5

Introduction to Simulation Techniques

W. Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ),
Universität Leipzig, Augustusplatz 10/11,
D-04109 Leipzig, Germany
wolfhard.janke@itp.uni-leipzig.de

These lectures give an introduction to Monte Carlo simulations of classical statistical physics systems and their statistical analysis. After briefly recalling a few elementary properties of phase transitions, the concept of importance sampling Monte Carlo methods is discussed and illustrated by a few standard local update algorithms (Metropolis, heat-bath, Glauber). Then emphasis is placed on thorough analyses of the generated data paying special attention to the choice of estimators, autocorrelation times and statistical error analysis. This leads to 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 discussed next. Then the general tool of reweighting techniques will be explained. This paves the way to introduce simulated and parallel tempering methods which are very useful for simulations of complex, possibly disordered systems. Finally, also the important alternative approach using multicanonical ensembles is briefly outlined.

5.1 Introduction

The statistical mechanics of complex physical systems poses many hard problems which are very difficult if not impossible to solve by purely analytical methods. Numerical simulation techniques will therefore be indispensable tools on our way to a better understanding of systems such as (spin) glasses and disordered magnets, or of the huge field of biologically motivated problems such as protein folding, to mention only a few important classical problems. Quantum statistical problems in condensed matter or the broad field of elementary particle physics and quantum gravity are other major applications.

The numerical tools commonly employed can be roughly divided into molecular dynamics (MD) and Monte Carlo (MC) simulations. With the still ongoing advances in computer technology – according to Moore's law, since about

W. Janke: *Introduction to Simulation Techniques*, Lect. Notes Phys. **716**, 207–260 (2007)
DOI 10.1007/3-540-69684-9_5

1950, every 5 years a factor of 10 is gained in computing speed – both approaches can be expected to gain even more importance in the future than they have already today. In the past few years the predictive power of especially the MC approach was in addition considerably enhanced by the discovery of greatly improved simulation algorithms. Not all of them are already well enough understood to be applicable to really complex physical systems. But, as a first step, it is gratifying to note that at least for relatively simple spin systems, orders of magnitude of computing time can be saved by these refinements. The purpose of these lecture notes is to give a concise introduction to what is feasible today. For further reading, there are quite a few recent textbooks [1–4] available which treat some of the material discussed here in more depth. In particular, in these books one can also find recent applications to physically relevant systems which are purposely omitted in this short introduction.

For illustration purposes, we shall rather confine ourselves to 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 concepts 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 conformal field theory, where they serve as idealized "matter" fields on Feynman graphs or fluctuating manifolds.

The rest of these lecture notes is organized as follows. In Sect. 5.2, the definitions of Ising and Potts models are recalled and some standard observables (specific heat, magnetization, susceptibility, correlation functions,...) are briefly discussed. Next the most characteristic properties of phase transitions, scaling properties and the definition of critical exponents are summarized. In Sect. 5.3, the basic method underlying all importance sampling Monte Carlo simulations is described. The following Sect. 5.4 is first devoted to a short discussion of the initial non-equilibrium period and ageing phenomena, and then in Sect. 5.5 a fairly detailed account of statistical error analysis in equilibrium is given which also includes temporal correlation effects. The latter highlight the problems of critical slowing down at a continuous phase transition and phase coexistence with exponentially large flipping times at a first-order transition. One very successful solution of the former problem are non-local cluster algorithms which are described in Sect. 5.6. In Sect. 5.7 we discuss reweighting techniques which quite naturally lead to the tempering update algorithms explained in Sect. 5.8. These algorithms may be viewed as dynamical reweighting methods that can circumvent exponentially large flipping times and proved to be very successful for the simulation of complex, disordered systems. The alternative method of multicanonical ensembles is only very briefly discussed in Sect. 5.9, with emphasis on similarities and

differences to tempering methods. Finally, Sect. 5.10 contains a few concluding remarks.

5.2 Models and Phase Transitions

5.2.1 Models and Observables

Most of the simulation techniques introduced below can be illustrated for the simple Ising spin model whose partition function is defined as [5]

$$Z = \sum_{\{\sigma_i\}} \exp(-H/k_B T) \; , \tag{5.1}$$

with

$$H = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \;\; , \qquad \sigma_i = \pm 1 \; . \tag{5.2}$$

Here T is the temperature and h is an external magnetic field, k_B is Boltzmann's constant, the spins σ_i are assumed (for simplicity) to live on the sites i of a D-dimensional cubic lattice of volume $V = L^D$, and the symbol $\langle ij \rangle$ indicates that the lattice sum runs over all 2D nearest-neighbor pairs. In all examples discussed below, periodic boundary conditions will be assumed.

Standard observables are the internal energy per site, $e = E/V$, with $E = -\mathrm{d} \ln Z / \mathrm{d}\beta \equiv \langle H \rangle$, and the specific heat,

$$C/k_B = \frac{\mathrm{d}e}{\mathrm{d}(k_B T)} = \beta^2 \left(\langle H^2 \rangle - \langle H \rangle^2 \right) / V \; , \tag{5.3}$$

where $\beta \equiv 1/k_B T$. In the following we always use units in which $k_B \equiv 1$ and $J \equiv 1$. On finite lattices the magnetization and susceptibility are usually defined as

$$m = M/V = \langle |\mu| \rangle \; , \qquad \mu = \sum_i \sigma_i / V \; , \tag{5.4}$$

$$\chi = \beta V \left(\langle \mu^2 \rangle - \langle |\mu| \rangle^2 \right) \; . \tag{5.5}$$

In the high-temperature phase one often employs the fact that the magnetization vanishes in the infinite volume limit and considers

$$\chi' = \beta V \langle \mu^2 \rangle \; . \tag{5.6}$$

Similarly, the spin-spin correlation function,

$$G(\boldsymbol{x}_i - \boldsymbol{x}_j) = \langle \sigma_i \, \sigma_j \rangle - \langle \sigma_i \rangle \, \langle \sigma_j \rangle \; , \tag{5.7}$$

then simplifies to $G(\boldsymbol{x}_i - \boldsymbol{x}_j) = \langle \sigma_i \, \sigma_j \rangle$, where $\boldsymbol{x}_i$ measures the position of the lattice sites i which are numbered, say, in a lexicographical order. At large

distances, $G(\boldsymbol{x}) \propto \exp(-|\boldsymbol{x}|/\xi)$ decays exponentially. Its decay rate defines the correlation length

$$\xi = -\lim_{|\boldsymbol{x}|\to\infty} |\boldsymbol{x}|/\ln G(\boldsymbol{x}) \;, \tag{5.8}$$

which strictly speaking depends on the (discrete) orientation of $\boldsymbol{x}$. For definiteness we will hence always consider correlations along one of the main lattice directions.

In vanishing external field the Ising model exhibits a continuous phase transition in temperature for all dimensions $D \geq 2$. The two-dimensional (2D) model is self-dual, relating its behaviour for high temperatures $T = 1/\beta$ to that at $T^* = 1/\beta^*$ in the low-temperature phase, where

$$\sinh(2\beta)\sinh(2\beta^*) = 1 \;. \tag{5.9}$$

Under the mild assumption of a *single* phase-transition point, this fixes already the critical temperature to be

$$\sinh(2\beta_c) = 1 \quad \text{or} \quad \beta_c = \ln(1+\sqrt{2})/2 = 0.440\,686\ldots \;, \quad T_c = 2.269\,185\ldots \;. \tag{5.10}$$

The exact solution by Onsager [6–8] in 1944 yields the free energy, internal energy, specific heat etc. in zero external field for the general case of anisotropic couplings J_x, J_y. Also the exact result for the magnetization below T_c in zero field (again for general J_x, J_y) was first announced by Onsager at a conference in Florence 1949 [9]. The first published derivation was given three years later by Yang [10] in 1952, for the special case $J_x = J_y = J$, and subsequently generalized to arbitrary J_x, J_y by Chang [11] in the same year. Even the correlation length in arbitrary directions is known analytically [7, 8]. Along the coordinate axes of a square lattice, the formula takes a surprisingly simple form,

$$\xi_d(\beta) = \frac{1}{2(\beta^* - \beta)} \qquad (\beta < \beta_c) \;, \tag{5.11}$$

$$\xi_o(\beta^*) = \xi_d(\beta)/2 \qquad (\beta^* > \beta_c) \;, \tag{5.12}$$

where β and β^* are the dual couplings defined in (5.9). The susceptibility, however, is still not exactly known, even though highly accurate approximations could be derived [7, 8, 12]. In 3D even for the simple Ising model, no exact solutions are available. Numerical work, high-temperature series expansions and field-theoretical considerations provide, however, very precise estimates. From 4D on, the so-called upper critical dimension, mean-field behaviour starts to be become qualitatively correct, albeit in 4D only up to multiplicative logarithmic corrections. The critical temperature is for all finite dimensions $D \geq 3$ only approximately known, approaching $T_c = 2D$ in the mean-field limit $D \to \infty$.

A simple generalization of the Ising model is the q-state Potts model [13] whose Hamiltonian in zero external field is given by

$$H_{\text{Potts}} = -J \sum_{\langle ij \rangle} \delta_{\sigma_i \sigma_j} , \qquad \sigma_i \in 1, \ldots, q , \tag{5.13}$$

which is equivalent to the Ising model for $q = 2$. The 2D Potts model is exactly known from self-duality, $(\exp(\beta) - 1)(\exp(\beta^*) - 1) = q$, to exhibit at $\beta_c = \ln(1 + \sqrt{q})$ a second-order transition for $q \leq 4$ and a first-order transition for all $q \geq 5$ [14,15]. At the transition point (but only there), a couple of exact results are available for both types of transitions, including the free energy, internal energy and specific heat [14,15] as well as the correlation length ξ_d in the disordered phase and the related interface tension σ_{od} between the ordered and disordered phase [16]. In 3D, numerical work suggests for all $q \geq 3$ a first-order transition which rapidly becomes stronger with increasing q.

5.2.2 Phase Transitions

The most interesting aspect of a system's phase diagram is the region where cooperation effects may cause a phase transition, e.g., from a disordered phase at high temperatures to an ordered phase at low temperatures as in the paradigmatic Ising model. To predict the properties of this most challenging region of a phase diagram as accurately as possible is one of the major objectives of all statistical mechanics approaches, including numerical computer simulation studies. The theory of phase transitions is a very broad subject described comprehensively in many textbooks (see, e.g., Refs. [17–20]). Here we only roughly classify them into *first-order* and *second-order* (or, more generally, continuous) phase transitions, and give a very brief summary of those properties that are most relevant for numerical simulations.

Some characteristic features of the thermodynamic behaviour at first- and second-order phase transitions are sketched in Fig. 5.1. Most phase transitions in Nature are of first order [21–24]. The best known example is the field-driven transition in magnets at temperatures below the Curie point, while the paradigm of a temperature-driven first-order transition experienced every day is ordinary melting [25,26]. Simple models sharing such a behaviour are the Ising and Potts models defined in (5.2) and (5.13). In general, first-order phase transitions are characterized by *discontinuities* in the order parameter (the jump Δm of the magnetization m in Fig. 5.1), or the energy (the latent heat Δe), or both. This reflects the fact that, at the transition temperature T_0, two (or more) phases can coexist. In the example of a magnet at low temperatures the coexisting phases are the phases with positive and negative magnetization, while at the melting transition they are the solid (ordered) and liquid (disordered) phases. The correlation length in the coexisting pure phases is usually finite. Consequently also the specific heat and the susceptibility do not diverge in the pure phases. Mathematically there are, however, superimposed delta-function like singularities associated with the jumps of e and m.

In these lecture notes we will mainly consider second-order phase transitions, which are characterized by a *divergent* correlation length at the transition point. The growth of correlations as one reaches the critical region from

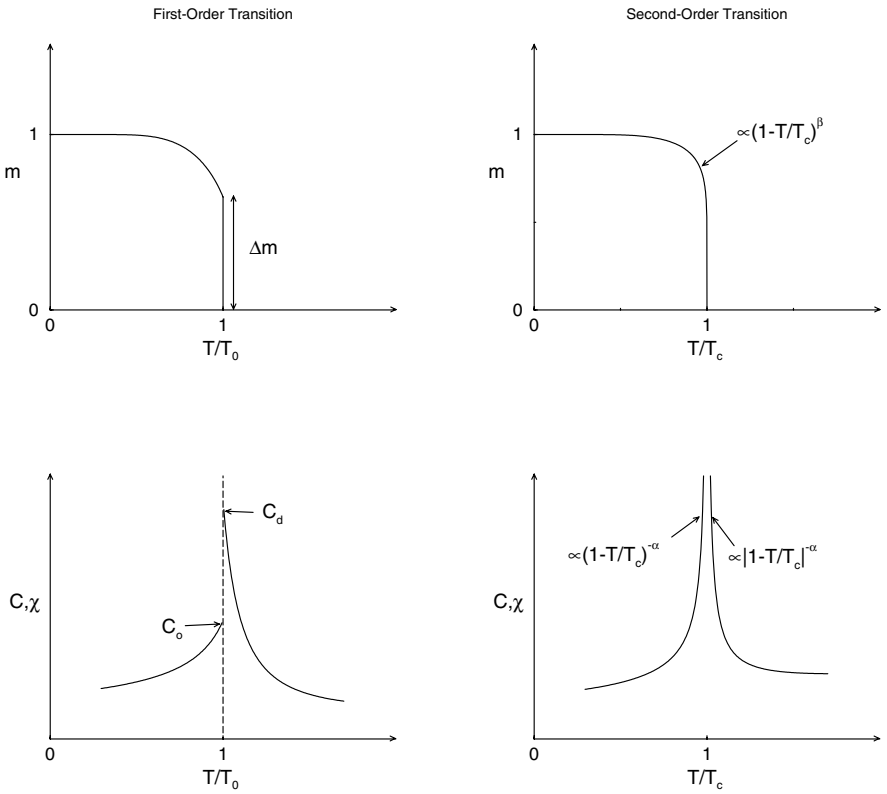

Fig. 5.1. The characteristic behaviour of the magnetization, m, specific heat, C, and susceptibility, χ, at first- and second-order phase transitions

high temperatures is illustrated in Fig. 5.2, where six typical spin configurations of the 2D Ising model on a 100×100 lattice are shown. One clearly observes the emerging larger and larger domains or clusters which eventually start percolating the system when the critical point is approached. While this apparently gives an intuitive picture of what happens near criticality, some care is necessary with the interpretation of such plots since the domains or clusters visible in Fig. 5.2 are so-called *geometrical* clusters, whose fractal and percolation properties do *not* encode the proper thermodynamic *critical* behaviour. Rather, they carry information on a closely related *tricritical* point [27]. The proper Fortuin-Kasteleyn clusters encoding the *critical* properties of the model can be constructed by a stochastic rule implied by the Fortuin-Kasteleyn representation of Potts models. These clusters, which are always smaller than the geometrical ones, form also the basis for cluster-update algorithms discussed later in Sect. 5.6.

