Chapter 3

Monte Carlo Simulations in Statistical Physics – From Basic Principles to Advanced Applications

Wolfhard Janke

Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ), Universität Leipzig, Postfach 100 920, 04009 Leipzig, Germany
wolfhard.janke@itp.uni-leipzig.de

This chapter starts with an overview of Monte Carlo computer simulation methodologies which are illustrated for the simple case of the Ising model. After reviewing importance sampling schemes based on Markov chains and standard local update rules (Metropolis, Glauber, heat-bath), nonlocal cluster-update algorithms are explained which drastically reduce the problem of critical slowing down at second-order phase transitions and thus improve the performance of simulations. How this can be quantified is explained in the section on statistical error analyses of simulation data including the effect of temporal correlations and autocorrelation times. Histogram reweighting methods are explained in the next section. Eventually, more advanced generalized ensemble methods (simulated and parallel tempering, multicanonical ensemble, Wang-Landau method) are discussed which are particularly important for simulations of first-order phase transitions and, in general, of systems with rare-event states. The setup of scaling and finite-size scaling analyses is the content of the following section. The chapter concludes with two advanced applications to complex physical systems. The first example deals with a quenched, diluted ferromagnet, and in the second application we consider the adsorption properties of macromolecules such as polymers and proteins to solid substrates. Such systems often require especially tailored algorithms for their efficient and successful simulation.

Contents

1. Introduction

Classical statistical physics is conceptually a well understood subject which poses, however, many difficult problems when specific properties of interacting systems are considered. In almost all non-trivial applications, analytical methods can only provide approximate answers. Experiments, on the other hand, are often plagued by side effects which are difficult to control. Numerical computer simulations are, therefore, an important third complementary method of modern physics. The relationship between theory, experiment, and computer simulation is sketched in Fig. 1. On the one hand a computer simulation allows one to assess the range of validity of approximate analytical work for generic models and on the other hand it can bridge the gap to experiments for real systems with typically fairly complicated interactions. Computer simulations are thus helpful on our way to a deeper understanding of complex physical systems such as disordered magnets and (spin) glasses or of biologically motivated problems such as protein folding and adsorption of macromolecules to solid substrates, to mention only a few. Quantum statistical problems in condensed matter or the broad field of elementary particle physics and quantum gravity are other major applications which, after suitable mappings, basically rely on the same simulation techniques.

This chapter provides an overview of computer simulations employing Monte Carlo methods based on Markov chain importance sampling. Most methods can be illustrated with the simple Ising spin model. Not all aspects can be discussed

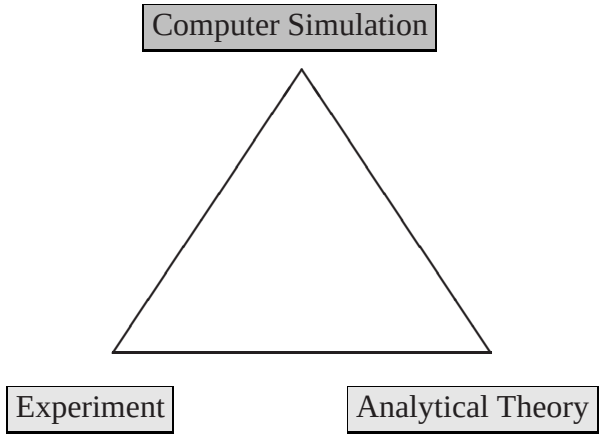

Fig. 1. Sketch of the relationship between theory, experiment and computer simulation.

in detail and for further study the reader is referred to recent textbooks,[1–4] where some of the material presented here is discussed in more depth. The rest of this chapter is organized as follows. In the next Sect. 2, first the definition of the standard Ising model is briefly recalled. Then the basic method underlying all importance sampling Monte Carlo simulations is described and some properties of local update algorithms (Metropolis, Glauber, heat-bath) are discussed. The following subsection is devoted to non-local cluster algorithms which in some cases can dramatically speed up the simulations. A fairly detailed account of statistical error analyses is given in Sect. 3. Here temporal correlation effects and autocorrelation times are discussed, which explain the problems with critical slowing down at a continuous phase transition and exponentially large flipping times at a first-order transition. Reweighting techniques are discussed in Sect. 4 which are particularly important for finite-size scaling studies. More advanced generalized ensemble simulation methods are briefly outlined in Sect. 5, focusing on simulated and parallel tempering, the multicanonical ensemble and the Wang-Landau method. In Sect. 6 suitable observables for scaling analyses (specific heat, magnetization, susceptibility, correlation functions, ...) are briefly discussed. Some characteristic properties of phase transitions, scaling laws, the definition of critical exponents and the method of finite-size scaling are summarized. In order to illustrate how all these techniques can be put to good use, in Sect. 7 two concrete applications are discussed: The phase diagram of a quenched, diluted ferromagnet

and the adsorption properties of polymers to solid substrates. Finally, in Sect. 8 this chapter closes with a few concluding remarks.

2. The Monte Carlo Method

The goal of Monte Carlo simulations is to estimate expectation values

$$\langle \mathcal{O} \rangle \equiv \sum_{\text{states } \sigma} \mathcal{O}(\sigma) e^{-\beta \mathcal{H}(\sigma)} / \mathcal{Z} \ , \tag{1}$$

where $\mathcal{O}$ stands for any quantity of the system defined by its Hamiltonian $\mathcal{H}$ and

$$\mathcal{Z} = e^{-\beta \mathcal{F}} = \sum_{\text{states } \sigma} e^{-\beta \mathcal{H}(\sigma)} = \sum_{E} \Omega(E) e^{-\beta E} \tag{2}$$

is the (canonical) partition function. The first sum runs over all possible microstates of the system and the second sum runs over all energies, where the density of states $\Omega(E)$ counts the number of microstates contributing to a given energy E. The state space may be discrete or continuous (where sums become integrals etc.). As usual $\beta \equiv 1/k_B T$ denotes the inverse temperature fixed by an external heat bath and k_B is Boltzmann's constant.

In the following most simulation methods will be illustrated for the minimalistic Ising model[5] where

$$\mathcal{H}(\sigma) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i \ , \qquad \sigma_i = \pm 1 \ . \tag{3}$$

Here J is a coupling constant which is positive for a ferromagnet ($J > 0$) and negative for an anti-ferromagnet ($J < 0$), h is an external magnetic field, and the symbol $\langle ij \rangle$ indicates that the lattice sum is restricted to all nearest-neighbor pairs of spins living at the lattice sites i. In the examples discussed below, usually D-dimensional simple-cubic lattices with $V = L^D$ spins subject to periodic boundary conditions are considered. From now on we will always assume natural units in which $k_B = 1$ and $J = 1$.

For any realistic number of degrees of freedom, complete enumeration of all microstates contributing to (1) or (2) is impossible. For the Ising model with only two states per site, enumeration still works up to a, say, 6×6 square lattice where $2^{36} \approx 6.9 \times 10^{10}$ microstates contribute. Since this yields the exact expectation value of *any* quantity, enumeration for very small systems is a useful exercise for comparison with the numerical methods discussed here. However, already for a moderate 10^3 lattice, the number of terms would be astronomically large:[a] $2^{1000} \approx 10^{300}$.

[a]For comparison, a standard estimate for the number of protons in the Universe is 10^{80}.

2.1. *Random sampling*

One way out is stochastic sampling of the huge state space. Simple random sampling, however, does not work for statistical systems with many degrees of freedom. Here the problem is that the region of state space that contributes significantly to canonical expectation values at a given temperature $T \ll \infty$ is extremely narrow and hence far too rarely hit by random sampling. In fact, random sampling corresponds to setting $\beta = 1/T = 0$, i.e., exploring mainly the typical microstates at infinite temperature. Of course, the low-energy states in the tails of this distribution contain theoretically (that is, for infinite statistics) all information about the system's properties at finite temperature, too, but this is of very little practical relevance since the probability to hit this tail in random sampling is by far too small. With finite statistics consisting of typically $10^9 - 10^{12}$ randomly drawn microstates, this tail region is virtually not sampled at all.

2.2. *Importance sampling*

The solution to this problem has been known since long as *importance sampling*[6,7] where a Markov chain[8–10] is set up to draw a microstate σ_i not at random but according to the given equilibrium distribution

$$\mathcal{P}_i^{\mathrm{eq}} \equiv \mathcal{P}^{\mathrm{eq}}(\sigma_i) = e^{-\beta\mathcal{H}(\sigma_i)}/\mathcal{Z} \ . \tag{4}$$

For definiteness, on the r.h.s. a canonical ensemble governed by the *Boltzmann weight* $e^{-\beta\mathcal{H}(\sigma_i)}$ was assumed, but this is not essential for most of the following.

A Markov chain is defined by the transition probability $W_{ij} \equiv W(\sigma_i \to \sigma_j)$ for a given microstate σ_i to "evolve" into another microstate σ_j (which may be again σ_i) subject to the condition that this probability *only* depends on the preceding state σ_i but *not* on the history of the whole trajectory in state space, i.e., the stochastic process is almost local in time. Mnemonically this can be depicted as

$$\ldots \xrightarrow{W} \sigma^{(k)} \xrightarrow{W} \sigma^{(k+1)} \xrightarrow{W} \sigma^{(k+2)} \xrightarrow{W} \ldots \ ,$$

where $\sigma^{(k)}$ is the current state of the system after the kth step of the Markov chain. To ensure that, after an initial transient or equilibration period, microstates occur with the given probability (4), the transition probability W_{ij} has to satisfy three

conditions:

$$i)\ \ W_{ij} \geq 0 \qquad \forall\, i,j\ , \tag{5}$$

$$ii)\ \ \sum_j W_{ij} = 1 \qquad \forall\, i\ , \tag{6}$$

$$iii)\ \ \sum_i W_{ij} \mathcal{P}_i^{\rm eq} = \mathcal{P}_j^{\rm eq} \qquad \forall\, j\ . \tag{7}$$

The first two conditions merely formalize that, for any initial state σ_i, W_{ij} should be a properly normalized probability distribution. The equal sign in (5) may occur and, in fact, does so for almost all pairs of microstates i,j in any realistic implementation of the Markov process. To ensure *ergodicity* one additionally has to require that starting from *any* given microstate σ_i *any* other σ_j can be reached in a *finite* number of steps, i.e., an integer $n < \infty$ must exist such that $(W^{n+1})_{ij} = \sum_{k_1,k_2,\ldots,k_n} W_{ik_1} W_{k_1k_2} \ldots W_{k_nj} > 0$. In other words, at least one (finite) path connecting σ_i and σ_j must exist in state space that can be realized with non-zero probability.[b]

The *balance condition* (7) implies that the transition probability W has to be chosen such that the desired equilibrium distribution (4) is a fixed point of W, i.e., an eigenvector of W with unit eigenvalue. The usually employed *detailed balance* is a stronger, sufficient condition:

$$W_{ij}\, \mathcal{P}_i^{\rm eq} = W_{ji}\, \mathcal{P}_j^{\rm eq}\ . \tag{8}$$

By summing over i and using the normalization condition (6), one easily proves the more general balance condition (7).

After an initial equilibration period, expectation values can be estimated as arithmetic mean over the Markov chain,

$$\langle \mathcal{O} \rangle = \sum_\sigma \mathcal{O}(\sigma) \mathcal{P}^{\rm eq}(\sigma) \approx \overline{\mathcal{O}} \equiv \frac{1}{N} \sum_{k=1}^{N} \mathcal{O}(\sigma^{(k)})\ , \tag{9}$$

where $\sigma^{(k)}$ stands for a microstate at "time" k.[c] Since in equilibrium $\langle \mathcal{O}(\sigma^{(k)}) \rangle = \langle \mathcal{O} \rangle$ at any "time" k, one immediately sees that $\langle \overline{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle$, showing that the mean value $\overline{\mathcal{O}}$ is a so-called *unbiased estimator* of the expectation value $\langle \mathcal{O} \rangle$. A more detailed exposition of the mathematical concepts underlying any Markov chain Monte Carlo algorithm can be found in many textbooks and reviews.[1–4,11–13]

[b] In practice, one may nevertheless observe "effective" ergodicity breaking when $(W^{n+1})_{ij}$ is so small that this event will typically not happen in finite simulation time.

[c] In Monte Carlo simulations, "time" refers to the stochastic evolution in state space and is *not* directly related to physical time as for instance in molecular dynamics simulations where the trajectories are determined by Newton's deterministic equation.

2.3. *Local update algorithms*

The Markov chain conditions (5)–(7) are rather general and can be satisfied with many different transition probabilities. A very flexible prescription is the original Metropolis algorithm,[14] which is applicable in practically all cases (lattice/off-lattice, discrete/continuous, short-range/long-range interactions, ...). Here one first proposes with *selection probability*

$$f_{ij} = f(\sigma_i \longrightarrow \sigma_j) \ , \quad f_{ij} \geq 0 \ , \quad \sum_j f_{ij} = 1 \ , \tag{10}$$

a potential update from the current "old" microstate $\sigma_{\rm o} = \sigma_i$ to some microstate σ_j. The proposed microstate σ_j is then accepted as the "new" state $\sigma_{\rm n} = \sigma_j$ with an *acceptance probability*

$$w_{ij} = w(\sigma_i \longrightarrow \sigma_j) = \min\left(1, \frac{f_{ji}}{f_{ij}} \frac{\mathcal{P}_j^{\rm eq}}{\mathcal{P}_i^{\rm eq}}\right) \ , \tag{11}$$

where $\mathcal{P}^{\rm eq}$ is the desired equilibrium distribution specified in (4). Otherwise the system remains in the old microstate, $\sigma_{\rm n} = \sigma_{\rm o}$, which may also trivially happen when $f_{ii} \neq 0$.

Keeping this in mind, one readily sees that the transition probability W_{ij} is given as

$$W_{ij} = \begin{cases} f_{ij} w_{ij} & j \neq i \\ f_{ii} + \sum_{j \neq i} f_{ij}(1 - w_{ij}) & j = i \end{cases} \ . \tag{12}$$

Since $f_{ij} \geq 0$ and $0 \leq w_{ij} \leq 1$, the first Markov condition $W_{ij} \geq 0$ follows immediately. Also the second condition (6) is easy to prove:

$$\begin{aligned} \sum_j W_{ij} &= W_{ii} + \sum_{j \neq i} W_{ij} \\ &= f_{ii} + \sum_{j \neq i} f_{ij}(1 - w_{ij}) + \sum_{j \neq i} f_{ij} w_{ij} = \sum_j f_{ij} = 1 \ . \end{aligned} \tag{13}$$

Finally we show that W_{ij} satisfies the detailed balance condition (8). We first consider the case $f_{ji}\mathcal{P}_j^{\rm eq} > f_{ij}\mathcal{P}_i^{\rm eq}$. Then, from (11), one immediately finds $W_{ij}\mathcal{P}_i^{\rm eq} = f_{ij}\mathcal{P}_i^{\rm eq}$ for the l.h.s. of (8). Since $W_{ji} = f_{ji}\min\left(1, \frac{f_{ij}}{f_{ji}} \frac{\mathcal{P}_i^{\rm eq}}{\mathcal{P}_j^{\rm eq}}\right)$, the r.h.s. of (8) becomes

$$W_{ji}\mathcal{P}_j^{\rm eq} = f_{ji} \frac{f_{ij}}{f_{ji}} \frac{\mathcal{P}_i^{\rm eq}}{\mathcal{P}_j^{\rm eq}} \mathcal{P}_j^{\rm eq} = f_{ij}\mathcal{P}_i^{\rm eq} \ , \tag{14}$$

which completes the proof. For the second case $f_{ji}\mathcal{P}_j^{\rm eq} < f_{ij}\mathcal{P}_i^{\rm eq}$, the proof proceeds precisely along the same lines.

The update prescription (10), (11) is still very general: (a) The selection probability may be asymmetric ($f_{ij} \neq f_{ji}$), (b) it has not yet been specified how to pick the trial state σ_j given $\sigma_{\rm o}$, and (c) $\mathcal{P}^{\rm eq}$ could be "some" arbitrary probability distribution. The last point (c) is obviously trivial, but the resulting formulas simplify when a Boltzmann weight as in (4) is assumed. Then

$$\frac{\mathcal{P}_j^{\rm eq}}{\mathcal{P}_i^{\rm eq}} = e^{-\beta\Delta E} \tag{15}$$

where $\Delta E = E_j - E_i = E_{\rm n} - E_{\rm o}$ is the energy difference between the proposed new and the old microstate. The second point (b), on the other hand, is of great practical relevance since an arbitrary proposal for $\sigma_{\rm n}$ would typically lead to a large ΔE and hence a high rejection rate if $\beta > 0$. One therefore commonly tries to update only one degree of freedom at a time. Then $\sigma_{\rm n}$ differs only *locally* from $\sigma_{\rm o}$. For short-range interactions this automatically has the additional advantage that only the local neighborhood of the selected degree of freedom contributes to ΔE, so that there is no need to compute the total energies in each update step. These two specializations are usually employed, but the selection probabilities may still be chosen asymmetrically. If this is the case, one refers to this update prescription as the *Metropolis-Hastings*[15] update algorithm. For a recent example with asymmetric f_{ij} in the context of polymer simulations see, e.g., Ref. 16.

2.3.1. *Metropolis algorithm*

In generic applications, however, the f_{ij} are symmetric. For instance, if we pick one of the V Ising spins at random and propose to flip it, then $f_{ij} = 1/V$ does not depend on i and j and hence is trivially symmetric. In this case the acceptance probability simplifies to

$$\begin{aligned} w_{ij} &= \min\left(1, \frac{\mathcal{P}_j^{\rm eq}}{\mathcal{P}_i^{\rm eq}}\right) = \min\left(1, e^{-\beta\Delta E}\right) \\ &= \begin{cases} 1 & E_{\rm n} < E_{\rm o} \\ \exp\left[-\beta(E_{\rm n} - E_{\rm o})\right] & E_{\rm n} \geq E_{\rm o} \end{cases} . \end{aligned} \tag{16}$$

This is the standard *Metropolis* update algorithm, which is very easy to implement.

If the proposed update lowers the energy, it is always accepted. On the other hand, when the new microstate has a higher energy, the update has still to be accepted with probability (16) in order to ensure the proper treatment of entropic contributions – in thermal equilibrium, it is the *free* energy $F = U - TS$ which has to be minimized and not the energy. Only in the limit of zero temperature, $\beta \to \infty$, the acceptance probability for new states with higher energy tends to

zero and the Metropolis method degenerates to a minimization algorithm for the energy functional. With some additional refinements, this is the basis for the *simulated annealing* technique,[17] which is often applied to hard optimization and minimization problems.

For the Ising model with only two states per spin, a spin flip is the only admissible local update proposal. Hence in this simple example there is no parameter available by which one could tune the *acceptance ratio*, which is defined as the fraction of trial moves that are accepted. For models with many states per spin (e.g., q-state Potts or Z_n clock models) or in continuous systems (e.g., Heisenberg spin model or off-lattice molecular systems), however, it is in the most cases not recommendable to propose the new state uniformly out of all available possibilities. Rather, one usually restricts the trial states to a neighborhood of the current "old" state. For example, in a continuous atomic system, a trial move may consist of displacing a randomly chosen atom by a random step size up to some maximum $S_{\max}$ in each Cartesian direction. If $S_{\max}$ is small, almost all attempted moves will be accepted and the acceptance ratio is close to unity, but the configuration space is explored slowly. On the other hand, if $S_{\max}$ is large, a successful move would make a large step in configuration space, but many trial moves would be rejected because configurations with low Boltzmann weight are very likely, yielding an acceptance ratio close to zero. As a compromise of these two extreme situations, one often applies the common rule of thumb that $S_{\max}$ is adjusted to achieve an acceptance ratio of 0.5.[18,19]

Empirically this value proves to be a reasonable but at best heuristically justified choice. In principle, one should measure the statistical error bars as a function of $S_{\max}$ for otherwise identical simulation conditions and then choose that $S_{\max}$ which minimises the statistical error. In general the optimal $S_{\max}$ depends on the model at hand and even on the considered observable, so finally some "best average" would have to be used. At any rate, the corresponding acceptance ratio would certainly not coincide with 0.5. Example computations of this type reported values in the range $0.4 - 0.6$ (Refs. 18,20) but for certain models also much smaller (or larger) values may be favourable. Incidentally, there appeared recently a proof in the mathematical literature[21] claiming an optimal acceptance ratio of 0.234 which, however, relies on assumptions[22] not met in a typical statistical physics simulation.[d]

Whether relying on the rule of thumb value 0.5 or trying to optimise $S_{\max}$, this should be done *before* the actual simulation run. Trying to maintain a given acceptance ratio automatically during the run by periodically updating $S_{\max}$ is at

[d]Thanks are due to Yuko Okamoto who pointed to this paper and to Bob Swendsen who immediately commented on it during the CompPhys11 Workshop in November 2011 in Leipzig.

least potentially dangerous.[19] The reason is that the accumulated average of the acceptance ratio and hence the updated $S_{\max}$ are dependent on the recent history of the Monte Carlo trajectory – and *not* only on the current configuration – what violates the Markovian requirement. Consequently the balance condition is no longer fulfilled which may lead to more or less severe systematic deviations (bias). As claimed already a while ago in Ref. 18 and reemphasized recently in Ref. 20, by following a carefully determined schedule for the adjustments of $S_{\max}$, the systematic error may be kept smaller than the statistical error in a controlled way, but to be on the safe side one should be very cautious with this type of refinements.

Finally a few remarks on the practical implementation of the Metropolis method. To decide whether a proposed update should be accepted or not, one draws a uniformly distributed random number $r \in [0,1)$, and if $r \leq w_{ij}$, the new state is accepted. Otherwise one keeps the old configuration and continues with the next spin. In computer simulations, random numbers are generated by means of "pseudo-random number generators" (RNGs), which produce – according to some deterministic rule – (more or less) uniformly distributed numbers whose values are "very hard" to predict.[23] 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 uncover the deterministic rule underlying their generation. The "goodness" of a RNG is thus assessed by the difficulty to derive its underlying deterministic rule. Related requirements are the absence of correlations and a very long period, what can be particularly important in high-statistics simulations. Furthermore, a RNG should be portable among different computer platforms and, very importantly, it should yield reproducible results for testing purposes. The design of RNGs is a science in itself, and many things can go wrong with them.[e] As a recommendation one should better not experiment too much with some fancy RNG picked up somewhere from the WWW, say, but rely on well-documented and well-tested subroutines.

