E \rangle^2}{V} = \beta^2 V \left(\langle e^2 \rangle - \langle e \rangle^2 \right) , \tag{4.4}$$

where we have set $\mathcal{H} \equiv E = eV$ with V denoting the number of lattice sites, i.e., the lattice volume. The magnetization per site $m = M/V$ and the susceptibility χ are defined as

$$M = \frac{1}{\beta} \frac{\mathrm{d} \ln \mathcal{Z}}{\mathrm{d}h} = V\langle \mu \rangle \, , \qquad \mu = \frac{1}{V} \sum_i \sigma_i \, , \tag{4.5}$$

and

$$\chi = \beta V \left(\langle \mu^2 \rangle - \langle \mu \rangle^2 \right) \, . \tag{4.6}$$

The correlation between spins σ_i and σ_j at sites labeled by i and j can be measured by considering correlation functions like the two-point spin-spin correlation $G(i, j)$, which is defined as

$$G(\boldsymbol{r}) = G(i, j) = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \, , \tag{4.7}$$

where $\boldsymbol{r} = \boldsymbol{r}_i - \boldsymbol{r}_j$ (assuming translational invariance). Away from criticality and at large distances $|\boldsymbol{r}| \gg 1$ (where we have assumed a lattice spacing $a = 1$), $G(\boldsymbol{r})$ decays exponentially

$$G(\boldsymbol{r}) \sim |\boldsymbol{r}|^\kappa \, \mathrm{e}^{-|\boldsymbol{r}|/\xi} \, , \tag{4.8}$$

where ξ is the correlation length and the exponent κ of the power-law prefactor depends in general on the dimension and on whether one studies the ordered or disordered phase. Some model (and simulation) specific details of the latter observables and further important quantities like various magnetization cumulants will be discussed later when dealing with concrete applications.

The Ising model is the paradigm model for systems exhibiting a continuous (or, roughly speaking, second-order) phase transition from an ordered low-temperature to a disordered high-temperature phase at some critical temperature T_c when the temperature is varied. In two dimensions (2D), the thermodynamic limit of this model in zero external field has been solved exactly by Onsager [11], and also for finite $L_x \times L_y$ lattices the exact partition function is straightforward to compute [12, 13]. For infinite lattices, even the correlation length is known in arbitrary lattice directions [14, 15]. The exact magnetization for $h = 0$, apparently already known to Onsager [16][1], was first derived by Yang [17, 18], and the susceptibility is known to very high precision [19, 20], albeit still not exactly. In 3D no exact solutions are available, but analytical and numerical results from various methods give a consistent and very precise picture.

The most characteristic feature of a second-order phase transition is the divergence of the correlation length at T_c. As a consequence thermal fluctuations are equally important on *all* length scales, and one therefore expects power-law singularities in thermodynamic functions. The leading divergence of the correlation length is usually parameterized in the high-temperature phase as

$$\xi = \xi_{0^+} |1 - T/T_c|^{-\nu} + \ldots \quad (T \geq T_c) \, , \tag{4.9}$$

where the ... indicate sub-leading analytical as well as confluent corrections. This defines the critical exponent $\nu > 0$ and the critical amplitude ξ_{0^+} on the high-temperature side of the transition. In the low-temperature phase one expects a similar behavior

---

[1] See also the historical remarks in Refs. [14, 15].

$$\xi = \xi_{0-} (1 - T/T_c)^{-\nu} + \dots \qquad (T \leq T_c) , \qquad (4.10)$$

with the same critical exponent ν but a different critical amplitude $\xi_{0-} \neq \xi_{0+}$.

An important consequence of the divergence of the correlation length is that qualitative properties of second-order phase transitions should not depend on short-distance details of the Hamiltonian. This is the basis of the universality hypothesis [21] which means that all (short-ranged) systems with the same symmetries and same dimensionality should exhibit similar singularities governed by one and the same set of critical exponents. For the amplitudes this is not true, but certain amplitude ratios are also universal.

The singularities of the specific heat, magnetization (for $T < T_c$), and susceptibility are similarly parameterized by the critical exponents α, β and γ, respectively,

$$
\begin{aligned}
C &= C_{\text{reg}} + C_0 |1 - T/T_c|^{-\alpha} + \dots , \\
m &= m_0 (1 - T/T_c)^{\beta} + \dots , \\
\chi &= \chi_0 |1 - T/T_c|^{-\gamma} + \dots ,
\end{aligned}
\qquad (4.11)
$$

where C_{reg} is a regular background term, and the amplitudes are again in general different on the two sides of the transition. Right at the critical temperature T_c, two further exponents δ and η are defined through

$$
\begin{aligned}
m &\propto h^{1/\delta} , \\
G(\boldsymbol{r}) &\propto r^{-D+2-\eta} .
\end{aligned}
\qquad (4.12)
$$

In the 1960's, Rushbrooke [22], Griffiths [23], Josephson [24, 25] and Fisher [26] showed that these six critical exponents are related via four inequalities. Subsequent experimental evidence indicated that these relations were in fact equalities, and they are now firmly established and fundamentally important in the theory of critical phenomena. With D representing the dimensionality of the system, the scaling relations are

$$
\begin{aligned}
D\nu &= 2 - \alpha & &\text{(Josephson's law)} , \\
2\beta + \gamma &= 2 - \alpha & &\text{(Rushbrooke's law)} , \\
\beta(\delta - 1) &= \gamma & &\text{(Griffiths' law)} , \\
\nu(2 - \eta) &= \gamma & &\text{(Fisher's law)} .
\end{aligned}
\qquad (4.13)
$$

In the conventional scaling scenario, Rushbrooke's and Griffiths' laws can be deduced from the Widom scaling hypothesis that the Helmholtz free energy is a homogeneous function [27, 28]. Widom scaling and the remaining two laws can in turn be derived from the Kadanoff block-spin construction [29] and ultimately from that of the renormalization group (RG) [30]. Josephson's law can also be derived from the hyperscaling hypothesis, namely that the free energy behaves near criticality as

Table 4.1. Critical exponents of the Ising model in two (2D) and three (3D) dimensions. All 2D exponents are exactly known [31, 32], while for the 3D Ising model the world-average for ν and γ calculated in [33] is quoted. The other exponents follow from the hyperscaling relation $\alpha = 2 - D\nu$, and the scaling relations $\beta = (2 - \alpha - \gamma)/2$, $\delta = \gamma/\beta + 1$, and $\eta = 2 - \gamma/\nu$

dimension	ν	α	β	γ	δ	η
$D = 2$	1	0 (log)	1/8	7/4	15	1/4
$D = 3$	0.630 05(18)	0.109 85	0.326 48	1.237 17(28)	4.7894	0.036 39

the inverse correlation volume: $f_\infty(t) \sim \xi_\infty^{-D}(t)$. Twice differentiating this relation one recovers Josephson's law (4.13). The critical exponents for the 2D and 3D Ising model [31, 32, 33] are collected in Table 4.1.

In any numerical simulation study, the system size is necessarily finite. While the correlation length may still become very large, it is therefore always finite. This implies that also the divergences in other quantities are rounded and shifted [34, 35, 36, 37]. How this happens is described by finite-size scaling (FSS) theory, which in a nut-shell may be explained as follows: Near T_c the role of ξ is taken over by the linear size L of the system. By rewriting (4.9) or (4.10) and replacing $\xi \to L$

$$|1 - T/T_c| \propto \xi^{-1/\nu} \longrightarrow L^{-1/\nu} , \tag{4.14}$$

it is easy to see that the scaling laws (4.11) are replaced by the FSS Ansätze,

$$C = C_{\text{reg}} + aL^{\alpha/\nu} + \dots ,$$
$$m \propto L^{-\beta/\nu} + \dots ,$$
$$\chi \propto L^{\gamma/\nu} + \dots . \tag{4.15}$$

As a mnemonic rule, a critical exponent x of the temperature scaling law is replaced by $-x/\nu$ in the corresponding FSS law. In general these scaling laws are valid in a neighborhood of T_c as long as the scaling variable

$$x = (1 - T/T_c)L^{1/\nu} \tag{4.16}$$

is kept fixed [34, 35, 36, 37]. This implies for the locations T_{max} of the (finite) maxima of thermodynamic quantities such as the specific heat or susceptibility, an FSS behavior of the form

$$T_{\text{max}} = T_c(1 - x_{\text{max}}L^{-1/\nu} + \dots) . \tag{4.17}$$

In this more general formulation the scaling law for, e.g., the susceptibility reads

$$\chi(T, L) = L^{\gamma/\nu} f(x) , \tag{4.18}$$

where $f(x)$ is a scaling function. By plotting $\chi(T, L)/L^{\gamma/\nu}$ versus the scaling variable x, one thus expects that the data for different T and L fall onto a master

curve described by $f(x)$. This is a nice visual method for demonstrating the scaling properties.

Similar considerations for first-order phase transitions [38, 39, 40, 41] show that here the δ-function like singularities in the thermodynamic limit, originating from phase coexistence, are also smeared out for finite systems [42, 43, 44, 45, 46]. They are replaced by narrow peaks whose height (width) grows proportional to the volume (1/volume) with a displacement of the peak location from the infinite-volume limit proportional to 1/volume [47, 48, 49, 50, 51, 52].

4.3 The Monte Carlo Method

Let us now discuss how the expectation values in (4.2) can be estimated in a Monte Carlo simulation. For any reasonable system size, a direct summation of the partition function is impossible, since already for the minimalistic Ising model with only two possible states per site the number of terms would be enormous: For a moderate 20×20 lattice, the state space consists already of $2^{400} \approx 10^{120}$ different spin configurations.[2] Also a naive random sampling of the spin configurations does not work. Here the problem is that the relevant region in the high-dimensional phase space is relatively narrow and hence too rarely hit by random sampling. The solution to this problem is known since long under the name importance sampling [53].

4.3.1 Importance Sampling

The basic idea of importance sampling is to set up a suitable Markov chain that draws configurations not at random but according to their Boltzmann weight

$$\mathcal{P}^{\mathrm{eq}}(\{\sigma_i\}) = \frac{\mathrm{e}^{-\beta\mathcal{H}(\{\sigma_i\})}}{\mathcal{Z}} \ . \tag{4.19}$$

A Markov chain defines stochastic rules for transitions from one state to another subject to the condition that the probability for the new configuration only depends on the preceding state but not on the history of the whole trajectory in state space, i.e., it is almost local in time. Symbolically this can be written as

$$\dots \xrightarrow{W} \{\sigma_i\} \xrightarrow{W} \{\sigma_i\}' \xrightarrow{W} \{\sigma_i\}'' \xrightarrow{W} \dots \ , \tag{4.20}$$

where the transition probability W has to satisfy the following conditions:

(i) $W(\{\sigma_i\} \longrightarrow \{\sigma_i\}') \geq 0$ for all $\{\sigma_i\}, \{\sigma_i\}'$,
(ii) $\sum_{\{\sigma_i\}'} W(\{\sigma_i\} \longrightarrow \{\sigma_i\}') = 1$ for all $\{\sigma_i\}$,
(iii) $\sum_{\{\sigma_i\}} W(\{\sigma_i\} \longrightarrow \{\sigma_i\}') P^{\mathrm{eq}}(\{\sigma_i\}) = P^{\mathrm{eq}}(\{\sigma_i\}')$ for all $\{\sigma_i\}'$.

---

[2] This number should be compared with the estimated number of protons in the Universe which is about 10^{80}.

From condition (iii) we see that the desired Boltzmann distribution P^{eq} is a fixed point of W (eigenvector of W with unit eigenvalue). A somewhat simpler sufficient condition is detailed balance,

$$\mathcal{P}^{\mathrm{eq}}(\{\sigma_i\})\,W(\{\sigma_i\} \longrightarrow \{\sigma_i\}') = \mathcal{P}^{\mathrm{eq}}(\{\sigma_i\}')\,W(\{\sigma_i\}' \longrightarrow \{\sigma_i\})\,. \quad (4.21)$$

By summing over $\{\sigma_i\}$ and using condition (ii), the more general condition (iii) follows. After an initial equilibration period (cf. Sect. 4.5.1), expectation values can be estimated as an arithmetic mean over the Markov chain of length N, e.g.,

$$E = \langle \mathcal{H} \rangle = \sum_{\{\sigma_i\}} \mathcal{H}(\{\sigma_i\})\mathcal{P}^{\mathrm{eq}}(\{\sigma_i\}) \approx \frac{1}{N}\sum_{j=1}^{N}\mathcal{H}(\{\sigma_i\}_j)\,, \quad (4.22)$$

where $\{\sigma_i\}_j$ denotes the spin configuration at "time" j. A more detailed exposition of the mathematical concepts underlying any Markov chain Monte Carlo algorithm can be found in many textbooks and reviews [1, 2, 3, 4, 34, 54, 55].

4.3.2 Local Update Algorithms

The Markov chain conditions (i)–(iii) are still quite general and can be satisfied by many different concrete update rules. In a rough classification one distinguishes between local and non-local algorithms. Local update algorithms discussed in this subsection are conceptually much simpler and, as the main merit, quite universally applicable. The main drawback is their relatively poor performance close to second-order phase transitions where the spins or fields of a typical configuration are strongly correlated over large spatial distances. Here non-local update algorithms based on multigrid methods or in particular self-adaptive cluster algorithms discussed later in Sect. 4.4 perform much better.

4.3.2.1 Metropolis Algorithm

The most flexible update rule is the classic Metropolis algorithm [56], which is applicable in practically all cases (lattice/off-lattice, discrete/continuous, short-range/long-range interactions, ...). Here one proposes an update for a single degree of freedom (spin, field, ...) and accepts this proposal with probability

$$W(\{\sigma_i\}_{\mathrm{old}} \longrightarrow \{\sigma_i\}_{\mathrm{new}}) = \begin{cases} 1 & E_{\mathrm{new}} < E_{\mathrm{old}} \\ \mathrm{e}^{-\beta(E_{\mathrm{new}}-E_{\mathrm{old}})} & E_{\mathrm{new}} \geq E_{\mathrm{old}} \end{cases}, \quad (4.23)$$

where E_{old} and E_{new} denote the energy of the old and new spin configuration $\{\sigma_i\}_{\mathrm{old}}$ and $\{\sigma_i\}_{\mathrm{new}}$, respectively, where $\{\sigma_i\}_{\mathrm{new}}$ differs from $\{\sigma_i\}_{\mathrm{old}}$ only locally by one modified degree of freedom at, say, $i = i_0$. More compactly, this may also be written as

$$W(\{\sigma_i\}_{\mathrm{old}} \longrightarrow \{\sigma_i\}_{\mathrm{new}}) = \min\{1, \mathrm{e}^{-\beta\Delta E}\}\,, \quad (4.24)$$

where $\Delta E = E_{\text{new}} - E_{\text{old}}$. If the proposed update lowers the energy, it is always accepted. On the other hand, when the new configuration has a higher energy, the update has still to be accepted with a certain probability in order to ensure the proper treatment of entropic contributions – in thermal equilibrium, it is the free energy $F = U - TS$ which has to be minimized and not the energy. Only in the limit of zero temperature, $\beta \to \infty$, the acceptance probability for this case tends to zero and the Metropolis method degenerates to a minimization algorithm for the energy functional. With some additional refinements, this is the basis for the simulated annealing technique [57], which is often applied to hard optimization and minimization problems.

The verification of the detailed balance condition (4.21) is straightforward. If $E_{\text{new}} < E_{\text{old}}$, then the l.h.s. of (4.21) becomes $\exp(-\beta E_{\text{old}}) \times 1 = \exp(-\beta E_{\text{old}})$. On the r.h.s. we have to take into account that the reverse move would increase the energy, $E_{\text{old}} > E_{\text{new}}$ (with E_{old} now playing the role of the new energy), such that now the second line of (4.23) with E_{old} and E_{new} interchanged is relevant. This gives $\exp(-\beta E_{\text{new}}) \times \exp(-\beta(E_{\text{old}} - E_{\text{new}})) = \exp(-\beta E_{\text{old}})$ on the r.h.s. of (4.21), completing the demonstration of detailed balance. In the opposite case with $E_{\text{new}} < E_{\text{old}}$, a similar reasoning leads to $\exp(-\beta E_{\text{old}}) \times \exp(-\beta(E_{\text{new}} - E_{\text{old}})) = \exp(-\beta E_{\text{new}}) = \exp(-\beta E_{\text{new}}) \times 1$. Admittedly, this proof looks a bit like a tautology. To uncover its non-trivial content, it is a useful exercise to replace the r.h.s. of the Metropolis rule (4.23) by some general function $f(E_{\text{new}} - E_{\text{old}})$ and repeat the above steps [58].

Finally a few remarks on the practical implementation of the Metropolis method are in order. To decide whether a proposed update should be accepted or not, one draws a uniformly distributed random number $r \in [0, 1)$, and if $W \leq r$, the new state is accepted. Otherwise one keeps the old configuration and continues with the next spin. In computer simulations, random numbers are generated by means of pseudo-random number generators (RNGs), which produce (more or less) uniformly distributed numbers whose values are very hard to predict – by using some deterministic rule (see [59] and references therein). In other words, given a finite sequence of subsequent pseudo-random numbers, it should be (almost) impossible to predict the next one or to even guess the deterministic rule underlying their generation. The goodness of an RNG is thus measured by the difficulty to derive its underlying deterministic rule. Related requirements are the absence of trends (correlations) and a very long period. Furthermore, an RNG should be portable among different computer platforms and, very importantly, it should yield reproducible results for testing purposes. The design of RNGs is a science in itself, and many things can go wrong with them. As a recommendation one should better not experiment too much with some fancy RNG one has picked up somewhere from the Web, say, but rely on well-tested and well-documented routines.

There are many different ways how the degrees of freedom to be updated can be chosen. They may be picked at random or according to a random permutation, which can be updated every now and then. But also a simple fixed lexicographical

(sequential) order is permissible.[3] In lattice models one may also update first all odd and then all even sites, which is the usual choice in vectorized codes. A so-called sweep is completed when on the average[4] for all degrees of freedom an update was proposed. The qualitative behavior of the update algorithm is not sensitive to these details, but its quantitative performance does depend on the choice of the update scheme.

4.3.2.2 Heat-Bath Algorithm

This algorithm is only applicable to lattice models and at least in its most straight-forward form only to discrete degrees of freedom with a few allowed states. The new value σ'_{i_0} at site i_0 is determined by testing all its possible states in the heat-bath of its (fixed) neighbors (e.g., four on a square lattice and six on a simple-cubic lattice with nearest-neighbor interactions):

$$W(\{\sigma_i\}_{\text{old}} \longrightarrow \{\sigma_i\}_{\text{new}}) = \frac{e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{new}})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{old}})}} = \frac{e^{-\beta\sigma'_{i_0} S_{i_0}}}{\sum_{\sigma_{i_0}} e^{-\beta\sigma_{i_0} S_{i_0}}} , \quad (4.25)$$

where $S_{i_0} = -\sum_j \sigma_j - h$ is an effective spin or field collecting all neighboring spins (in their old states) interacting with the spin at site i_0 and h is the external magnetic field. Note that this decomposition also works in the case of vectors ($\sigma_i \to \boldsymbol{\sigma_i}$, $h \to \boldsymbol{h}$, $S_{i_0} \to \boldsymbol{S_{i_0}}$), interacting via the usual dot product ($\sigma'_{i_0} S_{i_0} \to \boldsymbol{\sigma'_{i_0}} \cdot \boldsymbol{S_{i_0}}$). As the last equality in (4.25) shows, all other contributions to the energy not involving σ'_{i_0} cancel due to the ratio in (4.25), so that for the update at each site i_0 only a small number of computations is necessary (e.g, about four for a square and six for a simple-cubic lattice of arbitrary size). Detailed balance (4.21) is obviously satisfied since

$$e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{old}})} \frac{e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{new}})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{new}})}} = e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{new}})} \frac{e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{old}})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\{\sigma_i\}_{\text{old}})}} . \quad (4.26)$$

How is the probability (4.25) realized in practice? Due to the summation over all local states, special tricks are necessary when each degree of freedom can take many different states, and only in special cases the heat-bath method can be efficiently generalized to continuous degrees of freedom. In many applications, however, the admissible local states of σ_{i_0} can be labeled by a small number of integers, say $n = 1, \ldots, N$. Since the probability in (4.25) is normalized to unity, the sequence $(P_1, P_2, \ldots, P_n, \ldots, P_N)$ decomposes the unit interval into segments of length $P_n = \exp(-\beta n S_{i_0})/\sum_{k=1}^{N} \exp(-\beta k S_{i_0})$. If one now draws a random number $R \in [0, 1)$ and compares the accumulated probabilities $\sum_{k=1}^{n} P_k$ with R, then the new state n_0 is given as the smallest integer for which $\sum_{k=1}^{n_0} P_k \geq R$. Clearly, for a large number of possible local states, the determination of n_0 can become quite time-consuming (in particular, if many small P_n are at the beginning

---

[3] Some special care is necessary, however, for one-dimensional spin chains.
[4] This is only relevant when the random update order is chosen.

of the sequence, in which case a clever permutation of the P_n-list can help a lot). The order of updating the individual variables can be chosen as for the Metropolis algorithm (random, sequential, ...).

In the special case of the Ising model with only two states per spin, $\sigma_i = \pm 1$, (4.25) reads explicitly as

$$W(\{\sigma_i\}_{\text{old}} \longrightarrow \{\sigma_i\}_{\text{new}}) = \frac{e^{-\beta\sigma'_{i_0} S_{i_0}}}{e^{\beta S_{i_0}} + e^{-\beta S_{i_0}}} \ . \tag{4.27}$$

And since $\Delta E = E_{\text{new}} - E_{\text{old}} = (\sigma'_{i_0} - \sigma_{i_0})S_{i_0}$, the probability for a spin flip, $\sigma'_{i_0} = -\sigma_{i_0}$, becomes [58]

$$W_{\sigma_{i_0} \to -\sigma_{i_0}} = \frac{e^{-\beta\Delta E/2}}{e^{\beta\Delta E/2} + e^{-\beta\Delta E/2}} \ . \tag{4.28}$$

The acceptance ratio (4.28) is plotted in Fig. 4.1 as a function of ΔE for various (inverse) temperatures and compared with the corresponding ratio (4.24) of the Metropolis algorithm. As we shall see in the next paragraph, for the Ising model, the Glauber and heat-bath algorithm are identical.