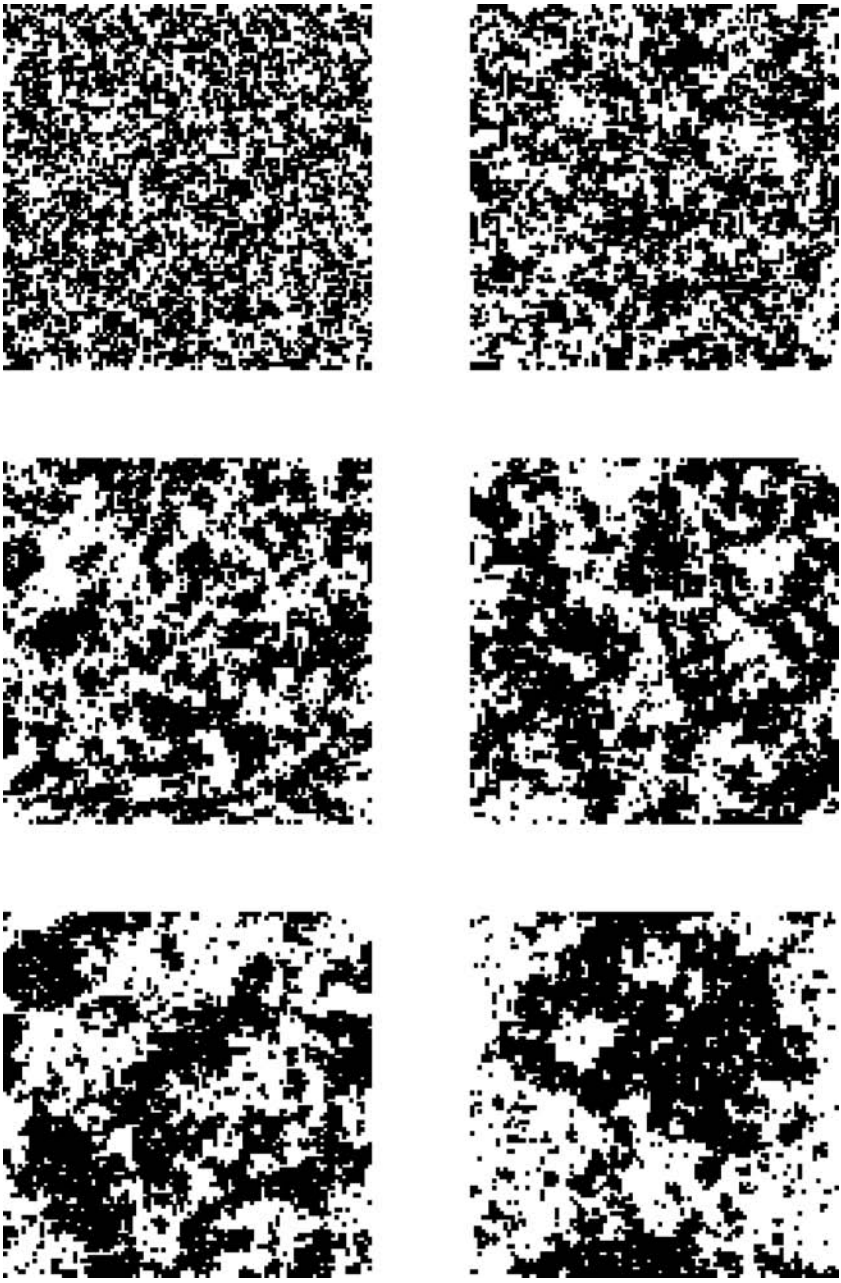

Fig. 5.2. From high temperatures (*upper left*) to the critical region (*lower right*), characterized by large spatial correlations. Shown are actual 2D Ising configurations for a 100×100 lattice at $\beta/\beta_c = 0.50$, 0.70, 0.85, 0.90, 0.95, and 0.98

Temperature Scaling

For an infinite correlation length, thermal fluctuations are equally important on all length scales, and one therefore expects power-law singularities in thermodynamic functions. The leading singularity of the correlation length is usually parameterized in the high-temperature phase as

$$\xi = \xi_{0^+} |1 - T/T_c|^{-\nu} + \dots \quad (T \geq T_c) \,, \tag{5.14}$$

where the ... indicate sub-leading corrections (analytical as well as confluent). This defines the critical exponent ν and the critical amplitude ξ_{0^+} on the high-temperature side of the transition. In the low-temperature phase one expects a similar behaviour,

$$\xi = \xi_{0^-} (1 - T/T_c)^{-\nu} + \dots \quad (T \leq T_c) \,, \tag{5.15}$$

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

An important feature of second-order phase transitions is that due to the divergence of ξ the short-distance details of the Hamiltonian should not matter. This is the basis of the *universality* hypothesis which states 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,

$$C = C_{\rm reg} + C_0 |1 - T/T_c|^{-\alpha} + \dots \,, \tag{5.16}$$

$$m = m_0 (1 - T/T_c)^{\beta} + \dots \,, \tag{5.17}$$

$$\chi = \chi_0 |1 - T/T_c|^{-\gamma} + \dots \,, \tag{5.18}$$

where $C_{\rm reg}$ is a regular background term, and the amplitudes are again in general different on the two sides of the transition, cf. Fig. 5.1. Right at the critical temperature T_c, two further exponents δ and η are defined through

$$m \propto h^{1/\delta} \,, \tag{5.19}$$

$$G(\boldsymbol{r}) \propto r^{-d+2-\eta} \,. \tag{5.20}$$

The critical exponents for the 2D and 3D Ising model and the 2D q-state Potts model with $q = 3$ and 4 are collected in Table 5.1.

Finite-Size Scaling

For systems of finite size, as in any numerical simulation, the correlation length cannot diverge, and also the divergences in all other quantities are

Table 5.1. Critical exponents of the 2D q-state Potts model with $q = 2, 3$ and 4, and the 3D Ising model. All 2D exponents are exactly known [14, 15], while for the 3D Ising model the "world-average" for ν and γ calculated in Ref. [28] 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 $\eta = 2 - \gamma/\nu$

Model	ν	α	β	γ	δ	η
2D Ising	1	0 (log)	1/8	7/4	15	1/4
3D Ising	0.630 05(18)	0.109 85	0.326 48	1.237 17(28)	4.7894	0.036 39
2D $q = 3$ Potts	5/6	1/3	1/9	13/9	14	4/15
2D $q = 4$ Potts	2/3	2/3	1/12	7/6	15	1/2

then rounded and shifted [29–32]. This is illustrated in Fig. 5.3, where the specific heat of the 2D Ising model on various $L \times L$ lattices is shown. The curves are computed from the exact solution of Kaufman [33] for any $L_x \times L_y$ lattice with periodic boundary conditions (see also Ferdinand and Fisher [34]).

Near T_c the role of ξ in the scaling formulas is then taken over by the linear size of the system, L. By rewriting

$$|1 - T/T_c| \propto \xi^{-1/\nu} \longrightarrow L^{-1/\nu} \; , \tag{5.21}$$

we see that at T_c the scaling laws (5.16)–(5.18) are replaced by the *finite-size scaling* (FSS) Ansätze,

$$C = C_{\rm reg} + aL^{\alpha/\nu} + \dots \; , \tag{5.22}$$

$$m \propto L^{-\beta/\nu} + \dots \; , \tag{5.23}$$

$$\chi \propto L^{\gamma/\nu} + \dots \; . \tag{5.24}$$

In general these scaling laws are valid in the vicinity of T_c as long as the scaling variable $x = (1 - T/T_c)L^{1/\nu}$ is kept fixed [29–32]. In particular this is true for the locations $T_{\rm max}$ of the (finite) maxima of thermodynamic quantities such as the specific heat or susceptibility, which are expected to scale with the system size as

$$T_{\rm max} = T_c(1 - x_{\rm max}L^{-1/\nu} + \dots) \; . \tag{5.25}$$

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

$$\chi(T, L) = L^{\gamma/\nu} f(x) \; . \tag{5.26}$$

By plotting $\chi(T, L)/L^{\gamma/\nu}$ vs the scaling variable x, one thus expects that the data for different T and L fall onto a kind of master curve. This is a nice way to demonstrate the scaling properties visually.

Similar considerations for first-order phase transitions show that also the delta function like singularities, originating from phase coexistence, are smeared out for finite systems [35–39]. They are replaced by narrow peaks

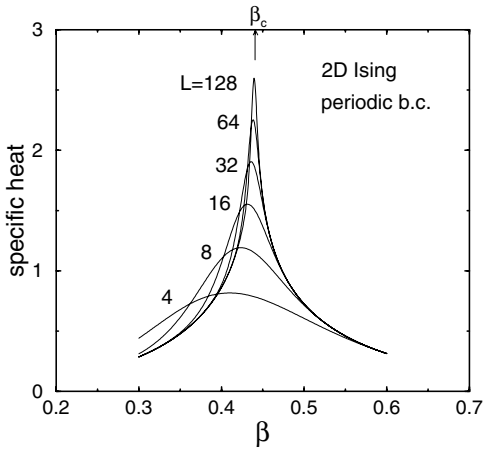

Fig. 5.3. Finite-size scaling behaviour of the specific heat of the 2D Ising model on $L \times L$ lattices. The critical point is indicated by the arrow on the top axis

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 [16, 40–44].

5.3 The Monte Carlo Method

Let us now discuss how the expectation values in (5.3)–(5.7) can be computed numerically. A direct summation of the partition function is impossible, since already for the Ising model with only two possible states per site the number of terms would be enormous: $2^{2500} \approx 10^{750}$ for a modestly large 50×50 lattice![1] 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: One has to use the *importance sampling* technique [45].

5.3.1 Importance Sampling

The basic idea of importance sampling is that one does not pick configurations at random, but draws them directly according to their Boltzmann weight,

$$P^{\mathrm{eq}}(\{\sigma_i\}) \propto \exp\left(-\beta H(\{\sigma_i\})\right) \ . \tag{5.27}$$

[1] For comparison, the estimated number of protons in the Universe is 10^{80}.

In more mathematical terms one sets up a Markov chain,

$$\cdots \xrightarrow{W} \{\sigma_i\} \xrightarrow{W} \{\sigma_i'\} \xrightarrow{W} \{\sigma_i''\} \xrightarrow{W} \cdots ,$$

with a transition probability W satisfying the conditions

$$\text{(a)}\quad W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) \geq 0 \quad \text{for all} \quad \{\sigma_i\}, \{\sigma_i'\} , \tag{5.28}$$

$$\text{(b)}\quad \sum_{\{\sigma_i'\}} W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = 1 \quad \text{for all} \quad \{\sigma_i\} , \tag{5.29}$$

$$\text{(c)}\quad \sum_{\{\sigma_i\}} W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) P^{\mathrm{eq}}(\{\sigma_i\}) = P^{\mathrm{eq}}(\{\sigma_i'\}) \quad \text{for all} \quad \{\sigma_i'\} . \tag{5.30}$$

From (5.30) we see that the desired Boltzmann distribution P^{eq} is a fixed point of W. A somewhat simpler sufficient condition is *detailed balance*,

$$P^{\mathrm{eq}}(\{\sigma_i\}) W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = P^{\mathrm{eq}}(\{\sigma_i'\}) W(\{\sigma_i'\} \longrightarrow \{\sigma_i\}) . \tag{5.31}$$

By summing over $\{\sigma_i\}$ and using (5.29), the more general condition (5.30) follows. After an initial equilibration period (cf. Sect. 5.4), expectation values can then be estimated as an arithmetic mean over the Markov chain, e.g.,

$$E = \langle H \rangle = \sum_{\{\sigma_i\}} H(\{\sigma_i\}) P^{\mathrm{eq}}(\{\sigma_i\}) \approx \frac{1}{N} \sum_{j=1}^{N} H(\{\sigma_i\}_j) , \tag{5.32}$$

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 [1–4, 29, 46, 47].

5.3.2 Local Update Algorithms

The required Markov chain properties can be satisfied with many different concrete update rules. These can be roughly divided into *local* and *non-local* algorithms. While non-local algorithms such as multigrid schemes or the cluster-update methods to be discussed later in Sect. 5.6 may considerably improve the performance of the simulations, they are more specialized and hence usually not automatically applicable to a given arbitrary physical system. This is why the conceptually much simpler local algorithms continue to be very important.

Metropolis Algorithm

The most flexible update prescription is the standard Metropolis algorithm [48] where the Markov chain is realized by *locally* updating the degrees of freedom step by step. This works for discrete and continuously varying degrees

of freedom, and for lattice and off-lattice formulations. Examples for lattice formulations range from our simple, paradigmatic Ising model, over freely rotating Heisenberg spins to field theories such as the Ginzburg-Landau model. Also all kinds of lattice gauge theories and even non-perturbative formulations of quantum gravity can be simulated with this method. Off-lattice formulations cover a huge range of physical phenomena. Prominent examples are simulations of fluids, polymers and proteins, to name only a few important applications. Depending on the problem at hand, the degrees of freedom may be spins, field values or gauge potentials, or particle positions in space. There is also no principle restriction on the form of the interactions which may be short- or long-ranged or even of mean-field type.

If E and E' denote the energy before and after the proposed local update, respectively, then the probability to accept this proposal is given by [48]

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \begin{cases} 1 & E' < E \\ \exp\left[-\beta(E'-E)\right] & E' \geq E\,, \end{cases} \tag{5.33}$$

where the proposed new spin configuration $\{\sigma_i'\}$ differs from $\{\sigma_i\}$ only by a single flipped spin. More compactly, this may also be written as

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \min\{1, \exp\left[-\beta(E'-E)\right]\}\,. \tag{5.34}$$

If the energy is lowered by the proposed update, it is thus always accepted. On the other hand, when the energy would be increased for the new configuration, the update has still to be accepted with a certain probability in order to ensure the proper treatment of entropic contributions – in thermal equilibrium, the *free* energy is minimized and not the energy. Only in the limit of zero temperature, $\beta \longrightarrow \infty$, this probability tends to zero and the MC algorithm degenerates to a minimization algorithm for the energy functional. With some additional refinements, this is the basis of the simulated annealing technique [49], which is often applied to hard optimization and minimization problems.

To show that the detailed balance condition (5.31) is indeed satisfied, we first consider the case that the proposed spin update lowers the energy, $E' < E$. In this case, the l.h.s. of (5.31) becomes $\exp(-\beta E)\times 1 = \exp(-\beta E)$. On the r.h.s. we have to take into account that the reverse move would increase the energy, $E > E'$, with E now playing the role of the "new" energy. Hence now the second line of (5.33) with E and E' interchanged is relevant, such that the r.h.s. of (5.31) becomes $\exp(-\beta E') \times \exp(-\beta(E-E')) = \exp(-\beta E)$, proving the equality of the two sides of the detailed balance condition. In the case that the proposed spin update increases the energy, $E' < E$, a similar reasoning leads to $\exp(-\beta E) \times \exp(-\beta(E'-E)) = \exp(-\beta E') = \exp(-\beta E') \times 1$.

Even though this "proof" looks rather like a tautology, it is indeed nontrivial, as one can easily convince oneself by replacing the r.h.s. of the Metropolis rule (5.33) by some general function $f(E'-E)$. The detailed balance condition then reads $\exp(-\beta E) f(E'-E) = \exp(-\beta E') f(E-E')$. With $\Delta E \equiv E'-E$ this can be recast into the form

$$g(\Delta E) \equiv \exp(\beta\Delta E/2) f(\Delta E) = \exp(-\beta\Delta E/2) f(-\Delta E) = g(-\Delta E) \ , \quad (5.35)$$

showing that $g(\Delta E) \equiv \exp(\beta\Delta E/2) f(\Delta E)$ can be quite a general function which, however, must be even in ΔE, $g(\Delta E) = g(-\Delta E)$. The simplest choice $g(\Delta E) = \text{const.}$, leads to $f(\Delta E) = \text{const.} \exp(-\beta\Delta E/2)$. While this would satisfy detailed balance, it is still not a permissible choice because the r.h.s. of (5.33) should admit the interpretation as a probability. For a given model, this can often be repaired by considering the allowed range of ΔE and introducing a suitable normalization factor [1]. Requiring thus that $0 \leq f(\Delta E) \leq 1$, we see that $0 \leq g(\Delta E) \leq \exp(\beta\Delta E/2)$. Choosing just the upper bound, $g(\Delta E) = \exp(\beta\Delta E/2)$ for $\Delta E < 0$ and applying a (non-differentiable) symmetrization to define $g(\Delta E)$ for $\Delta E \geq 0$, we end up with

$$g(\Delta E) = \exp(\beta\Delta E/2) \min\{1, \exp(-\beta\Delta E)\} = \begin{cases} \exp(\beta\Delta E/2) & \Delta E < 0 \\ \exp(-\beta\Delta E/2) & \Delta E \geq 0 \ , \end{cases} \quad (5.36)$$

implying $f(\Delta E) = \min\{1, \exp(-\beta\Delta E)\}$ – which is nothing but the Metropolis rule (5.34).

How is the Metropolis update rule (5.33) implemented in practice? Since the possible values of the transition probability W are restricted to values between 0 and 1, one first draws a uniformly distributed random number $r \in [0, 1)$. Then, if $W \leq r$, the proposed update is accepted, and otherwise it is rejected and one continues with the next spin. In words this is easy to state. In practice, however, "drawing a random number" in a computer program is a pretty involved mathematical problem [50]. Since in most applications a huge number of random numbers is required (for, say, 1 million sweeps through a 2D Ising lattice of size $1000 \times 1000 = 10^6$ already 10^{12}) and each random number usually occupies 8 Bytes, it is neither practical nor feasible to store physically generated, "truly" random (whatever that really means ...) events on a hard disk. Also, reading them from the hard disk into the computer memory would be far too slow. Therefore, one uses in MC computer simulations so-called "pseudo-random number generators", or short RNGs, which use deterministic rules to produce (more or less) uniformly distributed numbers, whose values are "very hard" to predict. 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 a 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, a RNG should be portable among different computer platforms, and it should yield reproducible results for testing purposes.

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. 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[2] for all degrees of freedom an update was proposed. The qualitative behaviour of the update algorithm is not sensitive to these details, but its quantitative performance does depend on the choice of update scheme.

The big merit of this simple algorithm is its flexibility which allows the application to a great variety of physical systems. The main drawback of this and most other *local* update algorithms (one exception is the overrelaxation method [51–54]) is that it is plagued by large autocorrelation times which severely limit the statistical accuracy achievable with a given computer budget as will be explained in detail in Sect. 5.5.