2.3.2. *Glauber algorithm*

As indicated earlier the Markov chain conditions (5)–(7) are rather general and the Metropolis rule (11) or (16) for the acceptance probability w_{ij} is not the only possible choice. For instance, when flipping a spin at site i_0 in the Ising model, w_{ij} can also be taken as[25]

$$w_{ij} = w(\sigma_{i_0} \to -\sigma_{i_0}) = \frac{1}{2}\left[1 - \sigma_{i_0} \tanh\left(\beta S_{i_0}\right)\right] \quad , \tag{17}$$

[e] A prominent example is the failure of the by then very prominent and apparently well-tested R250 generator when applied to the single-cluster algorithm.[24]

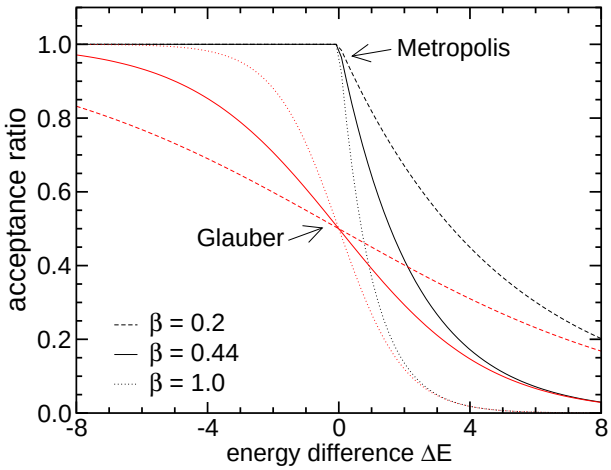

Fig. 2. Comparison of the acceptance ratio for a spin flip in the two-dimensional Ising model with the Glauber (or equivalently heat-bath) and Metropolis update algorithm for three different inverse temperatures β.

where $S_{i_0} = \sum_k \sigma_k + h$ is an effective spin or field collecting all neighboring spins (in their "old" states) interacting with the spin at site i_0 and h is the external magnetic field. This is the *Glauber* update algorithm. Detailed balance is straightforward to prove. Rewriting $\sigma_{i_0} \tanh(\beta S_{i_0}) = \tanh(\beta \sigma_{i_0} S_{i_0})$ (making use of $\sigma_{i_0} = \pm 1$ and the point symmetry of the hyperbolic tangent) and noting that $\Delta E = E_\mathrm{n} - E_\mathrm{o} = 2\sigma_{i_0} S_{i_0}$ (where σ_{i_0} is the "old" spin value and $(-\sigma_{i_0})$ the "new" one), Eq. (17) becomes

$$w(\sigma_{i_0} \to -\sigma_{i_0}) = \frac{1}{2}\left[1 - \tanh(\beta \Delta E/2)\right] = \frac{e^{-\beta \Delta E/2}}{e^{\beta \Delta E/2} + e^{-\beta \Delta E/2}} \ , \qquad (18)$$

showing explicitly that the acceptance probability of the Glauber algorithm also only depends on the total energy change as in the Metropolis case. In this form it is thus possible to generalize the Glauber update rule from the Ising model with only two states per spin to any general model that can be simulated with the Metropolis procedure. The acceptance probability (18) is plotted in Fig. 2 as a function of ΔE for various (inverse) temperatures and compared with the corresponding probability (16) of the Metropolis algorithm. Note that for all values of ΔE and temperature, the Metropolis acceptance probability is higher than that of the Glauber algorithm. As we shall see in the next paragraph, for the Ising model, the Glauber and heat-bath algorithms are identical.

The Glauber update algorithm for the Ising model is also theoretically of

interest since for the one-dimensional case the dynamics of the Markov chain can be calculated analytically. For the relaxation time of the magnetisation one finds the remarkably simple result[25] $m(t) = m(0)\exp(-t/\tau_{\rm relax})$ with $\tau_{\rm relax} = 1/[1-\tanh(2\beta)]$. For two and higher dimensions, however, no exact solutions are known.

2.3.3. *Heat-bath algorithm*

The heat-bath algorithm is different from the two previous update algorithms in that it does not follow the previous scheme "update proposal plus accept/reject step". Rather, the new value of σ_{i_0} at a randomly selected site i_0 is determined by testing all its possible states in the "heat bath" of its (fixed) neighbors (e.g., 4 on a square lattice and 6 on a simple-cubic lattice with nearest-neighbor interactions). For models with a finite number of states per degree of freedom the transition probability reads

$$w(\sigma_{\rm o}\to\sigma_{\rm n}) = \frac{e^{-\beta\mathcal{H}(\sigma_{\rm n})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\sigma_{\rm o})}} = \frac{e^{-\beta\sum_k H_{i_0 k}}}{\sum_{\sigma_{i_0}} e^{-\beta\sum_k H_{i_0 k}}} \quad , \tag{19}$$

where $\sum_k H_{i_0 k}$ collect all terms involving the spin σ_{i_0}. All other contributions to the energy not involving σ_{i_0} cancel due to the ratio in (19), so that for the update at each site i_0 only a small number of computations is necessary (e.g, about 4 for a square and 6 for a simple-cubic lattice of arbitrary size). Detailed balance (8) is obviously satisfied since

$$e^{-\beta\mathcal{H}(\sigma_{\rm o})}\frac{e^{-\beta\mathcal{H}(\sigma_{\rm n})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\sigma_{\rm n})}} = e^{-\beta\mathcal{H}(\sigma_{\rm n})}\frac{e^{-\beta\mathcal{H}(\sigma_{\rm o})}}{\sum_{\sigma_{i_0}} e^{-\beta\mathcal{H}(\sigma_{\rm o})}} \quad . \tag{20}$$

How is the probability (19) realized in practice? Due to the summation over all local states, special tricks are necessary when each degree of freedom can take many different states, and only in special cases the heat-bath method can be efficiently generalized to continuous degrees of freedom. In many applications, however, the admissible local states of σ_{i_0} can be labeled by a small number of integers, say $n = 1,\ldots,N$, which occur with probabilities p_n according to (19). Since this probability distribution is normalized to unity, the sequence $(p_1, p_2, \ldots, p_n, \ldots, p_N)$ decomposes the unit interval into segments of length $\propto p_n$. If one now draws a random number $R \in [0,1)$ and compares the accumulated probabilities $\sum_{k=1}^{n} p_k$ with R, then the new state n is the smallest upper bound that satisfies $\sum_{k=1}^{n} p_k \geq R$. Clearly, for a large number of possible local states, the determination of n can become quite time-consuming (in particular, if many small p_n are at the beginning of the sequence, in which case a clever

permutation of the p_n by relabeling the admissible local states can improve the performance).

In the special case of the Ising model with only two states per spin, $\sigma_i = \pm 1$, (19) simplifies to

$$w(\sigma_{\rm o} \to \sigma_{\rm n}) = \frac{e^{\beta \sigma_{i_0} S_{i_0}}}{e^{\beta S_{i_0}} + e^{-\beta S_{i_0}}} \ , \tag{21}$$

where σ_{i_0} is the *new* spin value and $S_{i_0} = \sum_k \sigma_k + h$ represents the effective spin interacting with σ_{i_0} as defined already below (17). And since $\Delta E = E_{\rm n} - E_{\rm o} = -(\sigma_{i_0} - (-\sigma_{i_0}))S_{i_0} = -2\sigma_{i_0} S_{i_0}$, the probability for a spin flip becomes[26]

$$w(-\sigma_{i_0} \to \sigma_{i_0}) = \frac{e^{-\beta \Delta E/2}}{e^{\beta \Delta E/2} + e^{-\beta \Delta E/2}} \ . \tag{22}$$

This is *identical* to the acceptance probability (18) for a spin flip in the Glauber update algorithm, that is, for the Ising model, the Glauber and heat-bath update rules give precisely the same results.

2.4. *Temporal correlations*

Data generated with a Markov chain method always exhibit temporal correlations which can be estimated from the autocorrelation function

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} \ , \tag{23}$$

where $\mathcal{O}$ denotes any measurable quantity, for example the energy or magnetization (technical issues and the way in which temporal correlations enter statistical error estimates will be discussed in more detail in Sect. 3.1.3). For large time separations k, $A(k)$ decays exponentially ($a = $ const),

$$A(k) \stackrel{k \to \infty}{\longrightarrow} a e^{-k/\tau_{\mathcal{O},\exp}} \ , \tag{24}$$

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

This is illustrated in Fig. 3 for the 2D Ising model on a rather small 16×16 square lattice with periodic boundary conditions at the infinite-volume critical point $\beta_c = \ln(1 + \sqrt{2})/2 = 0.440\,686\,793\ldots$. The spins were updated in sequential order by proposing always to flip a spin and accepting or rejecting this proposal according to (16). The raw data of the simulation are collected in a time-series file, storing $1\,000\,000$ measurements of the energy and magnetization taken after each sweep over the lattice, after discarding (quite generously) the first $200\,000$ sweeps for equilibrating the system from a disordered start configuration.

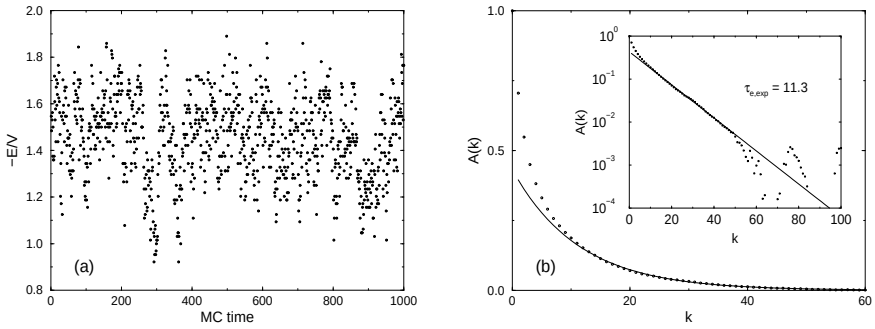

Fig. 3. (a) Part of the time evolution of the energy $e = E/V$ for the 2D Ising model on a 16×16 lattice at $\beta_c = \ln(1+\sqrt{2})/2 = 0.440\,686\,793\ldots$ and (b) the resulting autocorrelation function. In the inset the same data are plotted on a logarithmic scale, revealing a fast initial drop for very small k and noisy behaviour for large k. The solid lines show a fit to the ansatz $A(k) = a\exp(-k/\tau_{e,\exp})$ in the range $10 \le k \le 40$ with $\tau_{e,\exp} = 11.3$ and $a = 0.432$.

The last 1000 sweeps of the time evolution of the energy are shown in Fig. 3(a). Using the complete time series the autocorrelation function was computed according to (23) which is shown in Fig. 3(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 (24), $\ln A(k) = \ln a - k/\tau_{e,\exp}$, in the range $10 \le k \le 40$ yields an estimate for the exponential autocorrelation time of $\tau_{e,\exp} \approx 11.3$. In the small k behaviour of $A(k)$ we observe an initial fast drop, corresponding to faster relaxing modes, before the asymptotic behaviour sets in. This is the generic behaviour of autocorrelation functions in realistic models where the small-k deviations are, in fact, often much more pronounced than for the 2D Ising model.

The influence of autocorrelation times is particular pronounced for phase transitions and critical phenomena.[27–30] For instance, close to a critical point, the autocorrelation time typically scales in the infinite-volume limit as

$$\tau_{\mathcal{O},\exp} \propto \xi^z \ , \tag{25}$$

where $z \ge 0$ is the so-called *dynamical critical exponent*. Since the spatial correlation length $\xi \propto |T - T_c|^{-\nu} \to \infty$ when $T \to T_c$, also the autocorrelation time $\tau_{\mathcal{O},\exp}$ diverges when the critical point is approached, $\tau_{\mathcal{O},\exp} \propto |T - T_c|^{-\nu z}$. This leads to the phenomenon of *critical slowing down* at a continuous phase transition which can be observed experimentally for instance in critical opalescence.[31] The reason is that local spin-flip Monte Carlo dynamics (or diffusion dynamics in a lattice-gas picture) describes at least qualitatively the true physical dynamics of a system in contact with a heat bath. In a finite system, the correlation length ξ is

limited by the linear system size L, so that the characteristic length scale is then L and the scaling law (25) is replaced by

$$\tau_{\mathcal{O},\exp} \propto L^z \quad . \tag{26}$$

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

$$z \approx 2 \quad , \tag{27}$$

which is only weakly dependent on the dimensionality and can be understood by a simple random-walk or diffusion argument in energy space. Non-local update algorithms such as multigrid schemes[32–36] or in particular the cluster methods discussed in the next section can reduce the value of the dynamical critical exponent z significantly, albeit in a strongly model-dependent fashion.

At a first-order phase transition, a completely different mechanism leads to an even more severe "slowing-down" problem.[37,38] Here, the keyword is "phase coexistence". A finite system close to the (pseudo-) transition point can flip between the coexisting pure phases by crossing a two-phase region. Relative to the weight of the pure phases, this region of state space is strongly suppressed by an additional Boltzmann factor $\exp(-2\sigma L^{d-1})$, where σ denotes the interface tension between the coexisting phases, L^{d-1} is the (projected) "area" of the interface and the factor 2 accounts for periodic boundary conditions, which enforce always an even number of interfaces for simple topological reasons. The time spent for crossing this highly suppressed rare-event region scales proportional to the inverse of this interfacial Boltzmann factor, implying that the autocorrelation time increases exponentially with the system size,

$$\tau_{\mathcal{O},\exp} \propto e^{2\sigma L^{d-1}} \quad . \tag{28}$$

In the literature, this behaviour is sometimes termed *supercritical slowing down*, even though, strictly speaking, nothing is "critical" at a first-order phase transition. Since this type of slowing-down problem is directly related to the shape of the probability distribution, it appears for all types of update algorithms, i.e., in contrast to the situation at a second-order transition, here it cannot be cured by employing multigrid or cluster techniques. It can be overcome, however, at least in part by means of multicanonical methods which are briefly discussed at the end of this chapter in Sect. 5.

2.5. *Cluster algorithms*

The critical slowing down at a second-order phase transition reflects that excitations on all length scales become important, leading to diverging spatial cor-

relations. This suggests that some sort of non-local update rules should be able to alleviate this problem. Natural candidates are rules where whole clusters or droplets of spins are flipped at a time. Still, it took until 1987 before Swendsen and Wang[39] proposed the first legitimate cluster update procedure satisfying detailed balance. For the Ising model this follows from the identity

$$Z = \sum_{\{\sigma_i\}} \exp\left(\beta \sum_{\langle ij\rangle} \sigma_i\sigma_j\right) \tag{29}$$

$$= \sum_{\{\sigma_i\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p) + p\delta_{\sigma_i,\sigma_j}\right] \tag{30}$$

$$= \sum_{\{\sigma_i\}} \sum_{\{n_{ij}\}} \prod_{\langle ij\rangle} e^{\beta}\left[(1-p)\delta_{n_{ij},0} + p\delta_{\sigma_i,\sigma_j}\delta_{n_{ij},1}\right] \ , \tag{31}$$

where

$$p = 1 - e^{-2\beta} \ . \tag{32}$$

Here the n_{ij} are bond occupation variables which can take the values $n_{ij} = 0$ or 1, interpreted as "deleted" or "active" bonds. The representation (30) follows from the observation that the product $\sigma_i\sigma_j$ of two Ising spins can only take the two values ± 1, so that $\exp(\beta\sigma_i\sigma_j) = x + y\delta_{\sigma_i\sigma_j}$ can easily be solved for x and y. And in the third line (31) we made use of the trivial (but clever) identity $a+b = \sum_{n=0}^{1}(a\delta_{n,0} + b\delta_{n,1})$. Going one step further and performing in (31) the summation over spins, one arrives at the so-called Fortuin-Kasteleyn representation.[40–43]

2.5.1. *Swendsen-Wang multiple-cluster algorithm*

According to (31) a cluster update sweep consists of two alternating steps. One first updates the bond variables n_{ij} for given spins and then updates the spins σ_i for a given bond configuration:

(1) If $\sigma_i \neq \sigma_j$, set $n_{ij} = 0$, or if $\sigma_i = \sigma_j$, assign values $n_{ij} = 1$ and 0 with probability p and $1-p$, respectively, cf. Fig. 4.
(2) Identify *stochastic* 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 efficient algorithms from percolation theory available for this task.[44–47]

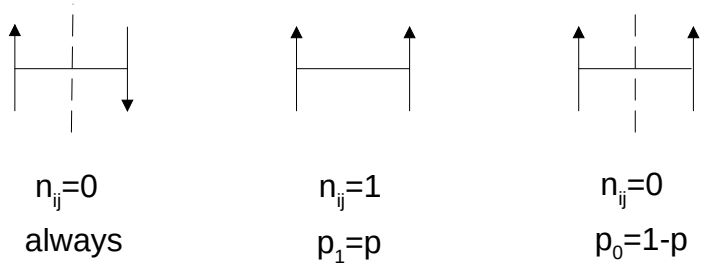

Fig. 4. Illustration of the bond variable update. The bond between unlike spins is always "deleted" as indicated by the dashed line. A bond between like spins is only "active" with probability $p = 1 - \exp(-2\beta)$. Only at zero temperature ($\beta \longrightarrow \infty$) stochastic and geometrical clusters coincide.

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 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. It is worth pointing out that at least for the 2D Ising and more generally 2D Potts models the *geometrical* clusters also do encode critical properties – albeit those of different but related (tricritical) models.[48]

As described above, the cluster algorithm is referred to as Swendsen-Wang (SW) or multiple-cluster update.[39] 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.

2.5.2. *Wolff single-cluster algorithm*

In the single-cluster algorithm of Wolff[49] one constructs only the one cluster connected with a randomly chosen site and then flips all spins of this cluster. Typical configuration plots before and after the cluster flip are shown in Fig. 5, which also illustrates the difference between *stochastic* and *geometrical* clusters mentioned in the last paragraph: The upper right plot clearly shows that, due to the randomly distributed inactive bonds between like spins, the stochastic cluster is much smaller than the underlying black geometrical cluster which connects *all* neighboring like spins.

In the single-cluster variant some care is necessary with the definition of the unit of "time" since the number of flipped spins varies from cluster to cluster. It

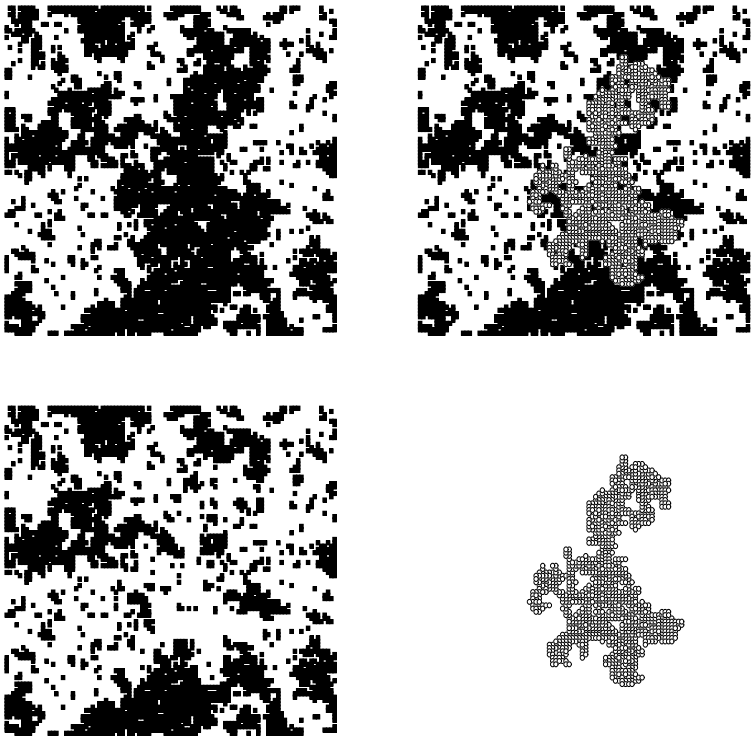

Fig. 5. Illustration of the Wolff single-cluster update for the 2D Ising model on a 100×100 square lattice at $0.97 \times \beta_c$. *Upper left:* Initial configuration. *Upper right:* The stochastic cluster is marked. Note how it is embedded into the larger geometric cluster connecting *all* neighboring like (black) spins. *Lower left:* Final configuration after flipping the spins in the cluster. *Lower right:* The flipped cluster.

also depends crucially on temperature since the average cluster size automatically adapts to the correlation length. With $\langle|C|\rangle$ denoting the average cluster size, a sweep is usually defined to consist of $V/\langle|C|\rangle$ single cluster steps, assuring that on the average V spins are flipped in one sweep. With this definition, autocorrelation times are directly comparable with results from the Swendsen-Wang or Metropolis algorithm. Apart from being somewhat easier to program, Wolff's single-cluster variant is usually 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.

2.5.3. *Embedded cluster algorithm*

While for q-state Potts models[50] with Hamiltonian $\mathcal{H}_{\rm Potts} = -\sum_{\langle ij\rangle} \delta_{\sigma_i\sigma_j}$, $\sigma_i = 1,\ldots,q$, the generalization of (29)–(32) is straightforward (because also the Potts spin-spin interaction $\delta_{\sigma_i\sigma_j}$ contributes only two possible values to the energy, as in the Ising model), for O(n) spin models with $n \geq 2$ defined by the Hamiltonian

$$\mathcal{H}_{O(n)} = -J \sum_{\langle ij\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \ , \quad \vec{\sigma}_i = (\sigma_{i,1}, \sigma_{i,2}, \ldots, \sigma_{i,n}) \ , \quad |\vec{\sigma}_i| = 1 \ , \tag{33}$$

one needs a new strategy.[49,51–53] The basic idea is to isolate Ising degrees of freedom by projecting the spins $\vec{\sigma}_i$ onto a randomly chosen unit vector $\vec{r}$,

$$\vec{\sigma}_i = \vec{\sigma}_i^{\parallel} + \vec{\sigma}_i^{\perp} \ , \quad \vec{\sigma}_i^{\parallel} = \epsilon_i \, |\vec{\sigma}_i \cdot \vec{r}| \, \vec{r} \ , \quad \epsilon_i = \mathrm{sign}(\vec{\sigma}_i \cdot \vec{r}) \ . \tag{34}$$

Inserting this in (33) one ends up with an effective Hamiltonian

$$\mathcal{H}_{O(n)} = -\sum_{\langle ij\rangle} J_{ij} \epsilon_i \epsilon_j + \mathrm{const} \ , \tag{35}$$

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

2.5.4. *Performance of cluster algorithms*

Beside the generalization to O(n)-symmetric spin models, cluster update algorithms have also been constructed for many other models.[36] Close to criticality, they clearly outperform local algorithms with dynamical critical exponent $z \approx 2$, that is, for both cluster variants much smaller values of z have been obtained in 2D and 3D.[36,54–59] For a rigorous lower bound for the autocorrelation time of the Swenden-Wang algorithm, see Ref. 60. In 2D, the efficiencies of Swendsen-Wang and Wolff cluster updates are comparable, whereas in 3D, the Wolff update is favourable.