4.3.2.3 Glauber Algorithm

The Glauber update prescription [60] is conceptually similar to the Metropolis algorithm in that also here a local update proposal is accepted with a certain probability
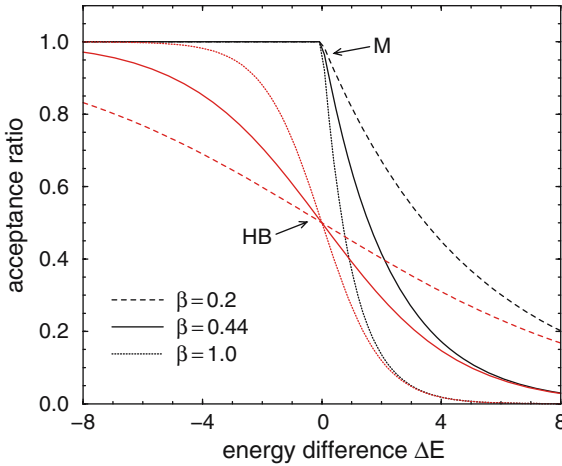

Fig. 4.1. Comparison of the acceptance ratio for a spin flip with the heat-bath (HB) (or Glauber) and Metropolis (M) algorithm in the Ising model for three different inverse temperatures β. Note that for all values of ΔE and temperature, the Metropolis acceptance ratio is higher than that of the heat-bath algorithm

or otherwise rejected. For the Ising model with spins $\sigma_i = \pm 1$ the acceptance probability can be written as

$$W_{\sigma_{i_0} \to -\sigma_{i_0}} = \frac{1}{2} \left[1 + \sigma_{i_0} \tanh\left(\beta S_{i_0}\right) \right] , \qquad (4.29)$$

where as before $\sigma_{i_0} S_{i_0}$ with $S_{i_0} = -\sum_j \sigma_j - h$ is the energy of the i_0^{th} spin in the current old state.

Due to the point symmetry of the hyperbolic tangent, one may rewrite $\sigma_{i_0} \tanh (\beta S_{i_0})$ as $\tanh (\sigma_{i_0} \beta S_{i_0})$. And since as before $\Delta E = E_{\text{new}} - E_{\text{old}} = -2\sigma_{i_0} S_{i_0}$, (4.29) becomes

$$W_{\sigma_{i_0} \to -\sigma_{i_0}} = \frac{1}{2} \left[1 - \tanh\left(\beta \Delta E/2\right) \right] , \qquad (4.30)$$

showing explicitly that the acceptance probability only depends on the total energy change as in the Metropolis case. In this form it is thus possible to generalize the Glauber update rule from the Ising model with only two states per spin to any general model that can be simulated with the Metropolis procedure. Also detailed balance is straightforward to prove. Finally by using trivial identities for hyperbolic functions, (4.30) can be further recast to read

$$\begin{aligned} W_{\sigma_{i_0} \to -\sigma_{i_0}} &= \frac{1}{2} \left[\frac{\cosh(\beta \Delta E/2) - \sinh(\beta \Delta E/2)}{\cosh(\beta \Delta E/2)} \right] \\ &= \frac{\mathrm{e}^{-\beta \Delta E/2}}{\mathrm{e}^{\beta \Delta E/2} + \mathrm{e}^{-\beta \Delta E/2}} , \end{aligned} \qquad (4.31)$$

which is just the flip probability (4.28) of the heat-bath algorithm for the Ising model, i.e., heat-bath updates for the special case of a 2-state model and the Glauber update algorithm are identical. In the general case with more than two states per spin, however, this is not the case.

The Glauber (or equivalently heat-bath) update algorithm for the Ising model is also theoretically of interest since in this case the dynamics of the Markov chain can be calculated analytically for a one-dimensional system [60]. For two and higher dimensions, however, no exact solutions are known.

4.3.3 Performance of Local Update Algorithms

Local update algorithms are applicable to a very wide class of models and the computer codes are usually quite simple and very fast. The main drawback are temporal correlations of the generated Markov chain which tend to become huge in the vicinity of phase transitions. They can be determined by analysis of autocorrelation functions

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} , \qquad (4.32)$$

where $\mathcal{O}$ denotes any measurable quantity, for example the energy or magnetization. More details and how temporal correlations enter into the statistical error analysis

will be discussed in Sect. 4.5.2.3. For large time separations k, $A(k)$ decays exponentially ($a = \text{const}$)

$$A(k) \xrightarrow{k \rightarrow \infty} a e^{-k/\tau_{\mathcal{O},\text{exp}}} , \tag{4.33}$$

which defines the exponential autocorrelation time $\tau_{\mathcal{O},\text{exp}}$. At smaller distances usually also other modes contribute and $A(k)$ behaves no longer purely exponentially.

This is illustrated in Fig. 4.2 for the 2D Ising model on a rather small 16×16 square lattice with periodic boundary conditions at the infinite-volume critical point $\beta_c = \ln(1 + \sqrt{2})/2 = 0.440\,686\,793\ldots$. The spins were updated in sequential order by proposing always to flip a spin and accepting or rejecting this proposal according to (4.23). The raw data of the simulation are collected in a time-series file, storing 1 000 000 measurements of the energy and magnetization taken after each sweep over the lattice, after discarding (quite generously) the first 200 000 sweeps for equilibrating the system from a disordered start configuration. The last 1 000 sweeps of the time evolution of the energy are shown in Fig. 4.2(a). Using the complete time series the autocorrelation function was computed according to (4.32) which is shown in Fig. 4.2(b). On the linear-log scale of the inset we clearly see the asymptotic linear behavior of $\ln A(k)$. A linear fit of the form (4.33), $\ln A(k) = \ln a - k/\tau_{e,\text{exp}}$, in the range $10 \leq k \leq 40$ yields an estimate for the exponential autocorrelation time of $\tau_{e,\text{exp}} \approx 11.3$. In the small k behavior of $A(k)$ we observe an initial fast drop, corresponding to faster relaxing modes, before the asymptotic behavior sets in. This is the generic behavior of autocorrelation functions in realistic models where the small-k deviations are, in fact, often much more pronounced than for the 2D Ising model.

Close to a critical point, in the infinite-volume limit, the autocorrelation time typically scales as

$$\tau_{\mathcal{O},\text{exp}} \propto \xi^z , \tag{4.34}$$

where $z \geq 0$ is the so-called dynamical critical exponent. Since the spatial correlation length $\xi \propto |T - T_c|^{-\nu} \rightarrow \infty$ when $T \rightarrow T_c$, also the autocorrelation time $\tau_{\mathcal{O},\text{exp}}$
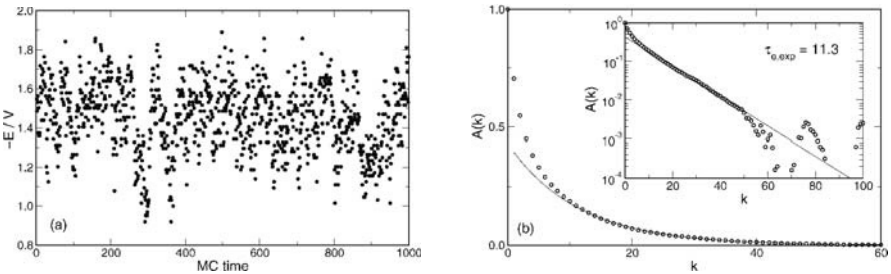

Fig. 4.2. (a) Part of the time evolution of the energy $e = E/V$ for the 2D Ising model on a 16×16 lattice at $\beta_c = \ln(1 + \sqrt{2})/2 = 0.440\,686\,793\ldots$ and (b) the resulting autocorrelation function. In the inset the same data are plotted on a logarithmic scale, revealing a fast initial drop for very small k and noisy behavior for large k. The solid lines show a fit to the ansatz $A(k) = a \exp(-k/\tau_{e,\text{exp}})$ in the range $10 \leq k \leq 40$ with $\tau_{e,\text{exp}} = 11.3$ and $a = 0.432$

diverges when the critical point is approached, $\tau_{\mathcal{O},\mathrm{exp}} \propto |T - T_c|^{-\nu z}$. This leads to the phenomenon of critical slowing down at a continuous phase transition. This is not in the first place a numerical artefact, but can also be observed experimentally for instance in critical opalescence, see Fig. 1.1 in [5]. The reason is that local spin-flip Monte Carlo dynamics (or diffusion dynamics in a lattice-gas picture) describes at least qualitatively the true physical dynamics of a system in contact with a heat-bath (which, in principle, enters stochastic elements also in molecular dynamics simulations). In a finite system, the correlation length ξ is limited by the linear system size L, and similar to the reasoning in (4.14) and (4.15), the scaling law (4.34) becomes

$$\tau_{\mathcal{O},\mathrm{exp}} \propto L^z \ . \tag{4.35}$$

For local dynamics, the critical slowing down effect is quite pronounced since the dynamical critical exponent takes a rather large value around

$$z \approx 2 \ , \tag{4.36}$$

which is only weakly dependent on the dimensionality and can be understood by a simple random-walk or diffusion argument in energy space. Non-local update algorithms such as multigrid schemes or in particular the cluster methods discussed in the next section can reduce the value of the dynamical critical exponent z significantly, albeit in a strongly model-dependent fashion.

At a first-order phase transition, a completely different mechanism leads to an even more severe slowing-down problem [47]. Here, the password is phase coexistence. A finite system close to the (pseudo-) transition point can flip between the coexisting pure phases by crossing a two-phase region. Relative to the weight of the pure phases, this region of state space is strongly suppressed by an additional Boltzmann factor $\exp(-2\sigma L^{d-1})$, where σ denotes the interface tension between the coexisting phases, L^{d-1} is the (projected) area of the interface and the factor two accounts for periodic boundary conditions, which enforce always an even number of interfaces for simple topological reasons. The time spent for crossing this highly suppressed rare-event region scales proportional to the inverse of this interfacial Boltzmann factor, implying that the autocorrelation time increases exponentially with the system size,

$$\tau_{\mathcal{O},\mathrm{exp}} \propto \mathrm{e}^{2\sigma L^{d-1}} \ . \tag{4.37}$$

In the literature, this behavior is sometimes termed supercritical slowing down, even though, strictly speaking, nothing is critical at a first-order phase transition. Since this type of slowing-down problem is directly related to the shape of the probability distribution, it appears for all types of update algorithms, i.e., in contrast to the situation at a second-order transition, here it cannot be cured by employing multigrid or cluster techniques. It can be overcome, however, at least in part by means of tempering and multicanonical methods also briefly discussed at the end of these notes in Sect. 4.8.

4.4 Cluster Algorithms

In this section we first concentrate on the problem of critical slowing down at a second-order phase transition which is caused by very large spatial correlations, reflecting that excitations become equally important on all length scales. It is therefore intuitively clear that some sort of non-local updates should be able to alleviate this problem. While it was realized since long that whole clusters or droplets of spins should play a central role in such an update, it took until 1987 before Swendsen and Wang [61] proposed a legitimate cluster update procedure first for q-state Potts models [62] with

$$\mathcal{H}_{\text{Potts}} = -J \sum_{\langle ij \rangle} \delta_{\sigma_i, \sigma_j} \ , \tag{4.38}$$

where $\sigma_i = 1, \ldots, q$. For $q = 2$ (and a trivial rescaling) the Ising model (4.3) is recovered. Soon after this discovery, Wolff [63] introduced the so-called single-cluster variant and developed a generalization to O(n)-symmetric spin models. By now cluster update algorithms have been constructed for many other models as well [64]. However, since in all constructions some model specific properties enter in a crucial way, they are still far less general applicable than local update algorithms of the Metropolis type. We therefore first concentrate again on the Ising model where (as for more general Potts models) the prescription for a cluster-update algorithm can be easily read off from the equivalent Fortuin-Kasteleyn representation [65, 66, 67, 68]

$$Z = \sum_{\{\sigma_i\}} e^{\beta \sum_{\langle ij \rangle} \sigma_i \sigma_j} \tag{4.39}$$

$$= \sum_{\{\sigma_i\}} \prod_{\langle ij \rangle} e^{\beta} \left[(1-p) + p \delta_{\sigma_i, \sigma_j} \right] \tag{4.40}$$

$$= \sum_{\{\sigma_i\}} \sum_{\{n_{ij}\}} \prod_{\langle ij \rangle} e^{\beta} \left[(1-p) \delta_{n_{ij}, 0} + p \delta_{\sigma_i, \sigma_j} \delta_{n_{ij}, 1} \right] \tag{4.41}$$

with

$$p = 1 - e^{-2\beta} \ . \tag{4.42}$$

Here the n_{ij} are bond occupation variables which can take the values $n_{ij} = 0$ or $n_{ij} = 1$, interpreted as deleted or active bonds. The representation (4.40) in the second line follows from the observation that the product $\sigma_i \sigma_j$ of two Ising spins can only take the two values ± 1, so that $\exp(\beta \sigma_i \sigma_j) = x + y \delta_{\sigma_i \sigma_j}$ can easily be solved for x and y. And in the third line (4.41) we made use of the trivial (but clever) identity $a + b = \sum_{n=0}^{1} (a \delta_{n,0} + b \delta_{n,1})$.

4.4.1 Swendsen-Wang Cluster

According to (4.41) a cluster update sweep consists of two alternating steps. First, updates of the bond variables n_{ij} for given spins, and second updates of the spins σ_i for a given bond configuration. In practice one proceeds as follows:

$$n_{ij}=0 \qquad\qquad n_{ij}=1 \qquad\qquad n_{ij}=0$$
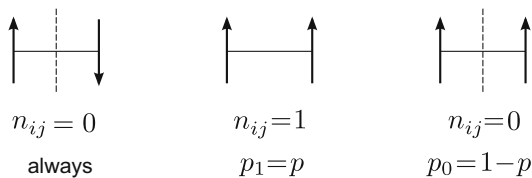$$\text{always} \qquad\qquad p_1=p \qquad\qquad p_0=1-p$$

Fig. 4.3. Illustration of the bond variable update. The bond between unlike spins is always deleted as indicated by the *dashed line*. A bond between like spins is only active with probability $p = 1 - \exp(-2\beta)$. Only at zero temperature ($\beta \to \infty$) stochastic and geometrical clusters coincide

  (i) Set $n_{ij} = 0$ if $\sigma_i \neq \sigma_j$, or assign values $n_{ij} = 1$ and 0 with probability p and $1 - p$, respectively, if $\sigma_i = \sigma_j$, cp. Fig. 4.3.
 (ii) Identify clusters of spins that are connected by active bonds ($n_{ij} = 1$).
(iii) Draw a random value ± 1 independently for each cluster (including one-site clusters), which is then assigned to all spins in a cluster.

Technically the cluster identification part is the most complicated step, but there are by now quite a few efficient algorithms available which can even be used on parallel computers. Vectorization, on the other hand, is only partially possible.

Notice the difference between the just defined stochastic clusters and geometrical clusters whose boundaries are defined by drawing lines through bonds between unlike spins. In fact, since in the stochastic cluster definition also bonds between like spins are deleted with probability $p_0 = 1 - p = \exp(-2\beta)$, stochastic clusters are smaller than geometrical clusters. Only at zero temperature ($\beta \to \infty$) p_0 approaches zero and the two cluster definitions coincide.

As described above, the cluster algorithm is referred to as Swendsen-Wang (SW) or multiple-cluster update [61]. The distinguishing point is that the *whole* lattice is decomposed into stochastic clusters whose spins are assigned a random value $+1$ or -1. In one sweep one thus attempts to update all spins of the lattice.

4.4.2 Wolff Cluster

Shortly after the original discovery of cluster algorithms, Wolff [63] proposed a somewhat simpler variant in which only a single cluster is flipped at a time. This variant is therefore sometimes also called single-cluster algorithm. Here one chooses a lattice site at random, constructs only the cluster connected with this site, and then flips all spins of this cluster. In principle, one could also here choose for all spins in the updated cluster a new value $+1$ or -1 at random, but then nothing at all would be changed if one hits the current value of the spins. Typical configuration plots before and after the cluster flip are shown in Fig. 4.4, which also nicely illustrates the difference between stochastic and geometrical clusters already stressed in the last paragraph. The upper right plot clearly shows that, due to the randomly distributed inactive bonds between like spins, the stochastic cluster is much smaller than the underlying black geometrical cluster which connects all neighboring like spins.
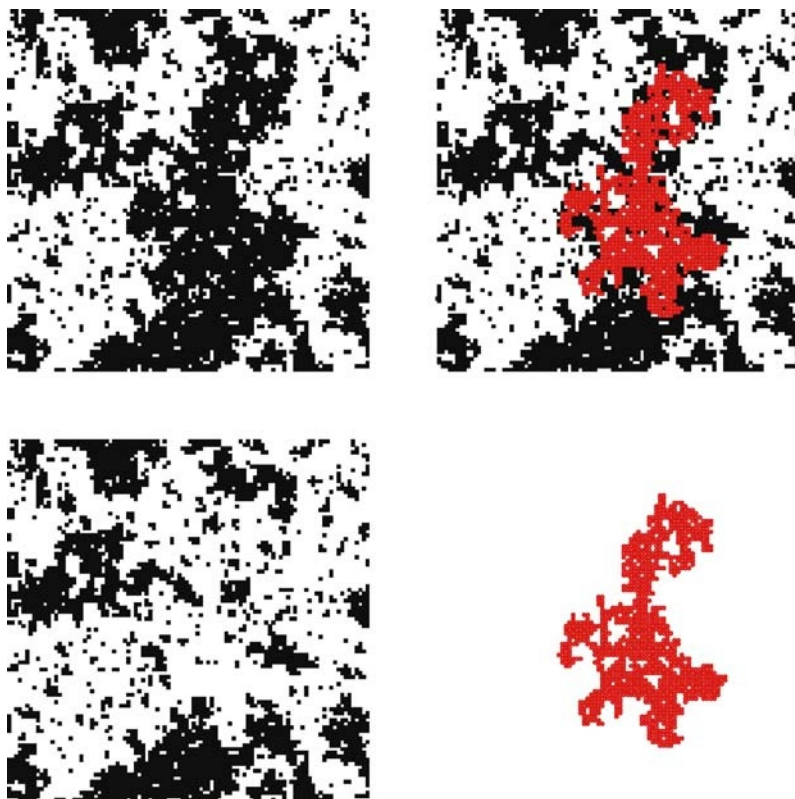
Fig. 4.4. Illustration of the Wolff cluster update, using actual simulation results for the 2D Ising model at $0.97\beta_c$ on a 100×100 lattice. **Upper left**: Initial configuration. **Upper right**: The stochastic cluster is marked. Note how it is embedded in the larger geometric cluster connecting all neighboring like (black) spins. **Lower left**: Final configuration after flipping the spins in the cluster. **Lower right**: The flipped cluster

In the single-cluster variant some care is necessary with the definition of the unit of time since the number of flipped spins varies from cluster to cluster. It also depends crucially on temperature since the average cluster size automatically adapts to the correlation length. With $\langle|C|\rangle$ denoting the average cluster size, a sweep is usually defined to consist of $V/\langle|C|\rangle$ single cluster steps, assuring that on the average V spins are flipped in one sweep. With this definition, autocorrelation times are directly comparable with results from the Swendsen-Wang or Metropolis algorithm. Apart from being somewhat easier to program, Wolff's single-cluster variant is usually even more efficient than the Swendsen-Wang multiple-cluster algorithm, especially in 3D. The reason is that with the single-cluster method, on the average, larger clusters are flipped.

4.4.3 Embedded Clusters

While it is straightforward to generalize the derivation (4.39)–(4.42) to q-state Potts models (because as in the Ising model each contribution to the energy, $\delta_{\sigma_i \sigma_j}$, can take only two different values), for $O(n)$ spin models with Hamiltonian

$$H = -J \sum_{\langle ij \rangle} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \;, \tag{4.43}$$

with $\boldsymbol{\sigma}_i = (\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,n})$ and $|\boldsymbol{\sigma}_i| = 1$, one needs a new strategy for $n \geq 2$ [63, 69, 70, 71] (the case $n = 1$ degenerates again to the Ising model). Here the basic idea is to isolate Ising degrees of freedom by projecting the spins $\boldsymbol{\sigma}_i$ onto a randomly chosen unit vector $\boldsymbol{r}$

$$\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i^{\parallel} + \boldsymbol{\sigma}_i^{\perp} \;,$$
$$\boldsymbol{\sigma}_i^{\parallel} = \epsilon \, |\boldsymbol{\sigma}_i \cdot \boldsymbol{r}| \, \boldsymbol{r} \;,$$
$$\epsilon = \text{sign}(\boldsymbol{\sigma}_i \cdot \boldsymbol{r}) \;. \tag{4.44}$$

If this is inserted in the original Hamiltonian one ends up with an effective Hamiltonian

$$H = -\sum_{\langle ij \rangle} J_{ij} \epsilon_i \epsilon_j + \text{const} \;, \tag{4.45}$$

with positive random couplings $J_{ij} = J |\boldsymbol{\sigma}_i \cdot \boldsymbol{r}| |\boldsymbol{\sigma}_j \cdot \boldsymbol{r}| \geq 0$, whose Ising degrees of freedom ϵ_i can be updated with a cluster algorithm as described above.

4.4.4 Performance of Cluster Algorithms

The advantage of cluster algorithms is most pronounced close to criticality where excitations on all length scales occur. A convenient performance measure is thus the dynamical critical exponent z (even though one should always check that the proportionality constant in $\tau \propto L^z$ is not exceedingly large, but this is definitely not the case here [72]). Some results on z are collected in Table 4.2, which allow us to conclude:

 (i) Compared to local algorithms with $z \approx 2$, z is dramatically reduced for both cluster variants in 2D and 3D [73, 74, 75].
 (ii) In 2D, Swendsen-Wang and Wolff cluster updates are equally efficient, while in 3D, the Wolff update is clearly favorable.
(iii) In 2D, the scaling with system size can hardly be distinguished from a very weak logarithmic scaling. Note that this is consistent with the Li-Sokal bound [76, 77] for the Swendsen-Wang cluster algorithm of $\tau_{\text{SW}} \geq C \, (= C_0 + A \ln L$ for the 2D Ising model), implying $z_{\text{SW}} \geq \alpha/\nu \, (= 0$ for the 2D Ising model).
(iv) Different observables (e.g., energy E and magnetization M) may yield quite different values for z when defined via the scaling behavior of the integrated autocorrelation time discussed below in Sect. 4.5.2.3.