Heat-Bath Algorithm

This algorithm is only applicable to lattice models and at least in its most straightforward form only to discrete degrees of freedom with a few allowed states. The new value of the selected variable at site i_0 is determined by testing all its possible states in the "heat-bath" of its (fixed) neighbors (i.e., 4 on a square lattice and 6 on a simple-cubic lattice with nearest-neighbor interactions):

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \frac{e^{-\beta H(\{\sigma_i'\}}}{\sum_{\sigma_{i_0}'} e^{-\beta H(\{\sigma_i'\}}} \;, \tag{5.37}$$

which obviously satisfies the detailed balance condition (5.31) since

$$e^{-\beta H(\{\sigma_i\}} \frac{e^{-\beta H(\{\sigma_i'\}}}{\sum_{\sigma_{i_0}'} e^{-\beta H(\{\sigma_i'\}}} = e^{-\beta H(\{\sigma_i'\}} \frac{e^{-\beta H(\{\sigma_i\}}}{\sum_{\sigma_{i_0}} e^{-\beta H(\{\sigma_i\}}} \;. \tag{5.38}$$

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 the special case of the Ising model with only two states per spin, (5.37) may be written more explicitly as

$$W(\{\sigma_i\} \longrightarrow \{\sigma_i'\}) = \frac{e^{-\beta \sigma_{i_0}' E_{i_0}}}{e^{\beta E_{i_0}} + e^{-\beta E_{i_0}}} \;, \tag{5.39}$$

where $\sigma_{i_0} E_{i_0}$ is the energy of the spin at site i_0 in the state σ_{i_0}, that is $E_{i_0} = -J \sum_j \sigma_j - h$, where j runs over all sites interacting with site i_0 and h is the external magnetic field. The energy difference $\Delta E = E_{\text{new}} - E_{\text{old}}$ can be expressed as $\Delta E = (\sigma_{i_0}' - \sigma_{i_0}) E_{i_0}$, since by definition no other interactions are affected by the spin value at site i_0.

Let us now assume that before the update $\sigma_{i_0} = +1$. The probability that the spin is flipped to $\sigma_{i_0}' = -1$ is then

[2] This is only relevant when the random update order is chosen.

$$W(\sigma_{i_0} \longrightarrow -\sigma_{i_0}) = \frac{e^{\beta E_{i_0}}}{e^{\beta E_{i_0}} + e^{-\beta E_{i_0}}} \ . \tag{5.40}$$

Since in this case $\Delta E = -2E_{i_0}$, the flip probability can be equivalently written as

$$W(\sigma_{i_0} \longrightarrow -\sigma_{i_0}) = \frac{e^{-\beta \Delta E/2}}{e^{\beta \Delta E/2} + e^{-\beta \Delta E/2}} \ . \tag{5.41}$$

This is also true in the other case where initially $\sigma_{i_0} = -1$. The heat-bath probability of a flip to $\sigma'_{i_0} = +1$ is then $e^{-\beta E_{i_0}} / \left(e^{\beta E_{i_0}} + e^{-\beta E_{i_0}}\right)$, but since the energy difference now reads $\Delta E = +2E_{i_0}$, we again arrive at the flip probability (5.41).

The order of updating the individual variables can be done as for the Metropolis algorithm (random, sequential, ...).

Glauber Algorithm

This update procedure [55], named after the 2005 Nobel Laureate Roy J. Glauber of Harvard University[3], is conceptually similar to the Metropolis algorithm in that one also here proposes locally a change for a single degree of freedom and then accepts this update proposal with a certain probability. For the Ising model with spins $\sigma_i = \pm 1$ this update rule is often written as

$$W(\sigma_i \longrightarrow -\sigma_i) = \frac{1}{2}\left[1 + \sigma_i \tanh\left(\beta E_i\right)\right] \ , \tag{5.42}$$

where as before $\sigma_i E_i$ is the energy of the ith spin in the current "old" state, that is $E_i = -J \sum_j \sigma_j - h$.

Since $\sigma_i = \pm 1$ and using the point symmetry of the tanh-function, one may rewrite $\sigma_i \tanh\left(\beta E_i\right)$ as $\tanh\left(\sigma_i \beta E_i\right)$. In a local spin flip $\sigma_i \longrightarrow -\sigma_i$, only the energy contributions collected in E_i are affected, and we obtain again $\Delta E = E_{\text{new}} - E_{\text{old}} = -2\sigma_i E_i$ for the total energy change due to the proposed flip. Hence we can rewrite (5.42) as

$$W(\sigma_i \longrightarrow -\sigma_i) = \frac{1}{2}\left[1 - \tanh\left(\beta \Delta E/2\right)\right] \ . \tag{5.43}$$

In this representation, the acceptance probability is explicitly seen to depend only on the total energy change – similar to the Metropolis case. In this form it is thus possible to generalize the Glauber update rule from the Ising model

[3] Half of the Nobel Prize in Physics 2005 was awarded to Roy J. Glauber for his outstanding theoretical contributions to what is called today "Quantum Optics", with his seminal papers dating back to the year 1963 [Phys. Rev. Lett. **10**, 84 (1963); Phys. Rev. **130**, 2529 (1963); *ibid.* **131**, 2766 (1963)], when also his paper [55] on dynamical properties of the Ising model appeared. The other half of the 2005 Prize is shared by John L. Hall of the University of Colorado and Theodor W. Hänsch of Ludwig-Maximilians-Universität Munich.

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.

By using trivial identities for hyperbolic functions, (5.43) can be further recast to read

$$W(\sigma_i \longrightarrow -\sigma_i) = \frac{1}{2}\left[\frac{\cosh(\beta\Delta E/2) - \sinh(\beta\Delta E/2)}{\cosh(\beta\Delta E/2)}\right] = \frac{e^{-\beta\Delta E/2}}{e^{\beta\Delta E/2} + e^{-\beta\Delta E/2}} \,. \tag{5.44}$$

Notice that this agrees with the flip probability (5.41) 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.

The Glauber (or equivalently heat-bath) update algorithm for the Ising model is also of theoretical interest since in this case the MC (pseudo-) dynamics can be calculated analytically – albeit only in one dimension [55]. For two and higher dimensions no exact solutions are known.

5.4 Initial Non-Equilibrium Period and Ageing

The initial equilibration or thermalization period, in general, is a non-trivial non-equilibrium process which is of interest it its own right. 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, we only need the methods introduced so far since most theoretical concepts assume a *local* spin-flip dynamics as realized by one the three 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 we choose instead to do the simulation right at T_c or at a temperature $T < T_c$, the situation is, however, quite different. In the latter two cases one speaks of a "quench", since 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. This mechanism is illustrated in Fig. 5.4 for the 2D Ising and 3-state Potts model, showing the time evolution after a quench to $T < T_c$ from an initially completely disordered state. This leads

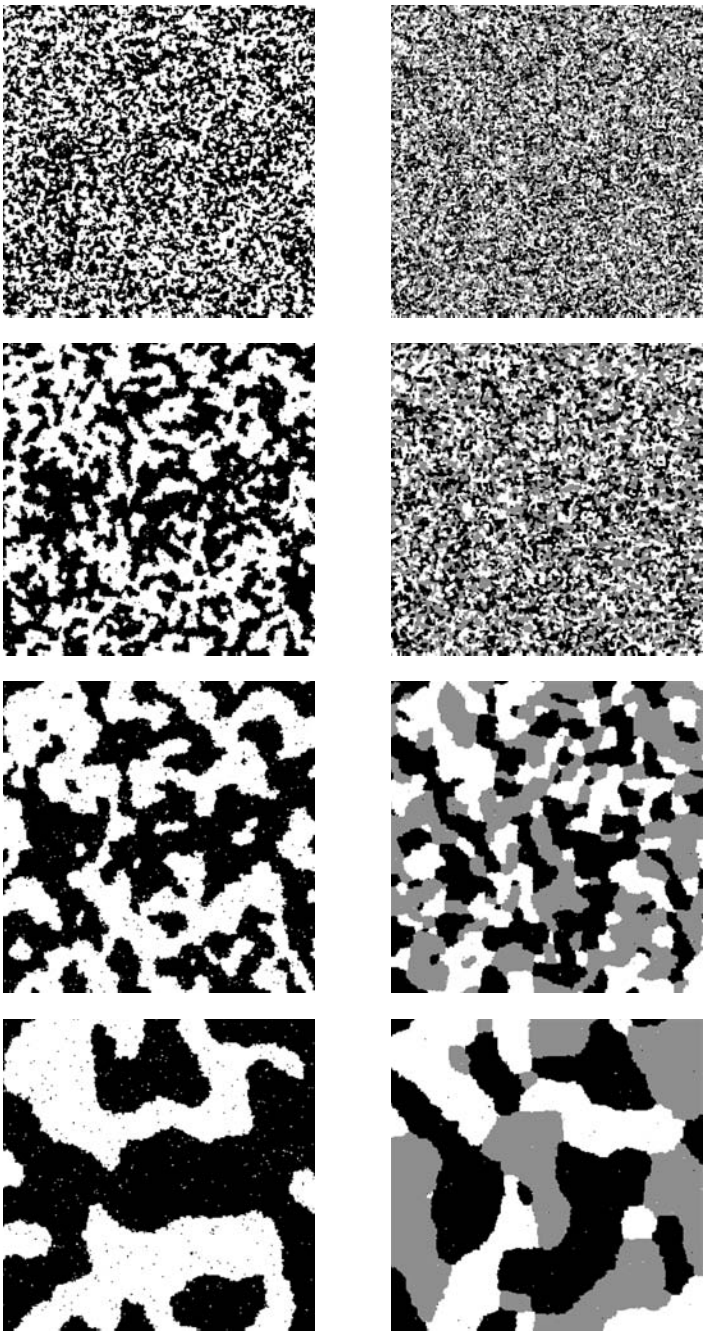

Fig. 5.4. Phase-ordering with progressing MC "time" (*from top to bottom*) of initially disordered spin configurations for the 2D Ising model at $T = 1.5 \approx 0.66\,T_c$ (*left*) and the 2D 3-state Potts model at $T = 0.4975 \approx T_c/2$ (*right*) (from Ref. [75])

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. In the case of a simple ferromagnet like the Ising- or q-state Potts model with a non-conserved scalar order parameter, the dynamical exponent can be found exactly as $z = 2$ [56], according to diffusion or random-walk arguments. Right at the transition temperature, critical dynamics (for a recent review, see Ref. [57]) plays the central role and the dynamical exponent takes the somewhat larger non-trivial value $z \approx 2.17$ [58]. To equilibrate the whole system, ξ must approach the system size L, so that the typical relaxation time for equilibration scales as

$$\tau_{\text{relax}} \sim L^z \quad . \tag{5.45}$$

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.

Commonly considered two-time quantities are the two-time autocorrelation function (in q-state Potts model notation)

$$C(t,s) = \frac{1}{q-1}\left(\frac{q}{V}\sum_{i=1}^{V}\left[\delta_{\sigma_i(t),\sigma_i(s)}\right]_{\text{av}} - 1\right) \tag{5.46}$$

and the two-time response function

$$R(t,s) = \left.\frac{\delta[\sigma_i(t)]_{\text{av}}}{\delta h(s)}\right|_{h=0} , \tag{5.47}$$

where $h(s)$ 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 (5.47)). In computer simulation studies it is more convenient to consider the integrated response or thermoremanent magnetization (TRM) [59],

$$\rho(t,s) = T\int_0^s du R(t,u) = \frac{T}{h}M_{\text{TRM}}(t,s) \; . \tag{5.48}$$

Dynamical scaling arguments predict the scaling forms (for reviews see, e.g., Refs. [60,61])

$$C(t,s) = s^{-b} f_C(t/s) \; , \qquad R(t,s) = s^{-1-a} f_R(t/s) \; , \tag{5.49}$$

with scaling functions f_C and f_R which approach for large values of the scaling variable $x \equiv t/s$ the power-law behaviour

$$f_C(x) \to x^{-\lambda_C/z} , \qquad f_R(x) \to x^{-\lambda_R/z} \quad (x \gg 1) . \tag{5.50}$$

In phase-ordering kinetics after a quench to $T < T_c$, $b = 0$ and $z = 2$ [56]. As the other exponents depend on the dimensionality of the considered system, we shall focus here on two dimensions only, where for the Ising model, it is commonly accepted that $\lambda_C = \lambda_R = 5/4$. The value of the remaining exponent a, however, is more controversial [62]. In the literature there are strong claims for $a = 1/z = 1/2$ [60,63], but also $a = 1/4$ [64] has been conjectured.

Extending the symmetry considerations to *local* scale invariance in analogy to conformal invariance [65], even the explicit form of the scaling function $f_R(x)$ has been predicted [66,67],

$$f_R(x) = r_0 x^{1+a-\lambda_R/z}(x-1)^{-1-a} , \tag{5.51}$$

where r_0 is a normalization constant. The integration over the response function in (5.48) leads for the thermoremanent magnetization to the scaling form [64,68–70]

$$\rho(t,s) = r_0 s^{-a} f_M(t/s) + r_1 s^{-\lambda_R/z} g_M(t/s) , \tag{5.52}$$

where some care is necessary in dealing with cross-over effects leading to the second term which can be argued to take the explicit form $g_M(x) \approx x^{-\lambda_R/z}$. The first term follows directly from the integration over the explicit expression for $f_R(x)$ in (5.51) which results in a hypergeometric function [67,69],

$$f_M(x) = x^{-\lambda_R/z} \, {}_2F_1(1+a, \lambda_R/z - a; \lambda_R/z - a + 1; 1/x) . \tag{5.53}$$

Due to the linear combination of scaling functions in (5.52) with s-dependent prefactors, the scaling properties cannot be tested easily. One therefore usually subtracts first the correction term $\propto g_M(x)$ and then considers $f_M(x)$. While the two-time autocorrelation function $C(t,s)$ is conceptually and in particular computationally the much simpler quantity, local scale invariance predictions are much harder to derive for $C(t,s)$ than for $R(t,s)$. The expression for $f_C(x)$ contains combinations of hypergeometric and incomplete Gamma functions, depending on three additional undetermined constants apart from a normalization factor [71].

In computer simulations one proceeds as follows. One prepares many independent disordered start configurations of the order of a few hundred to a few thousand and monitors for each of them the time evolution after the quench to $T < T_c$ or, in critical relaxation [57], to $T = T_c$. Here it is important to make sure that the time evolutions are statistically independent of each other. In practice this means that different random number sequences have to be used

for each sample.[4] The final result (for each time s and t) is then an average over these samples.

For the 2D and 3D Ising model, extensive numerical tests of the scaling predictions have been performed by Henkel, Pleimling and collaborators [69–71], showing a very good agreement with the almost parameter-free analytical expressions. To check the generality of the scaling arguments, we extended this work in a recent MC study [75,76] to more general q-state Potts models in two dimensions. Figures 5.5 and 5.6 compare our numerical results for the 2D Ising ($q = 2$) and 3-state Potts model after a quench to $T = 1.5 \approx 0.66\,T_c$ (in Ising model normalization) and $T = 0.4975 \approx T_c/2$, respectively (assuming in both cases $\lambda_C = \lambda_R$ with $\lambda_C \approx 1.25$ [75, 76] and $a = 1/z = 1/2$). We see that the two models behave very similarly during ageing, i.e., also for the 3-state Potts model the scaling predictions (5.49) are well satisfied. Moreover, the explicit analytical predictions for $f_M(t/s)$ in (5.53) and (the more complicated one) for $C(t, s)$ as given in Ref. [71] relying on local scale invariance are both in excellent agreement with the MC data. For details of the numerical set-up, see Refs. [75, 76], where also additional simulations of the 2D 8-state Potts model are described that give similarly good results.

5.5 Statistical Analysis of Monte Carlo Data

About a decade ago most of the statistical analysis methods discussed in this section were still quite cumbersome since due to disk-space limitations they usually had to be applied "on the flight" during the simulation. In particular dynamical aspects of a given model are usually not easy to predict beforehand such that the guess of reasonable analysis parameters was quite difficult. The situation has changed dramatically when it became affordable to store hundreds of megabytes on hard-disks. Since then a simulation study can clearly be separated into "raw data generation" and "data analysis" parts. The interface between these two parts should consist of time series of measurements of the relevant physical observables taken during the actual simulations. In principle there are no limitations on the choice of observables $\mathcal{O}$ which could be, for example, the energy H or the magnetization μ. Once the system is in equilibrium (which, in general, is non-trivial to assure), we simply save $\mathcal{O}_j \equiv \mathcal{O}[\{\sigma_i\}_j]$ where j labels the measurements. Given these data files one can perform detailed error analyses; in particular adapting parameters to a specific situation is now straightforward and very fast.