2.5.5. *Improved estimators*

The intimate relationship of cluster algorithms with the correlated percolation representation of Fortuin and Kasteleyn[40–43] leads to another quite important improvement which is not directly related with the dynamical properties discussed so far. Within the percolation picture, it is quite natural to introduce alternative estimators ("measurement prescriptions") for most standard quantities which turn out to be so-called "improved estimators". By this one means measurement prescriptions that yield the same expectation value as the standard ones but have a smaller statistical variance which helps to reduce the statistical errors.

Suppose we want to estimate 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, $\hat{\mathcal{O}}' = \hat{\mathcal{O}} + \hat{\mathcal{X}}$, as long as the added estimator $\hat{\mathcal{X}}$ has zero expectation, $\langle\hat{\mathcal{X}}\rangle = 0$. The variance of the estimator $\hat{\mathcal{O}}'$, however, can be quite different and is not necessarily related to any physical quantity (contrary to the standard mean-value estimator of the energy, for instance, whose variance is proportional to the specific heat). It is exactly this freedom in the choice of $\hat{\mathcal{O}}$ 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, $G(\vec{x}_i - \vec{x}_j) \equiv \langle\vec{\sigma}_i \cdot \vec{\sigma}_j\rangle$, is given by[53]

$$\hat{G}(\vec{x}_i - \vec{x}_j) = n\frac{V}{|C|}\vec{r}\cdot\vec{\sigma}_i \; \vec{r}\cdot\vec{\sigma}_j \; \Theta_C(\vec{x}_i)\Theta_C(\vec{x}_j) \;\; , \tag{36}$$

where $\vec{r}$ is the normal of the mirror plane used in the construction of the cluster of size $|C|$ and $\Theta_C(\vec{x})$ is its characteristic function (=1 if $\vec{x} \in C$ and 0 otherwise). In the Ising case ($n = 1$), this simplifies to

$$\hat{G}(\vec{x}_i - \vec{x}_j) = \frac{V}{|C|}\Theta_C(\vec{x}_i)\Theta_C(\vec{x}_j) \;\; , \tag{37}$$

i.e., to the test whether the two sites $\vec{x}_i$ and $\vec{x}_j$ belong to same stochastic cluster or not. Only in the former case, the average over clusters is incremented by one, otherwise nothing is added. This implies that $\hat{G}(\vec{x}_i - \vec{x}_j)$ is strictly positive which is not the case for the standard estimator $\vec{\sigma}_i \cdot \vec{\sigma}_j$, where ± 1 contributions have to average to a positive value. It is therefore at least intuitively clear that the cluster (or percolation) estimator has a smaller variance and is thus indeed an improved estimator, in particular for large separations $|\vec{x}_i - \vec{x}_j|$. For the Fourier transform, $\tilde{G}(\vec{k}) = \sum_{\vec{x}} G(\vec{x}) \exp(-i\vec{k}\cdot\vec{x})$, Eq. (36) implies the improved estimator

$$\hat{\tilde{G}}(\vec{k}) = \frac{n}{|C|}\left[\left(\sum_{i\in C}\vec{r}\cdot\vec{\sigma}_i \cos\vec{k}\vec{x}_i\right)^2 + \left(\sum_{i\in C}\vec{r}\cdot\vec{\sigma}_i \sin\vec{k}\vec{x}_i\right)^2\right] \;\; , \tag{38}$$

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

$$\hat{\tilde{G}}(\vec{0}) = \hat{\chi}'/\beta = \frac{n}{|C|}\left(\sum_{i\in C}\vec{r}\cdot\vec{\sigma}_i\right)^2 \;\; . \tag{39}$$

For the Ising model ($n = 1$) this reduces to $\chi'/\beta = \langle|C|\rangle$, i.e., the improved estimator of the susceptibility is just the average cluster size of the single-cluster

update algorithm. For the XY ($n = 2$) and Heisenberg ($n = 3$) models one finds empirically that in two as well as in three dimensions $\langle|C|\rangle \approx 0.81\chi'/\beta$ for $n = 2$ (Refs. 51,58) and $\langle|C|\rangle \approx 0.75\chi'/\beta$ for $n = 3$ (Refs. 53,59), respectively.

Close to criticality, the average cluster size becomes large, growing in a finite system of linear length L (cf. Sect. 6) as $\chi' \propto L^{\gamma/\nu} \simeq L^2$, since $\gamma/\nu = 2 - \eta$ with η usually small, and the advantage of cluster estimators diminishes. In fact, in particular for short-range quantities such as the energy (the next-neighbor correlation) it may even degenerate into a "deproved" or "deteriorated" estimator, while long-range quantities such as $G(\vec{x}_i - \vec{x}_j)$ for large distances $|\vec{x}_i - \vec{x}_j|$ usually still profit from it. A significant reduction of variance by means of the estimators (36)–(39) can, however, always be expected outside the critical region where the average cluster size is small compared to the volume of the system.

3. Statistical Analysis of Monte Carlo Data

3.1. *Statistical errors and autocorrelation times*

3.1.1. *Estimators*

When discussing the importance sampling idea in Sect. 2.2 we already saw in Eq. (9) that within Markov chain Monte Carlo simulations, the expectation value $\langle\mathcal{O}\rangle$ of some quantity $\mathcal{O}$, for instance the energy, can be estimated as arithmetic mean,

$$\langle\mathcal{O}\rangle = \sum_\sigma \mathcal{O}(\sigma) P^{\mathrm{eq}}(\sigma) \approx \overline{\mathcal{O}} = \frac{1}{N}\sum_{k=1}^{N} \mathcal{O}_k \ , \tag{40}$$

where the "measurement" $\mathcal{O}_k = \mathcal{O}(\sigma^{(k)})$ is obtained from the kth microstate $\sigma^{(k)}$ and N is the number of measurement sweeps. Of course, this is only valid after a sufficiently long thermalization period without measurements, which is needed to equilibrate the system after starting the Markov chain in an arbitrarily chosen initial configuration.

Conceptually it is important to distinguish between the expectation value $\langle\mathcal{O}\rangle$, an ordinary number representing the exact result (which is usually unknown, of course), and the mean value $\overline{\mathcal{O}}$, which is a so-called *estimator* of the former. In contrast to $\langle\mathcal{O}\rangle$, the estimator $\overline{\mathcal{O}}$ is a *random* variable which for finite N fluctuates around the theoretically expected value. Certainly, from a single Monte Carlo simulation with N measurements, we obtain only a single number for $\overline{\mathcal{O}}$ at the end of the day. For estimating the statistical uncertainty due to the fluctuations, i.e., the statistical error, it seems at first sight that one would have to repeat the

whole simulation many times. Fortunately, this is not so because one can express the variance of $\overline{\mathcal{O}}$,

$$\sigma^2_{\overline{\mathcal{O}}} = \langle[\overline{\mathcal{O}} - \langle\overline{\mathcal{O}}\rangle]^2\rangle = \langle\overline{\mathcal{O}}^2\rangle - \langle\overline{\mathcal{O}}\rangle^2 \ , \tag{41}$$

in terms of the statistical properties of the individual measurements $\mathcal{O}_k, k = 1,\ldots,N$, of a single Monte Carlo run.

3.1.2. *Uncorrelated measurements*

Inserting (40) into (41) gives

$$\begin{aligned}\sigma^2_{\overline{\mathcal{O}}} &= \langle\overline{\mathcal{O}}^2\rangle - \langle\overline{\mathcal{O}}\rangle^2 \\ &= \frac{1}{N^2}\sum_{k=1}^{N}\left(\langle\mathcal{O}_k^2\rangle - \langle\mathcal{O}_k\rangle^2\right) + \frac{1}{N^2}\sum_{k\neq l}^{N}\left(\langle\mathcal{O}_k\mathcal{O}_l\rangle - \langle\mathcal{O}_k\rangle\langle\mathcal{O}_l\rangle\right) \ ,\end{aligned} \tag{42}$$

where we have collected diagonal and off-diagonal terms. The second, off-diagonal term encodes the "temporal" correlations between measurements at "times" k and l and thus vanishes for completely uncorrelated data (which is, of course, never really the case for importance sampling Monte Carlo simulations). Assuming equilibrium, the variances $\sigma^2_{\mathcal{O}_k} = \langle\mathcal{O}_k^2\rangle - \langle\mathcal{O}_k\rangle^2$ of individual measurements appearing in the first, diagonal term do not depend on "time" k, such that $\sigma^2_{\mathcal{O}_k} = \sigma^2_{\mathcal{O}}$ and (42) simplifies to

$$\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}}/N \ . \tag{43}$$

Whatever form the distribution $\mathcal{P}(\mathcal{O}_k)$ assumes (which, in fact, is often close to Gaussian because the $\mathcal{O}_k$ are usually already lattice averages over many degrees of freedom), by the central limit theorem the distribution of the mean value is Gaussian, at least for weakly correlated 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 $[\langle\mathcal{O}\rangle - \sigma_{\overline{\mathcal{O}}}, \langle\mathcal{O}\rangle + \sigma_{\overline{\mathcal{O}}}]$.[61] 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%.

3.1.3. *Correlated measurements and autocorrelation times*

For correlated data the off-diagonal term in (42) does not vanish and things become more involved.[62–65] Using the symmetry $k \leftrightarrow l$ to rewrite the summation

$\sum_{k\neq l}^{N}$ as $2\sum_{k=1}^{N}\sum_{l=k+1}^{N}$, reordering the summation, and using time-translation invariance in equilibrium, one obtains[66]

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N}\left[\sigma^2_{\mathcal{O}} + 2\sum_{k=1}^{N}\Big(\langle\mathcal{O}_1\mathcal{O}_{1+k}\rangle - \langle\mathcal{O}_1\rangle\langle\mathcal{O}_{1+k}\rangle\Big)\left(1-\frac{k}{N}\right)\right] \ , \qquad (44)$$

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

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{\sigma^2_{\mathcal{O}}}{N}\, 2\tau_{\mathcal{O},\mathrm{int}} \ , \qquad (45)$$

where we have introduced the *integrated* autocorrelation time

$$\tau_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k)\left(1-\frac{k}{N}\right) \ , \qquad (46)$$

with

$$A(k) \equiv \frac{\langle\mathcal{O}_1\mathcal{O}_{1+k}\rangle - \langle\mathcal{O}_1\rangle\langle\mathcal{O}_{1+k}\rangle}{\sigma^2_{\mathcal{O}}} \overset{k\to\infty}{\longrightarrow} a e^{-k/\tau_{\mathcal{O},\mathrm{exp}}} \qquad (47)$$

being the normalized autocorrelation function ($A(0)=1$). In any meaningful simulation study one chooses $N \gg \tau_{\mathcal{O},\mathrm{exp}}$, so that $A(k)$ is already exponentially small before the correction term $(1-k/N)$ in (46) becomes important. It is therefore often omitted for simplicity.

As far as the accuracy of Monte Carlo data is concerned, the important point of Eq. (45) 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 Monte Carlo 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 $\epsilon_{\overline{\mathcal{O}}} = \sqrt{\sigma^2_{\mathcal{O}}/N_{\mathrm{eff}}}$, but now with a parameter

$$N_{\mathrm{eff}} = N/2\tau_{\mathcal{O},\mathrm{int}} \leq N \ , \qquad (48)$$

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 (25), (26) or (28), it is obvious that without extra care this effective sample size may become very small close to a continuous or first-order phase transition, respectively.

3.1.4. *Bias*

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

$$\hat{\sigma}^2_{\mathcal{O}} = \overline{\mathcal{O}^2} - \overline{\mathcal{O}}^2 = \overline{(\mathcal{O} - \overline{\mathcal{O}})^2} = \frac{1}{N}\sum_{k=1}^{N} \left(\mathcal{O}_k - \overline{\mathcal{O}}\right)^2 \quad . \tag{49}$$

Subtracting and adding $\langle\overline{\mathcal{O}}\rangle^2$, one finds for the *expected* value of $\hat{\sigma}^2_{\mathcal{O}}$,

$$\langle\hat{\sigma}^2_{\mathcal{O}}\rangle = \langle\overline{\mathcal{O}^2} - \overline{\mathcal{O}}^2\rangle = \left(\langle\overline{\mathcal{O}^2}\rangle - \langle\overline{\mathcal{O}}\rangle^2\right) - \left(\langle\overline{\mathcal{O}}^2\rangle - \langle\overline{\mathcal{O}}\rangle^2\right) = \sigma^2_{\mathcal{O}} + \sigma^2_{\overline{\mathcal{O}}} \ . \tag{50}$$

Using (45) this gives

$$\langle\hat{\sigma}^2_{\mathcal{O}}\rangle = \sigma^2_{\mathcal{O}}\left(1 - \frac{2\tau_{\mathcal{O},\mathrm{int}}}{N}\right) = \sigma^2_{\mathcal{O}}\left(1 - \frac{1}{N_{\mathrm{eff}}}\right) \neq \sigma^2_{\mathcal{O}} \ . \tag{51}$$

The estimator $\hat{\sigma}^2_{\mathcal{O}}$ in (49) thus systematically underestimates the true value by a term of the order of $\tau_{\mathcal{O},\mathrm{int}}/N$. 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, the bias may be quite large. Since for local update algorithms $\tau_{\mathcal{O},\mathrm{int}}$ scales quite strongly with the system size, some care is necessary when choosing the run time N. Otherwise the system-size dependence of the specific heat or susceptibility may be systematically influenced by temporal correlations.[67] Any serious simulation should therefore provide at least a rough order-of-magnitude estimate of autocorrelation times.

3.1.5. *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. Similar to the estimator (49) for the variance a (weakly biased) estimator $\hat{A}(k)$ for the autocorrelation function is obtained by replacing in (47) the expectation values (ordinary

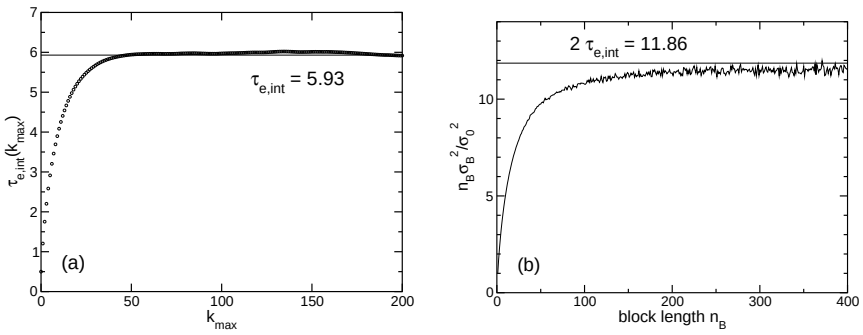

Fig. 6. (a) Integrated autocorrelation time approaching $\tau_{e,\mathrm{int}} \approx 5.93$ for large upper cutoff $k_{\max}$ and (b) binning analysis for the energy of the 2D Ising model on a 16×16 lattice at β_c, using the same data as in Fig. 3. The horizontal line in (b) shows $2\tau_{e,\mathrm{int}}$ with $\tau_{e,\mathrm{int}}$ read off from (a).

numbers) by mean values (random variables), e.g., $\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle$ by $\overline{\mathcal{O}_1 \mathcal{O}_{1+k}}$. 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 (45), it is useful to record the "running" autocorrelation time estimator

$$\hat{\tau}_{\mathcal{O},\mathrm{int}}(k_{\max}) = \frac{1}{2} + \sum_{k=1}^{k_{\max}} \hat{A}(k) \ , \tag{52}$$

which approaches $\tau_{\mathcal{O},\mathrm{int}}$ in the limit of large $k_{\max}$ where, however, the statistical error rapidly increases. As an example, Fig. 6(a) shows results for the 2D Ising model from an analysis of the same raw data as in Fig. 3.

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,[34,35,63]

$$\epsilon_{\tau_{\mathcal{O},\mathrm{int}}} = \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{2(2k_{\max}+1)}{N}} \approx \tau_{\mathcal{O},\mathrm{int}} \sqrt{\frac{12}{N_{\mathrm{eff}}}} \ . \tag{53}$$

For a 5% relative accuracy one thus needs at least $N_{\mathrm{eff}} \approx 5\,000$ or $N \approx 10\,000\,\tau_{\mathcal{O},\mathrm{int}}$ measurements. For an order of magnitude estimate consider the 2D Ising model on a square lattice with $L = 100$ simulated with a local update algorithm. Close to criticality, the integrated autocorrelation time for this example is of the order of $L^z \approx L^2 \approx 100^2$ (ignoring an unknown prefactor of "order unity" which depends on the considered quantity), implying $N \approx 10^8$. Since in each sweep L^2 spins have to be updated and assuming that each spin update takes

about 0.1 μsec, we end up with a total time estimate of about 10^5 seconds ≈ 1 CPU-day to achieve this accuracy.

An alternative is to approximate the tail end of $A(k)$ by a single exponential as in (24). Summing up the small k part exactly, one finds[68]

$$\tau_{\mathcal{O},\mathrm{int}}(k_{\max}) = \tau_{\mathcal{O},\mathrm{int}} - ce^{-k_{\max}/\tau_{\mathcal{O},\mathrm{exp}}} \ , \tag{54}$$

where c is a constant. The latter expression may be used for a numerical estimate of both the exponential and integrated autocorrelation times.[68]

3.2. *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 analyses 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 n_B (such that[f] $N = N_B n_B$), one forms a new, shorter time series of block averages,

$$\mathcal{O}_j^{(B)} \equiv \frac{1}{n_B}\sum_{i=1}^{n_B} \mathcal{O}_{(j-1)n_B+i} \ , \qquad j = 1,\ldots,N_B \ , \tag{55}$$

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

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \sigma_B^2/N_B = \frac{1}{N_B(N_B-1)}\sum_{j=1}^{N_B}(\mathcal{O}_j^{(B)} - \overline{\mathcal{O}^{(B)}})^2 \ . \tag{56}$$

By comparing with (45) we see that $\sigma_B^2/N_B = 2\tau_{\mathcal{O},\mathrm{int}}\sigma_{\mathcal{O}}^2/N$. Recalling the definition of the block length $n_B = N/N_B$, this shows that one may also use

$$2\tau_{\mathcal{O},\mathrm{int}} = n_B\sigma_B^2/\sigma_{\mathcal{O}}^2 \tag{57}$$

for the estimation of $\tau_{\mathcal{O},\mathrm{int}}$. This is demonstrated in Fig. 6(b). Estimates of $\tau_{\mathcal{O},\mathrm{int}}$ obtained in this way are often referred to as "blocking τ" or "binning τ".

A simple toy model (bivariate time series), where the behaviour of the "blocking τ" and also of $\tau_{\mathcal{O},\mathrm{int}}(k_{\max})$ for finite n_B resp. $k_{\max}$ can be worked out exactly, is discussed in Ref. 26. These analytic formulas are very useful for validating the computer implementations.

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

3.3. *Jackknife analysis*

Even if the data are completely uncorrelated in time, one still has to handle the problem of error estimation for quantities that are not "directly" measured in the simulation but are computed as a non-linear combination of "basic" observables such as $\langle\mathcal{O}\rangle^2$ or $\langle\mathcal{O}_1\rangle/\langle\mathcal{O}_2\rangle$. This problem can either be solved by error propagation or by using the Jackknife method,[69,70] where instead of considering rather small blocks of length n_B and their fluctuations as in the binning analysis, one forms N_B large Jackknife blocks $\mathcal{O}_j^{(J)}$ containing all data but the jth block of the previous binning method,

$$\mathcal{O}_j^{(J)} = \frac{N\overline{\mathcal{O}} - n_B\mathcal{O}_j^{(B)}}{N - n_B} \ , \qquad j = 1, \ldots, N_B \ , \tag{58}$$

cf. the schematic sketch in Fig. 7. Each of the Jackknife blocks thus consists of $N - n_B = N(1 - 1/N_B)$ 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 $1/(N - n_B)$ compared to $1/N$). The N_B Jackknife blocks are, of course, trivially correlated because one and the same original data is re-used in $N_B - 1$ different Jackknife blocks. This trivial correlation caused by re-using the original data over and over again has nothing to do with temporal correlations. As a consequence, the Jacknife block variance σ_J^2 will be much smaller than the variance estimated in the binning method. Because of the trivial nature of the correlations, however, this reduction can be corrected by multiplying σ_J^2 with a factor $(N_B - 1)^2$, leading to

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{N_B - 1}{N_B} \sum_{j=1}^{N_B} (\mathcal{O}_j^{(J)} - \overline{\mathcal{O}^{(J)}})^2 \ . \tag{59}$$

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

4. Reweighting Techniques

The physics underlying reweighting techniques[71,72] is extremely simple and the basic idea has been known since long (see the list of references in Ref. 72), but

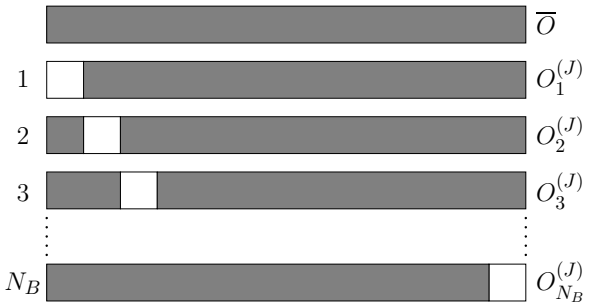

Fig. 7. Sketch of the organization of Jackknife blocks. The grey part of the N data points is used for calculating the total and the Jackknife block averages. The white blocks enter into the more conventional binning analysis using non-overlapping blocks.

their power in practice has been realized only relatively late in 1988. The important observation by Ferrenberg and Swendsen[71,72] was that the best performance is achieved *near* criticality where histograms are usually broad. In this sense reweighting techniques are complementary to improved estimators, which usually perform best *off* criticality.

4.1. *Single-histogram technique*

The single-histogram reweighting technique[71] is based on the following very simple observation. Denoting the number of states (spin configurations) that have the same energy $e = E/V$ by $\Omega(e)$, the partition function at the simulation point $\beta_0 = 1/k_B T_0$ can always be written as[g]

$$Z(\beta_0) = \sum_\sigma e^{-\beta_0 \mathcal{H}(\sigma)} = \sum_e \Omega(e) e^{-\beta_0 E} \propto \sum_e P_{\beta_0}(e) \ , \tag{60}$$

where we have introduced the unnormalized energy histogram (density)

$$P_{\beta_0}(e) \propto \Omega(e) e^{-\beta_0 E} \ . \tag{61}$$

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

[g]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.