Table 4.2. Dynamical critical exponents z for the 2D and 3D Ising model ($\tau \propto L^z$). The subscripts indicate the observables and method used (exp resp. int: exponential resp. integrated autocorrelation time, rel: relaxation, dam: damage spreading)

algorithm	2D	3D	observable	authors
Metropolis	2.1667(5)	–	$z_{M,\text{exp}}$	Nightingale and Blöte [78, 79]
	–	2.032(4)	z_{dam}	Grassberger [80, 81]
	–	2.055(10)	$z_{M,\text{rel}}$	Ito et al. [82]
Swendsen-Wang cluster	0.35(1)	0.75(1)	$z_{E,\text{exp}}$	Swendsen and Wang [61]
	0.27(2)	0.50(3)	$z_{E,\text{int}}$	Wolff [72]
	0.20(2)	0.50(3)	$z_{\chi,\text{int}}$	Wolff [72]
	0(log L)	–	$z_{M,\text{exp}}$	Heermann and Burkitt [83]
	0.25(5)	–	$z_{M,\text{rel}}$	Tamayo [84]
Wolff cluster	0.26(2)	0.28(2)	$z_{E,\text{int}}$	Wolff [72]
	0.13(2)	0.14(2)	$z_{\chi,\text{int}}$	Wolff [72]
	0.25(5)	0.3(1)	$z_{E,\text{rel}}$	Ito and Kohring [85]

4.4.5 Improved Estimators

The intimate relationship of cluster algorithms with the correlated percolation representation of Fortuin and Kasteleyn leads to another quite important improvement which is not directly related with the dynamical properties discussed so far. Within the percolation picture, it is quite natural to introduce alternative estimators (measurement prescriptions) for most standard quantities which turn out to be so-called improved estimators. By this one means measurement prescriptions that yield the same expectation value as the standard ones but have a smaller statistical variance which helps to reduce the statistical errors. Suppose we want to measure the expectation value $\langle \mathcal{O} \rangle$ of an observable $\mathcal{O}$. Then any estimator $\widehat{\mathcal{O}}$ satisfying $\langle \widehat{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle$ is permissible. This does not determine $\widehat{\mathcal{O}}$ uniquely since there are infinitely many other possible choices $\widehat{\mathcal{O}}' = \widehat{\mathcal{O}} + \widehat{\mathcal{X}}$, as long as the added estimator $\widehat{\mathcal{X}}$ has zero expectation $\langle \widehat{\mathcal{X}} \rangle = 0$. The variance of the estimator $\widehat{\mathcal{O}}'$, however, can be quite different and is not necessarily related to any physical quantity (contrary to the standard mean-value estimator of the energy, for instance, whose variance is proportional to the specific heat). It is exactly this freedom in the choice of $\widehat{\mathcal{O}}$ which allows the construction of improved estimators.

For the single-cluster algorithm an improved cluster estimator for the spin-spin correlation function in the high-temperature phase $G(\boldsymbol{x}_i - \boldsymbol{x}_j) \equiv \langle \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \rangle$ is given by [71]

$$\widehat{G}(\boldsymbol{x}_i - \boldsymbol{x}_j) = n\frac{V}{|C|}(\boldsymbol{r} \cdot \boldsymbol{\sigma}_i)\ (\boldsymbol{r} \cdot \boldsymbol{\sigma}_j)\ \Theta_C(\boldsymbol{x}_i)\Theta_C(\boldsymbol{x}_j)\ , \tag{4.46}$$

where $\boldsymbol{r}$ is the normal of the mirror plane used in the construction of the cluster of size $|C|$ and $\Theta_C(\boldsymbol{x})$ is its characteristic function ($= 1$ if $\boldsymbol{x} \in C$ and zero otherwise). In the Ising case ($n = 1$), this simplifies to

$$\widehat{G}(\boldsymbol{x}_i - \boldsymbol{x}_j) = \frac{V}{|C|}\Theta_C(\boldsymbol{x}_i)\Theta_C(\boldsymbol{x}_j)\ , \tag{4.47}$$

i.e., to the test whether the two sites $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ belong to same stochastic cluster or not. Only in the former case, the average over clusters is incremented by one, otherwise nothing is added. This implies that $\widehat{G}(\boldsymbol{x}_i - \boldsymbol{x}_j)$ is strictly positive which is not the case for the standard estimator $\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j$, where ± 1 contributions have to average to a positive value. It is therefore at least intuitively clear that the cluster (or percolation) estimator has a smaller variance and is thus indeed an improved estimator, in particular for large separations $|\boldsymbol{x}_i - \boldsymbol{x}_j|$.

For the Fourier transform $\widetilde{G}(\boldsymbol{k}) = \sum_{\boldsymbol{x}} G(\boldsymbol{x})\exp(-i\boldsymbol{k} \cdot \boldsymbol{x})$, (4.46) implies the improved estimator

$$\widehat{\widetilde{G}}(\boldsymbol{k}) = \frac{n}{|C|}\left[\left(\sum_{i \in C}\boldsymbol{r} \cdot \boldsymbol{\sigma}_i \cos \boldsymbol{k}\boldsymbol{x}_i\right)^2 + \left(\sum_{i \in C}\boldsymbol{r} \cdot \boldsymbol{\sigma}_i \sin \boldsymbol{k}\boldsymbol{x}_i\right)^2\right]\ , \tag{4.48}$$

which, for $\boldsymbol{k} = \boldsymbol{0}$, reduces to an improved estimator for the susceptibility $\chi' = \beta V \langle m^2 \rangle$ in the high-temperature phase

$$\widehat{\widetilde{G}}(\boldsymbol{0}) = \widehat{\chi}'/\beta = \frac{n}{|C|}\left(\sum_{i \in C}\boldsymbol{r} \cdot \boldsymbol{\sigma}_i\right)^2\ . \tag{4.49}$$

For the Ising model ($n = 1$) this reduces to $\chi'/\beta = \langle |C| \rangle$, i.e., the improved estimator of the susceptibility is just the average cluster size of the single-cluster update algorithm. For the XY ($n = 2$) and Heisenberg ($n = 3$) model one finds empirically that in two as well as in three dimensions $\langle |C| \rangle \approx 0.81\chi'/\beta$ for $n = 2$ [69, 86] and $\langle |C| \rangle \approx 0.75\chi'/\beta$ for $n = 3$ [71, 87], respectively.

Close to criticality, the average cluster size becomes large, growing $\propto \chi' \propto L^{\gamma/\nu} \simeq L^2$ (since $\gamma/\nu = 2 - \eta$ with η usually small) and the advantage of cluster estimators diminishes. In fact, in particular for short-range quantities such as the energy (the next-neighbor correlation) it may even degenerate into a depraved or deteriorated estimator, while long-range quantities such as $G(\boldsymbol{x}_i - \boldsymbol{x}_j)$ for large distances $|\boldsymbol{x}_i - \boldsymbol{x}_j|$ usually still profit from it. A significant reduction of variance by means of the estimators (4.46)–(4.49) can, however, always be expected outside the FSS region where the average cluster size is small compared to the volume of the system.

Finally it is worth pointing out that at least for 2D Potts models also the geometrical clusters do encode critical properties – albeit those of different but related (tricritical) models [88, 89, 90, 91, 92][5].

---

[5] See also the extensive list of references to earlier work given therein.

4.5 Statistical Analysis of Monte Carlo Data

4.5.1 Initial Non-Equilibrium Period and Ageing

When introducing the importance sampling technique in Sect. 4.3.1 it was already indicated in (4.22) that within Markov chain Monte Carlo simulations, the expectation value $\langle \mathcal{O} \rangle$ of some quantity $\mathcal{O}$, for instance the energy, can be estimated as arithmetic mean

$$\langle \mathcal{O} \rangle = \sum_{\{\sigma_i\}} \mathcal{O}(\{\sigma_i\}) P^{\text{eq}}(\{\sigma_i\}) \approx \overline{\mathcal{O}} = \frac{1}{N} \sum_{j=1}^{N} \mathcal{O}_j \ , \qquad (4.50)$$

where $\mathcal{O}_j = \mathcal{O}(\{\sigma_i\}_j)$ is the measured value for the jth configuration and N is the number of measurement sweeps. Also a warning was given that this is only valid after a sufficiently long equilibration period without measurements, which is needed by the system to approach equilibrium after starting the Markov chain in an arbitrary initial configuration.

This initial equilibration or thermalization period, in general, is a non-trivial non-equilibrium process which is of interest in its own right and no simple general recipe determining how long one should wait before starting measurements can be given. Long suspected to be a consequence of the slow dynamics of glassy systems only, the phenomenon of ageing for example has also been found in the phase-ordering kinetics of simple ferromagnets such as the Ising model. To study this effect numerically, one only needs the methods introduced so far since most theoretical concepts assume a local spin-flip dynamics as realized by one of the three local update algorithms discussed above. Similarly to the concept of universality classes in equilibrium, all three algorithms should yield qualitatively similar results, being representatives of what is commonly referred to as dynamical Glauber universality class.

Let us assume that we pick as the initial configuration of the Markov chain a completely disordered state. If the simulation is run at a temperature $T > T_c$, equilibration will, in fact, be fast and nothing spectacular happens. If, however, we choose instead to perform the simulation right at T_c or at a temperature $T < T_c$, the situation is quite different. In the latter two cases one speaks of a quench, since now the starting configuration is in a statistical sense far away from a typical equilibrium configuration at temperature T. This is easiest to understand for temperatures $T < T_c$, where the typical equilibrium state consists of homogeneously ordered configurations. After the quench, local regions of parallel spins start forming domains or clusters, and the non-equilibrium dynamics of the system is governed by the movement of the domain walls. In order to minimize their surface energy, the domains grow and straighten their surface. A typical time evolution for the 2D Ising model is illustrated in Fig. 4.5, showing spin configurations after a quench to $T < T_c$, starting from an initially completely disordered state.

This leads to a growth law for the typical correlation length scale of the form $\xi \sim t^{1/z}$, where t is the time (measured in units of sweeps) elapsed since the quench
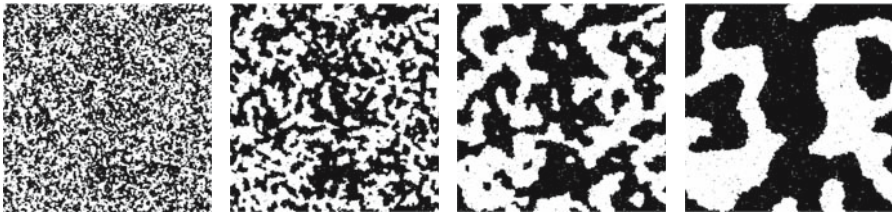
Fig. 4.5. Phase-ordering with progressing Monte Carlo time (from left to right) of an initially disordered spin configuration for the 2D Ising model at $T = 1.5 \approx 0.66\,T_c$ [93]

and z is the dynamical critical exponent already introduced in Sect. 4.3.2. In the case of a simple ferromagnet like the Ising- or q-state Potts model with a non-conserved scalar order parameter, below T_c the dynamical exponent can be found exactly as $z = 2$ [94], according to diffusion or random-walk arguments. Right at the transition temperature, critical dynamics (for a recent review, see [95]) plays the central role and the dynamical exponent of, e.g., the 2D Ising model takes the somewhat larger non-trivial value $z \approx 2.17$ [78, 79] cf. Table 4.2. To equilibrate the whole system, ξ must approach the system size L, so that the typical relaxation time for equilibration scales as

$$\tau_{\text{relax}} \sim L^z \ . \tag{4.51}$$

Note that this implies in the infinite-volume limit $L \to \infty$ that true equilibrium can never be reached.

Since $1/z < 1$, the relaxation process after the quench happens on a growing time scale. This can be revealed most clearly by measurements of two-time quantities $f(t, s)$ with $t > s$, which no longer transform time-translation invariantly as they would do for small perturbations in equilibrium, where f would be a function of the time difference $t - s$ only. Instead, in phase-ordering kinetics, two-time quantities depend non-trivially on the ratio t/s of the two times. The dependence of the relaxation on the so-called waiting time s is the notional origin of ageing: Older samples respond more slowly.

For the most commonly considered two-time quantities, dynamical scaling forms can be theoretically predicted (for recent reviews see, e.g., [96, 97]). Well studied are the two-time autocorrelation function (here in q-state Potts model notation)

$$C(t, s) = \frac{1}{q-1} \left(\frac{q}{V} \sum_{i=1}^{V} \left[\delta_{\sigma_i(t), \sigma_i(s)} \right]_{\text{av}} - 1 \right) = s^{-b} f_C(t/s) \ , \tag{4.52}$$

with the asymptotic behavior $f_C(x) \to x^{-\lambda_C/z}$ $(x \gg 1)$, and the two-time response function

$$R(t, s) = \left. \frac{\delta \left[\sigma_i(t) \right]_{\text{av}}}{\delta h_i(s)} \right|_{h=0} = s^{-1-a} f_R(t/s) \ , \tag{4.53}$$

where $f_R(x) \to x^{-\lambda_R/z}$ $(x \gg 1)$. Here $h(s)$ is the amplitude of a small spatially random external field which is switched off after the waiting time s and $[\ldots]_{\text{av}}$

denotes an average over different random initial configurations (and random fields in (4.53)). In phase-ordering kinetics after a quench to $T < T_c$, in general $b = 0$ (and $z = 2$) [94], but all other exponents depend on the dimensionality of the considered system. In the simplest case of the Ising model in two dimensions, it is commonly accepted that $\lambda_C = \lambda_R = 5/4$. The value of the remaining exponent a, however, is more controversial [98, 99], with strong claims for $a = 1/z = 1/2$ [96, 100], but also $a = 1/4$ [101, 102] has been conjectured. In computer simulation studies the two-time response function is rather difficult to handle and it is more convenient to consider the integrated response or thermoremanent magnetization (TRM) [103],

$$\rho(t, s) = T \int_0^s \mathrm{d}u R(t, u) = \frac{T}{h} M_{\mathrm{TRM}}(t, s) \ . \tag{4.54}$$

By extending dynamical scaling to local scale invariance (LSI) in analogy to conformal invariance [104], even explicit expressions of the scaling functions $f_C(x)$ and $f_R(x)$ have been predicted [105] (for a recent review, see [106]). For the 2D and 3D Ising model, extensive numerical tests of the LSI predictions have been performed by Henkel, Pleimling and collaborators [107, 108, 109], showing a very good agreement with the almost parameter-free analytical expressions. Recently this could be confirmed also for more general q-state Potts models with $q = 3$ and $q = 8$ in two dimensions [93, 110].

If one is primarily interested in equilibrium properties of the considered statistical system, there is, of course, no need to study the initial equilibration period in such a great detail. It is, however, advisable to watch the time evolution of the system and to make sure that no apparent trends are still visible when starting the measurements. If estimates of the autocorrelation or relaxation time are available, a good a priori estimate is to wait at least about $20\,\tau_{\mathcal{O},\exp}$. Finally, as a (not further justified) rule of thumb, most practicers of Monte Carlo simulations spend at least about 10% of the total computing time on the equilibration or thermalization period.

4.5.2 Statistical Errors and Autocorrelation Times

4.5.2.1 Estimators

As already indicated in (4.50), conceptually it is important to distinguish between the expectation value $\langle \mathcal{O} \rangle$ and the mean value $\overline{\mathcal{O}}$, which is an estimator for the former. While $\langle \mathcal{O} \rangle$ is an ordinary number and represents the exact result (which is usually unknown, of course), the estimator $\overline{\mathcal{O}}$ is still a random number which for finite N fluctuates around the theoretically expected value. Of course, from a single Monte Carlo simulation with N measurements, we obtain only a single number for $\overline{\mathcal{O}}$ at the end of the day. To estimate the statistical uncertainty due to the fluctuations, i.e., the statistical error bar, it seems at first sight that one would have to repeat the whole simulation many times. Fortunately, this is not necessary since one can estimate the variance of $\overline{\mathcal{O}}$,

$$\sigma_{\overline{\mathcal{O}}}^2 = \langle [\overline{\mathcal{O}} - \langle \overline{\mathcal{O}} \rangle]^2 \rangle = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 \ , \qquad (4.55)$$

from the statistical properties of individual measurements $\mathcal{O}_i, i = 1, \dots, N$, in a single Monte Carlo run.

4.5.2.2 Uncorrelated Measurements and Central-Limit Theorem

Inserting (4.50) into (4.55) gives

$$\sigma_{\overline{\mathcal{O}}}^2 = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \mathcal{O}_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \ , \qquad (4.56)$$

and by collecting diagonal and off-diagonal terms one arrives at [111]

$$\sigma_{\overline{\mathcal{O}}}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right) \ . \qquad (4.57)$$

Assuming equilibrium, the individual variances $\sigma_{\mathcal{O}_i}^2 = \langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2$ do not depend on "time" i, such that the first term gives $\sigma_{\mathcal{O}_i}^2 / N$. The second term with $\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle = \langle (\mathcal{O}_i - \langle \mathcal{O}_i \rangle)(\mathcal{O}_j - \langle \mathcal{O}_j \rangle) \rangle$ records the correlations between measurements at times i and j. For completely uncorrelated data (which is, of course, an unrealistic assumption for importance sampling Monte Carlo simulations), the second term would vanish and (4.57) simplifies to

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \sigma_{\mathcal{O}_i}^2 / N \ . \qquad (4.58)$$

This result is true for any distribution $\mathcal{P}(\mathcal{O}_i)$. In particular, for the energy or magnetization, distributions of the individual measurements are often plotted as physically directly relevant (N independent) histograms (see, e.g., Fig. 4.8(b) below) whose squared width ($= \sigma_{\mathcal{O}_i}^2$) is proportional to the specific heat or susceptibility, respectively.

Whatever form the distribution $\mathcal{P}(\mathcal{O}_i)$ assumes (which, in fact, is often close to Gaussian because the $\mathcal{O}_i$ are usually already lattice averages over many degrees of freedom), by the central limit theorem the distribution of the mean value is Gaussian, at least for uncorrelated data in the asymptotic limit of large N. The variance of the mean, $\sigma_{\overline{\mathcal{O}}}^2$, is the squared width of this (N dependent) distribution which is usually taken as the one-sigma squared error, $\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2$, and quoted together with the mean value $\overline{\mathcal{O}}$. Under the assumption of a Gaussian distribution for the mean, the interpretation is that about 68% of all simulations under the same conditions would yield a mean value in the range $[\overline{\mathcal{O}} - \sigma_{\overline{\mathcal{O}}}, \overline{\mathcal{O}} + \sigma_{\overline{\mathcal{O}}}]$ [113]. For a two-sigma interval which also is sometimes used, this percentage goes up to about 95.4%, and for a three-sigma interval which is rarely quoted, the confidence level is higher than 99.7%.

4.5.2.3 Correlated Measurements and Autocorrelation Times

For correlated data the second term in (4.57) does not vanish and things become more involved [114, 115, 116]. Using the symmetry $i \leftrightarrow j$ to reduce the summation $\sum_{i \neq j}^{N}$ to $2 \sum_{i=1}^{N} \sum_{j=i+1}^{N}$, reordering the summation, and using time-translation invariance in equilibrium, one finally obtains [111]

$$\sigma_{\overline{\mathcal{O}}}^2 = \frac{1}{N} \left[\sigma_{\mathcal{O}_i}^2 + 2 \sum_{k=1}^{N} \left(\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle \right) \left(1 - \frac{k}{N} \right) \right] , \qquad (4.59)$$

where, due to the last factor $(1 - k/N)$, the $k = N$ term may be trivially kept in the summation. Factoring out $\sigma_{\mathcal{O}_i}^2$, this can be written as

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \frac{\sigma_{\mathcal{O}_i}^2}{N} 2\tau'_{\mathcal{O},\text{int}} , \qquad (4.60)$$

where we have introduced the (proper) integrated autocorrelation time

$$\tau'_{\mathcal{O},\text{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \left(1 - \frac{k}{N} \right) , \qquad (4.61)$$

with

$$A(k) \equiv \frac{\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle}{\sigma_{\mathcal{O}_i}^2} \xrightarrow{k \to \infty} a e^{-k/\tau_{\mathcal{O},\text{exp}}} \qquad (4.62)$$

being the normalized autocorrelation function ($A(0) = 1$) already introduced in (4.32). Since in any meaningful simulation study $N \gg \tau_{\mathcal{O},\text{exp}}$, $A(k)$ in (4.61) is already exponentially small before the correction term in parentheses becomes important. For simplicity this correction is hence usually omitted (as is the prime of $\tau'_{\mathcal{O},\text{int}}$ in (4.61)) and one employs the following definition for the integrated autocorrelation time

$$\tau_{\mathcal{O},\text{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) . \qquad (4.63)$$

The notion "integrated" derives from the fact that this may be interpreted as a trapezoidal discretization of the (approximate) integral $\tau_{\mathcal{O},\text{int}} \approx \int_0^N dk\, A(k)$. Notice that, in general, $\tau_{\mathcal{O},\text{int}}$ (and also $\tau'_{\mathcal{O},\text{int}}$) is different from $\tau_{\mathcal{O},\text{exp}}$. In fact, one can show [117] that $\tau_{\mathcal{O},\text{int}} \leq \tau_{\mathcal{O},\text{exp}}$ in realistic models. Only if $A(k)$ is a pure exponential, the two autocorrelation times, $\tau_{\mathcal{O},\text{int}}$ and $\tau_{\mathcal{O},\text{exp}}$, coincide (up to minor corrections for small $\tau_{\mathcal{O},\text{int}}$ [58, 111]).

As far as the accuracy of Monte Carlo data is concerned, the important point of (4.60) is that due to temporal correlations of the measurements the statistical error $\epsilon_{\overline{\mathcal{O}}} \equiv \mathcal{O} \Rightarrow \sqrt{\sigma_{\overline{\mathcal{O}}}^2}$ on the Monte Carlo estimator $\overline{\mathcal{O}}$ is enhanced by a factor of $\sqrt{2\tau_{\mathcal{O},\text{int}}}$. This can be rephrased by writing the statistical error similar to the uncorrelated case as $\epsilon_{\overline{\mathcal{O}}} = \sqrt{\sigma_{\mathcal{O}_j}^2 / N_{\text{eff}}}$, but now with a parameter

$$N_{\text{eff}} = N/2\tau_{\mathcal{O},\text{int}} \leq N , \tag{4.64}$$

describing the effective statistics. This shows more clearly that only every $2\tau_{\mathcal{O},\text{int}}$ iterations the measurements are approximately uncorrelated and gives a better idea of the relevant effective size of the statistical sample. In view of the scaling behavior of the autocorrelation time in (4.34), (4.35) or (4.37), it is obvious that without extra care this effective sample size may become very small close to a continuous or first-order phase transition, respectively.