[4] If the same sequence of random numbers, i.e., the same dynamics, would be used for all samples with different start configurations, then one would study the phenomenon of "damage spreading" [72–74], where one basically asks how likely it is that two initially different configurations merge after some time into the same state when evolving under exactly the *same* dynamics.

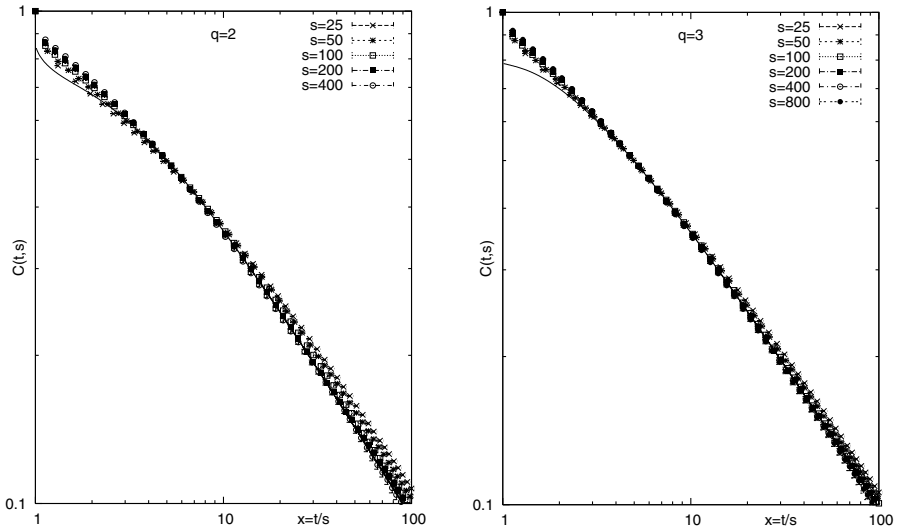

Fig. 5.5. Two-time autocorrelation function for the 2D Ising ($q = 2$) (*left*) and 3-state Potts (*right*) models [75,76]. The different data symbols correspond to different waiting times s. The *solid lines* show the fits to the scaling prediction [71] based on local scale invariance

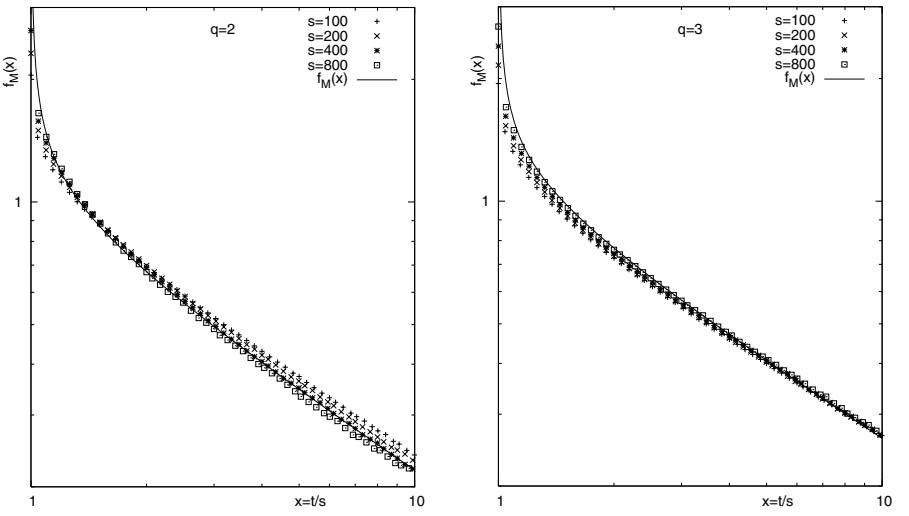

Fig. 5.6. Leading contribution $f_M(x)$ to the thermoremanent magnetization (5.48) and (5.52) for the 2D Ising ($q = 2$) (*left*) and 3-state Potts (*right*) models [75,76]. The different data symbols correspond to different waiting times s. The *solid lines* show the local scale invariance prediction (5.53) which depends only a single normalization parameter

5.5.1 Estimators

If the time series data result from an importance sampling MC simulation, the expectation value $\langle \mathcal{O} \rangle$ can be estimated as a simple arithmetic mean over the Markov chain,

$$\overline{\mathcal{O}} = \frac{1}{N} \sum_{j=1}^{N} \mathcal{O}_j \ , \tag{5.54}$$

where we assume that the time series collects, after an appropriate equilibration period, a total of N measurements. 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 thermal average (which is only for very few models known), the estimator $\overline{\mathcal{O}}$ is still a *random* number, fluctuating around the theoretically expected value. Of course, in practice this is a "virtual" concept as one does not probe the fluctuations of the mean value directly since this would require repeating the whole MC simulation many times. However, one can estimate its variance,

$$\sigma^2_{\overline{\mathcal{O}}} = \langle [\overline{\mathcal{O}} - \langle \overline{\mathcal{O}} \rangle]^2 \rangle = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 \ , \tag{5.55}$$

from the statistical properties of individual measurements $\mathcal{O}_j$ in a single MC run.

5.5.2 Uncorrelated Measurements and Central-Limit Theorem

For simplicity, let us first make the unrealistic assumption that the N subsequent measurements $\mathcal{O}_j$ are all completely uncorrelated (as would be true only in simple sampling). Then the relation between the two variances would simply be

$$\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_j} / N \ , \tag{5.56}$$

where $\sigma^2_{\mathcal{O}_j} = \langle \mathcal{O}_j^2 \rangle - \langle \mathcal{O}_j \rangle^2$ is the variance of the individual measurements. A further, milder assumption is, of course, that the simulation is already in equilibrium so that time-translation invariance over the Markov chain is satisfied. Equation (5.56) is true for any distribution $\mathcal{P}(\mathcal{O}_j)$ of the $\mathcal{O}_j$. For the energy or magnetization the latter distributions are often plotted as physically directly relevant histograms (see, e.g., Fig. 5.14(b) below) whose squared width ($= \sigma^2_{\mathcal{O}_j}$) is proportional to the specific heat or susceptibility, respectively.

Whatever form the distribution $\mathcal{P}(\mathcal{O}_j)$ assumes (which, in fact, is already often close to Gaussian because the $\mathcal{O}_j$ are usually 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, $\sigma^2_{\overline{\mathcal{O}}}$, is the squared width of this (N dependent) distribution which is usually taken as the "one-sigma" squared error, $\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}}$, 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 $[\overline{\mathcal{O}} - \sigma_{\overline{\mathcal{O}}}, \overline{\mathcal{O}} + \sigma_{\overline{\mathcal{O}}}]$. 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%.

5.5.3 Correlated Measurements and Autocorrelation Times

In "real life" things become more involved since when using importance sampling update algorithms subsequent measurements are necessarily correlated in time [77–79]. Inserting (5.54) into (5.55), one obtains

$$\sigma^2_{\overline{\mathcal{O}}} = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \mathcal{O}_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \ , \tag{5.57}$$

and by collecting diagonal and off-diagonal terms one arrives at

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right) \ . \tag{5.58}$$

The first term is identified as the variance of the individual measurements multiplied with $1/N$. In the second sum we first use the symmetry $i \leftrightarrow j$ to reduce the summation to $\sum_{i \neq j}^{N} = 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N}$. Reordering the summation and using time-translation invariance (assuming that equilibrium has already been reached, cf. the previous Sect. 5.4) we finally get

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N} \left[\sigma^2_{\mathcal{O}_i} + 2 \sum_{k=1}^{N} \left(\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle \right) \left(1 - \frac{k}{N} \right) \right] \ , \tag{5.59}$$

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

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{\sigma^2_{\mathcal{O}_i}}{N} \, 2\tau'_{\mathcal{O},\mathrm{int}} \ , \tag{5.60}$$

where we have introduced the (proper) *integrated* autocorrelation time

$$\tau'_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \left(1 - \frac{k}{N} \right) \ , \tag{5.61}$$

with

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} \tag{5.62}$$

denoting the normalized autocorrelation function ($A(0) = 1$).

For large time separations k the autocorrelation function decays exponentially ($a = \text{const.}$),

$$A(k) \stackrel{k\to\infty}{\longrightarrow} a e^{-k/\tau_{\mathcal{O},\text{exp}}} , \tag{5.63}$$

which defines the *exponential* autocorrelation time $\tau_{\mathcal{O},\text{exp}}$. Since in any meaningful simulation study $N \gg \tau_{\mathcal{O},\text{exp}}$, $A(k)$ in (5.61) 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 (5.61)) and one employs the following definition for the *integrated* autocorrelation time:

$$\tau_{\mathcal{O},\text{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) . \tag{5.64}$$

The notion "integrated" derives from the fact that this may be interpreted as a trapezoidal discretization of the (approximate) integral $\tau_{\mathcal{O},\text{int}} \approx \int_0^N \mathrm{d}k A(k)$. 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 [80] that $\tau_{\mathcal{O},\text{int}} \le \tau_{\mathcal{O},\text{exp}}$ 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 $\tau_{\mathcal{O},\text{int}}$, see Eq. (5.86) below) [79].

Close to a critical point, the autocorrelation time scales for an infinite system typically as

$$\tau_{\mathcal{O},\text{int}} \propto \tau_{\mathcal{O},\text{exp}} \propto \xi^z , \tag{5.65}$$

where z is the dynamical critical exponent. For local algorithms, $z \approx 2$, which can be understood by a random-walk argument. Since $\xi \propto |T - T_c|^{-\nu} \to \infty$ when $T \to T_c$, also τ diverges when the critical point is approached. This leads to the phenomenon of *critical slowing down* at a continuous phase transition. In a finite system with extent L, ξ is basically replaced by L and

$$\tau_{\mathcal{O},\text{int}} \propto \tau_{\mathcal{O},\text{exp}} \propto L^z . \tag{5.66}$$

Non-local update algorithms such as multigrid schemes or in particular the cluster methods discussed later in Sect. 5.6 can reduce the value of the dynamical critical exponent z significantly, albeit in a model-dependent fashion.

At a first-order phase transition the "slowing-down" problem is even more severe, but the mechanism is completely different. Here, 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 2 accounts for periodic boundary conditions, which enforce for simple topological reasons always an even number of interfaces [16]. Whatever update algorithm is used, the time spent for crossing this highly suppressed rare-event region scales proportional to the inverse of this interfacial Boltzmann factor, i.e., the autocorrelation time behaves as

$$\tau \propto e^{2\sigma L^{d-1}} \ . \tag{5.67}$$

This exponential increase of autocorrelations with system size at a first-order phase transition is often described in the literature as *supercritical slowing down* (even though, strictly speaking, nothing is "critical" here). This type of slowing-down problem can be overcome in part by means of tempering and multicanonical methods also discussed later in Sects. 5.8 and 5.9.

As far as the accuracy of MC data is concerned, the important point of Eq. (5.60) is that due to temporal correlations of the measurements the statistical error $\epsilon_{\overline{\mathcal{O}}} \equiv \sqrt{\sigma^2_{\overline{\mathcal{O}}}}$ on the MC estimator $\overline{\mathcal{O}}$ is enhanced by a factor of $\sqrt{2\tau_{\mathcal{O},\mathrm{int}}}$. This can be rephrased by writing the statistical error similar to the uncorrelated case as $\epsilon_{\overline{\mathcal{O}}} = \sqrt{\sigma^2_{\mathcal{O}_j}/N_{\mathrm{eff}}}$, but now with a parameter

$$N_{\mathrm{eff}} = N/2\tau_{\mathcal{O},\mathrm{int}} \leq N \ , \tag{5.68}$$

describing the *effective* statistics. This shows more clearly that only every $2\tau_{\mathcal{O},\mathrm{int}}$ 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 behaviour of the autocorrelation time in (5.65) or (5.66) respectively (5.67), it is obvious that without extra care this effective sample size may become very small close to a continuous or first-order phase transition. Since some quantities (e.g., the specific heat or susceptibility) can severely be underestimated if the effective statistics is too small [81], any serious simulation should therefore provide at least a rough order-of-magnitude estimate of autocorrelation times.

5.5.4 Bias

For a better understanding of the latter point, let us consider as a specific example the specific heat, $C = \beta^2 V \left(\langle e^2\rangle - \langle e\rangle^2\right) = \beta^2 V \sigma^2_{e_i}$. The standard estimator for the variance is

$$\hat{\sigma}^2_{e_i} = \overline{e^2} - \overline{e}^2 = \overline{(e-\overline{e})^2} = \frac{1}{N}\sum_{i=1}^{N} (e_i - \overline{e})^2 \ . \tag{5.69}$$

What is the *expected* value of $\hat{\sigma}^2_{e_i}$? To answer this question, we subtract and add $\langle\overline{e}\rangle^2$,

$$\langle\hat{\sigma}^2_{e_i}\rangle = \langle\overline{e^2} - \overline{e}^2\rangle = \langle\overline{e^2}\rangle - \langle\overline{e}\rangle^2 - \left(\langle\overline{e}^2\rangle - \langle\overline{e}\rangle^2\right) \ , \tag{5.70}$$

and then use the previously derived result: The first two terms on the r.h.s. of (5.70) just give $\sigma^2_{e_i}$, and the second two terms in parentheses yield $\sigma^2_{\overline{e}} = \sigma^2_{e_i} 2\tau_{e,\mathrm{int}}/N$, as calculated in (5.60). Combining these two results we arrive at

$$\langle\hat{\sigma}^2_{e_i}\rangle = \sigma^2_{e_i}\left(1 - \frac{2\tau_{e,\mathrm{int}}}{N}\right) = \sigma^2_{e_i}\left(1 - \frac{1}{N_{\mathrm{eff}}}\right) \neq \sigma^2_{e_i} \ . \tag{5.71}$$

The estimator $\hat{\sigma}^2_{e_i}$ as defined in (5.69) thus systematically underestimates the true value by a term of the order of $\tau_{e,\mathrm{int}}/N$. 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 $\tau_{e,\mathrm{int}}$ scales quite strongly with the system size for local update algorithms, some care is necessary in choosing the run time N. Otherwise the FSS of the specific heat and thus the determination of the *static* critical exponent α/ν could be completely spoiled by the temporal correlations!

As a side remark we note that even in the completely uncorrelated case the estimator (5.69) is biased, $\langle \hat{\sigma}^2_{e_i} \rangle = \sigma^2_{e_i}\,(1 - 1/N)$, since with our conventions in this case $\tau_{e,\mathrm{int}} = 1/2$ (some authors use a different convention in which τ more intuitively vanishes in the uncorrelated case; but this has certain disadvantages in other formulas). In this case one can (and usually does) define a bias-corrected estimator,

$$\hat{\sigma}^2_{e_i,\mathrm{corr}} = \frac{N}{N-1}\hat{\sigma}^2_{e_i} = \frac{1}{N-1}\sum_{i=1}^{N}(e_i - \overline{e})^2 \;, \tag{5.72}$$

which obviously satisfies $\langle \hat{\sigma}^2_{e_i,\mathrm{corr}} \rangle = \sigma^2_{e_i}$. For the squared error on the mean value, this leads to the error formula $\epsilon^2_{\overline{e}} = \hat{\sigma}^2_{\overline{e},\mathrm{corr}} = \hat{\sigma}^2_{e_i,\mathrm{corr}}/N = \frac{1}{N(N-1)}\sum_{i=1}^{N}(e_i - \overline{e})^2$, i.e., to the celebrated replacement of one of the $1/N$-factors by $1/(N-1)$ "due to one missing degree of freedom". Note that in the case of correlated data, a similar construction is at best approximately possible since the bias in (5.71) depends on the a priori unknown autocorrelation time $\tau_{e,\mathrm{int}}$.

5.5.5 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 $\hat{A}(k)$ for the autocorrelation function is obtained by replacing in (5.62) 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 $\hat{A}(k)$ 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 (5.60), it is useful to record the "running" autocorrelation time estimator

$$\hat{\tau}_{\mathcal{O},\mathrm{int}}(k_{\max}) = \frac{1}{2} + \sum_{k=1}^{k_{\max}} \hat{A}(k) \;, \tag{5.73}$$

which approaches $\tau_{\mathcal{O},\mathrm{int}}$ in the limit of large $k_{\max}$ where, however, its statistical error increases rapidly. 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 $k_{\max} \geq 6\hat{\tau}_{\mathcal{O},\mathrm{int}}(k_{\max})$, where $A(k) \approx e^{-6} \approx 10^{-3}$. In this case an a priori error estimate is available [79, 82, 83],

$$\epsilon_{\tau_{\mathcal{O},\mathrm{int}}} = \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{2(2k_{\max}+1)}{N}} \approx \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{12}{N_{\mathrm{eff}}}} \;. \tag{5.74}$$

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. As an order of magnitude estimate consider the 2D Ising model with $L = 100$ simulated with a local update algorithm. The integrated autocorrelation time for this example is of the order of $L^2 \approx 100^2$ (ignoring an priori unknown prefactor of "order unity" which depends on the considered quantity), thus 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 seconds ≈ 1 CPU-day to achieve this accuracy.