β_0 and thus know $P_{\beta_0}(e)$. It is then easy to see that

$$P_\beta(e) \propto \Omega(e)e^{-\beta E} = \Omega(e)e^{-\beta_0 E}e^{-(\beta-\beta_0)E} \propto P_{\beta_0}(e)e^{-(\beta-\beta_0)E} \ , \qquad (62)$$

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,

$$\langle f(e)\rangle(\beta) = \sum_e f(e)P_\beta(e) / \sum_e P_\beta(e) \ , \qquad (63)$$

the normalization of $P_\beta(e)$ indeed cancels. This gives for instance the energy $\langle e\rangle(\beta)$ and the specific heat $C(\beta) = \beta^2 V[\langle e^2\rangle(\beta) - \langle e\rangle(\beta)^2]$, in principle, as a continuous function of β from a single Monte Carlo simulation at β_0, where $V = L^D$ is the system size.

As an example of this reweighting procedure, using actual Swendsen-Wang cluster simulation data (with 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 reweighted data points for the specific heat $C(\beta)$ are shown in Fig. 8(a) and compared with the continuous curve obtained from the exact Kaufman solution[73,74] for finite $L_x \times L_y$ lattices. Note that the location of the peak maximum is slightly displaced from the infinite-volume transition point β_c due to the rounding and shifting of $C(\beta)$ caused by finite-size effects discussed in more detail in Sect. 6. This comparison clearly demonstrates that, in practice, the β-range over which reweighting can be trusted is limited. The reason for this limitation are unavoidable statistical errors in the numerical determination of P_{β_0} using a Monte Carlo simulation. In the tails of the histograms the relative statistical errors are largest, and the tails are exactly the regions that contribute most when multiplying $P_{\beta_0}(e)$ with the exponential reweighting factor to obtain $P_\beta(e)$ for β-values far off the simulation point β_0. This is illustrated in Fig. 8(b) where the simulated histogram at $\beta_0 = \beta_c$ is shown together with the reweighted histograms at $\beta = 0.375 \approx \beta_0 - 0.065$ and $\beta = 0.475 \approx \beta_0 + 0.035$, respectively. For the 2D Ising model the quality of the reweighted histograms can be judged by comparing with the curves obtained from Beale's[75] exact expression for $\Omega(e)$.

4.1.1. *Reweighting range*

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

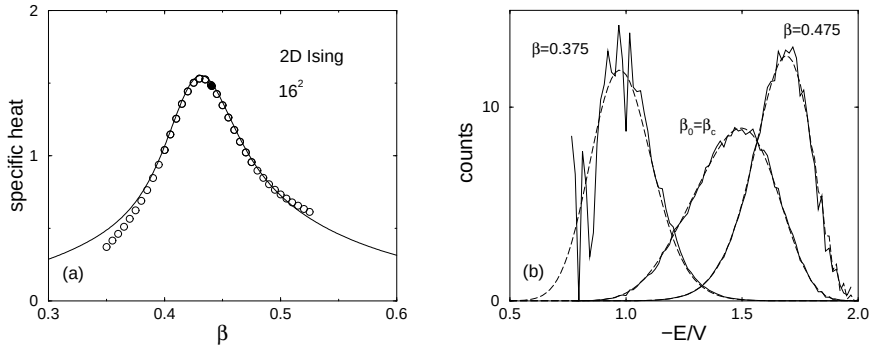

Fig. 8. (a) The specific heat of the 2D Ising model on a 16×16 square lattice computed by reweighting from a single Monte Carlo simulation at $\beta_0 = \beta_c$, marked by the filled data symbol. The continuous line shows for comparison the exact solution of Kaufman.[73,74] (b) The corresponding energy histogram at β_0, and reweighted to $\beta = 0.375$ and $\beta = 0.475$. The dashed lines show for comparison the exact histograms obtained from Beale's expression.[75]

this range is wide enough to locate from a single simulation, e.g., the specific-heat maximum by employing a standard maximization subroutine to the continuous 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 splines 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 maximum value),

$$|\langle e\rangle(T) - \langle e\rangle(T_0)| \le \Delta e(T_0) \ , \tag{64}$$

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

$$\frac{|T - T_0|}{T_0} \le \frac{1}{\sqrt{V}}\frac{1}{\sqrt{C(T_0)}} \ . \tag{65}$$

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. 8 with $V = 16 \times 16$, $\beta_0 = \beta_c \approx 0.44$ and $C(T_0) \approx 1.5$, this estimate yields $|\beta - \beta_0|/\beta_0 \approx |T - T_0|/T_0 \le 0.05$, i.e., $|\beta - \beta_0| \le 0.02$ or $0.42 \le \beta \le 0.46$. By comparison with the exact solution

we see that this is indeed a fairly conservative estimate of the reliable reweighting range.

If we only want to know the scaling 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

$$\frac{|T - T_0|}{T_0} \propto V^{-1/2} = L^{-D/2} \ . \tag{66}$$

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

$$\frac{|T - T_0|}{T_0} \propto L^{-1/\nu} \ , \tag{67}$$

i.e., the typical scaling behaviour of pseudo-transition temperatures in the finite-size scaling regime of a second-order phase transition.[76] 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 $C(T_0) \propto V = L^D$. This yields

$$\frac{|T - T_0|}{T_0} \propto V^{-1} = L^{-D} \ , \tag{68}$$

which is again the typical finite-size scaling behaviour of pseudo-transition temperatures close to a first-order phase transition.[38]

4.1.2. *Reweighting of non-conjugate observables*

If we also want to reweight other quantities such as the magnetization $m = \langle \mu \rangle$ we have to go one step further. The conceptually simplest way would be to store two-dimensional histograms $P_{\beta_0}(e, \mu)$ where $e = E/V$ is the energy and $\mu = \sum_i \sigma_i / V$ the 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 $\exp(\beta h \sum_i \sigma_i) = \exp(\beta V h \mu)$. 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 $g(\mu)$, e.g.,

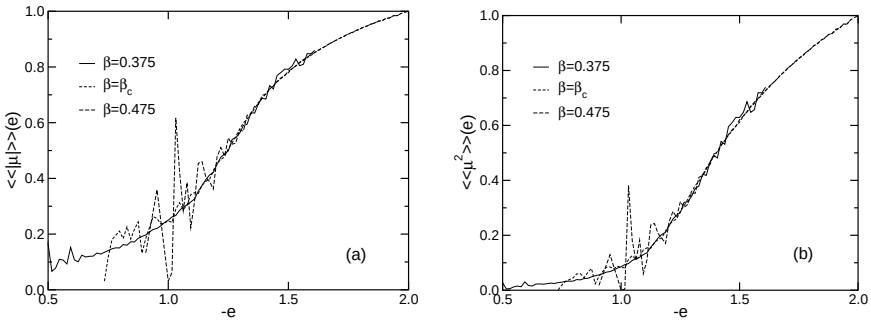

Fig. 9. Microcanonical expectation values for (a) the absolute magnetization and (b) the magnetization squared obtained from the 2D Ising model simulations shown in Fig. 8.

$g(\mu) = \mu^k$, we can write

$$\langle g(\mu)\rangle = \sum_\sigma g(\mu(\sigma)) e^{-\beta_0 \mathcal{H}(\sigma)}/Z(\beta_0) = \sum_{e,\mu} \Omega(e,\mu) g(\mu) e^{-\beta_0 E}/Z(\beta_0)$$
$$= \sum_e \frac{\sum_\mu \Omega(e,\mu) g(\mu)}{\sum_\mu \Omega(e,\mu)} \sum_\mu \Omega(e,\mu) e^{-\beta_0 E}/Z(\beta_0) \ . \tag{69}$$

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

$$\langle\langle g(\mu)\rangle\rangle(e) \equiv \frac{\sum_\mu \Omega(e,\mu) g(\mu)}{\sum_\mu \Omega(e,\mu)} \ , \tag{70}$$

we arrive at

$$\langle g(\mu)\rangle = \sum_e \langle\langle g(\mu)\rangle\rangle(e) P_{\beta_0}(e) \ . \tag{71}$$

Identifying $\langle\langle g(\mu)\rangle\rangle(e)$ with $f(e)$ in Eq. (63), the actual reweighting procedure is precisely as before. An example for computing $\langle\langle|\mu|\rangle\rangle(e)$ and $\langle\langle\mu^2\rangle\rangle(e)$ using the data of Fig. 8 is shown in Fig. 9. Mixed quantities, e.g. $\langle e^k \mu^l\rangle$, can be treated similarly. One caveat of this method is that one has to decide beforehand which "lists" $\langle\langle g(\mu)\rangle\rangle(e)$ one wants to store during the simulation, e.g., which powers k in $\langle\langle \mu^k\rangle\rangle(e)$ are relevant.

An alternative and more flexible method is based on time series. Suppose we have performed a Monte Carlo simulation at β_0 and stored the time series of N measurements $e_1, e_2, \ldots, e_N$ and $\mu_1, \mu_2, \ldots, \mu_N$. Then the most general

expectation values at another inverse temperature β can simply be obtained from

$$\langle f(e,\mu)\rangle = \sum_{i=1}^{N} f(e_i,\mu_i) e^{-(\beta-\beta_0)E_i} / \sum_{i=1}^{N} e^{-(\beta-\beta_0)E_i} \quad , \tag{72}$$

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

$$\langle f(e,\mu)\rangle = \langle f(e,\mu) e^{-(\beta-\beta_0)E}\rangle_0 / \langle e^{-(\beta-\beta_0)E}\rangle_0 \quad , \tag{73}$$

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.

4.2. *Multi-histogram technique*

The basic idea of the multi-histogram technique[77] can be summarized as follows:

i) Perform m Monte Carlo simulations at $\beta_1, \beta_2, \ldots, \beta_m$ with N_i, $i = 1, \ldots, m$, measurements,
ii) reweight all runs to a common reference point β_0,
iii) combine at β_0 all information by computing error weighted averages,
iv) reweight the "combined histogram" to any other β.

Since a weighted combination of several histograms enters this method it is also referred to as "weighted histogram analysis method" or "WHAM".[78,79] In fact, in chemistry and biochemistry the multi-histogram method is basically only known under this acronym.

To proceed we first note that the exact normalized energy distribution at $\beta = \beta_i$ can we written as

$$P_i(e) \equiv P_{\beta_i}(e) = \frac{\Omega(e) e^{-\beta_i E}}{Z_i} \quad , \tag{74}$$

where $Z_i \equiv Z(\beta_i)$ so that $\sum_e P_i(e) = 1$. This can be estimated by the empirical histogram $H_i(e)$ obtained from the simulation at β_i,

$$\hat{P}_i(e) = \frac{H_i(e)}{N_i} \ , \tag{75}$$

which also satisfies the normalization constraint $\sum_e \hat{P}_i(e) = 1$. Rearranging (74) and replacing the exact $P_i(e)$ by its estimator $\hat{P}_i(e)$ yields an estimator for the density of states (this corresponds to choosing the common reference point as $\beta_0 = 0$):

$$\hat{\Omega}_i(e) = Z_i e^{\beta_i E} \frac{H_i(e)}{N_i} \ . \tag{76}$$

Notice that we have introduced a subscript i to label the m estimators $\hat{\Omega}_i(e)$. The expectation value of each $\hat{\Omega}_i(e)$ should be the exact $\Omega(e)$, but being random variables their statistical properties are different as can be quantified by estimating their variance. This is simplest done by interpreting the histogram entries $H_i(e)$ as result of measuring $\mathcal{O} = \delta_{e_t,e}$ where e_t denotes the energy after the t's sweep of the simulation at β_i:

$$\frac{H_i(e)}{N_i} = \overline{\delta_{e_t,e}} = \frac{1}{N_i} \sum_{t=1}^{N_i} \delta_{e_t,e} \ . \tag{77}$$

As in (40) and (41) the expected value is $\langle H_i(e)/N_i \rangle = (1/N_i) \sum_{t=1}^{N_i} \langle \delta_{e_t,e} \rangle = P_i(e)$ and, neglecting temporal correlations for the moment,

$$\begin{aligned} \left\langle \left(\frac{H_i(e)}{N_i} \right)^2 \right\rangle &= \left\langle \frac{1}{N_i^2} \sum_{t,t'=1}^{N_i} \delta_{e_t,e} \delta_{e_{t'},e} \right\rangle \\ &= \frac{1}{N_i^2} \left[N_i(N_i-1) \langle \delta_{e_t,e} \rangle \langle \delta_{e_{t'},e} \rangle + N_i \langle \delta_{e_t,e} \delta_{e_{t'},e} \rangle \right] \\ &= P_i(e)^2 + \frac{1}{N_i} P_i(e)[1 - P_i(e)] \ , \end{aligned} \tag{78}$$

such that

$$\sigma^2_{H_i(e)/N_i} = \left\langle \left(\frac{H_i(e)}{N_i} - \left\langle \frac{H_i(e)}{N_i} \right\rangle \right)^2 \right\rangle = \frac{1}{N_i} P_i(e)[1 - P_i(e)] \ . \tag{79}$$

For sufficiently many energy bins, the normalized probabilities $P_i(e)$ are much smaller than unity, such that the second term $[1 - P_i(e)]$ can usually be neglected. Taking autocorrelations into account, as in (45) the variance (79) would be enhanced by a factor $2\tau_{\mathrm{int},i}(e)$. Recall that the subscript i of $\tau_{\mathrm{int},i}(e)$ refers to the

simulation point and the argument e to the energy bin. Note that the autocorrelation times of the histogram bins are usually much smaller than the autocorrelation time $\tau_{\mathrm{int},e}$ of the mean energy. For the following it is useful to define the effective statistics parameter $N_{\mathrm{eff},i}(e) = N_i/2\tau_{\mathrm{int},i}(e)$. Recalling (76), the variance of the m estimators $\hat{\Omega}_i(e)$ can then be written as

$$\sigma^2_{\hat{\Omega}_i(e)} = \frac{Z_i^2 e^{2\beta_i E}}{N_{\mathrm{eff},i}(e)} P_i(e) = \frac{Z_i e^{\beta_i E}}{N_{\mathrm{eff},i}(e)} \Omega(e) \ . \tag{80}$$

As usual the error weighted average

$$\hat{\Omega}_{\mathrm{opt}}(e) = \frac{\sum_{i=1}^m w_i(e)\hat{\Omega}_i(e)}{\sum_{i=1}^m w_i(e)} \tag{81}$$

with $w_i(e) = 1/\sigma^2_{\hat{\Omega}_i(e)}$ is an optimised estimator with minimal variance $\sigma^2_{\hat{\Omega}_{\mathrm{opt}}(e)} = 1/\sum_{i=1}^m w_i(e)$. This can be simplified to

$$\hat{\Omega}_{\mathrm{opt}}(e) = \frac{\sum_{i=1}^m H_i(e)/2\tau_{\mathrm{int},i}(e)}{\sum_{i=1}^m N_{\mathrm{eff},i}(e) Z_i^{-1} e^{-\beta_i E}} \tag{82}$$

and

$$\sigma^2_{\hat{\Omega}_{\mathrm{opt}}(e)}/\Omega^2(e) = \frac{1}{\sum_{i=1}^m \langle H_i(e)\rangle/2\tau_{\mathrm{int},i}(e)} \ . \tag{83}$$

So far the partition function values $Z_i \equiv Z(\beta_i)$ have been assumed to be exact (albeit usually unknown) parameters which are now self-consistently determined from

$$Z_j = \sum_e \hat{\Omega}_{\mathrm{opt}}(e) e^{-\beta_j E} = \sum_e \frac{\sum_{i=1}^m H_i(e)/2\tau_{\mathrm{int},i}(e)}{\sum_{i=1}^m (N_i/2\tau_{\mathrm{int},i}(e)) Z_i^{-1} e^{-\beta_i E}} e^{-\beta_j E} \ , \tag{84}$$

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 (84). In realistic simulations the statistics is of course limited and the remaining recursions average this uncertainty to get a self-consistent set of Z_i. In order to work in practice, the histograms at neighboring β-values must have sufficient overlap, i.e., the spacings of the simulation points must be chosen according to the estimates (66)–(68). The issue of optimal convergence of the WHAM equations (84) has recently been discussed in detail in Ref. 80.

Multiple-histogram reweighting has been employed in a wide spectrum of applications. In many applications the influence of autocorrelations has been neglected since it is quite cumbersome to estimate the $\tau_{\mathrm{int},i}(e)$ for each of the m

simulations and *all* energy bins. For work dealing with autocorrelations in this context see, e.g., Refs. 81,82. Note that, even when ignoring the $\tau_{\mathrm{int},i}(e)$, the error weighted average in (81) does still give a correct estimator for $\Omega(e)$ – it is only no longer properly optimised. Moreover, since for each energy bin typically only the histograms at neighboring simulation points contribute significantly, the two or three $\tau_{\mathrm{int},i}(e)$ values relevant for each energy bin e are close to each other. And since an overall constant drops out of the WHAM equation (84), the influence of autocorrelations on the final result turns out to be very minor anyway.

Alternatively[59] one may also compute from each of the m independent simulations by reweighting all quantities of interest as a function of β, together with their proper statistical errors including autocorrelation effects as discussed in Sect. 3.1.3. As a result 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 $e(\beta) \pm \Delta e$, where

$$e(\beta) = \left(\frac{e_1(\beta)}{(\Delta e_1)^2} + \frac{e_2(\beta)}{(\Delta e_2)^2} + \cdots + \frac{e_m(\beta)}{(\Delta e_m)^2} \right) (\Delta e)^2 \quad , \tag{85}$$

and

$$\frac{1}{(\Delta e)^2} = \frac{1}{(\Delta e_1)^2} + \frac{1}{(\Delta e_2)^2} + \cdots + \frac{1}{(\Delta e_m)^2} \quad . \tag{86}$$

Notice that by this method the average for each quantity can be individually optimised.

5. Generalized Ensemble Methods

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

5.1. *Simulated tempering*

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

In *simulated* tempering simulations[83,84] one starts from a joint partition function (*expanded* ensemble)

$$\mathcal{Z}_{\rm ST} = \sum_{i=1}^{m} e^{g_i} \sum_{\sigma} e^{-\beta_i \mathcal{H}(\sigma)} \quad , \tag{87}$$

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 update algorithm, a proposed move from $\beta = \beta_i$ to β_j with σ fixed is accepted with probability

$$w = \min\left\{1, \exp[-(\beta_j - \beta_i)\mathcal{H}(\sigma) + g_j - g_i]\right\} \quad . \tag{88}$$

Similar to multi-histogram reweighting (and also to multicanonical simulations discussed below), the free-energy parameters g_i are *a priori* unknown and have to be adjusted iteratively. To assure a reasonable acceptance rate for the β-update moves (usually between neighboring β_i-values), the histograms at β_i and β_{i+1}, $i = 1, \ldots, m-1$, must overlap. An estimate for a suitable spacing $\Delta\beta = \beta_{i+1} - \beta_i$ of the simulation points β_i is hence immediately given by the results (66)–(68) for the reweighting range,

$$\Delta\beta \propto \begin{cases} L^{-D/2} & \text{off-critical} \ , \\ L^{-1/\nu} & \text{critical} \ , \\ L^{-D} & \text{first-order} \ . \end{cases} \tag{89}$$

Overall the simulated tempering method shows some similarities to the "avoiding rare events" variant of multicanonical simulations briefly discussed in subsection 5.3.

5.2. *Parallel tempering*

In *parallel* tempering or *replica exchange* or *multiple Markov chain* Monte Carlo simulations,[85–88] the starting point is a product of partition functions (*extended* ensemble),

$$\mathcal{Z}_{\rm PT} = \prod_{i=1}^{m} \mathcal{Z}(\beta_i) = \prod_{i=1}^{m} \sum_{\sigma_i} e^{-\beta_i \mathcal{H}(\sigma_i)} \quad , \tag{90}$$

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 a parallel tempering simulation[88] which is a special case of the earlier replica exchange Monte Carlo method[85] proposed in the context of spin-glass simulations (to some extent the focus on this special application hides the general aspects of the method as becomes clearer in Ref. 86). After a certain number of sweeps, exchanges of the current configurations σ_i and σ_j are attempted (equivalently, the β_i may be exchanged, as is done in most implementations). Adapting the Metropolis criterion (16) to the present situation, the proposed exchange will be accepted with probability

$$w = \min\{1, \exp[(\beta_j - \beta_i)(E_j - E_i)]\} \tag{91}$$

where $E_i \equiv E(\sigma_i)$. To assure a reasonable acceptance rate, usually only "nearest-neighbor" exchanges ($j = i \pm 1$) are attempted and, as a first rough guess, the β_i could again be spaced by $\Delta\beta$ given in (89). By carefully monitoring the dynamics of the algorithm, recently much more refined prescriptions for the optimal choice of the simulation points β_i have been proposed.[89,90] In most applications, the smallest inverse temperature β_1 is chosen in the high-temperature phase where the autocorrelation time is expected to be very short and the system decorrelates rapidly. Conceptually this approach follows again the "avoiding rare events" strategy.

Notice that in parallel tempering no free-energy parameters have to be adjusted. The method is thus very robust and moreover can be almost trivially parallelized. For instance it it straightforward to implement this algorithm on a graphics card and perform "parallel tempering GPU computations".[91]

5.3. *Multicanonical ensembles*

To conclude this introduction to simulation techniques, at least a very brief outline of multicanonical ensembles[92,93] shall be given. For more details, in particular on practical implementations, see the earlier reviews[94–97] and the textbook by Berg.[4] Similarly to the tempering methods of the last section, multicanonical simulations may also be interpreted as a dynamical multi-histogram reweighting method. This interpretation is stressed by the notation used in the original papers by Berg and Neuhaus[92,93] and explains the name "*multi*canonical". At the same time, this method may also be viewed as a specific realization of non-Boltzmann sampling[98] which has been known since long to be a legitimate alternative to the more standard Monte Carlo approaches.[99] The practical significance of non-Boltzmann

sampling was first realized in the so-called "umbrella sampling" method,[100] but it took many years before the introduction of the multicanonical ensemble 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. By now the applications of the method range from physics and chemistry to biophysics, biochemistry and biology to engineering problems.

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[38,97] and the very rugged free-energy landscapes of spin glasses.[101–104] This allows a direct study of properties of the rare-event states such as, e.g., interface tensions or more generally free energy barriers, which would be very difficult (or practically impossible) with canonical simulations and also with the tempering methods described in Sects. 5.1 and 5.2.