4.5.2.4 Bias

A too small effective sample size does not only affect the error bars, but for some quantities even the mean values can be severely underestimated. This happens for so-called biased estimators, as is for instance the case for the specific heat and susceptibility. The specific heat can be computed as $C = \beta^2 V \left(\langle e^2 \rangle - \langle e \rangle^2 \right) = \beta^2 V \sigma_{e_i}^2$, with the standard estimator for the variance

$$\widehat{\sigma}_{e_i}^2 = \overline{e^2} - \overline{e}^2 = \overline{(e - \overline{e})^2} = \frac{1}{N} \sum_{i=1}^{N} (e_i - \overline{e})^2 . \tag{4.65}$$

Subtracting and adding $\langle \overline{e} \rangle^2$, one finds for the expectation value

$$\langle \widehat{\sigma}_{e_i}^2 \rangle = \langle \overline{e^2} - \overline{e}^2 \rangle = \left(\langle \overline{e^2} \rangle - \langle \overline{e} \rangle^2 \right) - \left(\langle \overline{e}^2 \rangle - \langle \overline{e} \rangle^2 \right) = \sigma_{e_i}^2 + \sigma_{\overline{e}}^2 . \tag{4.66}$$

Using (4.60) this gives

$$\langle \widehat{\sigma}_{e_i}^2 \rangle = \sigma_{e_i}^2 \left(1 - \frac{2\tau_{e,\text{int}}}{N} \right) = \sigma_{e_i}^2 \left(1 - \frac{1}{N_{\text{eff}}} \right) \neq \sigma_{e_i}^2 . \tag{4.67}$$

The estimator $\widehat{\sigma}_{e_i}^2$ in (4.65) thus systematically underestimates the true value by a term of the order of $\tau_{e,\text{int}}/N$. Such an estimator is called weakly biased (weakly because the statistical error $\propto 1/\sqrt{N}$ is asymptotically larger than the systematic bias; for medium or small N, however, also prefactors need to be carefully considered).

We thus see that for large autocorrelation times or equivalently small effective statistics N_{eff}, the bias may be quite large. Since $\tau_{e,\text{int}}$ scales quite strongly with the system size for local update algorithms, some care is necessary when choosing the run time N. Otherwise the FSS of the specific heat or susceptibility and thus the determination of the static critical exponent α/ν or γ/ν could be completely spoiled by the temporal correlations [118]! Any serious simulation study should therefore provide at least a rough order-of-magnitude estimate of autocorrelation times.

4.5.3 Numerical Estimation of Autocorrelation Times

The above considerations show that not only for the error estimation but also for the computation of static quantities themselves, it is important to have control over

autocorrelations. Unfortunately, it is very difficult to give reliable a priori estimates, and an accurate numerical analysis is often too time consuming. As a rough estimate it is about ten times harder to get precise information on dynamic quantities than on static quantities like critical exponents. A (weakly biased) estimator $\widehat{A}(k)$ for the autocorrelation function is obtained as usual by replacing in (4.32) the expectation values (ordinary numbers) by mean values (random variables), e.g., $\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle$ by $\overline{\mathcal{O}_i \mathcal{O}_{i+k}}$. With increasing separation k the relative variance of $\widehat{A}(k)$ diverges rapidly. To get at least an idea of the order of magnitude of $\tau_{\mathcal{O},\mathrm{int}}$ and thus the correct error estimate (4.60), it is useful to record the running autocorrelation time estimator

$$\widehat{\tau}_{\mathcal{O},\mathrm{int}}(k_{\max}) = \frac{1}{2} + \sum_{k=1}^{k_{\max}} \widehat{A}(k) \,, \tag{4.68}$$

which approaches $\tau_{\mathcal{O},\mathrm{int}}$ in the limit of large $k_{\max}$ where, however, its statistical error increases rapidly. As an example, Fig. 4.6(a) shows results for the 2D Ising model from an analysis of the same raw data as in Fig. 4.2.

As a compromise between systematic and statistical errors, an often employed procedure is to determine the upper limit $k_{\max}$ self-consistently by cutting off the summation once $k_{\max} \geq 6\, \widehat{\tau}_{\mathcal{O},\mathrm{int}}(k_{\max})$, where $A(k) \approx \mathrm{e}^{-6} \approx 10^{-3}$. In this case an a priori error estimate is available [116, 119, 120]

$$\epsilon_{\tau_{\mathcal{O},\mathrm{int}}} = \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{2(2k_{\max}+1)}{N}} \approx \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{12}{N_{\mathrm{eff}}}} \,. \tag{4.69}$$

For a 5% relative accuracy one thus needs at least $N_{\mathrm{eff}} \approx 5\,000$ or $N \approx 10\,000\,\tau_{\mathcal{O},\mathrm{int}}$ measurements. For an order of magnitude estimate consider the 2D Ising model on a square lattice with $L = 100$ simulated with a local update algorithm. Close to criticality, the integrated autocorrelation time for this example is of the order of $L^z \approx L^2 \approx 100^2$ (ignoring an priori unknown prefactor of order unity which
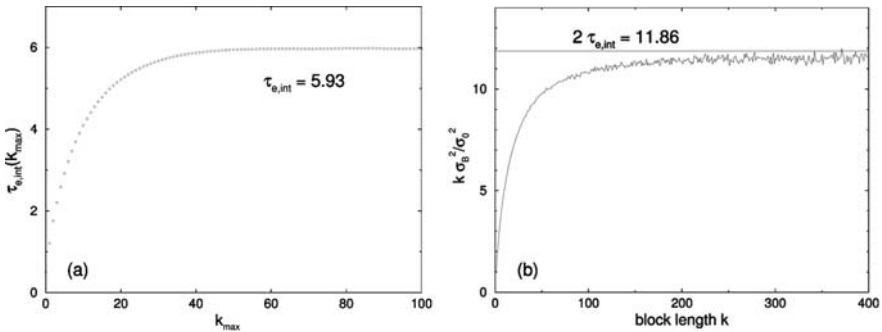

Fig. 4.6. (a) Integrated autocorrelation time approaching $\tau_{e,\mathrm{int}} \approx 5.93$ for large upper cutoff $k_{\max}$ and (b) binning analysis for the energy of the 2D Ising model on a 16×16 lattice at β_c, using the same data as in Fig. 4.2. The *horizontal line* in (b) shows $2\tau_{e,\mathrm{int}}$ with $\tau_{e,\mathrm{int}}$ read off from (a)

depends on the considered quantity), implying $N \approx 10^8$. Since in each sweep L^2 spins have to be updated and assuming that each spin update takes about 0.1 μsec, we end up with a total time estimate of about 10^5 s $\approx$ 1 CPU-day to achieve this accuracy.

An alternative is to approximate the tail end of $A(k)$ by a single exponential as in (4.33). Summing up the small k part exactly, one finds [121]

$$\tau_{\mathcal{O},\text{int}}(k_{\max}) = \tau_{\mathcal{O},\text{int}} - c\,e^{-k_{\max}/\tau_{\mathcal{O},\exp}} , \tag{4.70}$$

where c is a constant. The latter expression may be used for a numerical estimate of both the exponential and integrated autocorrelation times [121].

4.5.4 Binning Analysis

It should be clear by now that ignoring autocorrelation effects can lead to severe underestimates of statistical errors. Applying the full machinery of autocorrelation analysis discussed above, however, is often too cumbersome. On a day by day basis the following binning analysis is much more convenient (though somewhat less accurate). By grouping the N original time-series data into N_B non-overlapping bins or blocks of length k (such that[6] $N = N_B k$), one forms a new, shorter time series of block averages

$$\mathcal{O}_j^{(B)} \equiv \frac{1}{k} \sum_{i=1}^{k} \mathcal{O}_{(j-1)k+i} \tag{4.71}$$

with $j = 1, \ldots, N_B$, which by choosing the block length $k \gg \tau$ are almost uncorrelated and can thus be analyzed by standard means. The mean value over all block averages obviously satisfies $\overline{\mathcal{O}^{(B)}} = \overline{\mathcal{O}}$ and their variance can be computed according to the standard (unbiased) estimator, leading to the squared statistical error of the mean value

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \sigma_B^2/N_B = \frac{1}{N_B(N_B-1)} \sum_{j=1}^{N_B} (\mathcal{O}_j^{(B)} - \overline{\mathcal{O}^{(B)}})^2 . \tag{4.72}$$

By comparing with (4.60) we see that $\sigma_B^2/N_B = 2\tau_{\mathcal{O},\text{int}}\sigma_{\mathcal{O}_i}^2/N$. Recalling the definition of the block length $k = N/N_B$, this shows that one may also use

$$2\tau_{\mathcal{O},\text{int}} = k\sigma_B^2/\sigma_{\mathcal{O}_i}^2 \tag{4.73}$$

for the estimation of $\tau_{\mathcal{O},\text{int}}$. This is demonstrated in Fig. 4.6(b). Estimates of $\tau_{\mathcal{O},\text{int}}$ obtained in this way are often referred to as blocking τ or binning τ.

A simple toy model (bivariate time series), where the behavior of the blocking τ and also of $\tau_{\mathcal{O},\text{int}}(k_{\max})$ for finite k resp. $k_{\max}$ can be worked out exactly, is discussed in [58]. These analytic formulas are very useful for validating the computer implementations.

---

[6] Here we assume that N was chosen cleverly. Otherwise one has to discard some of the data and redefine N.

4.5.5 Jackknife Analysis

Even if the data are completely uncorrelated in time, one still has to handle the problem of error estimation for quantities that are not directly measured in the simulation but are computed as a non-linear combination of basic observables. This problem can either be solved by error propagation or by using the Jackknife method [122, 123], where instead of considering rather small blocks of length k and their fluctuations as in the binning method, one forms N_B large Jackknife blocks $\mathcal{O}_j^{(J)}$ containing all data but the jth block of the previous binning method,

$$\mathcal{O}_j^{(J)} = \frac{N\overline{\mathcal{O}} - k\mathcal{O}_j^{(B)}}{N - k} \tag{4.74}$$

with $j = 1, \ldots, N_B$, cf. the schematic sketch in Fig. 4.7.

Each of the Jackknife blocks thus consists of $N - k$ data, i.e., it contains almost as many data as the original time series. When non-linear combinations of basic variables are estimated, the bias is hence comparable to that of the total data set (typically $1/(N - k)$ compared to $1/N$). The N_B Jackknife blocks are, of course, trivially correlated because one and the same original data enter in $N_B - 1$ different Jackknife blocks. This trivial correlation caused by re-using the original data over and over again has nothing to do with temporal correlations. As a consequence, the Jacknife block variance σ_J^2 will be much smaller than the variance estimated in the binning method. Because of the trivial nature of the correlations, however, this reduction can be corrected by multiplying σ_J^2 with a factor $(N_B - 1)^2$, leading to

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \frac{N_B - 1}{N_B} \sum_{j=1}^{N_B} (\mathcal{O}_j^{(J)} - \overline{\mathcal{O}^{(J)}})^2 \ . \tag{4.75}$$

To summarize this section, any realization of a Markov chain Monte Carlo update algorithm is characterized by autocorrelation times which enter directly into the statistical errors of Monte Carlo estimates. Since temporal correlations always increase the statistical errors, it is thus a very important issue to develop Monte Carlo

Fig. 4.7. Schematic sketch of the organization of Jackknife blocks. The *grey part* of the N data points is used for calculating the total and the Jackknife block averages. The white blocks enter into the more conventional binning analysis using non-overlapping blocks

update algorithms that keep autocorrelation times as small as possible. This is the reason why cluster and other non-local algorithms are so important.

4.6 Reweighting Techniques

The physics underlying reweighting techniques [124, 125] is extremely simple and the basic idea has been known since long (see the list of references in [125]), but their power in practice has been realized only relatively late in 1988. The important observation by Ferrenberg and Swendsen [124, 125] was that the best performance is achieved near criticality where histograms are usually broad. In this sense reweighting techniques are complementary to improved estimators, which usually perform best off criticality.

4.6.1 Single-Histogram Technique

The single-histogram reweighting technique [124] is based on the following very simple observation. If we denote the number of states (spin configurations) that have the same energy E by $\Omega(E)$, the partition function at the simulation point $\beta_0 = 1/k_B T_0$ can always be written as[7]

$$Z(\beta_0) = \sum_{\{s\}} e^{-\beta_0 H(\{s\})} = \sum_E \Omega(E) e^{-\beta_0 E} \propto \sum_E P_{\beta_0}(E) , \qquad (4.76)$$

where we have introduced the unnormalized energy histogram (density)

$$P_{\beta_0}(E) \propto \Omega(E) e^{-\beta_0 E} . \qquad (4.77)$$

If we would normalize $P_{\beta_0}(E)$ to unit area, the r.h.s. would have to be divided by $\sum_E P_{\beta_0}(E) = Z(\beta_0)$, but the normalization will be unimportant in what follows. Let us assume we have performed a Monte Carlo simulation at inverse temperature β_0 and thus know $P_{\beta_0}(E)$. It is then easy to see that

$$P_{\beta}(E) \propto \Omega(E) e^{-\beta E} = \Omega(E) e^{-\beta_0 E} e^{-(\beta-\beta_0)E} \propto P_{\beta_0}(E) e^{-(\beta-\beta_0)E} , \quad (4.78)$$

i.e., the histogram at any point β can be derived, in principle, by reweighting the simulated histogram at β_0 with the exponential factor $\exp[-(\beta-\beta_0)E]$. Notice that in reweighted expectation values

$$\langle f(E) \rangle(\beta) = \frac{\sum_E f(E) P_{\beta}(E)}{\sum_E P_{\beta}(E)} , \qquad (4.79)$$

the normalization of $P_{\beta}(E)$ indeed cancels. This gives, for instance, the energy $\langle e \rangle(\beta) = \langle E \rangle(\beta)/V$ and the specific heat $C(\beta) = \beta^2 V[\langle e^2 \rangle(\beta) - \langle e \rangle(\beta)^2]$, in

principle, as a continuous function of β from a single Monte Carlo simulation at β_0, where $V = L^D$ is the system size.

As an example of this reweighting procedure, using actual Swendsen-Wang cluster simulation data (with 5 000 sweeps for equilibration and 50 000 sweeps for measurements) of the 2D Ising model at $\beta_0 = \beta_c = \ln(1 + \sqrt{2})/2 = 0.440\,686\ldots$ on a 16×16 lattice with periodic boundary conditions, the specific heat $C(\beta)$ is shown in Fig. 4.8(a) and compared with the curve obtained from the exact Kaufman solution [12, 13] for finite $L_x \times L_y$ lattices. This clearly demonstrates that, in practice, the β-range over which reweighting can be trusted is limited. The reason for this limitation are unavoidable statistical errors in the numerical determination of P_{β_0} using a Monte Carlo simulation. In the tails of the histograms the relative statistical errors are largest, and the tails are exactly the regions that contribute most when multiplying $P_{\beta_0}(E)$ with the exponential reweighting factor to obtain $P_\beta(E)$ for β-values far off the simulation point β_0. This is illustrated in Fig. 4.8(b) where the simulated histogram at $\beta_0 = \beta_c$ is shown together with the reweighted histograms at $\beta = 0.375 \approx \beta_0 - 0.065$ and $\beta = 0.475 \approx \beta_0 + 0.035$, respectively. For the 2D Ising model the quality of the reweighted histograms can be judged by comparing with the curves obtained from Beale's [112] exact expression for $\Omega(E)$.

4.6.1.1 Reweighting Range

As a rule of thumb, the range over which reweighting should produce accurate results can be estimated by requiring that the peak location of the reweighted his-
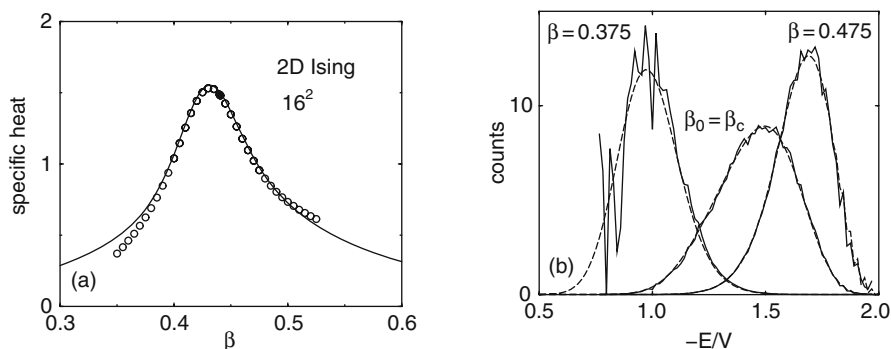

Fig. 4.8. (a) The specific heat of the 2D Ising model on a 16×16 square lattice computed by reweighting from a single Monte Carlo simulation at $\beta_0 = \beta_c$, marked by the filled data symbol. The *continuous line* shows for comparison the exact solution of Kaufman [12, 13]. **(b)** The corresponding energy histogram at β_0, and reweighted to $\beta = 0.375$ and $\beta = 0.475$. The *dashed lines* show for comparison the exact histograms obtained from Beale's expression [112]

---

[7] For simplicity we consider here only models with discrete energies. If the energy varies continuously, sums have to be replaced by integrals, etc. Also lattice size dependences are suppressed to keep the notation short.

togram should not exceed the energy value at which the input histogram had de-
creased to about one half or one third of its maximum value. In most applications
this range is wide enough to locate from a single simulation, e.g., the specific-heat
maximum by employing a standard maximization subroutine to the continuous func-
tion $C(\beta)$. This is by far more convenient, accurate and faster than the traditional
way of performing many simulations close to the peak of $C(\beta)$ and trying to deter-
mine the maximum by spline or least-squares fits.

For an analytical estimate of the reweighting range we now require that the peak
of the reweighted histogram is within the width $\langle e \rangle(T_0) \pm \Delta e(T_0)$ of the input
histogram (where a Gaussian histogram would have decreased to $\exp(-1/2) \approx$
0.61 of its the maximum value)

$$|\langle e \rangle(T) - \langle e \rangle(T_0)| \le \Delta e(T_0) \,, \tag{4.80}$$

where we have made use of the fact that for a not too asymmetric histogram $P_{\beta_0}(E)$
the maximum location approximately coincides with $\langle e \rangle(T_0)$. Recalling that the half
width Δe of a histogram is related to the specific heat via $(\Delta e)^2 \equiv \langle (e - \langle e \rangle)^2 \rangle =$
$\langle e^2 \rangle - \langle e \rangle^2 = C(\beta_0)/\beta_0^2 V$ and using the Taylor expansion $\langle e \rangle(T) = \langle e \rangle(T_0) +$
$C(T_0)(T - T_0) + \ldots$, this can be written as $C(T_0)|T - T_0| \le T_0\sqrt{C(T_0)/V}$ or

$$\frac{|T - T_0|}{T_0} \le \frac{1}{\sqrt{VC(T_0)}} \,. \tag{4.81}$$

Since $C(T_0)$ is known from the input histogram this is quite a general estimate of the
reweighting range. For the example in Fig. 4.8 with $V = 16 \times 16$, $\beta_0 = \beta_c \approx 0.44$
and $C(T_0) \approx 1.5$, this estimate yields $|\beta - \beta_0|/\beta_0 \approx |T - T_0|/T_0 \le 0.04$, i.e.,
$|\beta - \beta_0| \le 0.02$ or $0.42 \le \beta \le 0.46$. By comparison with the exact solution we see
that this is indeed a fairly conservative estimate of the reliable reweighting range.

If we only want to know the scaling behavior with system size $V = L^D$, we can
go one step further by considering three generic cases:

(i) *Off-criticality*, where $C(T_0) \approx$ const, such that

$$\frac{|T - T_0|}{T_0} \propto V^{-1/2} = L^{-D/2} \,. \tag{4.82}$$

(ii) *Criticality*, where $C(T_0) \simeq a_1 + a_2 L^{\alpha/\nu}$, with a_1 and a_2 being constants, and
α and ν denoting the standard critical exponents of the specific heat and cor-
relation length, respectively. For $\alpha > 0$, the leading scaling behavior becomes
$|T - T_0|/T_0 \propto L^{-D/2}L^{-\alpha/2\nu}$. Assuming hyperscaling ($\alpha = 2 - D\nu$) to be
valid, this simplifies to

$$\frac{|T - T_0|}{T_0} \propto L^{-1/\nu} \,, \tag{4.83}$$

i.e., the typical scaling behavior of pseudo-transition temperatures in the finite-
size scaling regime of a second-order phase transition [126]. For $\alpha < 0$, $C(T_0)$
approaches asymptotically a constant and the leading scaling behavior of the
reweighting range is as in the off-critical case.

(iii) *First-order transitions*, where $C(T_0) \propto V$. This yields

$$\frac{|T - T_0|}{T_0} \propto V^{-1} = L^{-D} , \qquad (4.84)$$

which is again the typical finite-size scaling behavior of pseudo-transition temperatures close to a first-order phase transition [47].

4.6.1.2 Reweighting of Non-Conjugate Observables

If we also want to reweight other quantities such as the magnetization $\langle m \rangle$ we have to go one step further. The conceptually simplest way would be to store two-dimensional histograms $P_{\beta_0}(E, M)$ where $M = Vm$ is the total magnetization. We could then proceed in close analogy to the preceding case, and even reweighting to non-zero magnetic field h would be possible, which enters via the Boltzmann factor $\exp(\beta h \sum_i s_i) = \exp(\beta h M)$. However, the storage requirements may be quite high (of the order of V^2), and it is often preferable to proceed in the following way. For any function $g(M)$, e.g., $g(M) = M^k$, we can write

$$\begin{aligned} \langle g(M) \rangle &= \sum_{\{s\}} g(M(\{s\})) \mathrm{e}^{-\beta_0 H} / Z(\beta_0) \\ &= \sum_{E,M} \Omega(E, M) g(M) \mathrm{e}^{-\beta_0 E} / Z(\beta_0) \\ &= \sum_E \frac{\sum_M \Omega(E, M) g(M)}{\sum_M \Omega(E, M)} \sum_M \Omega(E, M) \mathrm{e}^{-\beta_0 E} / Z(\beta_0) . \end{aligned} \qquad (4.85)$$

Recalling that $\sum_M \Omega(E, M) \mathrm{e}^{-\beta_0 E} / Z(\beta_0) = \Omega(E) \mathrm{e}^{-\beta_0 E} / Z(\beta_0) = P_{\beta_0}(E)$ and defining the microcanonical expectation value of $g(M)$ at fixed energy E (sometimes denoted as a list)

$$\langle\langle g(M) \rangle\rangle(E) \equiv \frac{\sum_M \Omega(E, M) g(M)}{\sum_M \Omega(E, M)} , \qquad (4.86)$$

we arrive at

$$\langle g(M) \rangle = \sum_E \langle\langle g(M) \rangle\rangle(E) P_{\beta_0}(E) . \qquad (4.87)$$

Identifying $\langle\langle g(M) \rangle\rangle(E)$ with $f(E)$ in (4.79), the actual reweighting procedure is precisely as before. An example for computing $\langle\langle |M| \rangle\rangle(E)$ and $\langle\langle M^2 \rangle\rangle(E)$ using the data of Fig. 4.8 is shown in Fig. 4.9. Mixed quantities, e.g. $\langle E^k M^l \rangle$, can be treated similarly. One caveat of this method is that one has to decide beforehand which lists $\langle\langle g(M) \rangle\rangle(E)$ one wants to store during the simulation, e.g., which powers k in $\langle\langle M^k \rangle\rangle(E)$ are relevant.