Another possibility is to approximate the tail end of $A(k)$ by a single exponential as in (5.63). Summing up the small k part exactly, one finds [84]

$$\tau_{\mathcal{O},\mathrm{int}}(k_{\max}) = \tau_{\mathcal{O},\mathrm{int}} - ce^{-k_{\max}/\tau_{\mathcal{O},\mathrm{exp}}} \;, \tag{5.75}$$

where c is a constant. The latter expression may be used for a numerical estimate of both the exponential and integrated autocorrelation times [84].

5.5.6 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[5] $N = N_B k$), one forms a new, shorter time series of block averages,

$$\mathcal{O}_j^{(B)} \equiv \frac{1}{k} \sum_{i=1}^{k} \mathcal{O}_{(j-1)k+i} \;, \qquad j = 1, \ldots, N_B \;, \tag{5.76}$$

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

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 unbiased estimator (5.72), leading to the squared statistical error of the mean value,

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \sigma^2_B/N_B = \frac{1}{N_B(N_B-1)} \sum_{j=1}^{N_B} (\mathcal{O}_j^{(B)} - \overline{\mathcal{O}^{(B)}})^2 \ . \tag{5.77}$$

By comparing with (5.60) we see that $\sigma^2_B/N_B = 2\tau_{\mathcal{O},\mathrm{int}}\sigma^2_{\mathcal{O}_i}/N$. Recalling the definition of the block length $k = N/N_B$, this shows that one may also use

$$2\tau_{\mathcal{O},\mathrm{int}} = k\sigma^2_B/\sigma^2_{\mathcal{O}_i} \tag{5.78}$$

for the estimation of $\tau_{\mathcal{O},\mathrm{int}}$. Estimates of $\tau_{\mathcal{O},\mathrm{int}}$ obtained in this way are often referred to as "blocking τ" or "binning τ".

5.5.7 Jackknife Analysis

But 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 [85,86] 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 j'th block of the previous binning method,

$$\mathcal{O}_j^{(J)} = \frac{N\overline{\mathcal{O}} - k\mathcal{O}_j^{(B)}}{N-k} \ , \qquad j = 1, \ldots, N_B \ . \tag{5.79}$$

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 Jackknife block variance σ^2_J 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^2_J with a factor $(N_B-1)^2$, leading to

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{N_B-1}{N_B} \sum_{j=1}^{N_B} (\mathcal{O}_j^{(J)} - \overline{\mathcal{O}^{(J)}})^2 \ . \tag{5.80}$$

To summarize this section, any realization of a Markov chain, i.e., MC update algorithm, is characterized by autocorrelation times which enter directly in the statistical errors of MC estimates. Since temporal correlations always increase the statistical errors, it is a very important issue to develop MC update algorithms that keep autocorrelation times as small as possible. This is the reason why cluster and other non-local algorithms are so important.

5.5.8 A Simplified Model: The Bivariate Gaussian Time Series

A useful "gauge model" for all the statistical analysis tools discussed so far is the bivariate Gaussian time series which allows for fairly simple exact solutions. Once the numerical routines reproduce the exact answers for this artificial time series, it is almost certain that they also work properly for "true" time series generated by a MC simulation. The bivariate Gaussian time series is generated by the recursion

$$\begin{aligned} e_0 &= e'_0 \ , \\ e_i &= \rho e_{i-1} + \sqrt{1-\rho^2}\, e'_i \ , \qquad i \geq 1 \ , \end{aligned} \tag{5.81}$$

where $0 \leq \rho < 1$ and the e'_i are *independent*, identically distributed (often abbreviated as "i.i.d.") Gaussian random variables satisfying $\langle e'_i \rangle = 0$ and $\langle e'_i e'_j \rangle = \delta_{ij}$. By iterating the recursion (5.81) it is then easy to see that $\langle e_i \rangle = 0$, $\langle e_i^2 \rangle = 1$ and

$$e_k = \rho e_{k-1} + \sqrt{1-\rho^2}\, e'_k = \rho^k e_0 + \sqrt{1-\rho^2} \sum_{l=1}^{k} \rho^{k-l} e'_l \ , \tag{5.82}$$

so that

$$A(k) = \langle e_0 e_k \rangle = \rho^k \equiv e^{-k/\tau_{\exp}} \ . \tag{5.83}$$

In this simplified model the autocorrelation function is thus a pure exponential with an exponential autocorrelation time given by

$$\tau_{\exp} = -1/\ln \rho \ . \tag{5.84}$$

It should be stressed that in realistic situations a purely exponential decay can only be expected asymptotically for large k where the slowest mode dominates. For smaller time separations usually also many other modes contribute whose autocorrelation time is smaller.

The visual appearance of uncorrelated and correlated data with $\tau_{\exp} = 10$ and 50 is depicted in Figs. 5.7(a)–(c) where in each case one percent of the total "MC time" evolution consisting of 100 000 consecutive "measurements" according to the rule (5.81) is shown. Despite the quite distinct temporal evolutions, histogramming the time series leads to the same Gaussian distribution within error bars, as it should, cf. Fig. 5.7(d). The corresponding autocorrelation functions $A(k)$ are shown in Fig. 5.8(a).

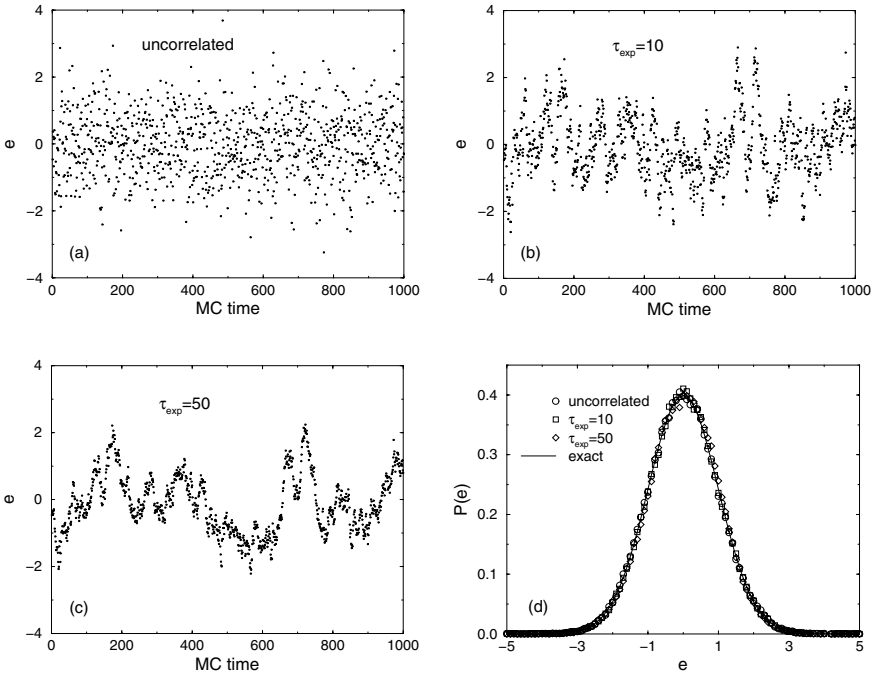

Fig. 5.7. "MC time" evolution according to the bivariate Gaussian process (5.81) (only the first percent shown) in (**a**) the uncorrelated case, (**b**) with $\tau_{\rm exp} = 10$, and (**c**) with $\tau_{\rm exp} = 50$. All three time evolutions with a total of 100 000 consecutive "measurements" lead to the same Gaussian histogram shown in (**d**)

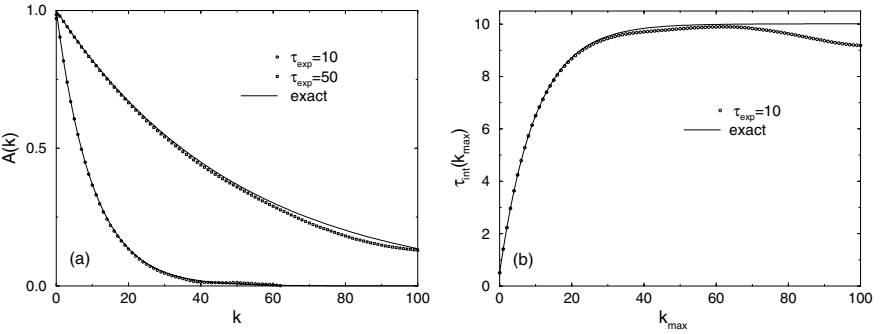

Fig. 5.8. (**a**) Autocorrelation functions and (**b**) integrated autocorrelation time for $\tau_{\rm exp} = 10$ on the basis of 100 000 "measurements" in comparison with exact results for the bivariate Gaussian model shown as the *solid lines*

The integrated autocorrelation time can also be calculated exactly,

$$\tau_{\rm int} = \frac{1}{2} + \sum_{k=1}^{\infty} A(k) = \frac{1}{2}\frac{1+\rho}{1-\rho} = \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp}) \tag{5.85}$$

$$= \tau_{\rm exp}\left[1 + \frac{1}{12\tau_{\rm exp}^2} + \mathcal{O}(1/\tau_{\rm exp}^4)\right] . \tag{5.86}$$

This shows that for a purely exponential autocorrelation function to a very good approximation, $\tau_{\rm int} \approx \tau_{\rm exp}$, which would immediately follow from $\tau_{\rm int} \approx \int_0^\infty dk A(k) = \tau_{\rm exp}$.

As explained in the last section, one usually truncates the summation in (5.85) self-consistently at about $k_{\rm max} = 6\tau_{\rm int}$ ($\approx 6\tau_{\rm exp}$) since $A(k)$ becomes very noisy for large time separations. Observing that (5.85) is nothing but a geometric series, also the resulting correction can be calculated exactly,

$$\tau_{\rm int}(k_{\rm max}) \equiv \frac{1}{2} + \sum_{k=1}^{k_{\rm max}} A(k) = \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp})\left[1 - \frac{2e^{-(k_{\rm max}+1)/\tau_{\rm exp}}}{1+e^{-1/\tau_{\rm exp}}}\right] \tag{5.87}$$

$$= \tau_{\rm int}\left\{1 - [1 - \tanh(1/2\tau_{\rm exp})]e^{-k_{\rm max}/\tau_{\rm exp}}\right\} \tag{5.88}$$

$$\approx \tau_{\rm int}\left[1 - \left(1 - \frac{1}{2\tau_{\rm exp}}\right)e^{-k_{\rm max}/\tau_{\rm exp}}\right] \qquad (\tau_{\rm exp} \gg 1) , \tag{5.89}$$

showing that with increasing $k_{\rm max}$ the asymptotic value of $\tau_{\rm int} \equiv \tau_{\rm int}(\infty)$ is approached exponentially fast. This is illustrated in Fig. 5.8(b) for the bivariate Gaussian time series with $\tau_{\rm exp} = 10$. Here we also see that for too large $k_{\rm max}$ the estimate for $\tau_{\rm int}(k_{\rm max})$ can deviate quite substantially from the exact value due to its divergent variance. The usually employed self-consistent cutoff would be around $6\tau_{\rm exp} = 60$ where $\tau_{\rm int}(k_{\rm max}) \approx 9.89$.

Let us now turn to the binning analysis by decomposing as in (5.76) the total number of measurements N into N_B non-overlapping blocks of length k ($N = N_B k$). In our simple example, the expected value of the block averages is, of course, zero, $\langle e_{B,n}\rangle = \frac{1}{k}\sum_{i=1}^{k}\langle e_{(n-1)k+i}\rangle = 0$. The variance of the block variables is hence just the expectation value of $e_{B,n}^2$,

$$\sigma_B^2 = \langle e_{B,n}^2\rangle = \frac{1}{k^2}\sum_{i,j=1}^{k}\rho^{|i-j|} = \frac{1}{k^2}\left[k + 2\sum_{i=1}^{k}\sum_{j=1}^{i-1}\rho^{i-j}\right]$$

$$= \frac{1}{k}\left[1 + \frac{2\rho}{1-\rho} - \frac{2\rho}{k}\frac{1-\rho^k}{(1-\rho)^2}\right] . \tag{5.90}$$

Recalling (5.85) this can be rewritten as

$$k\sigma_B^2 = 2\tau_{\rm int}\left[1 - \frac{\tau_{\rm int}}{k}\left(1 - e^{-k/\tau_{\rm exp}}\right)/\cosh^2(1/2\tau_{\rm exp})\right] \tag{5.91}$$

$$\approx 2\tau_{\rm int}\left[1 - \frac{\tau_{\rm exp}}{k}\left(1 - e^{-k/\tau_{\rm exp}}\right)\right] \qquad (\tau_{\rm exp} \gg 1) , \tag{5.92}$$

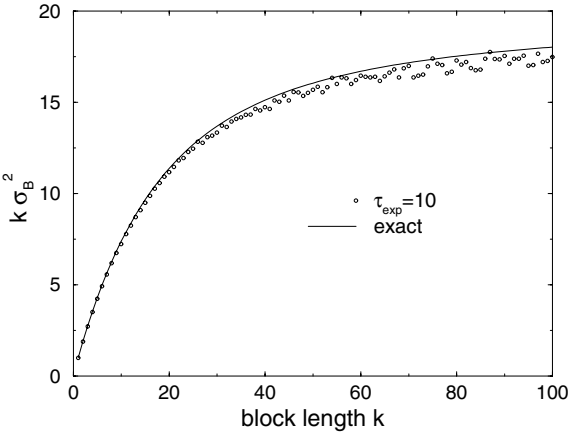

Fig. 5.9. Binning analysis of 100 000 "measurements" in the bivariate Gaussian model with $\tau_{\rm exp} = 10$. The *solid line* shows the exact result

showing that with increasing block length k the asymptotic value $2\tau_{\rm int}$ is approached according to a power law. For an illustration see Fig. 5.9.

5.5.9 Applications to the 2D Ising Model

In this section the autocorrelation and error analysis is illustrated for the 2D Ising model which albeit still very simple exhibits already some effects also seen in more complicated systems. The simulations are done with the Metropolis update algorithm for a 16×16 square lattice with periodic boundary conditions at the infinite-volume critical point $\beta_c = \ln(1+\sqrt{2})/2 \approx 0.440\,686\,793\,4\ldots$. The spins were updated in sequential order by proposing always a spin flip[6] and accepting or rejecting this proposal according to (5.33). The raw data of the simulation are collected in a time-series file, storing 1000000 measurements of the energy and magnetization taken after each sweep over the lattice, after discarding (quite generously) the first 200000 sweeps to equilibrate the system.

The last 500 sweeps of the time evolution of the energy are shown in Fig. 5.10(a), which should be compared with the Gaussian model time series in Figs. 5.7(b) and (c). Using the complete time series the autocorrelation functions were computed according to (5.62). The only difference to the analysis of the simplified model is that instead of using the Gaussian data one now reads in the Ising model time series – the analysis program is exactly the

[6] If the spins are updated in sequential order, but a spin flip is proposed with only 50% probability, the temporal correlations are much larger ($\tau_{e,\rm int} \approx 27$) [87]. This quite unusual update procedure was (inadvertently) chosen in Ref. [87], because always proposing a spin flip with sequential update order does *not* work properly for the 1D model with its only two nearest neighbours.