In general the idea goes as follows. With σ representing generically the degrees of freedom (discrete spins or continuous field variables), the canonical Boltzmann distribution

$$\mathcal{P}_{\rm can}(\sigma) \propto e^{-\beta\mathcal{H}(\sigma)} \tag{92}$$

is replaced by an auxiliary multicanonical distribution

$$\mathcal{P}_{\rm muca}(\sigma) \propto W(Q(\sigma))e^{-\beta\mathcal{H}(\sigma)} \equiv e^{-\beta\mathcal{H}_{\rm muca}(\sigma)} \ , \tag{93}$$

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

$$\langle\mathcal{O}\rangle_{\rm can} = \langle\mathcal{O}W^{-1}(Q)\rangle_{\rm muca}/\langle W^{-1}(Q)\rangle_{\rm muca} \ , \tag{94}$$

similarly to Eq. (73). The goal is now to find a suitable weight factor W such that the dynamics of the multicanonical simulation profits most.

To be specific, let us assume in the following that the relevant macroscopic observable is the energy E itself. This is for instance the case at a temperature driven

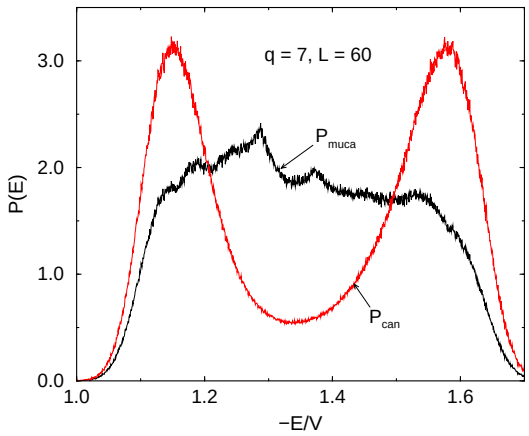

Fig. 10. The canonical energy density $P_{\rm can}(E)$ of the 2D 7-state Potts model on a 60×60 lattice at inverse temperature $\beta_{{\rm eqh},L}$, where the two peaks are of equal height, together with the multicanonical energy density $P_{\rm muca}(E)$, which is approximately constant between the two peaks.

first-order phase transition, where the canonical energy distribution $P_{\rm can}(E)$ develops a characteristic double-peak structure.[38] As an illustration, simulation data for the 2D 7-state Potts model[105] are shown in Fig. 10. With increasing system size, the region between the two peaks becomes more and more suppressed by the interfacial Boltzmann factor $\propto \exp(-2\sigma_{od}L^{D-1})$, where σ_{od} is the (reduced) interface tension, L^{D-1} the cross-section of a D-dimensional system, and the factor 2 accounts for the fact that with the usually employed periodic boundary condition at least two interfaces are present due to topological reasons. The time needed to cross this strongly suppressed rare-event two-phase region thus grows *exponentially* with the system size L, i.e., the autocorrelation time scales as $\tau \propto \exp(+2\sigma_{od}L^{D-1})$. In the literature, this is sometimes termed "supercritical slowing down" (even though nothing is "critical" here). Given such a situation, one usually adjusts $W = W(E)$ such that the multicanonical distribution $P_{\rm muca}(E)$ is approximately constant between the two peaks of $P_{\rm can}(E)$, thus aiming at a random-walk (pseudo-) dynamics of the Monte Carlo process,[106,107] cf. Fig. 10.

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

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

A more sophisticated recursion, in which the updated weight factor, or more conveniently the ratio $R(E) = W(E + \Delta E)/W(E)$, is computed from *all* available data accumulated so far, works as follows:[97,108–110]

1. Perform a simulation with $R_n(E)$ to obtain the nth histogram $H_n(E)$.
2. Compute the statistical weight of the nth run:
$$p(E) = H_n(E)H_n(E + \Delta E)/[H_n(E) + H_n(E + \Delta E)] \ . \tag{95}$$
3. Accumulate statistics:
$$p_{n+1}(E) = p_n(E) + p(E) \ , \tag{96}$$
$$\kappa(E) = p(E)/p_{n+1}(E) \ . \tag{97}$$
4. Update weight ratios:
$$R_{n+1}(E) = R_n(E) \left[H_n(E)/H_n(E + \Delta E)\right]^{\kappa(E)} \ . \tag{98}$$

   Goto 1.

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

In most applications local update algorithms have been employed, but for certain classes of models also non-local multigrid methods[34,35,111] are applicable.[68,112] 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[113–115] has been developed. For a recent application to improved finite-size scaling studies of second-order phase transitions, see Ref. 116. If $P_{\rm muca}$ was completely flat and the Monte Carlo update moves would perform an ideal random walk, one would expect that after V^2 local updates the system has travelled on average a distance V in total energy. Since one lattice sweep consists of V local updates, the autocorrelation time should scale in this idealized picture as $\tau \propto V$. Numerical tests for various models with a first-order phase transition have shown that in practice the data are at best consistent with a 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,[113–115] employing local update algorithms, typical fit results are $\alpha \approx 1.1 - 1.3$, and due to the limited accuracy of the data even a weak exponential growth law cannot be excluded.

In fact, at least for the field-driven first-order transition of the 2D Ising model below T_c, where one works with the magnetization instead of the energy (sometimes called "multimagnetical" simulations), it has been demonstrated recently[117] that even for a perfectly flat multicanonical distribution there are two "hidden" free energy barriers (in directions "orthogonal" to the magnetization) which lead to an exponential growth of τ with lattice size, which is albeit much weaker than the leading "supercritical slowing down" of the canonical simulation. Physically the two barriers are related to the nucleation of a large droplet of the "wrong phase" (say "$-$" spins in the background of "$+$" spins)[118–123] and the transition of this large, more or less spherical droplet to the strip phase (coexisting strips of "$-$" and "$+$" spins, separated by two straight interfaces) around $m = 0$.[124]

5.4. *Wang-Landau method*

Another more recently proposed method deals directly with estimators $\Omega(E)$ of the density of states.[125] By flipping spins randomly, the transition probability from energy level E_1 to E_2 is

$$w(E_1 \to E_2) = \min\left[1, \frac{\Omega(E_1)}{\Omega(E_2)}\right] \quad . \tag{99}$$

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

$$\Omega(E) \to f\,\Omega(E) \quad , \tag{100}$$

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

$$f_{n+1} = \sqrt{f_n}\,, \qquad n = 0, 1, \ldots \quad , \tag{101}$$

and the energy histogram reset to zero until some small value such as $f = e^{10^{-8}} \approx 1.00000001$ is reached.

For the 2D Ising model this procedure converges very rapidly towards the exactly known density of states, and also for other applications a fast convergence has been reported. Since the procedure violates the Markovian requirement and hence does not satisfy the balance condition (7), some care is necessary in setting up a proper protocol for the recursion (this is similar in spirit to the automatic updating of the optimal step size $S_{\max}$ in the Metropolis update algorithm discussed in Sect. 2.3.1). Most authors who employ the obtained density of states directly

to extract canonical expectation values by standard reweighting argue that, once f is close enough to unity, systematic deviations become negligible. While this claim can be verified empirically for the 2D Ising model (where exact results are available for judgement), possible systematic deviations are difficult to assess in the general case. A safe way would be to consider the recursion (99)–(101) as an alternative method to determine the multicanonical weights, and then to perform a usual multicanonical simulation employing these *fixed* weights. As emphasized earlier, any deviations of multicanonical weights from their optimal shape do not show up in the final canonical expectation values; they rather only influence the dynamics of the multicanonical simulations.

6. Scaling Analyses

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

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

$$C = \frac{\mathrm{d}u}{\mathrm{d}T} = \beta^2 \left(\langle E^2\rangle - \langle E\rangle^2\right)/V = \beta^2 V \left(\langle e^2\rangle - \langle e\rangle^2\right) \quad , \qquad (102)$$

where we have set $\mathcal{H} \equiv E = eV$ with V denoting the number of lattice sites, i.e., the "lattice volume". In simulations one usually employs the variance definition (since any discretized numerical differentiation would introduce some systematic error). The magnetization per site $m = M/V$ and the susceptibility χ are defined

as[h]

$$m = \langle |\mu| \rangle \ , \qquad \mu = \frac{1}{V} \sum_i \sigma_i \ , \tag{103}$$

and

$$\chi = \beta V \left(\langle \mu^2 \rangle - \langle |\mu| \rangle^2 \right) \ . \tag{104}$$

In the disordered phase for $T > T_c$, where $m = \langle \mu \rangle = 0$ by symmetry, one often works with the definition

$$\chi' = \beta V \langle \mu^2 \rangle \ . \tag{105}$$

The correlation between spins σ_i and σ_j at sites labeled by i and j can be measured by considering correlation functions like the two-point spin-spin correlation

$$G(\vec{r}) = G(i,j) = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \ , \tag{106}$$

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

$$G(\vec{r}) \sim |\vec{r}|^{-\kappa} e^{-|\vec{r}|/\xi} \ , \tag{107}$$

where ξ is the spatial correlation length and the exponent κ of the power-law prefactor depends in general on the dimension and on whether one studies the ordered or disordered phase. Strictly speaking ξ depends on the direction of $\vec{r}$.

6.1. *Critical exponents and scaling relations*

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

[h]Notice that here and in the following formulas, $|\mu|$ is used instead of μ as would follow from the formal definition of the zero-field magnetization $m(\beta) = (1/V\beta) \lim_{h\to 0} \partial \ln \mathcal{Z}(\beta,h)/\partial h$. The reason is that for a symmetric model on finite lattices one obtains $\langle \mu \rangle(\beta) = 0$ for all temperatures due to symmetry. Only in the proper infinite-volume limit, that is $\lim_{h\to 0} \lim_{V\to\infty}$, spontaneous symmetry breaking can occur below T_c. In a simulation on finite lattices, this is reflected by a symmetric double-peak structure of the magnetization distribution (provided the runs are long enough). By averaging μ one thus gets zero by symmetry, while the peak locations $\pm m_0(L)$ are close to the spontaneous magnetization so that the average of $|\mu|$ is a good estimator. Things become more involved for slightly asymmetric models, where this recipe would produce a systematic error and thus cannot be employed. For strongly asymmetric models, on the other hand, one peak clearly dominates and the average of μ can usually be measured without too many problems.

in thermodynamic functions. The leading divergence of the correlation length is usually parameterized in the high-temperature phase as

$$\xi = \xi_{0^+} |1 - T/T_c|^{-\nu} + \dots \quad (T \geq T_c) \ , \tag{108}$$

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

$$\xi = \xi_{0^-} (1 - T/T_c)^{-\nu} + \dots \quad (T \leq T_c) \ , \tag{109}$$

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

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,

$$C = C_{\rm reg} + C_0 |1 - T/T_c|^{-\alpha} + \dots \ , \tag{110}$$

$$m = m_0 (1 - T/T_c)^{\beta} + \dots \ , \tag{111}$$

$$\chi = \chi_0 |1 - T/T_c|^{-\gamma} + \dots \ , \tag{112}$$

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

$$m \propto h^{1/\delta} \qquad (T = T_c) \ , \tag{113}$$

$$G(\vec{r}) \propto r^{-D+2-\eta} \qquad (T = T_c) \ . \tag{114}$$

An important consequence of the divergence of the correlation length is that qualitative properties of second-order phase transitions should not depend on short-distance details of the Hamiltonian. This is the basis of the *universality* hypothesis[126] which means that all (short-ranged) systems with the same symmetries and same dimensionality should exhibit similar singularities governed by one and the same set of critical exponents. For the amplitudes this is not true, but certain amplitude ratios such as ξ_{0^+}/ξ_{0^-} or χ_{0^+}/χ_{0^-} are also universal.

In the 1960s, Rushbrooke,[127] Griffiths,[128] Josephson,[129] and Fisher[130] showed that the six critical exponents defined above are related via four inequalities. Subsequent experimental evidence indicated that these scaling relations were in fact equalities which are now firmly established by renormalization group (RG) con-

siderations and fundamentally important in the theory of critical phenomena:

$$2\beta + \gamma = 2 - \alpha \quad \text{(Rushbrooke's law)} \ , \tag{115}$$

$$\beta(\delta - 1) = \gamma \quad \text{(Griffiths' law)} \ , \tag{116}$$

$$\nu(2 - \eta) = \gamma \quad \text{(Fisher's law)} \ . \tag{117}$$

The fourth equality involves the dimension D. It is therefore a (somewhat weaker) so-called hyperscaling relation:

$$D\nu = 2 - \alpha \quad \text{(Josephson's law)} \ . \tag{118}$$

In the conventional scaling scenario, Rushbrooke's and Griffiths' laws can be deduced from the Widom scaling hypothesis that the Helmholtz free energy is a homogeneous function.[131] Widom scaling and the remaining two laws can in turn be derived from the Kadanoff block-spin construction[132] and ultimately from RG considerations.[133] Josephson's law can also be derived from the hyperscaling hypothesis, namely that the free-energy density behaves near criticality as the inverse correlation volume: $f \sim \xi^{-D}$. Twice differentiating this relation and inserting the scaling law (110) for the specific heat gives immediately (118).

The paradigm model for systems exhibiting a continuous (or, roughly speaking, second-order) phase transition is the Ising model. When the temperature is varied the system passes at T_c from an ordered low-temperature to a disordered high-temperature phase. In two dimensions (2D), the thermodynamic limit of this model in zero external field has been solved exactly by Onsager,[134] and even for finite $L_x \times L_y$ lattices the exact partition function is known.[73,74] Also the exact density of states can be calculated by means of computer algebra up to reasonably large lattice sizes.[75] This provides a very useful testing ground for any new algorithmic idea in computer simulations. For infinite lattices, the correlation length has been calculated in arbitrary lattice directions.[135,136] The exact magnetization for $h = 0$, apparently already known to Onsager,[137] was first derived by Yang[138] and later generalized by Chang.[139] The only quantity which up to date is not truly exactly known is the susceptibility. However, its properties have been characterized to very high precision[140–142] (for both, low- and high-temperature series expansions, 2000 terms are known exactly[141]). In three dimensions (3D) *no* exact solutions are available, but analytical and numerical results from various methods give a consistent and very precise picture. In four dimensions (4D) the so-called upper critical dimension D_u is reached and for $D \geq D_u = 4$ the critical exponents take their mean-field values (in 4D up to multiplicative logarithmic corrections[143]). The critical exponents of the Ising model are collected in Table 1.[144–146]

Table 1. Critical exponents of the Ising model. All 2D exponents are exactly known.[144,145] For the 3D Ising model the "world-average" for ν and γ calculated in Ref. 146 is quoted. The other exponents follow from hyperscaling ($\alpha = 2 - D\nu$) and scaling ($\beta = (2 - \alpha - \gamma)/2$, $\delta = \gamma/\beta + 1$, $\eta = 2 - \gamma/\nu$) relations. For all $D \geq D_u = 4$ the mean-field exponents are valid (in 4D up to multiplicative logarithmic corrections).

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

6.2. *Finite-size scaling (FSS)*

In computer simulation studies, the (linear) system size L is always necessarily finite. The correlation length may hence become large (of the order of L) but never diverges in a mathematical sense. For the divergences in other quantities this implies that they are also rounded and shifted.[11,147–149] How this happens is described by finite-size scaling (FSS) theory, which in a nut-shell may be explained as follows: Near T_c the role of ξ is taken over by the linear size L of the system. By rewriting (108) or (109) and replacing ξ by L, it is easy to see that

$$|1 - T/T_c| \propto \xi^{-1/\nu} \longrightarrow L^{-1/\nu} \ . \tag{119}$$

It follows that the scaling laws (110)–(112) have to be replaced by the *finite-size scaling* (FSS) ansatz,

$$C = C_{\rm reg} + aL^{\alpha/\nu} + \dots \ , \tag{120}$$

$$m \propto L^{-\beta/\nu} + \dots \ , \tag{121}$$

$$\chi \propto L^{\gamma/\nu} + \dots \ , \tag{122}$$

where $C_{\rm reg}$ is a regular, smooth background term and a a constant. As a mnemonic rule, a critical exponent x in a temperature scaling law is replaced by $-x/\nu$ in the corresponding FSS law. This describes the rounding of the singularities quantitatively.

In general these scaling laws are valid in a vicinity of T_c as long as the scaling variable

$$x = (1 - T/T_c)L^{1/\nu} \tag{123}$$

is kept fixed.[11,147–149] In this more general formulation the scaling law for, e.g., the susceptibility reads

$$\chi(T, L) = L^{\gamma/\nu} f(x) \ , \tag{124}$$

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

For given L the maximum of $\chi(T,L)$ as a function of temperature happens at some $x_{\max}$. For the location $T_{\max}$ of the maximum this implies a FSS behaviour of the form

$$T_{\max} = T_c(1 - x_{\max}L^{-1/\nu} + \dots) = T_c + cL^{-1/\nu} + \dots \ . \tag{125}$$

This quantifies the shift of so-called pseudo-critical points which depends on the observables considered. Only in the thermodynamic limit $L \to \infty$ all quantities diverge at the same temperature T_c.

Further useful quantities in FSS analyses are the energetic fourth-order parameter

$$V(\beta) = 1 - \frac{\langle e^4\rangle}{3\langle e^2\rangle^2} \ , \tag{126}$$

the magnetic cumulants (Binder parameters)

$$U_2(\beta) = 1 - \frac{\langle \mu^2\rangle}{3\langle |\mu|\rangle^2} \ , \tag{127}$$

$$U_4(\beta) = 1 - \frac{\langle \mu^4\rangle}{3\langle \mu^2\rangle^2} \ , \tag{128}$$

and their slopes

$$\begin{aligned}\frac{\mathrm{d}U_2(\beta)}{\mathrm{d}\beta} &= \frac{V}{3\langle|\mu|\rangle^2}\left[\langle\mu^2\rangle\langle e\rangle - 2\frac{\langle\mu^2\rangle\langle|\mu|e\rangle}{\langle|\mu|\rangle} + \langle\mu^2 e\rangle\right] \\ &= V(1-U_2)\left[\langle e\rangle - 2\frac{\langle|\mu|e\rangle}{\langle|\mu|\rangle} + \frac{\langle\mu^2 e\rangle}{\langle\mu^2\rangle}\right] \ , \end{aligned} \tag{129}$$

$$\frac{\mathrm{d}U_4(\beta)}{\mathrm{d}\beta} = V(1-U_4)\left[\langle e\rangle - 2\frac{\langle\mu^2 e\rangle}{\langle\mu^2\rangle} + \frac{\langle\mu^4 e\rangle}{\langle\mu^4\rangle}\right] \ . \tag{130}$$

The Binder parameters scale according to

$$U_{2p} = f_{U_{2p}}(x)[1 + \dots] \ , \tag{131}$$

i.e., for constant scaling variable x, U_{2p} takes approximately the same value for all lattice sizes, in particular $U_{2p}^* \equiv f_{U_{2p}}(0)$ at T_c. Applying the differentiation to this scaling representation, one picks up a factor of $L^{1/\nu}$ from the scaling function,

$$\frac{\mathrm{d}U_{2p}}{\mathrm{d}\beta} = (\mathrm{d}x/\mathrm{d}\beta)f'_{U_{2p}}[1 + \dots] = L^{1/\nu} f_{U'_{2p}}(x)[1 + \dots] \ . \tag{132}$$

As a function of temperature the Binder parameters for different L hence cross around (T_c, U_{2p}^*) with slopes $\propto L^{1/\nu}$, apart from corrections-to-scaling collected in $[1 + \ldots]$ explaining small systematic deviations. From a determination of this crossing point, one thus obtains a basically unbiased estimate of T_c, the critical exponent ν, and U_{2p}^*. Note that in contrast to the truly universal critical exponents, U_{2p}^* is only *weakly* universal. By this one means that the infinite-volume limit of such quantities does depend in particular on the boundary conditions and geometrical shape of the considered lattice, e.g., on the aspect ratio $r = L_y/L_x$.[150–157]

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

$$\frac{\mathrm{d}\langle|\mu|\rangle}{\mathrm{d}\beta} = V\left(\langle|\mu|e\rangle - \langle|\mu|\rangle\langle e\rangle\right) \ , \tag{133}$$

$$\frac{\mathrm{d}\ln\langle|\mu|\rangle}{\mathrm{d}\beta} = V\left(\frac{\langle|\mu|e\rangle}{\langle|\mu|\rangle} - \langle e\rangle\right) \ , \tag{134}$$

$$\frac{\mathrm{d}\ln\langle\mu^2\rangle}{\mathrm{d}\beta} = V\left(\frac{\langle\mu^2 e\rangle}{\langle\mu^2\rangle} - \langle e\rangle\right) \ . \tag{135}$$

These latter five quantities are good examples for expectation values depending on both e and μ. By applying the differentiation to the scaling form of $\langle|\mu|\rangle$, one reads off that

$$\frac{\mathrm{d}\langle|\mu|\rangle}{\mathrm{d}\beta} = L^{(1-\beta)/\nu} f_{\mu'}(x)[1 + \ldots] \ , \tag{136}$$

$$\frac{\mathrm{d}\ln\langle|\mu|^p\rangle}{\mathrm{d}\beta} = L^{1/\nu} f_{\mathrm{d}\mu^p}(x)[1 + \ldots] \ . \tag{137}$$

For first-order phase transitions similar considerations show[37,38,158–160] that there the delta function like singularities in the thermodynamic limit, originating from phase coexistence, are smeared out for finite systems as well.[161–165] They are replaced by narrow peaks whose height grows proportional to the volume $V = L^D$, analogously to (120) or (122), with a peak width decreasing as $1/V$ and a shift of the peak location from the infinite-volume transition temperature proportional to $1/V$, analogously to (125).[37,38,166–170]

6.3. *Organisation of the analysis*

To facilitate most flexibility in the analysis, it is advisable to store during data production the time series of measurements. Standard quantities are the energy and magnetization, but depending on the model at hand it may be useful to record also other observables. In this way the full dynamical information can be extracted still after the actual simulation runs and error estimation can be easily performed.

For example it is no problem to experiment with the size and number of Jackknife bins. Since a reasonable choice depends on the *a priori* unknown autocorrelation time, it is quite cumbersome to do a reliable error analysis "on the flight" during the simulation. Furthermore, basing data reweighting on time-series data is more efficient since histograms, if needed or more convenient, can still be produced from this data but working in the reverse direction is obviously impossible.

For some models it is sufficient to perform for each lattice size a single long run at some coupling β_0 close to the critical point β_c. This is, however, not always the case and also depends on the observables of interest. In this more general case, one may use several simulation points β_i and combine the results by the multi-histogram reweighting method or may apply a recently developed finite-size adapted generalized ensemble method.[116,171] In both situations, one can compute the relevant quantities from the time series of the energies $e = E/V$ (if E happens to be integer valued, this should be stored of course) and $\mu = \sum_i \sigma_i/V$ by reweighting.