An alternative and more flexible method is based on time series. Suppose we have performed a Monte Carlo simulation at β_0 and stored the time series of N
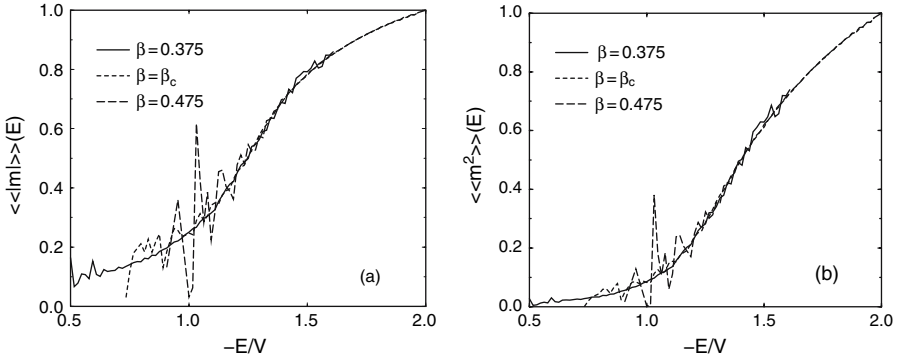
Fig. 4.9. Microcanonical expectation values for **(a)** the absolute magnetization and **(b)** the magnetization squared obtained from the 2D Ising model simulations shown in Fig. 4.8

measurements $E_1, E_2, \ldots, E_N$ and $M_1, M_2, \ldots, M_N$. Then the most general expectation values at another inverse temperature β can simply be obtained from

$$\langle f(E, M) \rangle = \frac{\sum_{i=1}^{N} f(E_i, M_i) \mathrm{e}^{-(\beta - \beta_0) E_i}}{\sum_{i=1}^{N} \mathrm{e}^{-(\beta - \beta_0) E_i}} \,, \qquad (4.88)$$

i.e., in particular all moments $\langle E^k M^l \rangle$ can be computed. Notice that this can also be written as

$$\langle f(E, M) \rangle = \frac{\langle f(E, M) \mathrm{e}^{-(\beta - \beta_0) E} \rangle_0}{\langle \mathrm{e}^{-(\beta - \beta_0) E} \rangle_0} \,, \qquad (4.89)$$

where the subscript zero refers to expectation values taken at β_0. Another very important advantage of the last formulation is that it works without any systematic discretization error also for continuously distributed energies and magnetizations.

As nowadays hard-disk space is no real limitation anymore, it is advisable to store time series in any case. This guarantees the greatest flexibility in the data analysis. As far as the memory requirement of the actual reweighting code is concerned, however, the method of choice is sometimes not so clear. Using directly histograms and lists, one typically has to store about $(6-8) V$ data, while working directly with the time series one needs $2N$ computer words. The cheaper solution (also in terms of CPU time) thus obviously depends on both, the system size V and the run length N. It is hence sometimes faster to generate from the time series first histograms and the required lists and then proceed with reweighting the latter quantities.

4.6.2 Multi-Histogram Technique

The basic idea of the multi-histogram technique [127] can be summarized as follows:

(i) Perform m Monte Carlo simulations at $\beta_1, \beta_2, \ldots, \beta_m$ with N_i, $i = 1, \ldots, m$, measurements.

(ii) Reweight all runs to a common reference point β_0.
(iii) Combine at β_0 all information by computing error weighted averages.
(iv) Reweight the combined histogram to any other β.

Here we shall assume that the histograms $P_{\beta_i}(E)$ are naturally normalized $\sum_E P_{\beta_i}(E) = N_i$, such that the statistical errors for each of the histograms $P_{\beta_i}(E)$ are approximately given by $\sqrt{P_{\beta_i}(E)}$. By choosing as reference point $\beta_0 = 0$ and working out the error weighted combined histogram one ends up with

$$\Omega(E) = \frac{\sum_{i=1}^{m} P_{\beta_i}(E)}{\sum_{i=1}^{m} N_i Z_i^{-1} e^{-\beta_i E}} , \tag{4.90}$$

where the unknown partition function values $Z_i \equiv Z(\beta_i)$ are determined self-consistently from

$$Z_i = \sum_E \Omega(E) e^{-\beta_i E} = \sum_E e^{-\beta_i E} \frac{\sum_{k=1}^{m} P_{\beta_k}(E)}{\sum_{k=1}^{m} N_k Z_k^{-1} e^{-\beta_k E}} , \tag{4.91}$$

up to an unimportant overall constant. A good starting point for the recursion is to fix, say, $Z_1 = 1$ and use single histogram reweighting to get an estimate of $Z_2/Z_1 = \exp[-(\widehat{F}_2 - \widehat{F}_1)]$, where $\widehat{F}_i = \beta_i F(\beta_i)$. Once Z_2 is determined, the same procedure can be applied to estimate Z_3 and so on. In the limit of infinite statistics, this would already yield the solution of (4.91). In realistic simulations the statistics is of course limited and the (very few) remaining recursions average this uncertainty to get a self-consistent set of Z_i. In order to work in practice, the histograms at neighboring β-values must have sufficient overlap, i.e., the spacings of the simulation points must be chosen according to the estimates (4.82)–(4.84).

Multiple-histogram reweighting has been widely applied in many different applications. Some problems of this method are that autocorrelations cannot properly be taken into account when computing the error weighted average (which is still correct but no longer optimized), the procedure for computing mixed quantities such as $\langle E^k M^l \rangle$ is difficult to justify (even though it does work as an ad hoc prescription quite well), and the statistical error analysis becomes quite cumbersome.

As an alternative one may compute by reweighting from each of the m simulations all quantities of interest as a function of β, including their statistical error bars which now also should take care of autocorrelations as discussed in Sect. 4.5.2.3. In this way one obtains, at each β-value, m estimates, e.g. $e_1(\beta) \pm \Delta e_1, e_2(\beta) \pm \Delta e_2, \ldots, e_m(\beta) \pm \Delta e_m$, which may be optimally combined according to their error bars to give $e(\beta) \pm \Delta e$. If the relative error $\Delta e / e(\beta)$ is minimized, this leads to [87]

$$e(\beta) = \left(\frac{e_1(\beta)}{(\Delta e_1)^2} + \frac{e_2(\beta)}{(\Delta e_2)^2} + \ldots + \frac{e_m(\beta)}{(\Delta e_m)^2} \right) (\Delta e)^2 , \tag{4.92}$$

with

$$\frac{1}{(\Delta e)^2} = \frac{1}{(\Delta e_1)^2} + \frac{1}{(\Delta e_2)^2} + \ldots + \frac{1}{(\Delta e_m)^2} . \tag{4.93}$$

Notice that in this way the average for each quantity can be individually optimized.

4.7 Finite-Size Scaling Analysis

Equipped with the various technical tools discussed above, the purpose of this section is to outline a typical FSS analysis of Monte Carlo simulations of second-order phase transitions. The described procedure is generally applicable but to keep the notation short, all formulas are formulated for Ising like systems. For instance for O(n) symmetric models, m should be replaced by $\boldsymbol{m}$ etc. The main results of such studies are usually estimates of the critical temperature and the critical exponents characterizing the universality class of the transition.

4.7.1 General Framework

To facilitate most flexibility in the analysis, it is advisable to store during data production the time series of measurements. Standard quantities are the energy and magnetization, but depending on the model at hand it may be useful to record also other observables. In this way the full dynamical information can be extracted still after the actual simulation runs and error estimation can be easily performed. For example it is no problem to experiment with the size and number of Jackknife bins. Since a reasonable choice depends on the a priori unknown autocorrelation time, it is quite cumbersome to do a reliable error analysis on the flight during the simulation. Furthermore, basing data reweighting on time-series data is more efficient since histograms, if needed or more convenient, can still be produced from this data but working in the reverse direction is obviously impossible.

For some models it is sufficient to perform for each lattice size a single long run at some coupling β_0 close to the critical point β_c. This is, however, not always the case and also depends on the observables of interest. In this more general case, one may use several simulation points β_i and combine the results by the multi-histogram reweighting method or may apply a very recently developed finite-size adapted generalized ensemble method [128]. In both situations, one can compute from the time series of the energies $e = E/V$ (if E happens to be integer valued, this should be stored of course) by reweighting the internal energy $\langle e \rangle(\beta)$, the specific heat $C(\beta) = \beta^2 V \left(\langle e^2 \rangle - \langle e \rangle^2 \right)$, and for instance also the energetic fourth-order parameter

$$V(\beta) = 1 - \frac{\langle e^4 \rangle}{3\langle e^2 \rangle^2} \tag{4.94}$$

as a function of temperature. Similarly, from measurements of the magnetization $m = M/V$ one can derive the temperature variation of the mean magnetization[8] $m(\beta) = \langle |m| \rangle$, the susceptibility $\chi(\beta) = \beta V \left(\langle m^2 \rangle - \langle |m| \rangle^2 \right)$ (or $\chi'(\beta) = \beta V \langle m^2 \rangle$ for $\beta \leq \beta_c$), the magnetic cumulants (Binder parameters)

---

[8] Notice that here and in the following formulas, $|m|$ is used instead of m as would follow from the formal definition (4.5) of the zero-field magnetization $m(\beta) = 1/(V\beta) \lim_{h \to 0} \partial \ln \mathcal{Z}(\beta, h)/\partial h$. The reason is that for a symmetric model on finite lattices one obtains $\langle m \rangle(\beta) = 0$ for all temperatures due to symmetry. Only in the proper infinite-volume limit, that is $\lim_{h \to 0} \lim_{V \to \infty}$, spontaneous symmetry breaking can occur below T_c. In a simulation on finite lattices, this is reflected by a symmetric double-peak

$$U_2(\beta) = 1 - \frac{\langle m^2 \rangle}{3\langle |m| \rangle^2} \ ,$$

$$U_4(\beta) = 1 - \frac{\langle m^4 \rangle}{3\langle m^2 \rangle^2} \ , \tag{4.95}$$

and their slopes

$$\frac{\mathrm{d}U_2(\beta)}{\mathrm{d}\beta} = \frac{V}{3\langle |m| \rangle^2} \left[\langle m^2 \rangle \langle e \rangle - 2\frac{\langle m^2 \rangle \langle |m|e \rangle}{\langle |m| \rangle} + \langle m^2 e \rangle \right]$$

$$= V(1 - U_2) \left[\langle e \rangle - 2\frac{\langle |m|e \rangle}{\langle |m| \rangle} + \frac{\langle m^2 e \rangle}{\langle m^2 \rangle} \right] \ ,$$

$$\frac{\mathrm{d}U_4(\beta)}{\mathrm{d}\beta} = V(1 - U_4) \left[\langle e \rangle - 2\frac{\langle m^2 e \rangle}{\langle m^2 \rangle} + \frac{\langle m^4 e \rangle}{\langle m^4 \rangle} \right] \ . \tag{4.96}$$

Further quantities with a useful FSS behavior are the derivatives of the magnetization,

$$\frac{\mathrm{d}\langle |m| \rangle}{\mathrm{d}\beta} = V\left(\langle |m|e \rangle - \langle |m| \rangle \langle e \rangle \right) \ ,$$

$$\frac{\mathrm{d}\ln\langle |m| \rangle}{\mathrm{d}\beta} = V\left(\frac{\langle |m|e \rangle}{\langle |m| \rangle} - \langle e \rangle \right) \ ,$$

$$\frac{\mathrm{d}\ln\langle m^2 \rangle}{\mathrm{d}\beta} = V\left(\frac{\langle m^2 e \rangle}{\langle m^2 \rangle} - \langle e \rangle \right) \ . \tag{4.97}$$

These latter five quantities are good examples for expectation values containing both, powers of e and m.

In the infinite-volume limit most of these quantities exhibit singularities at the transition point. As already discussed in Sect. 4.2, in finite systems the singularities are smeared out and the standard observables scale in the critical region according to

$$C = C_{\text{reg}} + L^{\alpha/\nu} f_C(x)[1 + \ldots] \ , \tag{4.98}$$

$$\langle |m| \rangle = L^{-\beta/\nu} f_m(x)[1 + \ldots] \ , \tag{4.99}$$

$$\chi = L^{\gamma/\nu} f_\chi(x)[1 + \ldots] \ , \tag{4.100}$$

where C_{reg} is a regular background term, α, ν, β (in the exponent of L) and γ are the usual critical exponents, and $f_i(x)$ are FSS functions with

---

structure of the magnetization distribution (provided the runs are long enough). By averaging m one thus gets zero by symmetry, while the peak locations $\pm m_0(L)$ are close to the spontaneous magnetization and the average of $|m|$ is a good estimator. Things become more involved for slightly asymmetric models, where this recipe would produce a systematic error and thus cannot be employed. For strongly asymmetric models, on the other hand, one peak clearly dominates and the average of m can usually be measured without too many problems.

$$x = (\beta - \beta_c)L^{1/\nu} \tag{4.101}$$

being the scaling variable (do not confuse the unfortunate double-meaning of β – here $\beta = 1/k_B T$). The brackets $[1+\ldots]$ indicate corrections-to-scaling terms which become unimportant for sufficiently large system sizes L.

A particular role play the magnetic cumulants or Binder parameters U_2 and U_4 which scale according to

$$U_{2p} = f_{U_{2p}}(x)[1+\ldots] , \tag{4.102}$$

i.e., for constant scaling variable x, they take approximately the same value for all lattice sizes, in particular $U_{2p}^* \equiv f_{U_{2p}}(0)$ at β_c. Their curves as function of temperature for different L hence cross around (β_c, U_{2p}^*) (with slopes $\propto L^{1/\nu}$), apart from corrections-to-scaling collected in $[1+\ldots]$ which explain small systematic deviations. From a determination of this crossing point, one thus obtains a basically unbiased estimate of β_c, the critical exponent ν, and U_{2p}^*. Note that in contrast to the truly universal critical exponents, U_{2p}^* is only weakly universal. By this one means that the infinite-volume limit of such quantities does depend in particular on the boundary conditions and geometrical shape of the considered lattice, e.g., on the aspect ratio $r = L_y/L_x$ [129, 130, 131, 132, 133, 134, 135, 136].

Differentiating U_{2p} with respect to β, one picks up an extra power of L from the scaling function, $dU_{2p}/d\beta = (dx/d\beta)f'_{U_{2p}} = L^{1/\nu}f'_{U_{2p}}$. This leads to

$$\frac{dU_{2p}}{d\beta} = L^{1/\nu}f'_{U_{2p}}(x)[1+\ldots] , \tag{4.103}$$

and similarly for the magnetization derivatives

$$\frac{d\langle|m|\rangle}{d\beta} = L^{(1-\beta)/\nu}f'_m(x)[1+\ldots] , \tag{4.104}$$

$$\frac{d\ln\langle|m|^p\rangle}{d\beta} = L^{1/\nu}f_{dm_p}(x)[1+\ldots] . \tag{4.105}$$

By applying standard reweighting techniques to the time-series data one first determines the temperature dependence of $C(\beta)$, $\chi(\beta)$, $\ldots$, in the neighborhood of the simulation point $\beta_0 \approx \beta_c$ (a reasonably good guess of β_0 can usually be obtained quite easily from a few short test runs). It should be stressed that in a serious study, by estimating the valid reweighting range, one should at any rate make sure that no systematic errors crept in by this procedure (which may be easily overlooked if one works too mechanically). Once the temperature dependence is known, one can determine the maxima, e.g., $C_{\max}(\beta_{\max_C}) \equiv \max_\beta C(\beta)$, by applying standard extremization routines: When reweighting is implemented as a subroutine, for instance $C(\beta)$ can be handled as a normal function with a continuously varying argument β, i.e., no interpolation or discretization error is involved when iterating towards the maximum. The locations of the maxima of C, χ, $dU_2/d\beta$, $dU_4/d\beta$, $d\langle|m|\rangle/d\beta$, $d\ln\langle|m|\rangle/d\beta$, and $d\ln\langle m^2\rangle/d\beta$ provide us with

seven sequences of pseudo-transition points $\beta_{\max_i}(L)$ which all should scale according to $\beta_{\max_i}(L) = \beta_c + a_i L^{-1/\nu} + \dots$. In other words, the scaling variable $x = (\beta_{\max_i}(L) - \beta_c)L^{1/\nu} = a_i + \dots$ should be constant, if we neglect the small higher-order corrections indicated by $\dots$.

Notice that while the precise estimates of a_i do depend on the value of ν, the qualitative conclusion that $x \approx$ const for each of the $\beta_{\max_i}(L)$ sequences does not require any a priori knowledge of ν or β_c. Using this information one thus has several possibilities to extract unbiased estimates of the critical exponents ν, α/ν, β/ν, and γ/ν from least-squares fits assuming the FSS behaviors (4.98)–(4.105).

4.7.2 A Practical Recipe

The typical procedure of an FSS analysis then proceeds as follows:

(i) Estimate the critical exponent ν by least-square fits to the scaling behavior (4.103) and (4.105). For this one may consider directly the maxima of $dU_{2p}/d\beta$ and $d\ln\langle|m|^p\rangle/d\beta$, $p = 1, 2$, or work with any other FSS sequence $\beta_{\max_i}(L)$.

Remarks: Considering only the asymptotic behavior, e.g., $d\ln\langle|m|\rangle/d\beta = aL^{1/\nu}$, and taking the logarithm, $\ln(d\ln\langle|m|\rangle/d\beta) = c + (1/\nu)\ln(L)$, one ends up with a linear two-parameter fit yielding estimates for the constant $c = \ln(a)$ and the exponent $1/\nu$. For small lattice sizes the asymptotic ansatz is, of course, not justified. Taking into account the (effective) correction term $[1 + bL^{-w}]$ would result in a non-linear four-parameter fit for a, b, $1/\nu$ and w. Even if we would fix w to some theoretically expected value (as is sometimes done), we would be still left with a non-linear fit which is usually much harder to control than a linear fit (where only a set of linear equations with a unique solution has to be solved, whereas a non-linear fit involves a numerical minimization of the χ^2-function, possessing possibly several local minima). The alternative method is to use the linear fit ansatz and to discard successively more and more small lattice sizes until the χ^2 per degree of freedom or the goodness-of-fit Q [113] has reached an acceptable value and does not show any further trend. Of course, all this relies heavily on correct estimates of the statistical error bars on the original data for $d\ln\langle|m|\rangle/d\beta$.

Furthermore, when combining the various fit results for ν to a final value, some care is necessary with the final statistical error estimate on ν, since the various fits for determining ν are of course correlated (since they use the data from one and the same simulation).

(ii) Once ν is estimated one can use the scaling form $\beta_{\max_i}(L) = \beta_c + a_i L^{-1/\nu} + \dots$ to extract β_c and a_i. As a useful check, one should repeat these fits at the error margins of ν, but usually this dependence turns out to be very weak.

Remark: Regarding the β_c fit alone, the uncertainty in the proper value of ν looks like a kind of systematic error or bias, whose origin, however, is also of statistical nature occurring in the first step.

(iii) As a useful cross-check one can determine β_c also from the Binder parameter crossings. For a first rough estimate, this is a very convenient and fast method.

Remarks: As a rule of thumb, an accuracy of about 3–4 digits for β_c can be obtained with this method without any elaborate infinite-volume extrapolations – the crossing points lie usually much closer to β_c than the various maxima locations. For high precision, however, it is quite cumbersome to control the necessary extrapolations and often more accurate estimates can be obtained by considering the scaling of the maxima locations. Also, error estimates of crossing points involve the data for two different lattice sizes which tends to be quite unhandy.

(iv) Next, similarly to ν, the ratios of critical exponents α/ν, β/ν, and γ/ν can be obtained from fits to (4.98)–(4.100), and (4.104). Again the maxima of these quantities or any of the FSS sequences $\beta_{\max_i}$ can be used. What concerns the fitting procedure the same remarks apply as for ν.

Remarks: The specific heat C usually plays a special role in that the exponent α is difficult to determine. The reason is that α is usually relatively small (3D Ising model: $\alpha \approx 0.1$), may be zero (logarithmic divergence as in the 2D Ising model) or even negative (as for instance in the 3D XY and Heisenberg models). In all these cases, the constant background contribution C_{reg} in (4.98) becomes important, which enforces a non-linear three-parameter fit with the just described problems. Also for the susceptibility χ, a regular background term cannot be excluded, but it is usually much less important since $\gamma \gg \alpha$. Therefore, in (4.99), (4.100), and (4.104), similar to the fits for ν, one may take the logarithm and work with much more stable linear fits.

(v) As a final step one may re-check the FSS behavior of C, χ, $\mathrm{d}U_2/\mathrm{d}\beta, \ldots$ at the numerically determined estimate of β_c. These fits should be repeated also at $\beta_c \pm \Delta\beta_c$ in order to estimate by how much the uncertainty in β_c propagates into the thus determined exponent estimates.

Remark: In (the pretty rare) cases where β_c is known exactly (e.g., through self-duality), this latter option is by far the most accurate one. This is the reason, why for such models numerically estimated critical exponents are usually also much more precise.