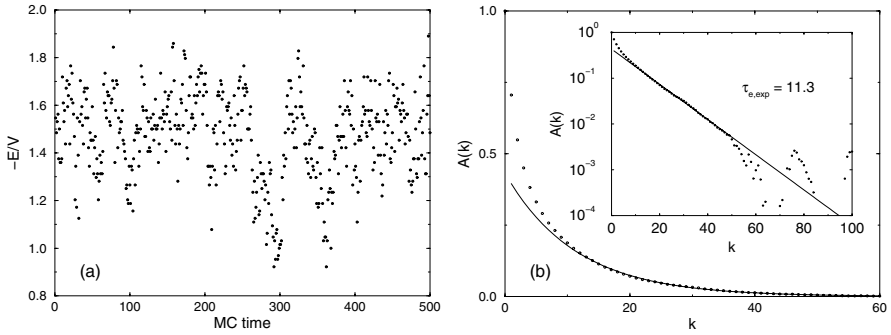

Fig. 5.10. (**a**) Part of the time evolution of the energy $e = E/V$ for the 2D Ising model on a 16×16 lattice at β_c and (**b**) the resulting autocorrelation function. The inset shows the same data on a logarithmic scale, revealing the fast initial drop for very small k and the noisy behaviour for large k. The *solid lines* show a fit to the ansatz $A(k) = a \exp(-k/\tau_{e,\mathrm{exp}})$ in the range $10 \leq k \leq 40$ with $\tau_{e,\mathrm{exp}} = 11.3$ and $a = 0.432$

same. The result for the energy autocorrelations is shown in Fig. 5.10(b). On the linear-log scale of the inset we clearly see the asymptotic linear behaviour of $\ln A(k)$. A linear fit of the form (5.63), $A(k) = a \exp(-k/\tau_{e,\mathrm{exp}})$, in the range $10 \leq k \leq 40$ yields an estimate for the *exponential* autocorrelation time of $\tau_{e,\mathrm{exp}} \approx 11.3$. Apart from the noise for large k, which is also present in the simplified model for finite statistics, the main difference to the artificial data of the simplified model lies in the small k behaviour. For the Ising model we clearly notice an initial fast drop, corresponding to faster relaxing modes, before the asymptotic behaviour sets in. This is, in fact, the generic behaviour of autocorrelation functions in realistic models.

Once the autocorrelation function is known, it is straightforward to sum up the integrated autocorrelation time. The result for the energy is depicted in Fig. 5.11(a), yielding an estimate of $\tau_{e,\mathrm{int}} \approx 5.93$. The binning analysis shown in Fig. 5.11(b) gives a consistent result as it should. Note that due to the initial fast drop of $A(k)$ the exponential autocorrelation time $\tau_{e,\mathrm{exp}} \approx 11.3$ is much larger than the integrated autocorrelation time $\tau_{e,\mathrm{int}} \approx 5.93$, which is in accord with the general inequality [80] quoted above.

5.6 Cluster Algorithms

The main drawback of local update algorithms is their pronounced critical slowing down at a continuous phase transition where temporal correlations diverge (thermodynamic limit) or become very large (finite-size scaling region): $\tau \propto \xi^z$ or $\propto L^z$ with $z \approx 2$. Since excitations on all length scales become important at T_c, it is intuitively clear that some sort of non-local updates should alleviate this problem. While it was clear since long that clusters or

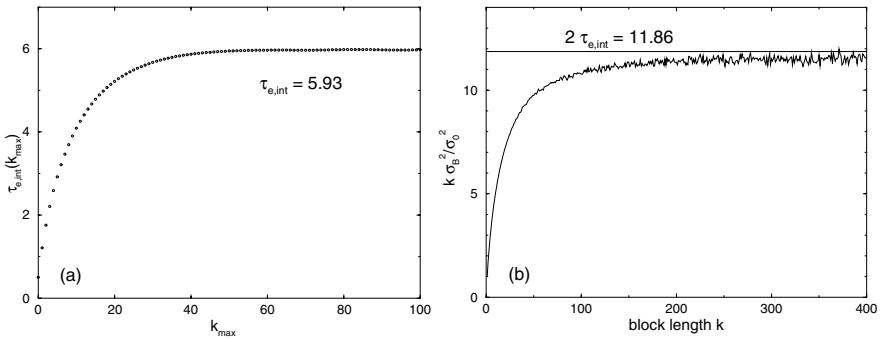

Fig. 5.11. (**a**) Integrated autocorrelation time approaching $\tau_{e,\mathrm{int}} \approx 5.93$ for large upper cutoff $k_{\max}$ and (**b**) binning analysis for the energy of the 2D Ising model on a 16×16 lattice at β_c. The *horizontal line* in (b) shows $2\tau_{e,\mathrm{int}}$ with $\tau_{e,\mathrm{int}}$ read off from (a)

droplets should play a central role in such an update, it took until 1987 before Swendsen and Wang [88] proposed a legitimate cluster update procedure for Potts models. Soon after Wolff [89] discovered the so-called single-cluster variant and developed a generalization to O(n)-symmetric spin models. By now cluster updates have been derived for many other models as well [90], but they are still less general applicable than local update algorithms of the Metropolis type. We therefore start again with 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 [91–94],

$$Z = \sum_{\{\sigma_i\}} \exp\left(\beta \sum_{\langle ij\rangle} \sigma_i\sigma_j\right) \tag{5.93}$$

$$= \sum_{\{\sigma_i\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p) + p\delta_{\sigma_i\sigma_j}\right] \tag{5.94}$$

$$= \sum_{\{\sigma_i\}} \sum_{\{n_{ij}\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p)\delta_{n_{ij},0} + p\delta_{\sigma_i\sigma_j}\delta_{n_{ij},1}\right] , \tag{5.95}$$

with

$$p = 1 - e^{-2\beta} . \tag{5.96}$$

Here the n_{ij} are bond variables which can take the values $n_{ij} = 0$ or 1, interpreted as "deleted" or "active" bonds. In the first line of this derivation we used the trivial fact 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 second line we made use of the "deep" identity $a + b = \sum_{n=0}^{1}(a\delta_{n,0} + b\delta_{n,1})$.

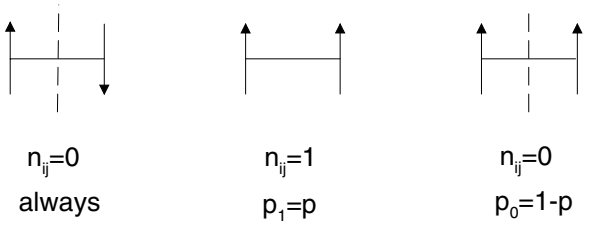

Fig. 5.12. Illustration of the bond variable update. The bond between unlike spins is always "deleted" as indicated by the *dashed line*. A bond between like spins is only "active" with probability $p = 1 - \exp(-2\beta)$. Only at zero temperature ($\beta \longrightarrow \infty$) stochastic and geometrical clusters coincide

Swendsen-Wang Cluster

According to (5.95) a cluster update sweep then consists of alternating updates of the bond variables n_{ij} for given spins with updates of the spins σ_i for a given bond configuration. In practice one proceeds as follows:

1. 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. 5.12.
2. Identify clusters of spins that are connected by "active" bonds ($n_{ij} = 1$).
3. 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 on the average smaller than geometrical clusters. Only at zero temperature ($\beta \longrightarrow \infty$) 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 [88]. 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.

Wolff Cluster

Shortly after the original discovery of cluster algorithms, Wolff [89] 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

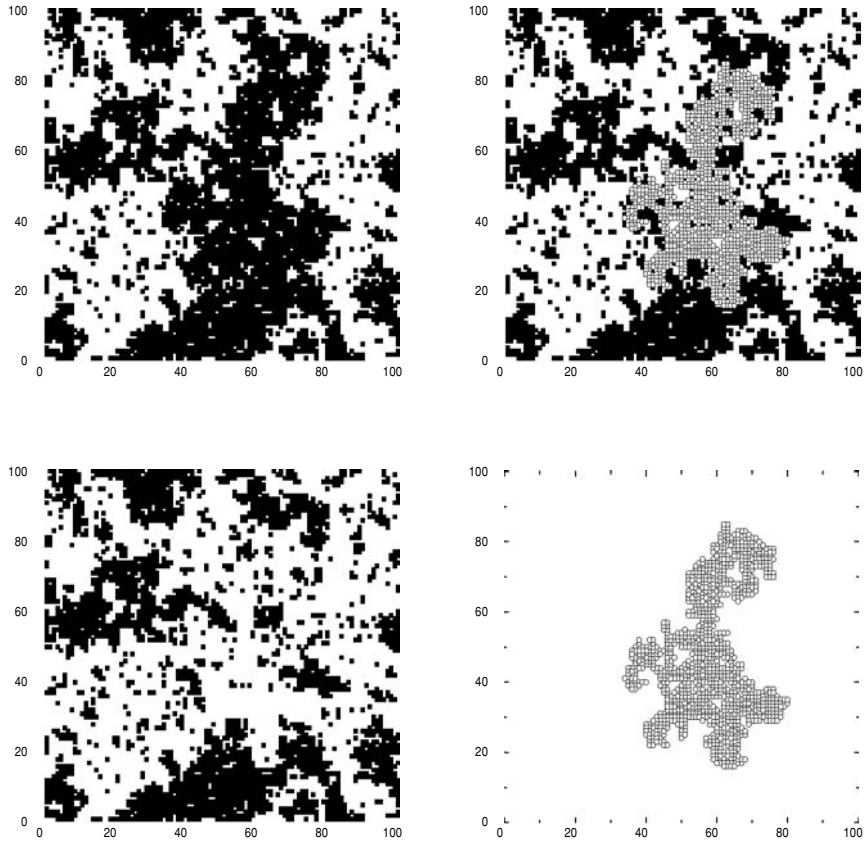

Fig. 5.13. Illustration of the Wolff cluster update, using actual simulation results for the 2D Ising model at $0.97 \times \beta_c$ on a 100×100 lattice. *Upper left*: Initial configuration. *Upper right*: The stochastic cluster is marked. *Lower left*: Final configuration after flipping the spins in the cluster. *Lower right*: The flipped cluster

one chooses a lattice site at random, constructs only the cluster connected with this site, and then flips all spins of this cluster. A typical example is shown in Fig. 5.13. In principle, one could also here choose for the new spin value $+1$ or -1 at random, but then nothing at all would be changed if one hits the current value of the spins.

Here a sweep consists of $V/\langle|C|\rangle$ single cluster steps, where $\langle|C|\rangle$ denotes the average cluster size. With this definition autocorrelation times are directly comparable with results from the Metropolis or Swendsen-Wang algorithm. Apart from being somewhat easier to program, Wolff's single-cluster variant is usually 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.

Figure 5.13 also nicely illustrates the difference between *geometrical* and *stochastic FK* clusters as already pointed out in Sect. 5.2 in connection with Fig. 5.2. In the upper right configuration plot one clearly sees that the stochastic cluster is much smaller than the underlying black geometrical one. It is worth to emphasize again that only the stochastic FK clusters encode in their fractal and percolation properties the *critical* behaviour of the thermodynamic system. In the 3D Ising model, geometrical clusters do not even percolate at the proper critical temperature (but already at a about 2% smaller temperature). This is one of the reasons why early attempts to construct "cluster updates" (working with geometrical clusters) were not successful. In 2D, also geometrical clusters *do* percolate at T_c, but they are still not useful for algorithmic purposes because their fractal properties are not directly related to the *critical* behaviour of the thermodynamic system at hand. Rather they encode the properties of a *tricritical* point in a related model (the diluted $q = 1$ Potts model) [27]. For instance, while at the end of this section it will be shown that the average FK cluster size $\langle |C| \rangle_{\mathrm{FK}}$ is a so-called improved estimator for the Ising susceptibility, $\langle |C| \rangle_{\mathrm{FK}} = \chi'/\beta \propto L^{\gamma/\nu}$, and hence scales in 2D with the proper Ising model exponents $\gamma = 7/4 = 1.75$ and $\nu = 1$, one finds for the average geometrical cluster size $\langle |C| \rangle_{\mathrm{geo}} \propto L^{\gamma_{\mathrm{geo}}/\nu_{\mathrm{geo}}}$ with the exact exponent ratio $\gamma_{\mathrm{geo}}/\nu_{\mathrm{geo}} = 91/48 = 1.8958\ldots$. Note that both cluster quantities can be measured in the same MC simulation run [27].

Performance for the Ising Model

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 [95]). Some results on z are collected in Table 5.2, which allow us to conclude:

(1) Compared to local algorithms with $z \approx 2$, z is dramatically reduced for both cluster variants in 2D and 3D.
(2) In 2D, Swendsen-Wang and Wolff cluster updates are equally efficient, while in 3D, the Wolff update is clearly favourable.
(3) 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 [96] for the Swendsen-Wang cluster algorithm of $\tau_{\mathrm{SW}} \geq C$ ($= C_0 + A \ln L$ for the 2D Ising model), implying $z_{\mathrm{SW}} \geq \alpha/\nu$ ($= 0$ for the 2D Ising model).
(4) Different observables (e.g., energy E and magnetization M) may yield quite different values for z when defined via the scaling behaviour of the integrated autocorrelation time.

Table 5.2. 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)	–	$z_{M,\mathrm{exp}}$	Nightingale and Blöte [58]
	–	2.032(4)	z_{dam}	Grassberger [97]
	–	2.055(10)	$z_{M,\mathrm{exp}}$	Ito *et al.* [98]
Swendsen-Wang cluster	0.35(1)	0.75(1)	$z_{E,\mathrm{exp}}$	Swendsen and Wang [88]
	0.27(2)	0.50(3)	$z_{E,\mathrm{int}}$	Wolff [95]
	0.20(2)	0.50(3)	$z_{\chi,\mathrm{int}}$	Wolff [95]
	$0(\log L)$	–	$z_{M,\mathrm{exp}}$	Heermann and Burkitt [99]
	0.25(5)	–	$z_{M,\mathrm{rel}}$	Tamayo [100]
Wolff cluster	0.26(2)	0.28(2)	$z_{E,\mathrm{int}}$	Wolff [95]
	0.13(2)	0.14(2)	$z_{\chi,\mathrm{int}}$	Wolff [95]
	0.25(5)	0.3(1)	$z_{E,\mathrm{rel}}$	Ito and Kohring [101]

Embedded Clusters

While it is quite easy to generalize the derivation (5.93)–(5.96) 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

$$H = -J \sum_{\langle ij \rangle} \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j \; ; \quad \boldsymbol{\sigma}_i = (\boldsymbol{\sigma}_{i,1}, \boldsymbol{\sigma}_{i,2}, \ldots, \boldsymbol{\sigma}_{i,n}) \; ; |\boldsymbol{\sigma}_i| = 1 \tag{5.97}$$

one needs a new strategy for $n \geq 2$ [89, 102–104] (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 $\boldsymbol{r}$,

$$\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i^{\parallel} + \boldsymbol{\sigma}_i^{\perp} \; ; \quad \boldsymbol{\sigma}_i^{\parallel} = \epsilon \, |\boldsymbol{\sigma}_i \cdot \boldsymbol{r}| \, \boldsymbol{r} ; \quad \epsilon = \mathrm{sign}(\boldsymbol{\sigma}_i \cdot \boldsymbol{r}) \; . \tag{5.98}$$

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

$$H = -\sum_{\langle ij \rangle} J_{ij} \epsilon_i \epsilon_j + \mathrm{const} \; , \tag{5.99}$$

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.

Improved Estimators

A further advantage of cluster algorithms is that they lead quite naturally to so-called improved estimators which are designed to further reduce the

statistical errors. Suppose we want to measure the expectation value $\langle \mathcal{O} \rangle$ of an observable $\mathcal{O}$. Then any estimator $\hat{\mathcal{O}}$ satisfying $\langle \hat{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle$ is permissible. This does not determine $\hat{\mathcal{O}}$ uniquely since there are infinitely many other possible choices, $\hat{\mathcal{O}}' = \hat{\mathcal{O}} + \hat{\mathcal{X}}$, where the added estimator $\hat{\mathcal{X}}$ is assumed to have zero expectation, $\langle \hat{\mathcal{X}} \rangle = 0$. The variances of the estimators $\hat{\mathcal{O}}'$, however, can be quite different and are not necessarily related to any physical quantity (contrary to the standard mean-value estimator of the energy whose variance is proportional to the specific heat). It is exactly this freedom in the choice of $\hat{\mathcal{O}}$ 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 [104]

$$\hat{G}(\boldsymbol{x}_i - \boldsymbol{x}_j) = n \frac{V}{|C|} \boldsymbol{r} \cdot \boldsymbol{\sigma}_i \; \boldsymbol{r} \cdot \boldsymbol{\sigma}_j \; \Theta_C(\boldsymbol{x}_i) \Theta_C(\boldsymbol{x}_j) \; , \tag{5.100}$$

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 0 otherwise). For the Fourier transform, $\tilde{G}(\boldsymbol{k}) = \sum_{\boldsymbol{x}} G(\boldsymbol{x}) \exp(-i \boldsymbol{k} \cdot \boldsymbol{x})$, this implies the improved estimator

$$\hat{\tilde{G}}(\boldsymbol{k}) = \frac{n}{|C|} \left[\left(\sum_{i \in C} \boldsymbol{r} \cdot \boldsymbol{\sigma}_i \cos \boldsymbol{k} \boldsymbol{x}_i \right)^2 + \left(\sum_{i \in C} \boldsymbol{r} \cdot \boldsymbol{\sigma}_i \sin \boldsymbol{k} \boldsymbol{x}_i \right)^2 \right] \; , \tag{5.101}$$

which, for $\boldsymbol{k} = \boldsymbol{0}$, reduces to an improved estimator for the susceptibility χ' in the high-temperature phase,

$$\hat{\tilde{G}}(\boldsymbol{0}) = \hat{\chi}'/\beta = \frac{n}{|C|} \left(\sum_{i \in C} \boldsymbol{r} \cdot \boldsymbol{\sigma}_i \right)^2 \; . \tag{5.102}$$

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 and Heisenberg model one finds empirically that in two as well as in three dimensions $\langle |C| \rangle \approx 0.81 \chi'/\beta$ for $n = 2$ ([102, 108]) and $\langle |C| \rangle \approx 0.75 \chi'/\beta$ for $n = 3$ ([104, 109]), respectively.