By using one of these techniques one first determines the temperature dependence of $C(\beta)$, $\chi(\beta)$, ..., in the neighborhood of the simulation point $\beta_0 \approx \beta_c$ (a reasonably "good" initial guess for β_0 is usually straightforward to obtain). Once the temperature dependence is known, one can determine the maxima, e.g., $C_{\max}(\beta_{\max_C}) \equiv \max_\beta C(\beta)$, by applying standard extremization routines: When reweighting is implemented as a subroutine, for instance $C(\beta)$ can be handled as a normal function with a continuously varying argument β, i.e., no interpolation or discretization error is involved when iterating towards the maximum. The locations of the maxima of C, χ, $\mathrm{d}U_2/\mathrm{d}\beta$, $\mathrm{d}U_4/\mathrm{d}\beta$, $\mathrm{d}\langle|\mu|\rangle/\mathrm{d}\beta$, $\mathrm{d}\ln\langle|\mu|\rangle/\mathrm{d}\beta$, and $\mathrm{d}\ln\langle\mu^2\rangle/\mathrm{d}\beta$ provide us with seven sequences of pseudo-transition points $\beta_{\max_i}(L)$ which all should scale according to $\beta_{\max_i}(L) = \beta_c + a_i L^{-1/\nu} + \ldots$. In other words, the scaling variable $x = (\beta_{\max_i}(L) - \beta_c)L^{1/\nu} = a_i + \ldots$ should be constant, if we neglect the small higher-order corrections indicated by ....

Notice that while the precise estimates of a_i do depend on the value of ν, the qualitative conclusion that $x \approx \mathrm{const}$ for each of the $\beta_{\max_i}(L)$ sequences does not require any *a priori* knowledge of ν or β_c. Using this information one thus has several possibilities to extract unbiased estimates of the critical exponents ν, α/ν, β/ν, and γ/ν from least-squares fits assuming the FSS behaviours (120), (121), (122), (132), (136), and (137).

Considering only the asymptotic behaviour, e.g., $\mathrm{d}\ln\langle|\mu|\rangle/\mathrm{d}\beta = aL^{1/\nu}$, and taking the logarithm, $\ln(\mathrm{d}\ln\langle|\mu|\rangle/\mathrm{d}\beta) = c + (1/\nu)\ln(L)$, one ends up with a linear two-parameter fit yielding estimates for the constant $c = \ln(a)$ and the exponent $1/\nu$. For small lattice sizes the asymptotic ansatz is, of course, not justified. Taking into account the (effective) correction term $[1 + bL^{-w}]$ would

result in a *non-linear* four-parameter fit for a, b, $1/\nu$ and w. Even if we would fix w to some "theoretically expected" value (as is sometimes done), we would still be left with a *non-linear* fit which is usually much harder to control than a linear fit (where only a set of linear equations with a unique solution has to be solved, whereas a non-linear fit involves a numerical minimization of the χ^2-function, possessing possibly several local minima). The alternative method is to use the linear fit ansatz and to discard successively more and more small lattice sizes until the χ^2 per degree-of-freedom or the goodness-of-fit parameter[61] Q has reached an acceptable value and does not show any further trend. Of course, all this relies heavily on correct estimates of the statistical error bars on the original data for $\mathrm{d}\ln\langle|\mu|\rangle/\mathrm{d}\beta$.

Once ν is estimated one can use the scaling form $\beta_{\max_i}(L) = \beta_c + a_i L^{-1/\nu} + \ldots$ to extract β_c and a_i. As a useful check, one should repeat these fits at the error margins of ν, but usually this dependence turns out to be very weak. As a useful cross-check one can determine β_c also from the Binder parameter crossings, which is the most convenient and fastest method for a first rough estimate. As a rule of thumb, an accuracy of about $3-4$ digits for β_c can be obtained with this method without any elaborate infinite-volume extrapolations – the crossing points lie usually much closer to β_c than the various maxima locations. For high precision, however, it is quite cumbersome to control the necessary extrapolations and often more accurate estimates can be obtained by considering the scaling of the maxima locations. Also, error estimates of crossing points involve the data for two different lattice sizes which tends to be quite unhandy.

Next, similarly to ν, the ratios of critical exponents α/ν, β/ν, and γ/ν can be obtained from fits to (120), (121), (122), and (136). Again the maxima of these quantities or any of the FSS sequences $\beta_{\max_i}$ can be used. What concerns the fitting procedure the same remarks apply as for ν. The specific heat C usually plays a special role in that the exponent α is difficult to determine. The reason is that α is usually relatively small (3D Ising model: $\alpha \approx 0.1$), may be zero (logarithmic divergence as in the 2D Ising model) or even negative (as for instance in the 3D XY and Heisenberg models). In all these cases, the constant background contribution C_{reg} in (120) becomes important, which enforces a non-linear three-parameter fit with the just described problems. Also for the susceptibility χ, a regular background term cannot be excluded, but it is usually much less important since $\gamma \gg \alpha$. Therefore, in (121), (122), and (136), similar to the fits for ν, one may take the logarithm and deal with much more stable linear fits.

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

into the thus determined exponent estimates. In (the pretty rare) cases where β_c is known exactly (e.g., through self-duality), this latter option is by far the most accurate one. This is the reason, why for such models numerically estimated critical exponents are usually quite precise.

When combining the various fit results for, e.g. β_c or ν, to a final average value, some care is necessary with the optimal weighted average and the final statistical error estimate, since the various fits for determining β_c or ν are of course correlated (since they all use the data from one and the same simulation). In principle this can be dealt with by applying a cross-correlation analysis.[172]

7. Applications

7.1. *Disordered ferromagnets*

Experiments on phase transitions in magnetic materials are usually subject to randomly distributed impurities. At continuous phase transitions, depending on the temperature resolution and the concentration of the impurities, the disorder may significantly influence measurements of critical exponents.[173] To emphasize this effect, in some experiments[174] non-magnetic impurities are introduced in a controlled way; see Fig. 11 for an example. Since the mobility of impurities is usually much smaller than the typical time scale of spin fluctuations, one may model the disorder effects in a completely "frozen", so-called "quenched" approximation. This limit is opposite to "annealed" disorder which refers to the case where the two relevant time scales are of the same order.

With the additional assumption that the quenched, randomly distributed impurities are completely uncorrelated, Harris[175] showed a long time ago under which conditions a *continuous* transition of an idealised pure material is modified by disorder coupling to the energy of the system. According to this so-called Harris criterion, the critical behaviour of the pure system around the transition temperature T_c is stable against quenched disorder when the critical exponent $\alpha_{\rm pure}$ of the specific heat, $C \propto |T-T_c|^{-\alpha_{\rm pure}}$, is negative. In renormalization-group language the perturbation is then "irrelevant" and the values of all critical exponents $\alpha, \beta, \gamma, \ldots$ remain unchanged. On the other hand, when $\alpha_{\rm pure} > 0$, then quenched disorder should be "relevant" and the renormalization-group flow approaches a new disorder fixed point governed by altered critical exponents. An example is the three-dimensional (3D) Ising model universality class with $\alpha_{\rm pure} \approx 0.110 > 0$. The intermediate situation $\alpha_{\rm pure} = 0$ is a special, "marginal" case where no easy predictions can be made. A typical example for the latter situation is the two-dimensional (2D) Ising model where quenched disorder is known to generate log-

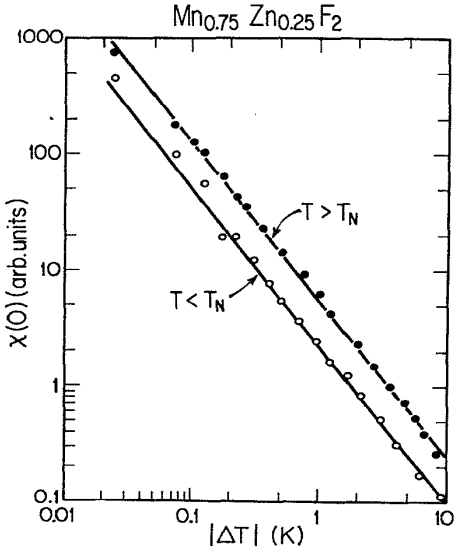

Fig. 11. Neutron scattering measurements of the susceptibility in $\mathrm{Mn}_{0.75}\mathrm{Zn}_{0.25}\mathrm{F}_2$ close to criticality, governed by the disorder fixed point of the Ising model, over the reduced temperature interval $4 \times 10^{-4} < |T/T_c - 1| < 2 \times 10^{-1}$. The solid lines show power-law fits with exponent $\gamma = 1.364(76)$ above and below T_c [after Mitchell *et al.* (Ref. 174)].

arithmic modifications.[176]

Figure 11 shows an experimental verification of the qualitative influence of disorder for a three-dimensional Ising-like system where the measured critical exponent $\gamma = 1.364(76)$ of the susceptibility $\chi \propto |T - T_c|^{-\gamma}$ is clearly different from that of the pure 3D Ising model, $\gamma_{\mathrm{pure}} = 1.2396(13)$. Theoretical results, on the other hand, remained relatively scarce in 3D until recently. Most analytical renormalization group and computer simulation studies focused on the Ising model,[177,178] usually assuming *site* dilution when working numerically. This motivated us to consider the case of *bond* dilution[179–181] which enables one to test the expected universality with respect to the type of disorder distribution and, in addition, facilitates a quantitative comparison with recent high-temperature series expansions.[182–184]

The Hamiltonian (in a Potts model normalisation) is given as

$$-\beta\mathcal{H} = \sum_{\langle i,j \rangle} K_{ij}\delta_{\sigma_i,\sigma_j} \quad , \tag{138}$$

where the spins take the values $\sigma_i = \pm 1$ and the sum goes over all nearest-neighbor pairs $\langle i, j \rangle$. The coupling strengths K_{ij} are drawn from the bimodal

distribution

$$\wp[K_{ij}] = \prod_{\langle i,j\rangle} P(K_{ij}) = \prod_{\langle i,j\rangle} [p\delta(K_{ij}-K) + (1-p)\delta(K_{ij}-RK)] \ . \quad (139)$$

Besides bond dilution ($R = 0$), which we will consider here, this also includes random-bond ferromagnets ($0 < R < 1$) and the physically very different class of spin glasses ($R = -1$) as special cases. For the case of bond dilution, the couplings are thus allowed to take two different values $K_{ij} = K \equiv J\beta \equiv J/k_BT$ and 0 with probabilities p and $1-p$, respectively, with $c = 1-p$ being the concentration of missing bonds, which play the role of the non-magnetic impurities. The pure case thus corresponds to $p = 1$. Below the bond-percolation threshold[185] $p_c = 0.248\,812\,6(5)$ one does not expect any finite-temperature phase transition since without a percolating (infinite) cluster of spins long-range order cannot develop.

The model (138), (139) with $R = 0$ was studied by means of large-scale Monte Carlo simulations using the Swendsen-Wang (SW) cluster algorithm[39] (which in the strongly diluted case is better suited than the single-cluster Wolff variant). To arrive at final results in the quenched case, for each dilution, temperature and lattice size, the Monte Carlo estimates for $\langle Q_{\{J\}}\rangle$ of thermodynamic quantities $Q_{\{J\}}$ for a given random distribution $\{J\}$ of diluted bonds (realized as usual by averages over the time series of measurements) have to be averaged over many different disorder realisations,

$$Q \equiv [\langle Q_{\{J\}}\rangle]_{\rm av} = \frac{1}{\#\{J\}} \sum_{\{J\}} \langle Q_{\{J\}}\rangle \ , \quad (140)$$

where $\#\{J\}$ is the number of realisations considered. Denoting the empirically determined distribution of $\langle Q_{\{J\}}\rangle$ by $\mathcal{P}(\langle Q_{\{J\}}\rangle)$, this so-called quenched average can also be obtained from

$$Q = \int \mathcal{D}J_{ij}\wp(J_{ij})\langle Q_{\{J\}}\rangle = \int {\rm d}\langle Q_{\{J\}}\rangle \mathcal{P}(\langle Q_{\{J\}}\rangle)\langle Q_{\{J\}}\rangle \ , \quad (141)$$

where a discretized evaluation of the integrals for finite $\#\{J\}$ is implicitly implied. While conceptually straightforward, the quenched average in (140) is computationally very demanding since the number of realisations $\#\{J\}$ usually must be large, often of the order of a few thousands. In fact, if this number is chosen too small one may observe *typical* rather than average values[186] which may differ significantly when the distribution $\mathcal{P}(\langle Q_{\{J\}}\rangle)$ exhibits a long tail (which in general is hard to predict beforehand).

To get a rough overview of the phase diagram we first studied the dependence of the susceptibility peaks on the dilution, where the susceptibility $\chi =$

$KV(\langle\mu^2\rangle - \langle|\mu|\rangle^2)$ with $\mu = (1/V)\sum_i \sigma_i$ is defined as usual. To this end we performed for $p = 0.95, 0.90, \ldots, 0.36$ and moderate system sizes SW cluster MC simulations with $N_{\rm MCS} = 2\,500$ MC sweeps (MCS) each. By performing quite elaborate analyses of autocorrelation times, this statistics was judged to be reasonable ($N_{\rm MCS} > 250\ \tau_e$). By applying single-histogram reweighting to the data for each of the $2\,500 - 5\,000$ disorder realisation and then averaging the resulting $\chi(K)$ curves, we finally arrived at the data shown in Fig. 12.

From the locations of the maxima one obtains the phase diagram of the model in the $p-T$ plane shown in Fig. 13 which turned out to be in excellent agreement with a "single-bond effective-medium" (EM) approximation,[187]

$$K_c^{\rm EM}(p) = \ln\left[\frac{(1-p_c)e^{K_c(1)} - (1-p)}{p-p_c}\right] \quad , \tag{142}$$

where $K_c(1) = J/k_BT_c(1) = 0.443\,308\,8(6)$ is the precisely known transition point of the pure 3D Ising model.[188] As an independent confirmation of (142), the phase diagram also coincides extremely well with recent results from high-temperature series expansions.[184]

The quality of the disorder averages can be judged as in Fig. 14 by computing running averages over the disorder realisations taken into account and looking at

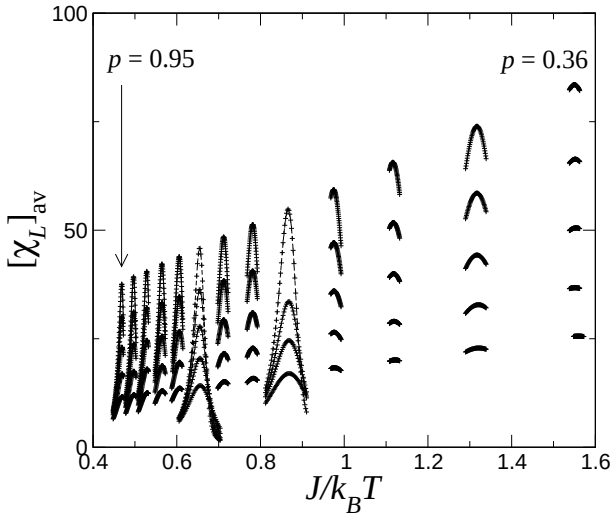

Fig. 12. The average magnetic susceptibility $[\chi_L]_{\rm av}$ of the 3D bond-diluted Ising model versus $K = J/k_BT$ for several concentrations p and $L = 8, 10, 12, 14, 16, 18$, and 20. For each value of p and each lattice size L, the curves are obtained by standard single-histogram reweighting of the simulation data at one value of K.

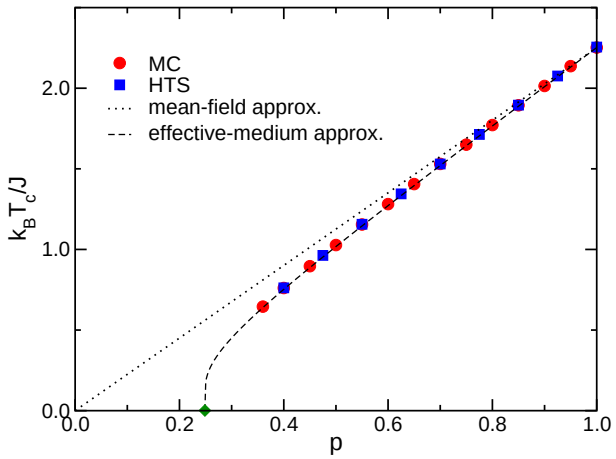

Fig. 13. Phase diagram of the bond-diluted Ising model on a three-dimensional simple cubic lattice in the dilution-temperature plane. The percolation point $p_c \approx 0.2488$ is marked by the diamond and $p = 1$ is the pure case without impurities. The results from the Monte Carlo (MC) simulations are compared with analyses of high-temperature series (HTS) expansions and with (properly normalized) mean-field and effective-medium approximations.

the distributions $\mathcal{P}(\chi_i)$. The plots show that the fluctuations in the running average disappear already after a few hundreds of realisations and that the dispersion of the χ_i values is moderate. The histogram also shows, however, that the distributions of physical observables typically do not become sharper with increasing system size at a finite-randomness disorder fixed point. Rather their relative widths stay constant, a phenomenon called non-self-averaging. More quantitatively, non-self-averaging can be checked by evaluating the normalized squared width $R_\chi(L) = V_\chi(L)/[\chi(L)]^2_{\rm av}$, where $V_\chi(L) = [\chi(L)^2]_{\rm av} - [\chi(L)]^2_{\rm av}$ is the variance of the susceptibility distribution. Figure 15 shows this ratio for three concentrations of the bond-diluted Ising model as a function of inverse lattice size. The fact that R_χ approaches a constant when L increases, as predicted by Aharony and Harris,[189] is the signature of a non-self-averaging system, in qualitative agreement with the results of Wiseman and Domany[190] for the site-diluted 3D Ising model.[i]

In order to study the critical behaviour in more detail, we concentrated on the three particular dilutions $p = 0.4$, 0.55, and 0.7. In a first set of simulations we focused on the FSS behaviour for lattice sizes up to $L = 96$. It is well known that *ratios* of critical exponents are almost equal for pure and disordered mod-

[i]Our estimate of R_χ is about an order of magnitude smaller since we worked with $\chi = KV(\langle\mu^2\rangle - \langle|\mu|\rangle^2)$ whereas in Ref. 190 the "high-temperature" expression $\chi' = KV\langle\mu^2\rangle$ was used.

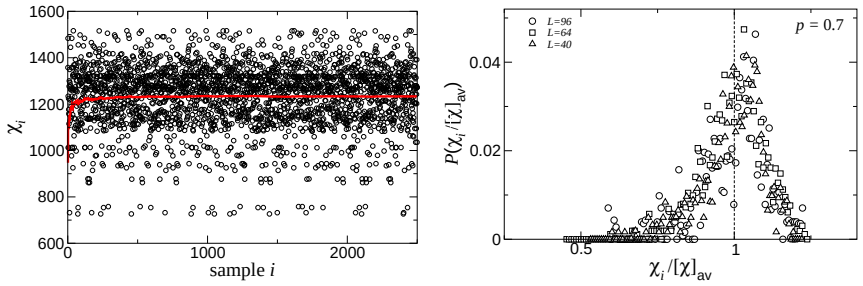

Fig. 14. *Left:* Susceptibility for the different disorder realisations of the three-dimensional bond-diluted Ising model for $L = 96$ and a concentration of magnetic bonds $p = 0.7$ at $K = 0.6535 \approx K_c(L)$. The running average over the samples is shown by the solid (red) line. *Right:* The resulting probability distribution of the susceptibility scaled by its quenched average $[\chi]_{\rm av}$, such that the results for the different lattice sizes $L = 40, 64$, and 96 collapse. The vertical dashed line indicates the average susceptibility $\chi_i/[\chi]_{\rm av} = 1$.

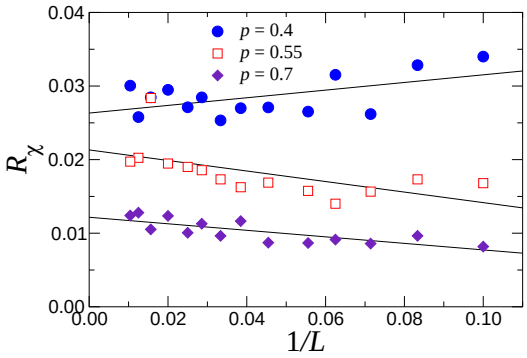

Fig. 15. Normalized squared width of the susceptibility distribution versus the inverse lattice size for the three concentrations $p = 0.4, 0.55$, and 0.7 at the effective critical coupling $K_c(L)$. The straight lines are linear fits used as guides to the eye.

els, e.g., $\gamma/\nu = 1.966(6)$ (pure[191]) and $\gamma/\nu = 1.963(5)$ (disordered[192]). The only distinguishing quantity is the correlation length exponent ν which can be extracted, e.g., from the derivative of the magnetisation versus inverse temperature, $\mathrm{d}\ln[m]_{\rm av}/\mathrm{d}K \propto L^{1/\nu}$, at K_c or the locations of the susceptibility maxima. Using the latter unbiased option and performing least-square fits including data from $L_{\rm min}$ to $L_{\rm max} = 96$ we obtained the effective critical exponents shown in Fig. 16. For the dilution closest to the pure model ($p = 0.7$), the system is influenced by the pure fixed point with $1/\nu = 1.5863(33)$. On the other hand, when the bond concentration is small ($p = 0.4$), the vicinity of the percolation fixed point where

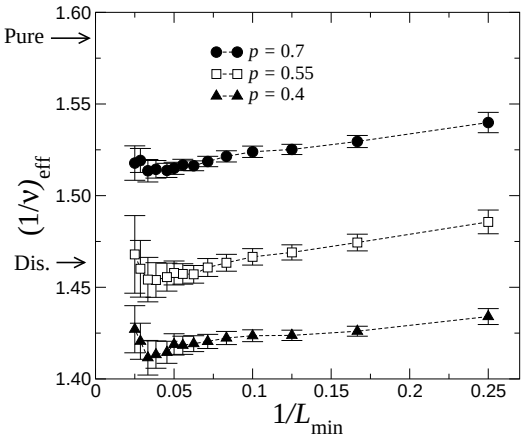

Fig. 16. Effective exponents $(1/\nu)_{\rm eff}$ as obtained from fits to the behaviour of $\mathrm{d}\ln[m]_{\rm av}/\mathrm{d}K \propto L^{1/\nu}$ as a function of $1/L_{\rm min}$ for $p = 0.4$, 0.55, and 0.7. The upper limit of the fit range is $L_{\rm max} = 96$.

$1/\nu \approx 1.12$ induces a decrease of $1/\nu$ below its expected disorder value. The dilution for which the cross-over effects are the least is around $p = 0.55$ which suggests that the scaling corrections should be rather small for this specific dilution.