4.7.3 Finite-Size Scaling at Work – An Example

The purpose of this subsection is to illustrate the above outlined recipe with actual data from recent simulations of a 2D Ising model with next-nearest neighbor interactions [137]. The Hamiltonian has the form

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - J_{\mathrm{d}} \sum_{(k,l)} \sigma_k \sigma_l \, , \tag{4.106}$$

where the spins can take the values $\sigma_i = \pm 1$, J denotes the nearest-neighbor (nn) coupling and J_{d} is the next-nearest-neighbor (nnn) coupling along the two diagonals

of a square lattice. The corresponding pairs of spins are denoted by the brackets $\langle i, j \rangle$ and (k, l), respectively. In [137] we restricted ourselves to that region of the phase diagram where the ground states show ferromagnetic order ($J \geq 0$, $J_{\mathrm{d}} \geq -J/2$), and always assumed periodic boundary conditions. Absorbing the nn coupling J into the inverse temperature β (i.e., formally putting $J = 1$), the remaining second parameter is the coupling-constant ratio $\alpha = J_{\mathrm{d}}/J$. In the following we will concentrate on the case $\alpha = 0.5$ [138]. The linear size of the lattices varies from $L = 10, 20, 40, \ldots$ up to 640. All simulations are performed with the single-cluster algorithm which is straightforward to adapt to nnn interactions by setting bonds also along the diagonals. Similarly to the standard nn model, the integrated autocorrelation time close to criticality is found for $\alpha = 1$ [137] to scale only weakly with lattice size: $\tau_{e,\mathrm{int}} \propto L^z$ with $z = 0.134(3)$.

Another example following closely the lines sketched above is provided by a Monte Carlo study of the 3D Ising model, albeit not on a regular but on Poissonnian random lattices of Voronoi-Delaunay type [139]. The random lattices are treated as quenched disorder in the local coordination numbers and hence necessitate an additional average over many realizations (in the study described in [139], for each lattice size 96 independent realizations were used). This introduces in all FSS formulas additional disorder averages which complicate some aspects of the analysis. The general concept of FSS analysis, however, does not depend on this special feature and it may be worthwhile to consult [139] for a supplementary example.

4.7.3.1 Critical Exponent ν

Having recorded the times series of the energy and magnetization, all quantities of the preceding paragraph can be computed in the FSS region. The scaling behavior of the maxima of $\mathrm{d}\ln\langle|m|^p\rangle/\mathrm{d}\beta$ and $\mathrm{d}U_{2p}/\mathrm{d}\beta$ for $p = 1$ and $p = 2$ is shown in
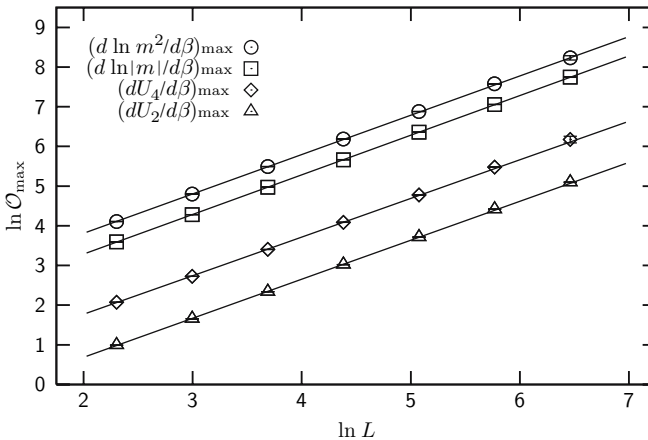
Fig. 4.10. FSS of the maxima of $\mathrm{d}\ln\langle|m|^p\rangle/\mathrm{d}\beta$ and $\mathrm{d}U_{2p}/\mathrm{d}\beta$ for $p = 1$ and $p = 2$ of the 2D nnn Ising model (4.106) with $\alpha = J_{\mathrm{d}}/J = 0.5$ and fits to extract $1/\nu$

Table 4.3. Fit results for the correlation length critical exponent ν of the 2D nnn Ising model with $\alpha = J_\mathrm{d}/J = 0.5$, and the weighted average of the four estimates. Also listed are the χ^2 per degree of freedom, χ^2/d.o.f., and the goodness-of-fit parameter Q [113]

	$\mathrm{d}\ln\langle|m|\rangle/\mathrm{d}\beta$	$\mathrm{d}\ln\langle m^2\rangle/\mathrm{d}\beta$	$\mathrm{d}U_2/\mathrm{d}\beta$	$\mathrm{d}U_4/\mathrm{d}\beta$	weighted av.
ν	1.0031(17)	1.0034(21)	1.0027(24)	1.0025(44)	1.0031(11)
χ^2/d.o.f	0.98	0.60	2.02	0.49	
Q	0.37	0.55	0.13	0.61	

the log-log plot of Fig. 4.10. From the parameters of the four linear fits over the data points with $L_\mathrm{min} > 40$ collected in Table 4.3, we obtain a weighted average of $\nu = 1.003\,1 \pm 0.001\,1$.

As the more detailed analysis in [139] clearly shows, considering all $4 \times 7 = 28$ possible FSS sequences (the four observables shown in Fig. 4.10 evaluated at the seven different β_{max_i} sequences) does not significantly improve the precision of the final estimate. The reason are the quite strong correlations between most of these 28 estimates. In a really large-scale simulation such a more detailed analysis may still be valuable, however, since it potentially helps to detect systematic trends which otherwise may remain unnoticed. Also here the weighted average is clearly dominated by the result from the $\mathrm{d}\ln\langle|m|\rangle/\mathrm{d}\beta$ fit, and correlations between the first and second pair of estimates are obvious. Therefore, to account for these correlations at least heuristically, we usually quote in our investigations the weighted average, but take the smallest contributing error estimate (here thus from the $\mathrm{d}\ln\langle|m|\rangle/\mathrm{d}\beta$ fit). This recipe then gives from Table 4.3 the final result

$$\nu = 1.003\,1 \pm 0.001\,7\,, \tag{4.107}$$

in good agreement with the 2D Ising universality class (cf. Table 4.1).

4.7.3.2 Critical Coupling β_c

Fixing the critical exponent ν at the numerically determined estimate (or in the present context at the exactly known value $\nu = 1$), it is now straightforward to obtain estimates of the critical coupling β_c from linear least-squares fits to

$$\beta_{\mathrm{max}_i} = \beta_c + a_i L^{-1/\nu}\,, \tag{4.108}$$

where β_{max_i} are the pseudo-transition points discussed earlier. Depending on the quantity considered, here we found a significant improvement of the fit quality if the smallest lattice sizes were excluded. This is illustrated in Table 4.4, where detailed results for various fit ranges are compiled.

As final result we quote the weighted average of the five estimates and again the smallest contributing error bar,

$$\beta_c = 0.262\,817\,4 \pm 0.000\,0017\,. \tag{4.109}$$

Table 4.4. FSS fits of the pseudo-transition points $\beta_{\max} = \beta_c + aL^{-1/\nu}$ of the nnn model (4.106) with $\alpha = 0.5$ for varying fit ranges, assuming $\nu = 1$. Here n is the number of data points, $L_{\min}$ denotes the smallest lattice size considered, and Q is the standard goodness-of-fit parameter. The selected fit ranges used for the final average are high-lighted in boldface. The last line labeled HTS gives a high-temperature series estimate [140] for comparison

observables	n	$L_{\min}$	β_c	Q
$\beta_{\max}^{C}$	7	10	0.262 699(13)	0.00
	6	20	0.262 766(15)	0.03
	5	**40**	**0.262 799(18)**	**0.88**
	4	80	0.262 807(22)	0.89
$\beta_{\inf}^{\lvert m \rvert}$	7	10	0.262 8706(36)	0.00
	6	20	0.262 8398(40)	0.00
	5	**40**	**0.262 8272(47)**	**0.16**
	4	80	0.262 8212(58)	0.38
$\beta_{\max}^{\chi}$	7	10	0.262 8253(12)	0.00
	6	20	0.262 8195(13)	0.00
	5	40	0.262 8178(14)	0.09
	4	**80**	**0.262 8153(17)**	**0.66**
$\beta_{\inf}^{\ln \lvert m \rvert}$	7	10	0.262 8437(62)	0.00
	6	**20**	**0.262 8243(68)**	**0.24**
	5	40	0.262 8183(77)	0.42
	4	80	0.262 8099(97)	0.70
$\beta_{\inf}^{\ln m^2}$	7	10	0.262 8684(94)	0.00
	6	**20**	**0.262 837(11)**	**0.43**
	5	40	0.262 837(13)	0.57
	4	80	0.262 818(17)	0.55
average			0.262 8204(144)	
weighted average			0.262 8174(16)	
final			0.262 8174(17)	
HTS (Oitmaa [140])			0.262 808	

The corrections to the asymptotic FSS behavior can be also visually inspected in Fig. 4.11, where the Monte Carlo data and fits are compared. One immediately notices a systematic trend that the $L = 10$ data deviate from the linear behavior. For larger L, however, the deviations are already so small that only a quantitative judgement in terms of the χ^2 per degree of freedom or goodness-of-fit parameter Q of the fits [113] can lead to a sensible conclusion.

4.7.3.3 Binder Parameters U_2 and U_4

Following our general recipe sketched above, the Binder parameter $U_4(L)$ is shown in Fig. 4.12 as a function of temperature. Even though the temperature range is much
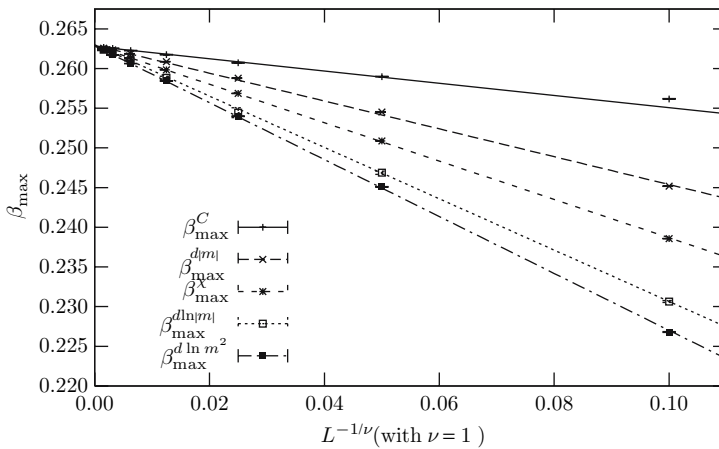
Fig. 4.11. FSS fits of the pseudo-transition points β_{max_i} with $\nu = 1.0$ fixed of the 2D nnn Ising model (4.106) with $\alpha = J_{\mathrm{d}}/J = 0.5$. The error weighted average of the FSS extrapolations yields $\beta_c = 0.262\ 817\ 4(16)$, cf. Table 4.4 for details

smaller than in the β_{max_i} plot of Fig. 4.11, a clear-cut crossing point can be observed. Already from the crossing of the two curves for the very modestly sized lattices with $L = 10$ and $L = 20$ (which can be obtained in a few minutes of computing time), one can read off that $\beta_c \approx 0.262\ 8$. This clearly demonstrates the power of this method, although it should be stressed that the precision is exceptionally good for this model.

On the scale of Fig. 4.12 one reads off that $U_4^* \approx 0.61$. Performing an extrapolation (on a very fine scale) to infinite size at $\beta = \beta_c$, one obtains the more accurate
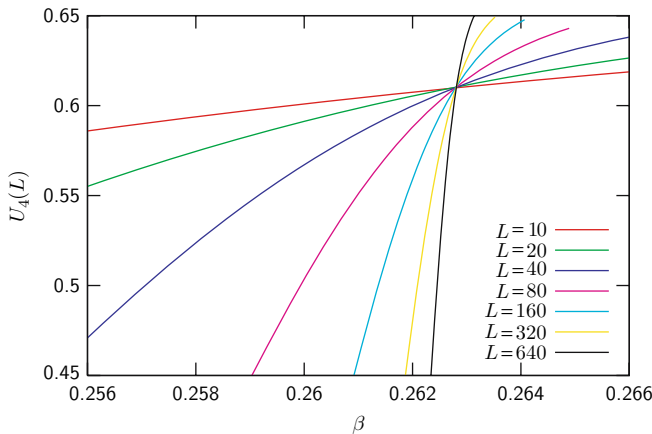
Fig. 4.12. Fourth-order Binder parameter U_4, exhibiting a sharp crossing point around $(\beta_c, U_4^*) \approx (0.262\ 82, 0.61)$. Note the much smaller temperature scale compared to Fig. 4.11

estimate of $U_4^* = 0.610\ 8(1)$. This result for the 2D nnn Ising model with $\alpha = 0.5$ is in perfect agreement with the very precisely known value for the standard square lattice nn Ising model with periodic boundary conditions from extrapolating exact transfer-matrix data for $L \leq 17$ [129], $U_4^* = 0.610\ 690\ 1(5)$, and a numerical evaluation of an exact expression [130], $U_4^* = 0.610\ 692(2)$. This illustrates the expected universality of U_4^* (and also U_2^*) for general isotropic interactions (e.g., also for $\alpha = 1$ one finds the same result within error bars [137]), as long as boundary conditions, lattice shapes etc. are the same. As emphasized already in Sect. 4.7.1, the cumulants are, however, only weakly universal in the sense that they do depend sensitively on the anisotropy of interactions, boundary conditions and lattice shapes (aspect ratios) [131, 132, 133, 134, 135, 136].

4.7.3.4 Critical Exponent γ

The exponent ratio γ/ν can be obtained from fits to the FSS behavior (4.100) of the susceptibility. By monitoring the quality of the fits, using all data starting from $L = 10$ is justified. The fits collected in Table 4.5 all have $Q \geq 0.15$.

Still it is fairly obvious, that the two fits with $Q < 0.2$ have some problems. Discarding them in the averages, one obtains from the weighted average (and again quoting the smallest contributing error estimate to heuristically take into account the correlations among the individual fits)

Table 4.5. Fit results for the critical exponents γ/ν, β/ν, and $(1-\beta)/\nu$. The fits for γ/ν and $(1-\beta)/\nu$ take all lattices with $L \geq 10$ into account while the fits for β/ν start at $L = 20$

at $K_{\max}$ of	γ/ν	Q	β/ν	Q	$(1-\beta)/\nu$	Q
C	1.7574(28)	0.87	0.12856(38)	0.00	0.8889(13)	0.00
χ	1.7407(10)	0.16	0.12480(32)	0.45	0.8710(24)	0.93
$dU_4/d\beta$	1.7700(50)	0.40	0.12481(39)	0.51	0.9154(99)	0.38
$dU_2/d\beta$	1.7417(12)	0.42	0.12562(32)	0.02	0.8815(35)	0.39
$d\langle|m|\rangle/d\beta$	1.7356(11)	0.19	0.12191(33)	0.00	0.8760(15)	0.82
$d\ln\langle|m|\rangle/d\beta$	1.7520(20)	0.62	0.12407(34)	0.02	0.8923(49)	0.57
$d\ln\langle m^2\rangle/d\beta$	1.7630(32)	0.76	0.12363(37)	0.01	0.9047(68)	0.81
average	1.7515(49)	≥ 0	0.12477(78)	≥ 0	0.8900(60)	≥ 0
weighted av.	1.7423(06)		0.12468(13)		0.8822(09)	
final	1.7423(10)		0.12468(32)		0.8822(13)	
average	1.7568(49)	≥ 0.2	0.12483(32)	≥ 0.02	0.8901(71)	≥ 0.3
weighted av.	1.7477(09)		0.12485(17)		0.8775(11)	
final	1.7477(12)		0.12485(32)		0.8775(15)	
exact	1.75		0.125		0.875	

$$\gamma/\nu = 1.747\ 7 \pm 0.001\ 2 \qquad (4.110)$$

to be compared with the exact result $7/4 = 1.75$. For the critical exponent η, the estimate (4.110) implies $\eta = 2 - \gamma/\nu = 0.252\ 3 \pm 0.001\ 2$, and, by inserting our value of $\nu = 1.003\ 1(17)$, one obtains $\gamma = 1.753\ 1 \pm 0.004\ 2$. Here and in the following we are quite conversative and always quote the maximal error, i.e., $\max\{(O_1 + \epsilon_1)(O_2 + \epsilon_2) - O_1 O_2, O_1 O_2 - (O_1 - \epsilon_1)(O_2 - \epsilon_2)\}$.

4.7.3.5 Critical Exponent β

The exponent ratio β/ν can be obtained either from the FSS behavior of $\langle|m|\rangle$ or $\mathrm{d}\langle|m|\rangle/\mathrm{d}\beta$, (4.99) or (4.104). In the first case, Table 4.5 shows that most $\beta_{\max_i}$ sequences yield poor Q values (≤ 0.1) even if the $L = 10$ lattice data is discarded. If one averages only the fits with $Q \geq 0.02$, the final result is

$$\beta/\nu = 0.124\ 85 \pm 0.000\ 32 , \qquad (4.111)$$

and, by using our estimate (4.107) for ν, $\beta = 0.125\ 23 \pm 0.000\ 54$, in very good agreement with the exact result $\beta/\nu = \beta = 1/8 = 0.125\ 00$ for the 2D Ising universality class. Assuming hyperscaling to be valid, the estimate (4.111) implies $\gamma/\nu = D - 2\beta/\nu = 1.750\ 30(64)$.

From the Q values in Table 4.5 one can conclude that the FSS of $\mathrm{d}\langle|m|\rangle/\mathrm{d}\beta$ is somewhat better behaved, so that one can keep again all lattice sizes $L \geq 10$ in the fits. By discarding only the fit for the $\beta_{\max_C}$ sequence, which has an exceptionally small Q value, one arrives at

$$(1 - \beta)/\nu = 0.877\ 5 \pm 0.001\ 5 , \qquad (4.112)$$

so that by inserting our estimate (4.107) for ν, $\beta/\nu = 0.119\ 4 \pm 0.003\ 2$, and finally $\beta = 0.119\ 8 \pm 0.003\ 0$.

4.7.3.6 Critical Exponent α

Due to the regular background term C_{reg} in the FSS behavior (4.98), the specific heat is usually among the most difficult quantities to analyze [141]. In the present example the critical exponent α is expected to be zero, as can be verified by using the hyperscaling relation $\alpha = 2 - D\nu = -0.0062(34)$. In such a situation it may be useful to test at least the consistency of a linear two-parameter fit with α/ν kept fixed. In the present case with $\alpha = 0$, this amounts to the special form $C = C_{\mathrm{reg}} + a \ln(L)$. As can be inspected in Fig. 4.13, the expected linear behavior is, in fact, satisfied over the whole range of lattice sizes.

To conclude this example analysis [138], it should be stressed that no particular care was taken to arrive at high-precision estimates for the critical exponents since in the original work [137] primarily the critical coupling was of interest. In applications aiming also at accurate exponent estimates, one may experiment more elaborately
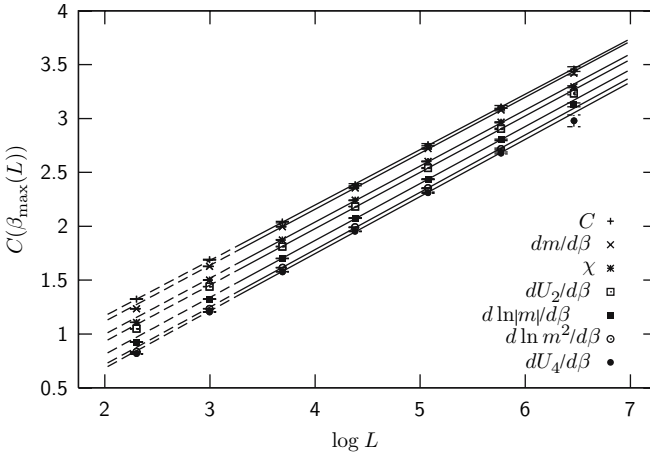
Fig. 4.13. FSS behavior of the specific heat evaluated at the various $\beta_{\max_i}$ sequences, assuming $\alpha = 0$, i.e., a logarithmic scaling $\propto \ln L$

with the fit ranges and averaging procedures. If (small) inconsistencies happen to persist, it is in particular also wise to re-check the extent of the reliable reweighting range, which often turns out to be the source of trouble in the first place (... which we have not seriously attempted to exclude in this example analysis).

4.7.4 Spatial Correlation Length

Since critical phenomena are intimately connected with diverging spatial correlations, it is in many applications important to also estimate the correlation length. In the high-temperature phase down to the critical point, we have $\langle \sigma_i \rangle = 0$ and the two-point correlation function (4.7) simplifies to

$$G(\boldsymbol{r}_i - \boldsymbol{r}_j) = \langle \sigma_i \sigma_j \rangle \ . \tag{4.113}$$

By summing over all lattice points one obtains the susceptibility (without β prefactor)

$$\chi'/\beta = \frac{1}{V} \sum_{\boldsymbol{r}_i, \boldsymbol{r}_j} G(\boldsymbol{r}_i - \boldsymbol{r}_j) = \sum_{\boldsymbol{r}} G(\boldsymbol{r})$$

$$= V \left\langle \left(\frac{1}{V} \sum_{\boldsymbol{r}} \sigma_i \right)^2 \right\rangle = V \langle m^2 \rangle \ . \tag{4.114}$$

Recall that above T_c, $\langle m \rangle = 0$. On D-dimensional periodic lattices with edge lengths $L_1 = L_2 = \ldots = L$, the correlation function can be decomposed into Fourier modes

$$G(\boldsymbol{r}_i - \boldsymbol{r}_j) = \frac{1}{V} \sum_{n_1,n_2,\ldots=0}^{L-1} \widehat{G}(\boldsymbol{k}) \mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{r}_i-\boldsymbol{r}_j)} \,, \tag{4.115}$$

where $\boldsymbol{k} \equiv (2\pi/L)(n_1, n_2, \ldots)$ are the discrete lattice momenta. In the high-temperature phase the amplitudes for long-wavelength modes ($|\boldsymbol{k}| \to 0$) are effectively given by

$$\widehat{G}(\boldsymbol{k}) = a \left[\sum_{i=1}^{D} 2(1 - \cos k_i) + m^2 \right]^{-1} \overset{|\boldsymbol{k}|\to 0}{\approx} a \frac{1}{\boldsymbol{k}^2 + m^2} \,, \tag{4.116}$$

with β dependent prefactor a and mass parameter m. Inserting this into (4.114), one finds for large distances $|\boldsymbol{r}| \gg 1$ (but $|\boldsymbol{r}| \ll L/2$ for finite periodic lattices)

$$G(\boldsymbol{r}) \propto |\boldsymbol{r}|^{-(D-1)/2} \mathrm{e}^{-m|\boldsymbol{r}|} \tag{4.117}$$

with ($|\boldsymbol{r}| \gg 1$), so that the inverse mass can be identified as the correlation length $\xi \equiv 1/m$.