It should be noted that by means of the estimators (5.100)–(5.102) a significant reduction of variance should only be expected outside the FSS region where the average cluster size is small compared to the volume of the system.

5.7 Reweighting Techniques

Even though the physics underlying reweighting techniques [110, 111] is extremely simple and the basic idea has been known since long (see the list of references in Ref. [111]), their power in practice has been realized only

relatively late in 1988. The important observation by Ferrenberg and Swendsen [110,111] was that the best performance is achieved *near* criticality where histograms are usually broad. In this sense reweighting techniques are complementary to improved estimators.

5.7.1 Single-Histogram Technique

The single-histogram reweighting technique [110] 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]

$$Z(\beta_0) = \sum_{\{s\}} e^{-\beta_0 H(\{s\})} = \sum_E \Omega(E) e^{-\beta_0 E} \propto \sum_E P_{\beta_0}(E) \ , \tag{5.103}$$

where we have introduced the unnormalized energy histogram (density)

$$P_{\beta_0}(E) \propto \Omega(E) e^{-\beta_0 E} \ . \tag{5.104}$$

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

$$P_\beta(E) \propto \Omega(E) e^{-\beta E} = \Omega(E) e^{-\beta_0 E} e^{-(\beta-\beta_0)E} \propto P_{\beta_0}(E) e^{-(\beta-\beta_0)E} \ , \tag{5.105}$$

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,

$$\langle f(E)\rangle(\beta) = \sum_E f(E) P_\beta(E) / \sum_E P_\beta(E) \ , \tag{5.106}$$

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 MC 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 5000 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. 5.14(a) and compared with the curve obtained from the exact Kaufman solution [33,34] for finite $L_x \times L_y$ lattices. This clearly

[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.

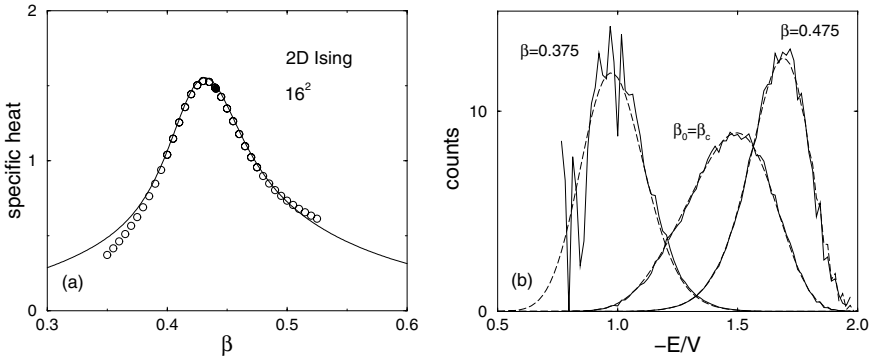

Fig. 5.14. (**a**) The specific heat of the 2D Ising model on a 16×16 square lattice computed by reweighting from a single MC simulation at $\beta_0 = \beta_c$, marked by the filled data symbol. The continuous line shows for comparison the exact solution of Kaufman [33,34]. (**b**) The corresponding 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 [112] expression

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 MC 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. 5.14(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. Here the quality of the histograms can be judged by comparing with the curves obtained from Beale's [112] exact expression for $\Omega(E)$.

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 histogram should not exceed the energy value at which the input histogram had decreased 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 standard maximization routines to the continuous function $C(\beta)$. 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 determine 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),

$$|\langle e\rangle(T) - \langle e\rangle(T_0)| \leq \Delta e(T_0) \ , \tag{5.107}$$

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 $C(T_0)|T - T_0| \leq T_0\sqrt{C(T_0)/V}$ or

$$\frac{|T - T_0|}{T_0} \leq \frac{1}{\sqrt{V}} \frac{1}{C(T_0)} \ . \tag{5.108}$$

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. 5.14 with $V = 16 \times 16$, $\beta_0 = \beta_c \approx 0.44$ and $C(T_0) \approx 1.5$, this estimate yields $|\beta - \beta_0|/\beta_0 \approx |T - T_0|/T_0 \leq 0.04$, i.e., $|\beta - \beta_0| \leq 0.02$ or $0.42 \leq \beta \leq 0.46$. 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 behaviour with system size $V = L^D$, we can go one step further by considering three generic cases:

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

$$\frac{|T - T_0|}{T_0} \propto V^{-1/2} = L^{-D/2} \ . \tag{5.109}$$

ii) *Critical*, 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 correlation length, respectively. For $\alpha > 0$, the leading scaling behaviour 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

$$\frac{|T - T_0|}{T_0} \propto L^{-1/\nu} \ , \tag{5.110}$$

i.e., the typical scaling behaviour of pseudo-transition temperatures in the finite-size scaling regime of a second-order phase transition [113]. For $\alpha < 0$, $C(T_0)$ approaches asymptotically a constant and the leading scaling behaviour of the reweighting range is as in the off-critical case.

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

$$\frac{|T - T_0|}{T_0} \propto V^{-1} = L^{-D} \ , \tag{5.111}$$

which is again the typical finite-size scaling behaviour of pseudo-transition temperatures close to a first-order phase transition [16].

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

$$\begin{aligned}\langle g(M) \rangle &= \sum_{\{s\}} g(M(\{s\})) e^{-\beta_0 H} / Z(\beta_0) = \sum_{E,M} \Omega(E, M) g(M) e^{-\beta_0 E} / Z(\beta_0) \\ &= \sum_E \frac{\sum_M \Omega(E, M) g(M)}{\sum_M \Omega(E, M)} \sum_M \Omega(E, M) e^{-\beta_0 E} / Z(\beta_0) \ . \end{aligned} \tag{5.112}$$

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

$$\langle\langle g(M) \rangle\rangle(E) \equiv \frac{\sum_M \Omega(E, M) g(M)}{\sum_M \Omega(E, M)} \ , \tag{5.113}$$

we arrive at

$$\langle g(M) \rangle = \sum_E \langle\langle g(M) \rangle\rangle(E) P_{\beta_0}(E) \ . \tag{5.114}$$

Identifying $\langle\langle g(M) \rangle\rangle(E)$ with $f(E)$ in Eq. (5.106), the actual reweighting procedure is precisely as before. 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 example for computing $\langle\langle |M| \rangle\rangle(E)$ and $\langle\langle M^2 \rangle\rangle(E)$ using the data of Fig. 5.14 is shown in Fig. 5.15.

An alternative and more flexible method is based on time series. Suppose we have performed a MC simulation at β_0 and stored the time series of N 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

$$\langle f(E, M) \rangle = \sum_{i=1}^{N} f(E_i, M_i) e^{-(\beta - \beta_0) E_i} / \sum_{i=1}^{N} e^{-(\beta - \beta_0) E_i} \ , \tag{5.115}$$

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

$$\langle f(E, M) \rangle = \langle f(E, M) e^{-(\beta - \beta_0) E} \rangle_0 / \langle e^{-(\beta - \beta_0) E} \rangle_0 \ , \tag{5.116}$$

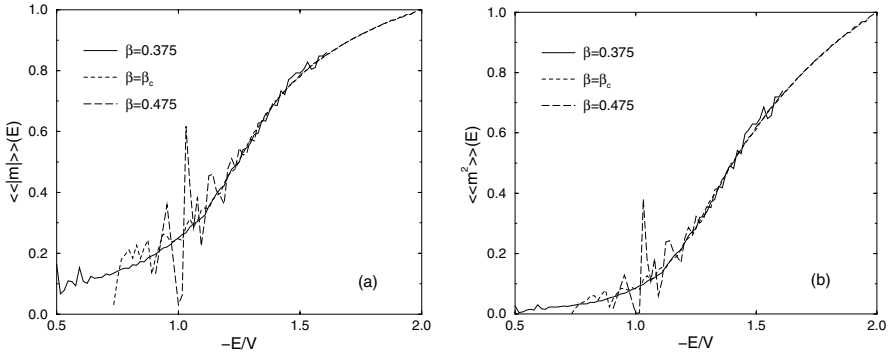

Fig. 5.15. Microcanonical expectation values for (**a**) the absolute magnetization and (**b**) the magnetization squared obtained from the 2D Ising model simulations shown in Fig. 5.14

where the subscript 0 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.

5.7.2 Multi-Histogram Technique

The basic idea of the multi-histogram technique [114] can be summarized as follows:

i) Perform m MC 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

$$\Omega(E) = \frac{\sum_{i=1}^{m} P_{\beta_i}(E)}{\sum_{i=1}^{m} N_i Z_i^{-1} e^{-\beta_i E}} \ , \tag{5.117}$$

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

$$Z_i = \sum_E \Omega(E) e^{-\beta_i E} = \sum_E e^{-\beta_i E} \frac{\sum_{k=1}^{m} P_{\beta_k}(E)}{\sum_{k=1}^{m} N_k Z_k^{-1} e^{-\beta_k E}} \ , \tag{5.118}$$

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 $Z_2/Z_1 = \exp[-(\hat{F}_2 - \hat{F}_1)]$, where $\hat{F}_i = \beta_i F(\beta_i)$. 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 (5.118). 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 neighbouring β-values must have sufficient overlap, i.e., the spacings of the simulation points must be chosen according to the estimates (5.109)–(5.111).

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 Subsect. 5.5.3. 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 [109]

$$e(\beta) = \left(\frac{e_1(\beta)}{(\Delta e_1)^2} + \frac{e_2(\beta)}{(\Delta e_2)^2} + \cdots + \frac{e_m(\beta)}{(\Delta e_m)^2} \right) (\Delta e)^2 \ , \tag{5.119}$$

with

$$\frac{1}{(\Delta e)^2} = \frac{1}{(\Delta e_1)^2} + \frac{1}{(\Delta e_2)^2} + \cdots + \frac{1}{(\Delta e_m)^2} \ . \tag{5.120}$$

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

5.8 Tempering Methods

Loosely speaking, tempering methods 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 $\beta_1 < \beta_2 < \cdots < \beta_m$ which here, however, are combined already during the simulation in a specific, dynamical way.

5.8.1 Simulated Tempering

In simulated tempering simulations [115,116] one starts from a joint partition function (expanded ensemble)

$$Z_{\rm ST} = \sum_{i=1}^{m} e^{g_i} \sum_{\{s\}} e^{-\beta_i H(\{s\})} \ , \tag{5.121}$$

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 takes place with probability $\min\left[1, \exp[-(\beta_j - \beta_i)H(\{s\})] + g_j - g_i\right]$. 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 neighbouring 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 (5.109)–(5.111) for the reweighting range,

$$\delta\beta \propto \begin{cases} L^{-D/2} & \text{off-critical} \ , \\ L^{-1/\nu} & \text{critical} \ , \\ L^{-D} & \text{first-order} \ . \end{cases} \tag{5.122}$$

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

5.8.2 Parallel Tempering

In parallel tempering (exchange Monte Carlo, multiple Markov chain Monte Carlo) simulations [117, 118] the starting point is the product of partition functions (extended ensemble),

$$Z_{\rm PT} = \prod_{i=1}^{m} Z(\beta_i) = \prod_{i=1}^{m} \sum_{\{s\}_i} e^{-\beta_i H(\{s\}_i)} \ , \tag{5.123}$$

and all m systems at different simulation points $\beta_1 < \beta_2 < \cdots < \beta_m$ 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 $\{s\}_i$ and $\{s\}_j$ are attempted (equivalently, the β_i may be exchanged, as is done in most implementations). Adapting the Metropolis criterion (5.34) to the present situation, the proposed exchange will be accepted with probability $W = \min(1, e^{\Delta})$, where $\Delta = (\beta_j - \beta_i)[E(\{s\}_j) - E(\{s\}_i)]$. To assure a reasonable acceptance rate, usually only "nearest-neighbour" exchanges ($j = i \pm 1$) are attempted and the β_i should again be spaced with the $\delta\beta$ given in (5.122). 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 rapidly decorrelates. Conceptually this approach follows again the "avoiding rare events" strategy.

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

5.9 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 [4,119–122]. Similar 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 [123,124] and explains the name "*multi*canonical". At the same time, this method may also be viewed as a specific realization of non-Boltzmann sampling [125] which has been known since long to be a legitimate alternative to the more standard MC approaches [126]. The practical significance of non-Boltzmann sampling was first realized in the so-called "umbrella sampling" method [127], but it took many years before the introduction of the multicanonical ensemble [123,124] 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 [16,122]. 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 discussed in Sect. 5.8.

In both multicanonical versions, the canonical Boltzmann distribution

$$\mathcal{P}_{\rm can}(\phi) \propto \exp(-\beta H(\phi)) \tag{5.124}$$

is replaced by an auxiliary distribution

$$\mathcal{P}_{\rm muca}(\phi) \propto W(\{Q_i\}) \exp(-\beta H(\phi)) \equiv \exp(-\beta H(\phi) - f(\{Q_i(\phi)\})) \ , \tag{5.125}$$

where ϕ denotes generically the degrees of freedom and Q_i stands for a macroscopic observable such as the energy or magnetization. With a suitably chosen reweighting factor $W(\{Q_i\})$, the probability distribution $P_{\rm muca}(\{Q_i\})$ of the macroscopic variables $\{Q_i\}$ can be tuned to take any desired form. Canonical expectation values can always be recovered exactly by inverse reweighting,

$$\langle \mathcal{O} \rangle_{\rm can} = \langle \mathcal{O} W^{-1}(\{Q_i\}) \rangle_{\rm muca} / \langle W^{-1}(\{Q_i\}) \rangle_{\rm muca} \ , \tag{5.126}$$

similar to Eq. (5.116).

The Monte Carlo sampling of $\mathcal{P}_{\rm muca}(\phi)$ proceeds in the usual way by comparing $\beta H(\phi) + f(\{Q_i(\phi)\})$ before and after a proposed update move of ϕ. In most applications local update algorithms have been employed, but for certain classes of models also non-local multigrid methods are applicable [84,128]. 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 [129–131] has been developed.

The performance of the simulation depends, however, in the first place on the choice of $\{Q_i\}$ and the reweighting factor $W(\{Q_i\})$, since for instance in the special case $W \equiv 1$ the troublesome canonical ensemble is recovered. The proper identification of the relevant set of Q_i's requires considerable physical intuition and insight into the specific system under study. While for disordered complex systems this may be a serious problem, in studies of first-order phase transitions the proper choice is clear since typically the energy E (temperature-driven transition) or magnetization M (field-driven transition) are the relevant variables. In both cases, the reweighting factor is usually chosen such that the multicanonical probability density $P_{\rm muca} = W P_{\rm can}$ is approximately flat between the two peaks of the canonical distribution. The most important technical point is the procedure for constructing the multicanonical weights, for which iterative procedures have been developed [4,119–122].

If $P_{\rm muca}$ was completely flat and the MC 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 energy or magnetization. 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 behaviour $\tau \propto V^{\alpha}$, with $\alpha \geq 1$. While for the temperature-driven transitions of 2D Potts models the multibondic variant seems to saturate the bound [129–131], 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 cannot really be excluded. In fact, at least for the field-driven first-order transition of the 2D Ising model, it has been demonstrated recently [132, 133] that even for a perfectly flat multicanonical distribution a "hidden" free energy nucleation barrier leads to an exponential growth of τ, which is, however, much weaker than in the corresponding canonical simulation.

5.10 Concluding Remarks

The intention of these lecture notes was to give an elementary introduction to the basic concepts of modern Monte Carlo simulations and to illustrate their usefulness by applications to the very simple Ising lattice spin model. The basic Monte Carlo methods based on local update rules are straightforward to apply to all models with discrete degrees of freedom and with some extra care also to continuous variables and off-lattice models. Some generalizations of cluster update methods have already been indicated. Also other models may be efficiently simulated by this non-local method, but there is no guarantee that for a given model a cluster update procedure can be developed. The statistical error analysis part is obviously completely general, and also reweighting, tempering and multicanonical methods can be adapted to almost every problem at hand.