The main problem of the FSS study is the competition between different fixed points (pure, disorder, percolation) in combination with corrections-to-scaling terms $\propto L^{-\omega}$, which we found hard to control for bond dilution. In contrast to recent claims for the site-diluted model that $\omega \approx 0.4$, we were not able to extract a reliable estimate of ω from our data for bond dilution.

In a second set of simulations we examined the temperature scaling of the magnetisation and susceptibility for lattice sizes up to $L = 40$. This data allows direct estimates of the exponents β and γ whose relative deviation from the pure model is comparable to that of ν, e.g. $\gamma = 1.2396(13)$ (pure[191]) and $\gamma = 1.342(10)$ (disordered[192]). As a function of the reduced temperature $\tau = (K_c - K)$ ($\tau < 0$ in the low-temperature (LT) phase and $\tau > 0$ in the high-temperature (HT) phase) and the system size L, the susceptibility is expected to scale as

$$[\chi(\tau, L)]_{\rm av} \sim |\tau|^{-\gamma} g_{\pm}(L^{1/\nu}|\tau|) \ , \tag{143}$$

where $g_{\pm}$ is a scaling function of the variable $x = L^{1/\nu}|\tau|$ and the subscript $\pm$ stands for the HT/LT phases. Assuming $[\chi(\tau)]_{\rm av} \propto |\tau|^{-\gamma_{\rm eff}}$ without any corrections-to-scaling terms, we can define a temperature dependent *effective* crit-

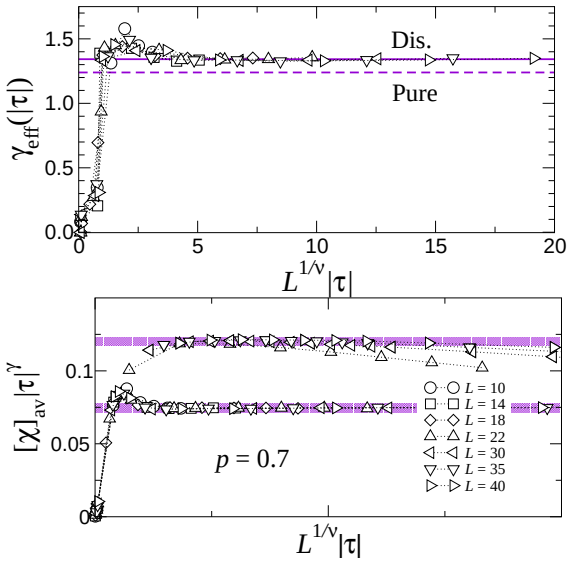

Fig. 17. *Top:* Variation of the temperature dependent effective critical exponent $\gamma_{\rm eff}(|\tau|) = -{\rm d}\ln[\chi]_{\rm av}/{\rm d}\ln|\tau|$ (in the low-temperature phase) as a function of the rescaled temperature $L^{1/\nu}|\tau|$ for the bond-diluted Ising model with $p = 0.7$ and several lattice sizes L. The horizontal solid and dashed lines indicate the site-diluted and pure values of γ, respectively. *Bottom:* The figure below shows the critical amplitudes $\Gamma_{\pm}$ above and below the critical temperature.

ical exponent $\gamma_{\rm eff}(|\tau|) = -{\rm d}\ln[\chi]_{\rm av}/{\rm d}\ln|\tau|$, which should converge towards the asymptotic critical exponent γ when $L \to \infty$ and $|\tau| \to 0$. Our results for $p = 0.7$ are shown in Fig. 17. For the greatest sizes, the effective exponent $\gamma_{\rm eff}(|\tau|)$ is stable around 1.34 when $|\tau|$ is not too small, i.e., when the finite-size effects are not too strong. The plot of $\gamma_{\rm eff}(|\tau|)$ vs. the rescaled variable $L^{1/\nu}|\tau|$ shows that the critical power-law behaviour holds in different temperature ranges for the different sizes studied. By analysing the temperature behaviour of the susceptibility, we also have directly extracted the power-law exponent γ using error weighted least-squares fits and choosing the temperature range that gives the smallest $\chi^2/{\rm d.o.f}$ for several system sizes. The results are consistent with $\gamma \approx 1.34 - 1.36$, cf. Table 2.

From the previous expression of the susceptibility as a function of the reduced temperature and size, it is instructive to plot the scaling function $g_{\pm}(x)$. For finite size and $|\tau| \neq 0$, the scaling functions may be Taylor expanded in powers of the inverse scaling variable $x^{-1} = (L^{1/\nu}|\tau|)^{-1}$, $[\chi_{\pm}(\tau, L)]_{\rm av} = |\tau|^{-\gamma}[g_{\pm}(\infty) + x^{-1}g'_{\pm}(\infty) + O(x^{-2})]$, where the amplitude $g_{\pm}(\infty)$ is usually denoted by $\Gamma_{\pm}$.

Table 2. Critical exponents and critical amplitude ratio of the susceptibility as measured with different techniques.

Technique	γ	Γ_+/Γ_-	ω	Ref.
Neutron scattering	1.44(6)	2.2	0.5	193[a]
	1.31(3)	2.8(2)		194,195[b]
	1.37(4)	2.40(2)		174[c]
RG		2.2		196
	1.318		0.39(4)	197,198[d]
	1.330(17)		0.25(10)	199[e]
MC	1.342(10)		0.37	192[f]
	1.34(1)	1.62(10)		200[g]
	1.342(7)			201[h]
	1.314(4)	1.67(15)		202[i]
HTS	1.305(5)			184[j]

[a] $Fe_{1-x}Zn_xF_2$, $x = 0.4$, 0.5, $|\tau| \sim 10^{-2}$.
[b] $Fe_{0.46}Zn_{0.54}F_2$, $1.5 \times 10^{-3} \leq |\tau| \leq 10^{-1}$.
[c] $ic5$ $Mn_{0.75}Zn_{0.25}F_2$, $4 \times 10^{-4} \leq |\tau| \leq 2 \times 10^{-1}$.
[d] 4 loop approximation.
[e] 6 loop approximation, fixed dimension.
[f] site dilution, $p = 0.4$ to 0.8.
[g] bond dilution, $p = 0.7$. The correction to scaling is too small to be determined.
[h] site dilution, $p = 0.8$. The observed correction to scaling could be the next-to-leading one.
[i] site dilution, $p = 0.8$.
[j] bond dilution, $p = 0.6$ to 0.7.

Multiplying by $|\tau|^\gamma$ leads to

$$[\chi_\pm(\tau, L)]_{\rm av}|\tau|^\gamma = g_\pm(x) = \Gamma_\pm + O(x^{-1}) \ . \tag{144}$$

When $|\tau| \to 0$ but with L still larger than the correlation length ξ, one should recover the critical behaviour given by $g_\pm(x) = O(1)$. The critical amplitudes $\Gamma_\pm$ follow, as shown in the lower plot of Fig. 17. Some experimental and numerical estimates are compiled in Table 2.

To summarize, this application is a good example for how large-scale Monte Carlo simulations employing the cluster update algorithm can be used to investigate the influence of quenched bond dilution on the critical properties of the 3D Ising. It also illustrates how scaling and finite-size scaling analyses can be applied to a non-trival problem.

7.2. *Polymer statistics: Adsorption phenomena*

Polymers in dilute solutions are found at high temperatures typically in extended random coil conformations.[203–205] Lowering the temperature, entropy becomes less important and due to the monomer-monomer attraction globular conformations gain weight until the polymer collapses at the so-called θ-point in a cooperative rearrangement of the monomers.[203–205] The globular conformations are relatively compact with little internal structure. Hence, entropy does still play some role, and a further freezing transition towards low-degenerate crystalline energy dominated states is expected and indeed observed.[206,207] For sufficiently short-range interactions these two transitions may fall together,[208] but in general they are clearly distinct.

The presence of an attractive substrate adds a second energy scale to the system which introduces several new features. Apart from the adsorption transition,[209,210] it also induces several low-temperature structural phases by the competition between monomer-monomer and monomer-surface attraction whose details depend on the exact number of monomers. Theoretical predictions may guide future experiments on such small scales which appear feasible due to recent advances of experimental techniques. Among such sophisticated techniques at the nanometer scale are, e.g., atomic force microscopy (AFM), where it is possible to measure the contour length and the end-to-end distance of individual polymers[211] or to quantitatively investigate the peptide adhesion on semiconductor surfaces.[212] Another experimental tool with an extraordinary resolution in positioning and accuracy in force measurements are optical tweezers.[213,214]

With this motivation we recently performed a careful classification of thermodynamic phases and phase transitions for a range of surface attraction strengths and temperatures and compared the results for end-grafted polymers[215] with those of non-grafted polymers[216] that can move freely within a simulation box.[217] In these studies we employed a bead-stick model of a linear polymer with three energy contributions:

$$E = 4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\left(r_{ij}^{-12} - r_{ij}^{-6}\right) + \frac{1}{4}\sum_{i=1}^{N-2}\left(1-\cos\vartheta_i\right) + \epsilon_s\sum_{i=1}^{N}\left(\frac{2}{15}z_i^{-9} - z_i^{-3}\right) \quad . \tag{145}$$

The first two terms are a standard 12-6 Lennard-Jones (LJ) potential and a weak bending energy describing the bulk behaviour. The distance between the monomers i and j is r_{ij} and $0 \leq \vartheta_i \leq \pi$ denotes the bending angle between the

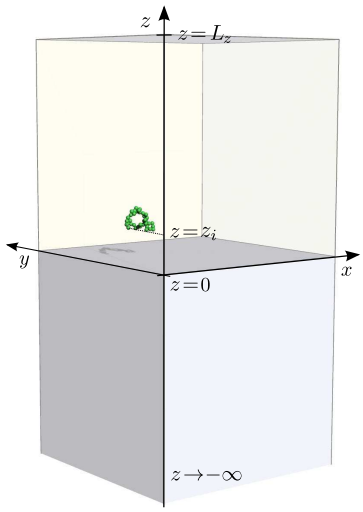

Fig. 18. Sketch of a single polymer subject to an attractive substrate at $z = 0$. The hard wall at $z = L_z$ prevents a non-grafted polymer from escaping.

ith, $(i + 1)$th, and $(i + 2)$th monomer. The third term is specific to an attractive substrate. This 9-3 LJ surface potential follows by integration over the continuous half-space $z < 0$ (cf. Fig. 18), where every space element interacts with each monomer by the usual 12-6 LJ expression.[218] The relative strength of the two LJ interactions is continuously varied by considering ϵ_s as a control parameter.

We employed parallel tempering simulations to a 40mer once grafted with one end to the substrate in the potential minimum and once freely moving in the space between the substrate and a hard wall a distance $L_z = 60$ away. There exist several attempts to optimise the choice of the simulation points β_i,[89,90] but usually one already gets a reasonable performance when observing the histograms and ensuring the acceptance probability to be around 50%, which approximately requires an equidistribution in β. We employed $64 - 72$ different replicas with 50 000 000 sweeps each, from which every 10th value was stored in a time series – the autocorrelation time in units of sweeps turned out to be of the order of thousands. Finally, all data are combined by the multi-histogram technique (using the variant of Ref. 219).

Apart from the internal energy and specific heat, a particular useful quantity for polymeric systems is the squared radius of gyration $R^2_{\rm gyr} = \sum_{i=1}^{N} (\vec{r}_i - \vec{r}_{\rm cm})^2$, with $\vec{r}_{\rm cm} = (x_{\rm cm}, y_{\rm cm}, z_{\rm cm}) = \sum_{i=1}^{N} \vec{r}_i/N$ being the center-of-mass of the polymer. In the presence of a symmetry breaking substrate, it

is useful to also monitor the tensor components parallel and perpendicular to the substrate, $R_{\parallel}^2 = \sum_{i=1}^{N}[(x_i - x_{\rm cm})^2 + (y_i - y_{\rm cm})^2]$ and $R_{\perp}^2 = \sum_{i=1}^{N}(z_i - z_{\rm cm})^2$. As an indicator for adsorption one may take the distance of the center-of-mass of the polymer to the surface. Additionally, we also analyzed the mean number of monomers docked to the surface n_s where for the continuous substrate potential we defined a monomer i to be docked if $z_i < z_c \equiv 1.5$.

The main results are summarized in the phase diagram shown in Fig. 19. It is constructed using the profile of several canonical fluctuations as shown for the specific heat in Fig. 20. For the non-grafted polymer this plot clearly reveals the freezing and adsorption transitions. Freezing leads to a pronounced peak near $T = 0.25$ (we use units in which $k_B = 1$) almost independently of the surface attraction strengths. That this is indeed the freezing transition is confirmed by the very rigid crystalline structures found below this temperature. To differentiate between the different crystalline structures, the radius of gyration, its tensor components parallel and perpendicular to the substrate, and the number of surface contacts were analyzed. This revealed that the crystalline phases arrange in a different number of layers to minimize the energy. For high surface attraction strengths, a single layer is favored (AC1), and for decreasing ϵ_s the number of layers increases until for the 40mer a maximal number of 4 layers is reached (AC4), cf. the representative conformations depicted in the right panel of Fig. 19. The fewer layers are involved in a layering transition, the more pronounced is that transition. Raising the temperature above the freezing temperature, polymers form adsorbed and still rather compact conformations. This is the phase of adsorbed globular (AG) conformations that can be subdivided into droplet-like globules for surface interactions ϵ_s that are not strong enough to induce a single layer below the freezing transition and more pancake-like flat conformations (AG1) at temperatures above the AC1 phase. At higher temperatures, two scenarios can be distinguished. For small adsorption strength ϵ_s, a non-grafted polymer first desorbs from the surface [from AG to the desorbed globular (DG) bulk phase] and disentangles at even higher temperatures [from DG to the desorbed expanded bulk phase (DE)]. For larger ϵ_s, the polymer expands while it is still adsorbed to the surface (from AG/AG1 to AE) and desorbs at higher temperatures (from AE to DE). The collapse transition in the adsorbed phase takes place at a lower temperature compared to the desorbed phase because the deformation at the substrate leads to an effective reduction of the number of contacts.

Grafting the polymer to the substrate mainly influences the adsorption transition. Figure 20(b), e.g., reveals that it is strongly weakened for all ϵ_s. Due to grafting, the translational entropy for desorbed chains is strongly reduced. As a consequence adsorption of finite grafted polymers appears to be continuous, in

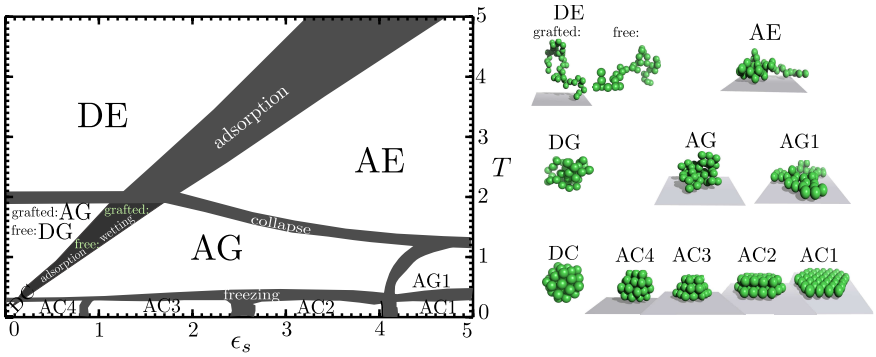

Fig. 19. The pseudo-phase diagram parametrized by adsorption strength ϵ_s and temperature T for a 40mer. The gray transition regions have a broadness that reflects the variation of the corresponding peaks of the fluctuations of canonical expectation values we investigated. Phases with an 'A/D' are adsorbed/desorbed. 'E', 'G' and 'C' denote phases with increasing order: expanded, globular and compact/crystalline. The right panel shows representative conformations of the individual phases.

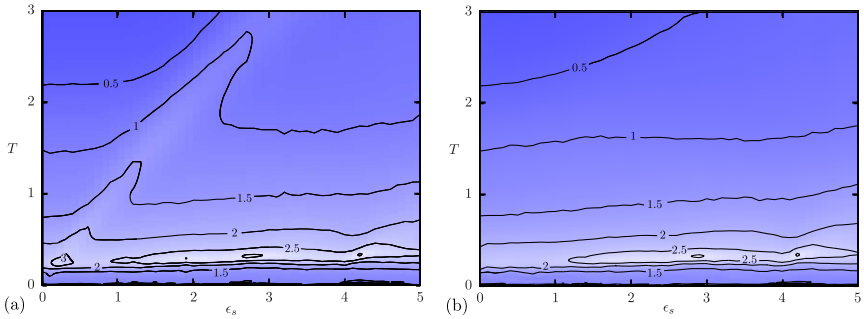

Fig. 20. Specific-heat profile, $c_V(\epsilon_s, T)$, for (a) the non-grafted and (b) the grafted polymer.

contrast to the non-grafted case where this behaviour becomes apparent for very long chains only. The reason is that *all* conformations of a grafted polymer are influenced by the substrate, because they cannot escape. Hence, the first-order-like conformational rearrangement of extended non-grafted polymers upon adsorption is not necessary and the adsorption is continuous.

The case of globular chains has to be discussed separately. While non-grafted globular chains adsorb continuously, for grafted globular chains it even is nontrivial to identify an adsorption transition. A globular chain attached to a substrate always has several surface contacts such that a "desorbed globule" stops to be a well-defined description here. For stronger surface attraction one might, however, identify the transition from attached globules that only have a few contacts

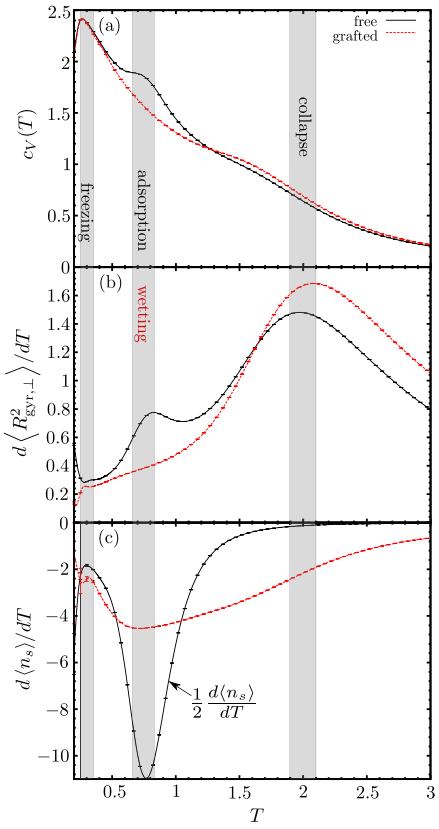

Fig. 21. (a) Specific heat $c_V(T)$, (b) fluctuation of the radius of gyration component perpendicular to the substrate $\mathrm{d}\left\langle R^2_{\mathrm{gyr},\perp}\right\rangle(T)/\mathrm{d}T$, and (c) fluctuation of the number of monomers in contact with the substrate $\mathrm{d}\langle n_s\rangle(T)/\mathrm{d}T$ for weak surface attraction, $\epsilon_s = 0.7$, where the adsorption occurs at a lower temperature than the collapse.

to docked conformations with the wetting transition. This roughly coincides with the position of the adsorption transition for the free chain between DG and AG in the phase diagram and is illustrated for $\epsilon_s = 0.7$ in Fig. 21. For a non-grafted polymer, at the adsorption transition a peak is visible in $c_V(T)$, $\mathrm{d}\left\langle R^2_{\mathrm{gyr},\perp}\right\rangle/\mathrm{d}T$ and $\mathrm{d}\langle n_s\rangle/\mathrm{d}T$. For the grafted polymer, on the other hand, the first two peaks disappear and with it the adsorption transition. Only a signal in the number of surface contacts is left. This change of surface contacts in an otherwise unchanged attached globule signals the wetting transition.

To summarize, this example was chosen to illustrate the application of extensive parallel tempering simulations to analyze and compare the whole phase

diagram of a generic off-lattice model for grafted and non-grafted polymers as a function of temperature and surface interaction strength. The main differences between the two cases were found at and above the adsorption transition where the restriction of translational degrees of freedom due to grafting becomes important.

8. Concluding Remarks

The aim of this chapter is to give an elementary introduction into the basic principles underlying modern Markov chain Monte Carlo simulations and to illustrate their usefulness by two advanced applications to quenched, disordered spin systems and adsorption phenomena of polymers.

The simulation algorithms employing local update rules are very generally applicable but suffer from critical slowing down at second-order phase transitions. Non-local cluster update methods are much more efficient but more specialized. Some generalizations from Ising to Potts and O(n) symmetric spin models have been indicated. In principle also other models may be efficiently simulated by cluster updates, but there does not exist a general strategy for their construction. Reweighting techniques and generalized ensemble ideas such as simulated and parallel tempering, the multicanonical ensemble and Wang-Landau method can be adapted to almost any statistical physics problem where rare-event states hamper the dynamics. Well known examples are first-order phase transitions and spin glasses, but also some macromolecular systems fall into this class. The performance of the various algorithms can be judged by statistical error analysis which is completely general. Finally, also the outlined scaling and finite-size scaling analyses can be applied to virtually any model exhibiting critical phenomena as was exemplified for a disordered spin system.

Acknowledgements

I thank Yurij Holovatch for his kind invitation to present one of the Ising Lectures at the Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine, Lviv, Ukraine.

I gratefully acknowledge the contributions to the work reviewed here by my collaborators, in particular Michael Bachmann, Bertrand Berche, Pierre-Emmanuel Berche, Elmar Bittner, Christophe Chatelain, Monika Möddel, Thomas Neuhaus, Andreas Nußbaumer, Stefan Schnabel, and Martin Weigel, and thank Bernd Berg, Kurt Binder, David Landau, Yuko Okamoto, and Bob Swendsen for many useful discussions.

This work was partially supported by DFG Grant No. JA 483/24-3, DFG

Forschergruppe FOR877 under Grant No. Ja 483/29-1, DFG Sonderforschungsbereich/Transregio SFB/TRR 102 Project B04, Graduate School of Excellence GSC 185 "BuildMoNa", DFH-UFA German-French Graduate School under Grant No. CDFA-02-07, and the computer time Grant No. hlz17 of NIC, Forschungszentrum Jülich.