4.7.4.1 Zero-Momentum Projected Correlation Function

In order to avoid the power-like prefactor in (4.117) and to increase effectively the statistics one actually measures in most applications a so-called projected (zero-momentum) correlation function defined by ($\boldsymbol{r} = (x_1, x_2, \ldots)$)

$$g(x_1 - x_1')$$
$$= \frac{1}{L^{D-1}} \sum_{x_2,x_3,\ldots=1}^{L} \sum_{x_2',x_3',\ldots=1}^{L} G(\boldsymbol{r}_i - \boldsymbol{r}_j)$$
$$= L^{D-1} \left\langle \left[\frac{1}{L^{D-1}} \sum_{x_2,x_3,\ldots=1}^{L} \sigma_{x_1,x_2,x_3,\ldots} \right] \left[\frac{1}{L^{D-1}} \sum_{x_2',x_3',\ldots=1}^{L} \sigma_{x_1',x_2',x_3',\ldots} \right] \right\rangle \,, \tag{4.118}$$

i.e., the correlations of line magnetizations $L^{-1} \sum_{x_2=1}^{L} \sigma_{x_1,x_2}$ for 2D systems or surface magnetizations $L^{-2} \sum_{x_2,x_3=1}^{L} \sigma_{x_1,x_2,x_3}$ for 3D systems at x_1 and x_1'. Notice that in all dimensions

$$\chi'/\beta = \frac{1}{2} g(0) + \sum_{i=1}^{L-1} g(i) + \frac{1}{2} g(L) \tag{4.119}$$

is given by the trapezoidal approximation to the area $\int_0^L g(x)\mathrm{d}x$ under the projected correlation function $g(x)$. Applying the summations in (4.118) to the Fourier decomposition of $G(\boldsymbol{r}_i - \boldsymbol{r}_j)$ and using

$$\frac{1}{L^{D-1}} \sum_{x_2, x_3, \ldots = 1}^{L} e^{ik_2 x_2 + ik_3 x_3 + \cdots} = \delta_{k_2, 0} \, \delta_{k_3, 0} \cdots , \qquad (4.120)$$

it is easy to see that

$$g(x_1 - x_1') = \frac{a}{L} \sum_{n_1 = 0}^{L-1} \frac{e^{ik_1 (x_1 - x_1')}}{2(1 - \cos k_1) + m^2} \qquad (4.121)$$

is the one-dimensional version of (4.115) and (4.116), since all but one momentum component are projected to zero in (4.120). This can be evaluated exactly as

$$\begin{aligned} g(x) &= \frac{a}{2 \sinh m^*} \frac{\cosh[m^* (L/2 - x)]}{\sinh(m^* L/2)} \\ &= \frac{a}{2 \sinh m^*} \left[e^{-m^* x} + \frac{2 e^{-m^* L}}{1 - e^{-m^* L}} \cosh(m^* x) \right] , \end{aligned} \qquad (4.122)$$

with m and m^* related by

$$\begin{aligned} \frac{m}{2} &= \sinh \left(\frac{m^*}{2} \right) , \\ \frac{m^*}{2} &= \ln \left[\frac{m}{2} + \sqrt{\left(\frac{m}{2} \right)^2 + 1} \right] . \end{aligned} \qquad (4.123)$$

For $\xi > 10$ ($m < 0.1$) the difference between ξ and $\xi^* \equiv 1/m^*$ is completely negligible, $(\xi^* - \xi)/\xi < 0.042\%$. Notice that there is no x-dependent prefactor in (4.122). And also note that $G(r)$ computed for r along one of the coordinate axes is a truly D-dimensional correlation function (albeit along some special direction), exhibiting the $r^{(D-1)/2}$ prefactor of (4.117).

Figure 4.14 shows as an example $g(x)$ for the standard nn Ising model at $T = 2.5 \approx 1.1 \, T_c$ on a 50×50 square lattice. By fitting the Monte Carlo data to the cosh-form (4.122), $m^* = 0.167 \, 9$ is obtained or $\xi^* = 5.957$. Inserting this value into (4.123), one obtains $\xi = 1/m = 5.950$. This is in very good agreement (at a 0.1-0.2% level) with the exactly known correlation length (of the two-dimensional correlation function) along one of the two main coordinate axes, $\xi_{||}^{(ex)} = -1/(\ln(\tanh(\beta)) + 2\beta) = 5.962 \, 376 \, 984 \ldots$ [14, 15].

4.7.4.2 Second-Moment Correlation Length

Alternatively, one may also measure directly the Fourier amplitudes

$$\widehat{G}(\mathbf{k}) = \sum_{\mathbf{r}} G(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} = \frac{1}{V} \langle |\widehat{\sigma}(\mathbf{k})|^2 \rangle , \qquad (4.124)$$

for a few long-wavelength modes $\widehat{\sigma}(\mathbf{k}) = \sum_{\mathbf{r}} \sigma(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}$, where the normalization is chosen such that $\widehat{G}(\mathbf{0}) = \chi'/\beta$. From (4.116) we read off that
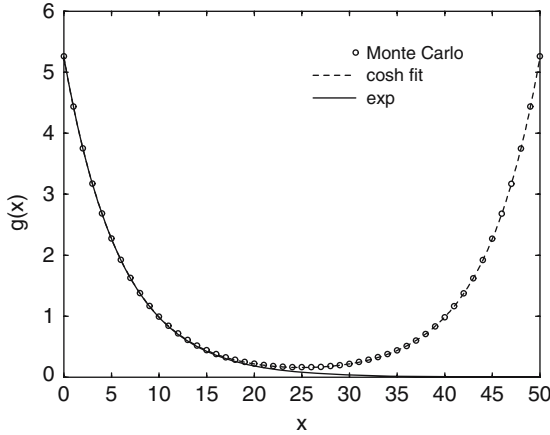
Fig. 4.14. Zero momentum projected correlation function $g(x)$ for the standard 2D nn Ising model at $T = 2.5 > T_c$. Also shown is a fit with the cosh-ansatz (4.122), yielding $m^* = 0.167\,9$ or $\xi^* = 5.957$, and the exponential approximation $\propto \exp(-m^*x)$

$$\widehat{G}(\boldsymbol{k})^{-1} = \frac{1}{a}\left(\sum_{i=1}^{D} 2(1 - \cos k_i) + m^2\right) \equiv c_1\kappa^2 + c_0 , \qquad (4.125)$$

with $c_1 = 1/a$ and $c_0 = m^2/a$, so that the squared correlation length

$$\xi^2 = 1/m^2 = c_1/c_0 \qquad (4.126)$$

can be extracted from a linear fit of $\widehat{G}(\boldsymbol{k})^{-1}$ versus $\kappa^2 = \sum_{i=1}^{D} 2(1 - \cos k_i) \approx \boldsymbol{k}^2$. In 2D, for instance, one may use $\boldsymbol{k} = 2\pi\boldsymbol{n}/L$ with $\boldsymbol{n} = (0,0), (1,0), (1,1), (2,0)$, and $(2,1)$, as done for the example in Fig. 4.15, which shows Monte Carlo data for $\widehat{G}(\boldsymbol{k})^{-1}$ from the same run used for Fig. 4.14 and a fit with (4.125). From the parameters c_1 and c_0 one then obtains $\xi = \sqrt{c_1/c_0} = 5.953$.

Even the simplest expression, using only $\boldsymbol{k} = \boldsymbol{0}$ and $\boldsymbol{k} = \boldsymbol{1} = (2\pi/L)(1,0,0,\ldots)$ and involving *no* fit at all, can be used:

$$\xi = \frac{1}{2\sin(\pi/L)}\left[\frac{\widehat{G}(\boldsymbol{0})}{\widehat{G}(\boldsymbol{1})} - 1\right]^{1/2} . \qquad (4.127)$$

This quantity, which is comparatively easy to measure in a Monte Carlo simulation, is usually referred to as second-moment correlation length. In the 2D Ising example, with $\widehat{G}(\boldsymbol{0}) = 62.66$ and $\widehat{G}(\boldsymbol{1}) = 1.768$ (cp. Fig. 4.15) and $L = 50$, (4.127) evaluates to $\xi = 5.965$, again in good agreement with the exact result for $\xi_{||}^{(\text{ex})}$. Finally note that the Fourier method gives directly ξ (and not ξ^*).
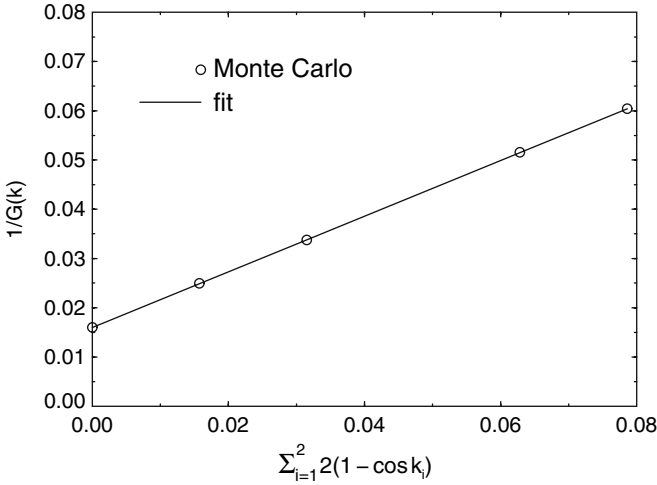
Fig. 4.15. Inverse long-wavelength Fourier components $\widehat{G}(\boldsymbol{k})^{-1}$ versus squared lattice momenta $\kappa^2 \equiv \sum_{i=1}^{2} 2(1 - \cos k_i) \approx \boldsymbol{k}^2$ for the 2D Ising model at $T = 2.5 > T_c$. The fit (4.125), $c_1\kappa^2 + c_0$, gives $c_1 = 0.565\ 5$ and $c_0 = 0.015\ 96$, and hence by (4.126), $\xi = \sqrt{c_1/c_0} = 5.953$

4.8 Generalized Ensemble Methods

All Monte Carlo methods described so far assumed a conventional canonical ensemble where the probability distribution of configurations is governed by a Boltzmann factor $\propto \exp(-\beta E)$. A simulation at some inverse temperature β_0 then covers a certain range of configuration space but not all (recall the discussion of the reweighting range). In principle a broader range can be achieved by patching several simulations at different temperatures using the multi-histogram method. Loosely speaking generalized ensemble methods aim at replacing this static patching by a single simulation in an appropriately defined generalized ensemble. The purpose of this section is to give at least a brief survey of the available methods.

4.8.1 Simulated Tempering

One approach are tempering methods which may be characterized as dynamical multi-histogramming. Similarly to the static reweighting approach, in simulated as well as in parallel tempering one considers m simulation points $\beta_1 < \beta_2 < \ldots < \beta_m$ which here, however, are combined already during the simulation in a specific, dynamical way.

In simulated tempering simulations [142, 143] one starts from a joint partition function (expanded ensemble)

$$\mathcal{Z}_{\text{ST}} = \sum_{i=1}^{m} e^{g_i} \sum_{\{\sigma\}} e^{-\beta_i \mathcal{H}(\{\sigma\})} \ , \tag{4.128}$$

where $g_i = \beta_i f(\beta_i)$ and the inverse temperature β is treated as an additional dynamical degree of freedom that can take the values $\beta_1, \ldots, \beta_m$. Employing a Metropolis algorithm, a proposed move from $\beta = \beta_i$ to β_j is accepted with probability

$$W = \min\left[1, \exp[-(\beta_j - \beta_i)\mathcal{H}(\{\sigma\})] + g_j - g_i\right] . \qquad (4.129)$$

Similar to multi-histogram reweighting (and also to multicanonical simulations), the free-energy parameters g_i are a priori unknown and have to be adjusted iteratively. To assure a reasonable acceptance rate for the β-update moves (usually between neighboring β_i-values), the histograms at β_i and β_{i+1}, $i = 1, \ldots, m-1$, must overlap. An estimate for a suitable spacing $\delta\beta = \beta_{i+1} - \beta_i$ of the simulation points β_i is hence immediately given by the results (4.82)–(4.84) for the reweighting range,

$$\delta\beta \propto \begin{cases} L^{-D/2} & \text{off-critical} \\ L^{-1/\nu} & \text{critical} \\ L^{-D} & \text{first-order} \end{cases} . \qquad (4.130)$$

Overall the simulated tempering method shows some similarities to the avoiding-rare-events variant of multicanonical simulations briefly discussed in the next subsection.

4.8.2 Parallel Tempering

In parallel tempering (replica exchange Monte Carlo, multiple Markov chain Monte Carlo) simulations [144, 145, 146], the starting point is a product of partition functions (extended ensemble),

$$\mathcal{Z}_{\text{PT}} = \prod_{i=1}^{m} \mathcal{Z}(\beta_i) = \prod_{i=1}^{m} \sum_{\{\sigma\}_i} e^{-\beta_i \mathcal{H}(\{\sigma\}_i)} , \qquad (4.131)$$

and all m systems at different simulation points $\beta_1 < \beta_2 < \ldots < \beta_m$ are simulated in parallel, using any legitimate update algorithm (Metropolis, cluster,...). This freedom in the choice of update algorithm is a big advantage of the parallel tempering method. After a certain number of sweeps, exchanges of the current configurations $\{\sigma\}_i$ and $\{\sigma\}_j$ are attempted (equivalently, the β_i may be exchanged, as is done in most implementations). Adapting the Metropolis criterion (4.24) to the present situation, the proposed exchange will be accepted with probability $W = \min(1, e^\Delta)$, where

$$\Delta = (\beta_j - \beta_i)\left[E(\{\sigma\}_j) - E(\{\sigma\}_i)\right] . \qquad (4.132)$$

To assure a reasonable acceptance rate, usually only nearest-neighbor exchanges ($j = i \pm 1$) are attempted and the β_i should again be spaced with the $\delta\beta$ given in (4.130). In most applications, the smallest inverse temperature β_1 is chosen in the high-temperature phase where the autocorrelation time is expected to be very short and the system decorrelates rapidly. Conceptually this approach follows again the avoiding-rare-events strategy.

Notice that in parallel tempering no free-energy parameters have to be adjusted. The method is thus very flexible and moreover can be almost trivially parallelized.

4.8.3 Multicanonical Ensembles

To conclude this introduction to simulation techniques, at least a very brief outline of multicanonical ensembles shall be given. For more details, in particular on practical implementations, see the recent reviews [4, 147, 148, 149, 150]. Similarly to the tempering methods of the last section, multicanonical simulations may also be interpreted as a dynamical multi-histogram reweighting method. This interpretation is stressed by the notation used in the original papers by Berg and Neuhaus [151, 152] and explains the name multicanonical. At the same time, this method may also be viewed as a specific realization of non-Boltzmann sampling [153] which has been known since long to be a legitimate alternative to the more standard Monte Carlo approaches [154]. The practical significance of non-Boltzmann sampling was first realized in the so-called umbrella-sampling method [155, 156, 157], but it took many years before the introduction of the multicanonical ensemble [151, 152] turned non-Boltzmann sampling into a widely appreciated practical tool in computer simulation studies of phase transitions. Once the feasibility of such a generalized ensemble approach was realized, many related methods and further refinements were developed.

Conceptually the method can be divided into two main strategies. The first strategy can be best described as avoiding rare events which is close in spirit to the alternative tempering methods. In this variant one tries to connect the important parts of phase space by easy paths which go around suppressed rare-event regions which hence cannot be studied directly. The second approach is based on enhancing the probability of rare event states, which is for example the typical strategy for dealing with the highly suppressed mixed-phase region of first-order phase transitions [47, 150]. This allows a direct study of properties of the rare-event states such as, e.g., interface tensions or more generally free energy barriers, which would be very difficult (or practically impossible) with canonical simulations and also with the tempering methods described in Sects. 4.8.1 and 4.8.2.

In general the idea is as follows. With $\{\sigma\}$ representing generically the degrees of freedom (discrete spins or continuous field variables), the canonical Boltzmann distribution

$$\mathcal{P}_{\mathrm{can}}(\{\sigma\}) \propto \mathrm{e}^{-\beta\mathcal{H}(\{\sigma\})} \tag{4.133}$$

is replaced by an auxiliary multicanonical distribution

$$\mathcal{P}_{\mathrm{muca}}(\{\sigma\}) \propto W(Q(\{\sigma\}))\mathrm{e}^{-\beta\mathcal{H}(\{\sigma\})} \equiv \mathrm{e}^{-\beta\mathcal{H}_{\mathrm{muca}}(\{\sigma\})} \,, \tag{4.134}$$

introducing a multicanonical weight factor $W(Q)$ where Q stands for any macroscopic observable such as the energy or magnetization. This defines formally $\mathcal{H}_{\mathrm{muca}} = \mathcal{H} - (1/\beta)\ln W(Q)$ which may be interpreted as an effective multicanonical Hamiltonian. The Monte Carlo sampling can then be implemented as usual by comparing $\mathcal{H}_{\mathrm{muca}}$ before and after a proposed update of $\{\sigma\}$, and canonical expectation values can be recovered exactly by inverse reweighting

$$\langle\mathcal{O}\rangle_{\mathrm{can}} = \langle\mathcal{O}W^{-1}(Q)\rangle_{\mathrm{muca}}/\langle W^{-1}(Q)\rangle_{\mathrm{muca}} \tag{4.135}$$

similarly to (4.89). The goal is now to find a suitable weight factor W such that the dynamics of the multicanonical simulation profits most.

To be specific, let us assume in the following that the relevant macroscopic observable is the energy E itself. This is for instance the case at a temperature driven first-order phase transition, where the canonical energy distribution $P_{can}(E)$ develops a characteristic double-peak structure [47]. As an illustration, simulation data for the 2D seven-state Potts model [158] are shown in Fig. 4.16. With increasing system size, the region between the two peaks becomes more and more suppressed ($\propto \exp(-2\sigma_{od}L^{D-1})$ where σ_{od} is the (reduced) interface tension, L^{D-1} the cross-section of a D-dimensional system, and the factor two accounts for the fact that with the usually employed periodic boundary condition at least two interfaces are present due to topological reasons) and the autocorrelation time thus grows exponentially with the system size L. In the literature, this is sometimes termed supercritical slowing down (even though nothing is critical here). Given such a situation, one usually adjusts $W = W(E)$ such that the multicanonical distribution $P_{muca}(E)$ is approximately constant between the two peaks of $P_{can}(E)$, thus aiming at a random-walk (pseudo-) dynamics of the Monte Carlo process, cf. Fig. 4.16.

The crucial non-trivial point is, of course, *how* this can be achieved. On a piece of paper, $W(E) \propto 1/P_{can}(E)$ – but we do not know $P_{can}(E)$ (otherwise there would be little need for the simulation . . .). The solution of this problem is a recursive computation. Starting with the canonical distribution, or some initial guess based on results for already simulated smaller systems together with finite-size scaling
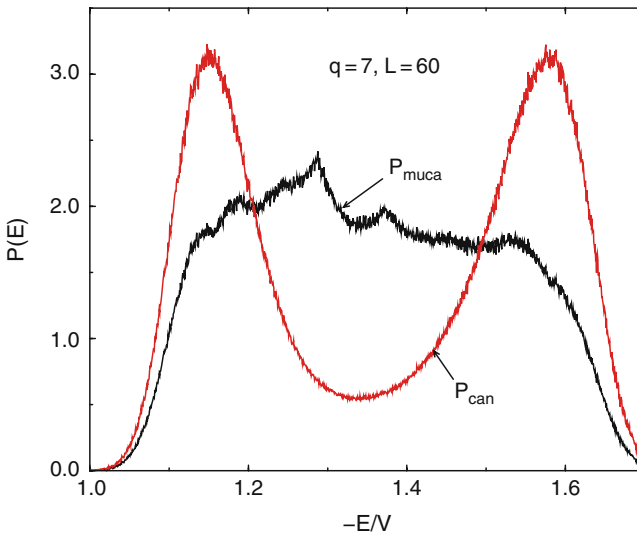

Fig. 4.16. The canonical energy density $P_{can}(E)$ of the 2D 7-state Potts model on a 60×60 lattice at inverse temperature $\beta_{eqh,L}$, where the two peaks are of equal height, together with the multicanonical energy density $P_{muca}(E)$, which is approximately constant between the two peaks

extrapolations, one performs a relatively short simulation to get an improved estimate of the canonical distribution. When this is inverted one obtains a new estimate of the multicanonical weight factor, which then is used in the next iteration and so on. In this naive variant only the simulation data of the last iteration are used in the construction of the improved weight factor.

A more sophisticated recursion, in which the updated weight factor, or more conveniently the ratio $R(E) = W(E+\Delta E)/W(E)$, is computed from *all* available data accumulated so far, works as follows [159]:

(i) Perform a simulation with $R_n(E)$ to obtain the nth histogram $H_n(E)$.
(ii) Compute the statistical weight of the nth run:

$$p(E) = H_n(E)H_n(E + \Delta E)/[H_n(E) + H_n(E + \Delta E)] . \qquad (4.136)$$

(iii) Accumulate statistics:

$$\begin{aligned} p_{n+1}(E) &= p_n(E) + p(E) , \\ \kappa(E) &= p(E)/p_{n+1}(E) . \end{aligned} \qquad (4.137)$$

(iv) Update weight ratios:

$$R_{n+1}(E) = R_n(E) \left[H_n(E)/H_n(E + \Delta E) \right]^{\kappa(E)} . \qquad (4.138)$$

Go to (i).

The recursion is initialized with $p_0(E) = 0$. To derive this recursion one assumes that (unnormalized) histogram entries $H_n(E)$ have an a priori statistical error $\sqrt{H_n(E)}$ and (quite crudely) that all data are uncorrelated. Due to the accumulation of statistics, this procedure is rather insensitive to the length of the nth run in the first step and has proved to be rather stable and efficient in practice.

In most applications local update algorithms have been employed, but for certain classes of models also non-local multigrid methods [119, 120, 160, 161] are applicable [121, 162]. A combination with non-local cluster update algorithms, on the other hand, is not straightforward. Only by making direct use of the random-cluster representation as a starting point, a multibondic variant [163, 164, 165] has been developed. For a recent application to improved finite-size scaling studies of second-order phase transitions, see [128]. If P_{muca} was completely flat and the Monte Carlo update moves would perform an ideal random walk, one would expect that after V^2 local updates the system has travelled on average a distance V in total energy. Since one lattice sweep consists of V local updates, the autocorrelation time should scale in this idealized picture as $\tau \propto V$. Numerical tests for various models with a first-order phase transition have shown that in practice the data are at best consistent with a behavior $\tau \propto V^\alpha$, with $\alpha \geq 1$. While for the temperature-driven transitions of 2D Potts models the multibondic variant seems to saturate the bound [163, 164, 165], employing local update algorithms, typical fit results are $\alpha \approx 1.1$–1.3, and due to the limited accuracy of the data even a weak exponential growth law cannot really be excluded.