Acknowledgements

Many people have influenced these lecture notes with their advice, questions, discussions and active contributions. In particular I wish to thank Michael Bachmann, Bertrand Berche, Pierre-Emmanuel Berche, Bernd A. Berg, Alain Billoire, Kurt Binder, Elmar Bittner, Christophe Chatelain, Malte Henkel, Desmond A. Johnston, Ralph Kenna, David P. Landau, Eric Lorenz, Thomas Neuhaus, Andreas Nußbaumer, Michel Pleimling, Adriaan Schakel, and Martin Weigel for sharing their insight and knowledge with me.

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grant Nos. JA 483/22-1 and JA 483/23-1, and the EU RTN-Network "ENRAGE": *Random Geometry and Random Matrices: From Quantum Gravity to Econophysics* under grant No. MRTN-CT-2004-005616.

References

1. M.E.J. Newman, G.T. Barkema: *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999)
2. D.P. Landau, K. Binder: *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000)
3. K. Binder, D.W. Heermann: *Monte Carlo Simulations in Statistical Physics: An Introduction*, 4th edition (Springer, Berlin, 2002)
4. B.A. Berg: *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004)
5. W. Lenz: Phys. Z. **21**, 613 (1920); E. Ising: Z. Phys. **31**, 253 (1925)
6. L. Onsager: Phys. Rev. **65**, 117 (1944)
7. B.M. McCoy, T.T. Wu: *The Two-Dimensional Ising Model* (Harvard University Press, Cambridge, 1973)
8. R.J. Baxter: *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982)
9. L. Onsager: Nuovo Cimento (Suppl.) **6**, 261 (1949); see also the historical remarks in Refs. [7, 8]
10. C.N. Yang: Phys. Rev. **85**, 808 (1952)
11. C.H. Chang: Phys. Rev. **88**, 1422 (1952)
12. W.P. Orrick, B.G. Nickel, A.J. Guttmann, J.H. Perk: Phys. Rev. Lett. **86**, 4120 (2001); J. Stat. Phys. **102**, 795 (2001)
13. R.B. Potts: Proc. Camb. Phil. Soc. **48**, 106 (1952)
14. F.Y. Wu: Rev. Mod. Phys. **54**, 235 (1982)
15. F.Y. Wu: Rev. Mod. Phys. **55**, 315(E) (1983)
16. W. Janke: *First-Order Phase Transitions*, in: *Computer Simulations of Surfaces and Interfaces*, NATO Science Series, II. Mathematics, Physics and Chemistry – Vol. **114**, Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, 9–20 September 2002, eds. B. Dünweg, D.P. Landau, A.I. Milchev (Kluwer, Dordrecht, 2003), pp. 111–135
17. H.E. Stanley: *Introduction to Phase Transitions and Critical Phenomena* (Oxford Press, Oxford, 1979)
18. J.J. Binney, N.J. Dowrick, A.J. Fisher, M.E.J. Newman: *The Theory of Critical Phenomena* (Oxford University Press, Oxford, 1992)
19. D.A. Lavis, G.M. Bell: *Statistical Mechanics of Lattice Systems 2* (Springer, Berlin, 1999)
20. See the volumes of review articles edited by C. Domb, J.L. Lebowitz (eds.): *Phase Transitions and Critical Phenomena* (Academic Press, New York)
21. J.D. Gunton, M.S. Miguel, P.S. Sahni: in: *Phase Transitions and Critical Phenomena*, Vol. **8**, eds. C. Domb, J.L. Lebowitz (Academic Press, New York, 1983)
22. K. Binder: Rep. Prog. Phys. **50**, 783 (1987)
23. H.J. Herrmann, W. Janke, F. Karsch (eds.): *Dynamics of First Order Phase Transitions* (World Scientific, Singapore, 1992)
24. W. Janke: *Recent Developments in Monte Carlo Simulations of First-Order Phase Transitions*, in: *Computer Simulations in Condensed Matter Physics VII*, eds. D.P. Landau, K.K. Mon, H.-B. Schüttler (Springer, Berlin, 1994), p. 29
25. H. Kleinert: *Gauge Fields in Condensed Matter*, Vol. II (World Scientific, Singapore, 1989)

26. W. Janke, H. Kleinert: Phys. Rev. **B33**, 6346 (1986)
27. W. Janke, A.M.J. Schakel: Nucl. Phys. **B700**, 385 (2004); Comp. Phys. Comm. **169**, 222 (2005); Phys. Rev. **E71**, 036703 (2005); Phys. Rev. Lett. **95**, 135702 (2005); and e-print cond-mat/0508734. See also the extensive list of references to earlier work given therein
28. M. Weigel, W. Janke: Phys. Rev. **B62**, 6343 (2000)
29. K. Binder: in: *Monte Carlo Methods in Statistical Physics*, ed. K. Binder (Springer, Berlin, 1979)
30. M.E. Barber: in: *Phase Transitions and Critical Phenomena*, Vol. **8**, eds. C. Domb, J.L. Lebowitz (Academic Press, New York, 1983), p. 146
31. V. Privman (ed.): *Finite-Size Scaling and Numerical Simulations of Statistical Systems* (World Scientific, Singapore, 1990)
32. K. Binder: in: *Computational Methods in Field Theory*, Schladming Lecture Notes, eds. H. Gausterer, C.B. Lang (Springer, Berlin, 1992), p. 59
33. B. Kaufman: Phys. Rev. **76**, 1232 (1949)
34. A.E. Ferdinand, M.E. Fisher: Phys. Rev. **185**, 832 (1969)
35. M.E. Fisher, A.N. Berker: Phys. Rev. **B26**, 2507 (1982)
36. V. Privman, M.E. Fisher: J. Stat. Phys. **33**, 385 (1983)
37. K. Binder, D.P. Landau: Phys. Rev. **B30**, 1477 (1984)
38. M.S.S. Challa, D.P. Landau, K. Binder: Phys. Rev. **B34**, 1841 (1986)
39. V. Privman, J. Rudnik: J. Stat. Phys. **60**, 551 (1990)
40. C. Borgs, R. Kotecky: J. Stat. Phys. **61**, 79 (1990)
41. J. Lee, J.M. Kosterlitz: Phys. Rev. Lett. **65**, 137 (1990)
42. C. Borgs, R. Kotecky, S. Miracle-Solé: J. Stat. Phys. **62**, 529 (1991)
43. C. Borgs, W. Janke: Phys. Rev. Lett. **68**, 1738 (1992)
44. W. Janke: Phys. Rev. **B47**, 14757 (1993)
45. J.M. Hammersley, D.C. Handscomb: *Monte Carlo Methods* (London, 1965)
46. D.W. Heermann: *Computer Simulation Methods in Theoretical Physics*, 2nd ed., (Springer, Berlin, 1990)
47. K. Binder (ed.): *The Monte Carlo Method in Condensed Matter Physics* (Springer, Berlin, 1992)
48. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller: J. Chem. Phys. **21**, 1087 (1953)
49. S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi: Science **220**, 671 (1983)
50. W. Janke: *Pseudo Random Numbers: Generation and Quality Checks*, invited lecture notes, in: Proceedings of the Euro Winter School *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, eds. J. Grotendorst, D. Marx, A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol. **10**, pp. 447–458 (2002), and references therein
51. M. Creutz: Phys. Rev. **D36**, 515 (1987)
52. S.L. Adler: Phys. Rev. **D37**, 458 (1988)
53. H. Neuberger: Phys. Lett. **B207**, 461 (1988)
54. R. Gupta, J. DeLapp, G.G. Battrouni, G.C. Fox, C.F. Baillie, J. Apostolakis: Phys. Rev. Lett. **61**, 1996 (1988)
55. R.J. Glauber: J. Math. Phys. **4**, 294 (1963)
56. A.D. Rutenberg, A.J. Bray: Phys. Rev. **E51**, 5499 (1995)
57. P. Calabrese, A. Gambassi: J. Phys. **A38**, R133–R193 (2005)
58. M.P. Nightingale, H.W.J. Blöte: Phys. Rev. Lett. **76**, 4548 (1996); Phys. Rev. **B62**, 1089 (2000)

59. A. Barrat: Phys. Rev. **E57**, 3629 (1998)
60. C. Godrèche, J.-M. Luck: J. Phys.: Condens. Matter **14**, 1589 (2002)
61. 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)
62. F. Corberi, E. Lippiello, M. Zannetti: Phys. Rev. Lett. **90**, 099601 (2003) [Comment]; M. Henkel, M. Pleimling: Phys. Rev. Lett. **90**, 099602 (2003) [Reply]
63. L. Berthier, J.L. Barrat, J. Kurchan: Eur. Phys. J. **B11**, 635 (1999)
64. F. Corberi, E. Lippiello, M. Zannetti: Eur. Phys. J. **B24**, 359 (2001); Phys. Rev. **E65**, 046136 (2003)
65. M. Henkel: *Conformal Invariance and Critical Phenomena* (Springer, Berlin, 1999)
66. M. Henkel, M. Pleimling, C. Godrèche, J.-M. Luck: Phys. Rev. Lett. **87**, 265701 (2001)
67. M. Henkel: Nucl. Phys. **B641**, 405 (2002)
68. W. Zippold, R. Kühn, H. Horner: Eur. Phys. J. **B13**, 531 (2000)
69. M. Henkel, M. Paessens, M. Pleimling: Europhys. Lett. **62**, 664 (2003)
70. M. Henkel, M. Pleimling: Phys. Rev. **E68**, 065101 (R) (2003)
71. M. Henkel, A. Picone, M. Pleimling: Europhys. Lett. **68**, 191 (2004)
72. F. Bagnoli: J. Stat. Phys. **85**, 151 (1996)
73. T. Vojta: Phys. Rev. **E55**, 5157 (1997)
74. H. Hinrichsen, E. Domany: Phys. Rev. **E56**, 94 (1997)
75. E. Lorenz: Diploma thesis, Universität Leipzig (2005) [`www.physik.uni-leipzig.de/~lorenz/diplom.pdf`]
76. E. Lorenz, W. Janke: *Numerical Tests of Local Scale Invariance in Ageing q-State Potts Models*, Leipzig preprint (2006), EPL **77**, 10003 (2007)
77. M.B. Priestley: *Spectral Analysis and Time Series*, 2 vols. (Academic, London, 1981), Chaps. 5–7
78. T.W. Anderson: *The Statistical Analysis of Time Series* (Wiley, New York, 1971)
79. N. Madras, A.D. Sokal: J. Stat. Phys. **50**, 109 (1988)
80. A.D. Sokal, L.E. Thomas: J. Stat. Phys. **54**, 797 (1989)
81. A.M. Ferrenberg, D.P. Landau, K. Binder: J. Stat. Phys. **63**, 867 (1991)
82. A.D. Sokal: *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne (1989)
83. A.D. Sokal: *Bosonic Algorithms*, in: *Quantum Fields on the Computer*, ed. M. Creutz (World Scientific, Singapore, 1992), p. 211
84. W. Janke, T. Sauer: J. Stat. Phys. **78**, 759 (1995)
85. B. Efron: *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics [SIAM], Philadelphia, 1982)
86. R.G. Miller: Biometrika **61**, 1 (1974)
87. W. Janke: *Statistical Analysis of Simulations: Data Correlations and Error Estimation*, invited lecture notes, in: Proceedings of the Euro Winter School *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, eds. J. Grotendorst, D. Marx, A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol. **10**, pp. 423–445 (2002)
88. R.H. Swendsen, J.-S. Wang: Phys. Rev. Lett. **58**, 86 (1987)
89. U. Wolff: Phys. Rev. Lett. **62**, 361 (1989)

90. W. Janke: *Nonlocal Monte Carlo Algorithms for Statistical Physics Applications*, Mathematics and Computers in Simulations **47**, 329–346 (1998)
91. P.W. Kasteleyn, C.M. Fortuin: J. Phys. Soc. Japan **26** (Suppl.), 11 (1969)
92. C.M. Fortuin, P.W. Kasteleyn: Physica **57**, 536 (1972)
93. C.M. Fortuin: Physica **58**, 393 (1972)
94. C.M. Fortuin: Physica **59**, 545 (1972)
95. U. Wolff: Phys. Lett. **A228**, 379 (1989)
96. X.-L. Li, A.D. Sokal: Phys. Rev. Lett. **63**, 827 (1989); *ibid.* **67**, 1482 (1991)
97. P. Grassberger: Physica **A214**, 547 (1995); **A217**, 227 (1995) (erratum)
98. N. Ito, K. Hukushima, K. Ogawa, Y. Ozeki: J. Phys. Soc. Japan **69**, 1931 (2000)
99. D.W. Heermann, A.N. Burkitt: Physica **A162**, 210 (1990)
100. P. Tamayo: Physica **A201**, 543 (1993)
101. N. Ito, G.A. Kohring: Physica **A201**, 547 (1993)
102. U. Wolff: Nucl. Phys. **B322**, 759 (1989)
103. M. Hasenbusch: Nucl. Phys. **B333**, 581 (1990)
104. U. Wolff: Nucl. Phys. **B334**, 581 (1990)
105. C.F. Baillie: Int. J. Mod. Phys. **C1**, 91 (1990)
106. R.H. Swendsen, J.-S. Wang, A.M. Ferrenberg: in: *The Monte Carlo Method in Condensed Matter Physics*, ed. K. Binder (Springer, Berlin, 1992)
107. M. Hasenbusch, S. Meyer: Phys. Lett. **B241**, 238 (1990)
108. W. Janke: Phys. Lett. **A148**, 306 (1990)
109. C. Holm, W. Janke: Phys. Rev. **B48**, 936 (1993)
110. A.M. Ferrenberg, R.H. Swendsen: Phys. Rev. Lett. **61**, 2635 (1988)
111. A.M. Ferrenberg, R.H. Swendsen: Phys. Rev. Lett. **63**, 1658(E) (1989)
112. P.D. Beale: Phys. Rev. Lett. **76**, 78 (1996)
113. N. Wilding: *Computer Simulation of Continuous Phase Transitions*, in: *Computer Simulations of Surfaces and Interfaces*, NATO Science Series, II. Mathematics, Physics and Chemistry – Vol. **114**, Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, 9–20 September 2002, eds. B. Dünweg, D.P. Landau, A.I. Milchev (Kluwer, Dordrecht, 2003), pp. 161–171
114. A.M. Ferrenberg, R.H. Swendsen: Phys. Rev. Lett. **63**, 1195 (1989)
115. E. Marinari, G. Parisi: Europhys. Lett. **19**, 451 (1992)
116. A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov: J. Chem. Phys. **96**, 1776 (1992)
117. C.J. Geyer: in: *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, ed. E.M. Keramidas (Interface Foundation, Fairfax, Virginia, 1991); pp. 156–163; C.J. Geyer, E.A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995)
118. K. Hukushima, K. Nemoto: J. Phys. Soc. Japan **65**, 1604 (1996)
119. B.A. Berg: Fields Inst. Comm. **26**, 1 (2000)
120. B.A. Berg: Comp. Phys. Comm. **104**, 52 (2002)
121. W. Janke: Physica **A254**, 164 (1998)
122. W. Janke: *Histograms and All That*, invited lectures, in: *Computer Simulations of Surfaces and Interfaces*, NATO Science Series, II. Mathematics, Physics and Chemistry – Vol. **114**, Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, 9–20 September 2002, eds. B. Dünweg, D.P. Landau, A.I. Milchev (Kluwer, Dordrecht, 2003), pp. 137–157
123. B.A. Berg, T. Neuhaus: Phys. Lett. **B267**, 249 (1991)

124. B.A. Berg, T. Neuhaus: Phys. Rev. Lett. **68**, 9 (1992)
125. W. Janke: Int. J. Mod. Phys. **C3**, 1137 (1992)
126. K. Binder: in *Phase Transitions and Critical Phenomena*, Vol. **5b**, eds. C. Domb, M.S. Green (Academic Press, New York, 1976), p. 1
127. G.M. Torrie, J.P. Valleau: Chem. Phys. Lett. **28**, 578 (1974); J. Comp. Phys. **23**, 187 (1977) 187; J. Chem. Phys. **66**, 1402 (1977)
128. W. Janke, T. Sauer: Phys. Rev. **E49**, 3475 (1994)
129. W. Janke, S. Kappler: Nucl. Phys. **B** (Proc. Suppl.) **42**, 876 (1995)
130. W. Janke, S. Kappler: Phys. Rev. Lett. **74**, 212 (1995)
131. M.S. Carroll, W. Janke, S. Kappler: J. Stat. Phys. **90**, 1277 (1998)
132. T. Neuhaus, J.S. Hager: J. Stat. Phys. **113** 47 (2003)
133. K. Leung, R.K.P. Zia: J. Phys. **A23**, 4593 (1990)