References

1. M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
2. D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
3. K. Binder and 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. N. Metropolis and S. Ulam, J. Americ. Stat. Ass. **44**, 335 (1949).
7. J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (London, 1965).
8. A. A. Markov, Izv. Adad. Nauk SPB VI, 61 (1907).
9. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
10. J. Honerkamp, *Statistical Physics – An Advanced Approach with Applications* (Springer, Berlin, 1998).
11. K. Binder, in *Monte Carlo Methods in Statistical Physics*, ed. K. Binder (Springer, Berlin, 1979).
12. D. W. Heermann, *Computer Simulation Methods in Theoretical Physics*, 2nd ed. (Springer, Berlin, 1990).
13. K. Binder (ed.), *The Monte Carlo Method in Condensed Matter Physics* (Springer, Berlin, 1992).
14. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
15. W. K. Hastings, Biometrika **57**, 97 (1970).
16. S. Schnabel, W. Janke, and M. Bachmann, J. Comp. Phys. **230**, 4454 (2011).
17. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Science **220**, 671 (1983).
18. D. Bouzida, S. Kumar, and R. H. Swendsen, Phys. Rev. A **45**, 8894 (1992).
19. M. A. Miller, L. M. Amon, W. P. Reinhardt, Chem. Phys. Lett. **331**, 278 (2000).
20. R. H. Swendsen, Physics Procedia **15**, 81 (2011).
21. G. O. Roberts, A. Gelman, and W. R. Gilks, Ann. Appl. Probab. **7**, 110 (1997).
22. M. Bédard, Stochastic Process. Appl. **118**, 2198 (2008).
23. 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, and A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol.

10, pp. 447–458 (2002), and references therein.
24. A. M. Ferrenberg, D. P. Landau, and Y. J. Wong, Phys. Rev. Lett. **69**, 3382 (1992).
25. R. J. Glauber, J. Math. Phys. **4**, 294 (1963).
26. W. Janke, in *Ageing and the Glass Transition*, ed. M. Henkel, M. Pleimling, and R. Sanctuary, Lect. Notes Phys. **716** (Springer, Berlin, Heidelberg, 2007), pp. 207–260.
27. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford Press, Oxford, 1979).
28. J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, *The Theory of Critical Phenomena* (Oxford University Press, Oxford, 1992).
29. D. A. Lavis and G. M. Bell, *Statistical Mechanics of Lattice Systems 2* (Springer, Berlin, 1999).
30. See the volumes of review articles edited by C. Domb and J. L. Lebowitz (eds.): *Phase Transitions and Critical Phenomena* (Academic Press, New York).
31. See, e.g., the photographs in Fig. 1.6 of Ref. 27.
32. J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986); Phys. Rev. D **40**, 2035 (1989).
33. R. G. Edwards, J. Goodman, and A. D. Sokal, Nucl. Phys. B **354**, 289 (1991).
34. A. D. Sokal, *Monte Carlo methods in statistical mechanics: Foundations and new algorithms*, Lecture Notes, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, Switzerland (1989).
35. A. D. Sokal, *Bosonic algorithms*, in *Quantum Fields on the Computer*, ed. M. Creutz (World Scientific, Singapore, 1992), p. 211.
36. W. Janke, *Nonlocal Monte Carlo algorithms for statistical physics applications*, Mathematics and Computers in Simulations **47**, 329–346 (1998).
37. 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, and H.-B. Schüttler (Springer, Berlin, 1994), p. 29.
38. W. Janke, *First-order phase transitions*, in *Computer Simulations of Surfaces and Interfaces*, NATO Science Series, II. Mathematics, Physics and Chemistry – Vol. **114**, eds. B. Dünweg, D. P. Landau, and A. I. Milchev (Kluwer, Dordrecht, 2003), pp. 111–135.
39. R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
40. P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Japan **26** (Suppl.), 11 (1969).
41. C. M. Fortuin and P. W. Kasteleyn, Physica **57**, 536 (1972).
42. C. M. Fortuin, Physica **58**, 393 (1972).
43. C. M. Fortuin, Physica **59**, 545 (1972).
44. D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor and Francis, London, 1992).
45. J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
46. P. L. Leath, Phys. Rev. B **14**, 5046 (1976).
47. M. E. J. Newman and R. M. Ziff, Phys. Rev. E **64**, 016706 (2001).
48. W. Janke and A. M. J. Schakel, Nucl. Phys. B **700**, 385 (2004); Comp. Phys. Comm. **169**, 222 (2005); Phys. Rev. E **71**, 036703 (2005); Phys. Rev. Lett. **95**, 135702 (2005); e-print `arXiv:cond-mat/0508734` and Braz. J. Phys. **36**, 708 (2006). For a review, see *Spacetime approach to phase transitions*, in *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory*, Vol. 2, ed. Y. Holovatch (World

Scientific, Singapore, 2007), pp. 123–180, and the extensive list of references to earlier work given therein.
49. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
50. R. B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).
51. U. Wolff, Nucl. Phys. B **322**, 759 (1989).
52. M. Hasenbusch, Nucl. Phys. B **333**, 581 (1990).
53. U. Wolff, Nucl. Phys. B **334**, 581 (1990).
54. U. Wolff, Phys. Lett. A **228**, 379 (1989).
55. C. F. Baillie, Int. J. Mod. Phys. C **1**, 91 (1990).
56. M. Hasenbusch and S. Meyer, Phys. Lett. B **241**, 238 (1990).
57. R. H. Swendsen, J.-S. Wang, and A. M. Ferrenberg, in *The Monte Carlo Method in Condensed Matter Physics*, ed. K. Binder (Springer, Berlin, 1992).
58. W. Janke, Phys. Lett. A **148**, 306 (1990).
59. C. Holm and W. Janke, Phys. Rev. B **48**, 936 (1993).
60. X.-L. Li and A. D. Sokal, Phys. Rev. Lett. **63**, 827 (1989); *ibid.* **67**, 1482 (1991).
61. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77 – The Art of Scientific Computing*, 2nd edition (Cambridge University Press, Cambridge, 1999).
62. H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. **8**, 1 (1973).
63. N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).
64. T. W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).
65. M. B. Priestley, *Spectral Analysis and Time Series*, 2 vols. (Academic, London, 1981), Chapters 5-7.
66. 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, and A. Muramatsu, John von Neumann Institute for Computing, Jülich, NIC Series, Vol. **10**, pp. 423–445 (2002).
67. A. M. Ferrenberg, D. P. Landau, and K. Binder, J. Stat. Phys. **63**, 867 (1991).
68. W. Janke and T. Sauer, J. Stat. Phys. **78**, 759 (1995).
69. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics [SIAM], Philadelphia, 1982).
70. R. G. Miller, Biometrika **61**, 1 (1974).
71. A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
72. A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1658(E) (1989).
73. B. Kaufman, Phys. Rev. **76**, 1232 (1949).
74. A. E. Ferdinand and M.E. Fisher, Phys. Rev. **185**, 832 (1969).
75. P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).
76. 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**, eds. B. Dünweg, D. P. Landau, and A. I. Milchev (Kluwer, Dordrecht, 2003), pp. 161–171.
77. A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
78. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comp. Chem. **13**, 1011 (1992).
79. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J.

Comp. Chem. **16**, 1339 (1995).
80. T. Bereau and R. H. Swendsen, J. Comp. Phys. **228**, 6119 (2009).
81. E. Gallicchio, M. Andrec, A. K. Felts, and R.M. Levy, J. Phys. Chem. B **109**, 6722 (2005).
82. J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, J. Chem. Theory Comput. **3**, 26 (2007).
83. E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
84. A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
85. R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
86. J.-S. Wang and R. H. Swendsen, Prog. Theor. Phys. Suppl. **157**, 317 (2005).
87. 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 and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
88. K. Hukushima and K. Nemoto, J. Phys. Soc. Japan **65**, 1604 (1996).
89. H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech. P03018 (2006).
90. E. Bittner, A. Nußbaumer, and W. Janke, Phys. Rev. Lett. **101**, 130603 (2008).
91. J. Gross, W. Janke, and M. Bachmann, Comp. Phys. Comm. **182**, 1638 (2011); Physics Procedia **15**, 29 (2011).
92. B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
93. B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
94. B. A. Berg, Fields Inst. Comm. **26**, 1 (2000).
95. B. A. Berg, Comp. Phys. Comm. **147**, 52 (2002).
96. W. Janke, Physica A **254**, 164 (1998).
97. 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**, eds. B. Dünweg, D. P. Landau, and A. I. Milchev (Kluwer, Dordrecht, 2003), pp. 137–157.
98. W. Janke, Int. J. Mod. Phys. C **3**, 1137 (1992).
99. K. Binder, in *Phase Transitions and Critical Phenomena*, Vol. 5b, eds. C. Domb and M. S. Green (Academic Press, New York, 1976), p. 1.
100. G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974); J. Comp. Phys. **23**, 187 (1977) 187; J. Chem. Phys. **66**, 1402 (1977).
101. B. A. Berg and W. Janke, Phys. Rev. Lett. **80**, 4771 (1998).
102. B. A. Berg, A. Billoire, and W. Janke, Phys. Rev. B **61**, 12143 (2000); Phys. Rev. E **65**, 045102 (RC) (2002); Physica A **321**, 49 (2003).
103. E. Bittner and W. Janke, Europhys. Lett. **74**, 195 (2006).
104. W. Janke (ed.), *Rugged Free Energy Landscapes: Common Computational Approaches to Spin Glasses, Structural Glasses and Biological Macromolecules*, Lecture Notes in Physics **736** (Springer, Berlin, 2008).
105. W. Janke, B. A. Berg, and M. Katoot, Nucl. Phys. B **382**, 649 (1992).
106. A. Nußbaumer, E. Bittner, and W. Janke, Europhys. Lett. **78**, 16004 (2007).
107. E. Bittner, A. Nußbaumer, and W. Janke, Nucl. Phys. B **820**, 694 (2009).
108. B. A. Berg, J. Stat. Phys **82**, 323 (1996).
109. B. A. Berg and W. Janke, unpublished notes (1996).

110. B. A. Berg, Comp. Phys. Comm. **153**, 397 (2003).
111. J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986); Phys. Rev. D **40**, 2035 (1989).
112. W. Janke and T. Sauer, Phys. Rev. E **49**, 3475 (1994).
113. W. Janke and S. Kappler, Nucl. Phys. B (Proc. Suppl.) **42**, 876 (1995).
114. W. Janke and S. Kappler, Phys. Rev. Lett. **74**, 212 (1995).
115. M. S. Carroll, W. Janke, and S. Kappler, J. Stat. Phys. **90**, 1277 (1998).
116. B. A. Berg and W. Janke, Phys. Rev. Lett. **98**, 040602 (2007).
117. T. Neuhaus and J. S. Hager, J. Stat. Phys. **113** 47 (2003).
118. K. Binder and M. H. Kalos, J. Stat. Phys. **22**, 363 (1980).
119. H. Furukawa and K. Binder, Phys. Rev. A **26**, 556 (1982).
120. M. Biskup, L. Chayes, and R. Kotecký, Europhys. Lett. **60**, 21 (2002); Comm. Math. Phys. **242**, 137 (2003); J. Stat. Phys. **116**, 175 (2003).
121. K. Binder, Physica A **319**, 99 (2003).
122. A. Nußbaumer, E. Bittner, T. Neuhaus, and W. Janke, Europhys. Lett. **75**, 716 (2006).
123. A. Nußbaumer, E. Bittner, and W. Janke, Phys. Rev. E **77**, 041109 (2008).
124. K. Leung and R. K. P. Zia, J. Phys. A **23**, 4593 (1990).
125. F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001).
126. R. B. Griffiths, Phys. Rev. Lett. **24**, 1479 (1970).
127. G. S. Rushbrooke, J. Chem. Phys. **39**, 842 (1963).
128. R. B. Griffiths, Phys. Rev. Lett. **14**, 623 (1965).
129. B. D. Josephson, Proc. Phys. Soc. **92**, 269 (1967); *ibid.* 276 (1967).
130. M. E. Fisher, Phys. Rev. **180**, 594 (1969).
131. L. P. Widom, J. Chem. Phys. **43**, 3892 (1965); *ibid.* 3898 (1965).
132. L. P. Kadanoff, Physics **2**, 263 (1966).
133. K. G. Wilson and J. Kogut, Phys. Rep. C **12**, 75 (1974).
134. L. Onsager, Phys. Rev. **65**, 117 (1944).
135. B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, Cambridge, 1973).
136. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).
137. L. Onsager, Nuovo Cimento (Suppl.) **6**, 261 (1949); see also the historical remarks in Refs. 135,136.
138. C. N. Yang, Phys. Rev. **85**, 808 (1952).
139. C. H. Chang, Phys. Rev. **88**, 1422 (1952).
140. B. Nickel, J. Phys. A: Math. Gen. **32**, 3889 (1999); J. Phys. A: Math. Gen. **33**, 1693 (2000); W. P. Orrick, B. G. Nickel, A. J. Guttmann, and J. H. H. Perk, Phys. Rev. Lett. **86**, 4120 (2001); J. Stat. Phys. **102**, 795 (2001).
141. S. Boukraa, A. J. Guttmann, S. Hassani, I. Jensen, J.-M. Maillard, B. Nickel, and N. Zenine, J. Phys. A: Math. Theor. **41**, 455202 (2008) [the coefficients of low- and high-T susceptibility series expansions up to order 2000 are given on `http://www.ms.unimelb.edu.au/~iwan/ising/Ising_ser.html`].
142. Y. Chan, A. J. Guttmann, B. G. Nickel, and J. H. H. Perk, J. Stat. Phys. **145**, 549 (2011) [e-print `arXiv:1012.5272v3` (cond-mat.stat-mech) with additional in-

formations].
143. R. Kenna, D. A. Johnston, and W. Janke, Phys. Rev. Lett. **96**, 115701 (2006); *ibid.* **97**, 155702 (2006) [Publisher's Note: *ibid.* **97**, 169901(E) (2006)].
144. F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
145. F. Y. Wu, Rev. Mod. Phys. **55**, 315(E) (1983).
146. M. Weigel and W. Janke, Phys. Rev. B **62**, 6343 (2000).
147. M. E. Barber, in *Phase Transitions and Critical Phenomena*, Vol. 8, eds. C. Domb and J. L. Lebowitz (Academic Press, New York, 1983), p. 146.
148. V. Privman (ed.), *Finite-Size Scaling and Numerical Simulations of Statistical Systems* (World Scientific, Singapore, 1990).
149. K. Binder, in *Computational Methods in Field Theory*, Schladming Lecture Notes, eds. H. Gausterer and C. B. Lang (Springer, Berlin, 1992), p. 59.
150. G. Kamieniarz and H. W. J. Blöte, J. Phys. A **26**, 201 (1993).
151. J. Salas and A. D. Sokal, J. Stat. Phys. **98**, 551 (2000).
152. X. S. Chen and V. Dohm, Phys. Rev. E **70**, 056136 (2004).
153. V. Dohm, J. Phys. A **39**, L259 (2006).
154. W. Selke and L. N. Shchur, J. Phys. A **38**, L739 (2005).
155. M. Schulte and C. Drope, Int. J. Mod. Phys. C **16**, 1217 (2005).
156. M. A. Sumour, D. Stauffer, M. M. Shabat, and A. H. El-Astal, Physica A **368**, 96 (2006).
157. W. Selke, Eur. Phys. J. B **51**, 223 (2006); J. Stat. Mech. P04008 (2007).
158. J. D. Gunton, M. S. Miguel, and P. S. Sahni, in *Phase Transitions and Critical Phenomena*, Vol. 8, eds. C. Domb and J. L. Lebowitz (Academic Press, New York, 1983).
159. K. Binder, Rep. Prog. Phys. **50**, 783 (1987).
160. H. J. Herrmann, W. Janke, and F. Karsch (eds.): *Dynamics of First Order Phase Transitions* (World Scientific, Singapore, 1992).
161. M. E. Fisher and A. N. Berker, Phys. Rev. B **26**, 2507 (1982).
162. V. Privman, M. E. Fisher, J. Stat. Phys. **33**, 385 (1983).
163. K. Binder and D. P. Landau, Phys. Rev. B **30**, 1477 (1984).
164. M. S. S. Challa, D. P. Landau, and K. Binder, Phys. Rev. B **34**, 1841 (1986).
165. V. Privman and J. Rudnik, J. Stat. Phys. **60**, 551 (1990).
166. C. Borgs and R. Kotecky, J. Stat. Phys. **61**, 79 (1990).
167. J. Lee and J. M. Kosterlitz, Phys. Rev. Lett. **65**, 137 (1990).
168. C. Borgs, R. Kotecky, and S. Miracle-Solé, J. Stat. Phys. **62**, 529 (1991).
169. C. Borgs and W. Janke, Phys. Rev. Lett. **68**, 1738 (1992).
170. W. Janke, Phys. Rev. B **47**, 14757 (1993).
171. E. Bittner and W. Janke, Phys. Rev. E **84**, 036701 (2011).
172. M. Weigel and W. Janke, Phys. Rev. Lett. **102**, 100601 (2009); Phys. Rev. E **81**, 066701 (2010).
173. J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996), Chap. 8.
174. P. W. Mitchell, R. A. Cowley, H. Yoshizawa, P. Böni, Y. J. Uemura, and R. J. Birgeneau, Phys. Rev. B **34**, 4719 (1986).
175. A. B. Harris, J. Phys. C **7**, 1671 (1974).
176. For a review, see B. Berche and C. Chatelain, *Phase transitions in two-dimensional random Potts models*, in *Order, Disorder and Criticality: Advanced Problems of*

Phase Transition Theory, Vol. 1, ed. Y. Holovatch (World Scientific, Singapore, 2004), pp. 147–199.
177. W. Selke, L. N. Shchur, and A. L. Talapov, in *Annual Reviews of Computational Physics I*, ed. D. Stauffer (World Scientific, Singapore, 1994), pp. 17–54.
178. For a recent overview, see R. Folk, Y. Holovatch, and T. Yavors'kii, Physics Uspiekhi **173**, 175 (2003) [e-print `arXiv:cond-mat/0106468`].
179. C. Chatelain, B. Berche, W. Janke, and P.-E. Berche, Phys. Rev. E **64**, 036120 (2001).
180. P.-E. Berche, C. Chatelain, B. Berche, and W. Janke, Comp. Phys. Comm. **147**, 427 (2002); Eur. Phys. J. B **38**, 463 (2004).
181. B. Berche, P.-E. Berche, C. Chatelain, and W. Janke, Condens. Matter Phys. **8**, 47 (2005).
182. M. Hellmund and W. Janke, Comp. Phys. Comm. **147**, 435 (2002); Nucl. Phys. B (Proc. Suppl.) **106&107**, 923 (2002).
183. M. Hellmund and W. Janke, Phys. Rev. E **67**, 026118 (2003).
184. M. Hellmund and W. Janke, Phys. Rev. B **74**, 144201 (2006).
185. C. D. Lorenz and R. M. Ziff, Phys. Rev. E **57**, 230 (1998).
186. B. Derrida, Phys. Rep. **103**, 29 (1984); A. Aharony and A. B. Harris, Phys. Rev. Lett. **77**, 3700 (1996); S. Wiseman and E. Domany, Phys. Rev. Lett. **81**, 22 (1998).
187. L . Turban, Phys. Lett. A **75**, 307 (1980); J. Phys. C **13**, L13 (1980).
188. A. L. Talapov and H. W. J. Blöte, J. Phys. A **29**, 5727 (1996).
189. A. Aharony and A. B. Harris, Phys. Rev. Lett. **77**, 3700 (1996).
190. S. Wiseman and E. Domany, Phys. Rev. Lett. **81**, 22 (1998); Phys. Rev. E **58**, 2938 (1998).
191. R. Guida and J. Zinn-Justin, J. Phys. A **31**, 8103 (1998).
192. H. G. Ballesteros, L. A. Fernández, V. Martín-Mayor, A. Muñoz Sudupe, G. Parisi, and J. J. Ruiz-Lorenzo, Phys. Rev. B **58**, 2740 (1998).
193. R. J. Birgeneau, R. A. Cowley, G. Shirane, and H. Yoshizawa, J. Stat. Phys. **34**, 817 (1984).
194. D. P. Belanger, A. R. King, and V. Jaccarino, Phys. Rev. B **34**, 452 (1986).
195. D. P. Belanger, Braz. J. Phys. **30**, 682 (2000).
196. S. A. Newlove, J. Phys. C: Solid State Phys. **16**, L423 (1983).
197. R. Folk, Yu. Holovatch, and T. Yavors'kii, J. Phys. Stud. **2**, 213 (1998).
198. R. Folk, Y. Holovatch, and T. Yavors'kii, Phys. Rev. B **61**, 15114 (2000).
199. A. Pelissetto and E. Vicari, Phys. Rev. B **62**, 6393 (2000).
200. P.-E. Berche, C. Chatelain, B. Berche, and W. Janke, Eur. Phys. J. B **38**, 463 (2004).
201. P. Calabrese, V. Martín-Mayor, A. Pelissetto, and E. Vicari, Phys. Rev. E **68**, 036136 (2003).
202. D. Ivaneyko, J. Ilnytskyi, B. Berche, and Yu. Holovatch, Condens. Matter Phys. **8**, 149 (2005).
203. P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca and London, 1979).
204. J. des Cloizeaux and G. Jannink, *Polymers in Solution* (Clarendon Press, Oxford, 1990).
205. A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (American Institute of Physics, New York, 1994).
206. W. Paul, T. Strauch, F. Rampf, and K. Binder, Phys. Rev. E **75**, 060801(R) (2007).

207. T. Vogel, M. Bachmann, and W. Janke, Phys. Rev. E **76**, 061803 (2007).
208. M. P. Taylor, W. Paul, and K. Binder, Phys. Rev. E **79**, 050801 (2009).
209. E. Eisenriegler, K. Kremer, and K. Binder, J. Chem. Phys. **77**, 6296 (1982).
210. E. Eisenriegler, *Polymers near Surfaces: Conformation Properties and Relation to Critical Phenomena* (World Scientific, Singapore, 1993).
211. F. Kuhner, M. Erdmann, and H. E. Gaub, Phys. Rev. Lett. **97**, 218301 (2006).
212. M. Bachmann, K. Goede, A. Beck-Sickinger, M. Grundmann, A. Irbäck, and W. Janke, Angew. Chem. Int. Ed. **122**, 9721 (2010).
213. D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, Nature **413**, 748 (2001).
214. K. Kegler, M. Salomo, and F. Kremer, Phys. Rev. Lett. **98**, 058304 (2007).
215. M. Möddel, W. Janke, and M. Bachmann, Macromolecules **44**, 9013 (2011).
216. M. Möddel, M. Bachmann, and W. Janke, J. Phys. Chem. B **113**, 3314 (2009).
217. M. Möddel, W. Janke, and M. Bachmann, Phys. Chem. Chem. Phys. **12**, 11548 (2010).
218. W. A. Steele, Surface Sci. **36**, 317 (1973).
219. M. K. Fenwick, J. Chem. Phys. **129**, 125106 (2008).