In fact, at least for the field-driven first-order transition of the 2D Ising model below T_c, where one works with the magnetization instead of the energy (sometimes called multimagnetical simulations), it has been demonstrated recently [166] that even for a perfectly flat multicanonical distribution there are two hidden free energy barriers (in directions orthogonal to the magnetization) which lead to an exponential growth of τ with lattice size, which is albeit much weaker than the leading supercritical slowing down of the canonical simulation. Physically the two barriers are related to the nucleation of a large droplet of the wrong phase (say down-spins in the background of up-spins) [167, 168, 169, 170, 171, 172, 173] and the transition of this large, more or less spherical droplet to the strip phase (coexisting strips of down- and up-spins, separated by two straight interfaces) around $m = 0$ [174].

4.8.4 Wang-Landau Recursion

Another more recently proposed method deals directly with estimators $\Omega(E)$ of the density of states [175, 176]. By flipping spins randomly, the transition probability from energy level E_1 to E_2 is

$$p(E_1 \rightarrow E_2) = \min\left[\frac{\Omega(E_1)}{\Omega(E_2)}, 1\right] . \qquad (4.139)$$

Each time an energy level is visited, the estimator is multiplicatively updated

$$\Omega(E) \rightarrow f \, \Omega(E) , \qquad (4.140)$$

where initially $\Omega(E) = 1$ and $f = f_0 = e^1$. Once the accumulated energy histogram is sufficiently flat, the factor f is refined

$$f_{n+1} = \sqrt{f_n} \qquad (4.141)$$

with $n = 0, 1, \ldots$, and the energy histogram reset to zero until some small value such as $f = e^{10^{-8}} \approx 1.000\,000\,01$ is reached.

For the 2D Ising model this procedure converges very rapidly towards the exactly known density of states, and also for other applications a fast convergence has been reported. Since the procedure is known to violate detailed balance, however, some care is necessary in setting up a proper protocol of the recursion. Most authors who employ the obtained density of states directly to extract canonical expectation values by standard reweighting argue that, once f is close enough to unity, systematic deviations become negligible. While this claim can be verified empirically for the 2D Ising model (where exact results are available for judgement), possible systematic deviations are difficult to assess in the general case. A safe way would be to consider the recursion (4.139)–(4.141) as an alternative method to determine the multicanonical weights, and then to perform a usual multicanonical simulation based on them. As emphasized earlier, any deviations of multicanonical weights from their optimal shape do not show up in the final canonical expectation values; they rather only influence the dynamics of the multicanonical simulations.

4.9 Concluding Remarks

The intention of these lecture notes was to give an elementary introduction to the concepts of modern Markov chain Monte Carlo simulations and to illustrate their usefulness by applications to the very simple Ising lattice spin model. The basic Monte Carlo methods employing local update rules are straightforward to generalize to all models with discrete degrees of freedom and, with small restrictions, also to all models with continuous variables and off-lattice systems. Non-local cluster update methods are much more efficient but also more specialized. Some generalizations to Potts and $O(n)$ symmetric spin models have been indicated and also further models may be efficiently simulated by this method, but there is no guarantee that for a given model a cluster update procedure can be developed. The statistical error analysis is obviously completely general, and also the example finite-size scaling analysis can be taken as a guideline for any model exhibiting a second-order phase transition. Finally, reweighting techniques and generalized ensemble ideas such as tempering methods, the multicanonical ensemble and Wang-Landau sampling can be adapted to almost every statistical physics problem at hand once the relevant macroscopic observables are identified.

Acknowledgements

References

1. M. Newman, G. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999) 80, 86
2. D. Landau, K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000) 80, 86
3. K. Binder, D. Heermann, *Monte Carlo Simulations in Statistical Physics: An Introduction*, 4th edn. (Springer, Berlin, 2002) 80, 86

136 W. Janke

4. B. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004) 80, 86, 131
5. H. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford Press, Oxford, 1979) 80, 92
6. J. Binney, N. Dowrick, A. Fisher, M. Newman, *The Theory of Critical Phenomena* (Oxford University Press, Oxford, 1992) 80
7. D. Lavis, G. Bell, *Statistical Mechanics of Lattice Systems 2* (Springer, Berlin, 1999) 80
8. C. Domb, J. Lebowitz (eds.), *Phase Transitions and Critical Phenomena* (Academic Press, New York, 1976) 80
9. W. Lenz, Phys. Z. **21**, 613 (1920) 81
10. E. Ising, Phys. Z. **31**, 253 (1925) 81
11. L. Onsager, Phys. Rev. **65**, 117 (1944) 82
12. B. Kaufman, Phys. Rev. **76**, 1232 (1949) 82, 109
13. A. Ferdinand, M. Fisher, Phys. Rev. **185**, 832 (1969) 82, 109
14. B. McCoy, T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, Cambridge, 1973) 82, 127
15. R. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982) 82, 127
16. L. Onsager, Nuovo Cimento **6**, 261 (1949) 82
17. C. Yang, Phys. Rev. **85**, 808 (1952) 82
18. C. Chang, Phys. Rev. **88**, 1422 (1952) 82
19. W. Orrick, B. Nickel, A. Guttmann, J. Perk, Phys. Rev. Lett. **86**, 4120 (2001) 82
20. W. Orrick, B. Nickel, A. Guttmann, J. Perk, J. Stat. Phys. **102**, 795 (2001) 82
21. R. Griffiths, Phys. Rev. Lett. **24**, 1479 (1970) 83
22. G. Rushbrooke, J. Chem. Phys. **39**, 842 (1963) 83
23. R. Griffiths, Phys. Rev. Lett. **14**, 623 (1965) 83
24. B. Josephson, Proc. Phys. Soc. **92**, 269 (1967) 83
25. B. Josephson, Proc. Phys. Soc. **92**, 276 (1967) 83
26. M. Fisher, Phys. Rev. **180**, 594 (1969) 83
27. L. Widom, J. Chem. Phys. **43**, 3892 (1965) 83
28. L. Widom, J. Chem. Phys. **43**, 3898 (1965) 83
29. L. Kadanoff, Physics **2**, 263 (1966) 83
30. K. Wilson, J. Kogut, Phys. Rep. **C12**, 75 (1974) 83
31. F. Wu, Rev. Mod. Phys. **54**, 235 (1982) 84
32. F. Wu, Rev. Mod. Phys. **55**, 315(E) (1983) 84
33. M. Weigel, W. Janke, Phys. Rev. **B62**, 6343 (2000) 84
34. K. Binder, in *Monte Carlo Methods in Statistical Physics* ed. by K. Binder (Springer, Berlin, 1979), p. 1 84, 86
35. M. Barber, in *Phase Transitions and Critical Phenomena*, Vol. 8 ed. by C. Domb, J. Lebowitz (Academic Press, New York, 1983), p. 146 84
36. V. Privman (ed.), *Finite-Size Scaling and Numerical Simulations of Statistical Systems* (World Scientific, Singapore, 1990) 84
37. K. Binder: in Computational Methods in Field Theory, Schladming Lecture Notes, eds. H. Gausterer, C.B. Lang (Springer, Berlin, 1992), P. 59 84
38. J. Gunton, M. Miguel, P. Sahni, in *Phase Transitions and Critical Phenomena*, Vol. 8, ed. by C. Domb, J. Lebowitz (Academic Press, New York, 1983) 85
39. K. Binder, Rep. Prog. Phys. **50**, 783 (1987) 85
40. H. Herrmann, W. Janke, F. Karsch (eds.), *Dynamics of First Order Phase Transitions* (World Scientific, Singapore, 1992) 85

41. W. Janke, in *Computer Simulations in Condensed Matter Physics*, Vol. VII, ed. by D. Landau, K. Mon, H.B. Schüttler (Springer, Berlin, 1994), p. 29 85
42. M. Fisher, A. Berker, Phys. Rev. B **26**, 2507 (1982) 85
43. V. Privman, M. Fisher, J. Stat. Phys. **33**, 385 (1983) 85
44. K. Binder, D. Landau, Phys. Rev. B **30**, 1477 (1984) 85
45. M. Challa, D. Landau, K. Binder, Phys. Rev. B **34**, 1841 (1986) 85
46. V. Privman, J. Rudnik, J. Stat. Phys. **60**, 551 (1990) 85
47. W. Janke, in *Computer Simulations of Surfaces and Interfaces*, *NATO Science Series, II. Mathematics, Physics and Chemistry* Vol. 114, ed. by B. Dünweg, D. Landau, A. Milchev (Kluwer, Dordrecht, 2003), pp. 111–135 85, 92, 111, 131, 132
48. C. Borgs, R. Kotecký J. Stat. Phys. **61**, 79 (1990) 85
49. J. Lee, J. Kosterlitz Phys. Rev. Lett. **65**, 137 (1990) 85
50. C. Borgs, R. Kotecký, S. Miracle-Solé, J. Stat. Phys. **62**, 529 (1991) 85
51. C. Borgs, W. Janke, Phys. Rev. Lett. **68**, 1738 (1992) 85
52. W. Janke, Phys. Rev. B **47**, 14757 (1993) 85
53. J. Hammersley, D. Handscomb, *Monte Carlo Methods* (Chapman and Hall, London, New York, 1964) 85
54. D. Heermann, *Computer Simulation Methods in Theoretical Physics*, 2nd edn. (Springer, Berlin, 1990) 86
55. K. Binder (ed.), *The Monte Carlo Method in Condensed Matter Physics* (Springer, Berlin, 1992) 86
56. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953) 86
57. S. Kirkpatrick, C. Gelatt. Jr., M. Vecchi, Science **220**, 671 (1983) 87
58. W. Janke, in *Ageing and the Glass Transition – Summer School, University of Luxembourg, September 2005*, *Lecture Notes in Physics*, Vol. 716, ed. by M. Henkel, M. Pleimling, R. Sanctuary (Springer, Berlin, Heidelberg, 2007), *Lecture Notes in Physics*, pp. 207–260 87, 89, 103, 106
59. W. Janke, in *Proceedings of the Euro Winter School Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, *NIC Series*, Vol. 10, ed. by J. Grotendorst, D. Marx, A. Muramatsu (John von Neumann Institute for Computing, Jülich, 2002), pp. 447–458 87
60. R. Glauber, J. Math. Phys. **4**, 294 (1963) 89, 90
61. R. Swendsen, J.S. Wang, Phys. Rev. Lett. **58**, 86 (1987) 93, 94, 97
62. R. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952) 93
63. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989) 93, 94, 96
64. W. Janke, Mathematics and Computers in Simulations **47**, 329 (1998) 93
65. P. Kasteleyn, C. Fortuin, J. Phys. Soc. Japan **26**, 11 (1969) 93
66. C. Fortuin, P. Kasteleyn, Physica **57**, 536 (1972) 93
67. C. Fortuin, Physica **58**, 393 (1972) 93
68. C. Fortuin, Physica **59**, 545 (1972) 93
69. U. Wolff, Nucl. Phys. **B322**, 759 (1989) 96, 98
70. M. Hasenbusch, Nucl. Phys. **B333**, 581 (1990) 96
71. U. Wolff, Nucl. Phys. **B334**, 581 (1990) 96, 97, 98
72. U. Wolff, Phys. Lett. A **228**, 379 (1989) 96, 97
73. C. Baillie, Int. J. Mod. Phys. C **1**, 91 (1990) 96
74. M. Hasenbusch, S. Meyer, Phys. Lett. B **241**, 238 (1990) 96
75. R. Swendsen, J.S. Wang, A. Ferrenberg, in *The Monte Carlo Method in Condensed Matter Physics* ed. by K. Binder (Springer, Berlin, 1992) 96

76. X.L. Li, A. Sokal, Phys. Rev. Lett. **63**, 827 (1989) 96
77. X.L. Li, A. Sokal, Phys. Rev. Lett. **67**, 1482 (1991) 96
78. M. Nightingale, H. Blöte, Phys. Rev. Lett. **76**, 4548 (1996) 97, 100
79. M. Nightingale, H. Blöte, Phys. Rev. B **62**, 1089 (2000) 97, 100
80. P. Grassberger, Physica A **214**, 547 (1995) 97
81. P. Grassberger, Physica A **217**, 227 (E) (1995) 97
82. N. Ito, K. Hukushima, K. Ogawa, Y. Ozeki, J. Phys. Soc. Japan **69**, 1931 (2000) 97
83. D. Heermann, A. Burkitt, Physica A **162**, 210 (1990) 97
84. P. Tamayo, Physica A **201**, 543 (1993) 97
85. N. Ito, G. Kohring, Physica A **201**, 547 (1993) 97
86. W. Janke, Phys. Lett. A **148**, 306 (1990) 98
87. C. Holm, W. Janke, Phys. Rev. B **48**, 936 (1993) 98, 113
88. W. Janke, A. Schakel, Nucl. Phys. **B700**, 385 (2004) 98
89. W. Janke, A. Schakel, Comp. Phys. Comm. **169**, 222 (2005) 98
90. W. Janke, A. Schakel, Phys. Rev. E **71**, 036703 (2005) 98
91. W. Janke, A. Schakel, Phys. Rev. Lett. **95**, 135702 (2005) 98
92. W. Janke, A. Schakel, in *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory*, Vol. 2, ed. by Y. Holovatch (World Scientific, Singapore, 2007), pp. 123–180 98
93. E. Lorenz, Ageing phenomena in phase-ordering kinetics in Potts models. Diploma thesis, Universität Leipzig (2005). www.physik.uni-leipzig.de/∼lorenz/ diplom.pdf 100, 101
94. A. Rutenberg, A. Bray, Phys. Rev. E **51**, 5499 (1995) 100, 101
95. P. Calabrese, A. Gambassi, J. Phys. A **38**, R133 (2005) 100
96. C. Godrèche, J.M. Luck, J. Phys.: Condens. Matter **14**, 1589 (2002) 100, 101
97. L.F. Cugliandolo: in Slow Relaxation and Non Equilibrium Dynamics in Condensed Matter, Les Houches Lectures, eds. J.-L. Barrat, J. Dalibard, J. Kurchan, M.V. Feigel'man (Springer, Berlin, 2003) 100
98. F. Corberi, E. Lippiello, M. Zannetti, Phys. Rev. Lett. **90**, 099601 (2003) 101
99. M. Henkel, M. Pleimling, Phys. Rev. Lett. **90**, 099602 (2003) 101
100. L. Berthier, J. Barrat, J. Kurchan, Europhys. J. B **11**, 635 (1999) 101
101. F. Corberi, E. Lippiello, M. Zannetti, Europhys. J. B **24**, 359 (2001) 101
102. F. Corberi, E. Lippiello, M. Zannetti, Phys. Rev. E **65**, 046136 (2003) 101
103. A. Barrat, Phys. Rev. E **57**, 3629 (1998) 101
104. M. Henkel, *Conformal Invariance and Critical Phenomena* (Springer, Berlin, 1999) 101
105. M. Henkel, M. Pleimling, C. Godrèche, J.M. Luck, Phys. Rev. Lett. **87**, 265701 (2001) 101
106. M. Henkel, Nucl. Phys. **B641**, 405 (2002) 101
107. M. Henkel, M. Paessens, M. Pleimling, Europhys. Lett. **62**, 664 (2003) 101
108. M. Henkel, M. Pleimling, Phys. Rev. E **68**, 065101 (R) (2003) 101
109. M. Henkel, A. Picone, M. Pleimling, Europhys. Lett. **68**, 191 (2004) 101
110. E. Lorenz, W. Janke, Europhys. Lett. **77**, 10003 (2007) 101
111. W. Janke, in *Proceedings of the Euro Winter School Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, *NIC Series*, Vol. 10, ed. by J. Grotendorst, D. Marx, A. Muramatsu (John von Neumann Institute for Computing, Jülich, 2002), pp. 423–445 102, 103
112. P. Beale, Phys. Rev. Lett. **76**, 78 (1996) 109
113. W. Press, S. Teukolsky, W. Vetterling, B. Flannery, *Numerical Recipes in Fortran 77 – The Art of Scientific Computing*, 2nd edn. (Cambridge University Press, Cambridge, 1999) 102, 117, 120, 121

114. M. Priestley, *Spectral Analysis and Time Series*, Vol. 2 (Academic, London, 1981). Chaps. 5–7 103
115. T. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971) 103
116. N. Madras, A. Sokal, J. Stat. Phys. **50**, 109 (1988) 103, 105
117. A. Sokal, L. Thomas, J. Stat. Phys. **54**, 797 (1989) 103
118. A. Ferrenberg, D. Landau, K. Binder, J. Stat. Phys. **63**, 867 (1991) 104
119. A. Sokal, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms* (Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, 1989) 105, 133
120. A. Sokal, in *Quantum Fields on the Computer*, ed. by M. Creutz (World Scientific, Singapore, 1992), p. 211 105, 133
121. W. Janke, T. Sauer, J. Stat. Phys. **78**, 759 (1995) 106, 133
122. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics [SIAM], Philadelphia, 1982) 107
123. R. Miller, Biometrika **61**, 1 (1974) 107
124. A. Ferrenberg, R. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988) 108
125. A. Ferrenberg, R. Swendsen, Phys. Rev. Lett. **63**, 1658(E) (1989) 108
126. N. Wilding, in *Computer Simulations of Surfaces and Interfaces*, *NATO Science Series, II. Mathematics, Physics and Chemistry* Vol. 114, ed. by B. Dünweg, D. Landau, A. Milchev (Kluwer, Dordrecht, 2003), pp. 161–171 110
127. A. Ferrenberg, R. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989) 112
128. B. Berg, W. Janke, Phys. Rev. Lett. **98**, 040602 (2007) 114, 133
129. G. Kamieniarz, H. Blöte, J. Phys. A **26**, 201 (1993) 116, 123
130. J. Salas, A. Sokal, J. Stat. Phys. **98**, 551 (2000) 116, 123
131. X. Chen, V. Dohm, Phys. Rev. E **70**, 056136 (2004) 116, 123
132. V. Dohm, J. Phys. A **39**, L259 (2006) 116, 123
133. W. Selke, L. Shchur, J. Phys. A **38**, L739 (2005) 116, 123
134. M. Schulte, C. Drope, Int. J. Mod. Phys. C **16**, 1217 (2005) 116, 123
135. M. Sumour, D. Stauffer, M. Shabat, A. El-Astal, Physica A **368**, 96 (2006) 116, 123
136. W. Selke, Europhys. J. B **51**, 223 (2006); preprint `http://arxiv.org/abs/cond-mat/0701515` 116, 123
137. A. Nußbaumer, E. Bittner, W. Janke, Europhys. Lett. **78**, 16004 (2007) 118, 119, 123, 124
138. E. Bittner, W. Janke, The pain of example analyses – a (self-)critical discussion. Unpublished results 119, 124
139. W. Janke, R. Villanova, Phys. Rev. B **66**, 134208 (2002) 119, 120
140. J. Oitmaa, J. Phys. A **14**, 1159 (1981) 121
141. C. Holm, W. Janke, Phys. Rev. Lett. **78**, 2265 (1997) 124
142. E. Marinari, G. Parisi, Europhys. Lett. **19**, 451 (1992) 129
143. A. Lyubartsev, A. Martsinovski, S. Shevkunov, P. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992) 129
144. C. Geyer, in *Proceedings of the 23rd Symposium on the Interface*, ed. by E. Keramidas (Interface Foundation, Fairfax, Virginia, 1991), pp. 156–163 130
145. C. Geyer, E. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995) 130
146. K. Hukushima, K. Nemoto, J. Phys. Soc. Japan **65**, 1604 (1996) 130
147. B. Berg, Fields Inst. Comm. **26**, 1 (2000) 131
148. B. Berg, Comp. Phys. Comm. **104**, 52 (2002) 131
149. W. Janke, Physica A **254**, 164 (1998) 131
150. W. Janke, in *Computer Simulations of Surfaces and Interfaces*, *NATO Science Series, II. Mathematics, Physics and Chemistry –Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, 9–20 September 2002*, Vol. 114, ed. by B. Dünweg, D. Landau, A. Milchev (Kluwer, Dordrecht, 2003) 131

151. B. Berg, T. Neuhaus, Phys. Lett. B **267**, 249 (1991) 131
152. B. Berg, T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992) 131
153. W. Janke, Int. J. Mod. Phys. C **3**, 1137 (1992) 131
154. K. Binder: in Phase Transitions and Critical Phenomena, Vol. 5b, eds. C. Domb, M.S. Green (Academic Press, New York, 1976), p. 1 131
155. G. Torrie, J. Valleau, Chem. Phys. Lett. **28**, 578 (1974) 131
156. G. Torrie, J. Valleau, J. Comp. Phys. **23**, 187 (1977) 131
157. G. Torrie, J. Valleau, J. Chem. Phys. **66**, 1402 (1977) 131
158. W. Janke, B. Berg, M. Katoot, Nucl. Phys. **B382**, 649 (1992) 132
159. B. Berg, W. Janke. Unpublished notes 133
160. J. Goodman, A. Sokal, Phys. Rev. Lett. **56**, 1015 (1986) 133
161. J. Goodman, A. Sokal, Phys. Rev. D **40**, 2035 (1989) 133
162. W. Janke, T. Sauer, Phys. Rev. E **49**, 3475 (1994) 133
163. W. Janke, S. Kappler, Nucl. Phys. (proc suppl.) **B42**, 876 (1995) 133
164. W. Janke, S. Kappler, Phys. Rev. Lett. **74**, 212 (1995) 133
165. M. Carroll, W. Janke, S. Kappler, J. Stat. Phys. **90**, 1277 (1998) 133
166. T. Neuhaus, J. Hager, J. Stat. Phys. **113**, 47 (2003) 134
167. K. Binder, M. Kalos, J. Stat. Phys. **22**, 363 (1980) 134
168. H. Furukawa, K. Binder, Phys. Rev. A **26**, 556 (1982) 134
169. M. Biskup, L. Chayes, R. Kotecký, Europhys. Lett. **60**, 21 (2002) 134
170. M. Biskup, L. Chayes, R. Kotecký, Comm. Math. Phys. **242**, 137 (2003) 134
171. M. Biskup, L. Chayes, R. Kotecký, J. Stat. Phys. **116**, 175 (2003) 134
172. K. Binder, Physica A **319**, 99 (2003) 134
173. A. Nußbaumer, E. Bittner, T. Neuhaus, W. Janke, Europhys. Lett. **75**, 716 (2006) 134
174. K. Leung, R. Zia, J. Phys. A **23**, 4593 (1990) 134
175. F. Wang, D. Landau, Phys. Rev. Lett. **86**, 2050 (2001) 134
176. F. Wang, D. Landau, Phys. Rev. E **64**, 056101 (2001) 134