John von Neumann Institute for Computing

Statistical Analysis of Simulations: Data Correlations and Error Estimation

Wolfhard Janke

published in

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes,
J. Grotendorst, D. Marx, A. Muramatsu (Eds.),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **10**, ISBN 3-00-009057-6, pp. 423-445, 2002.

© 2002 by John von Neumann Institute for Computing
Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

http://www.fz-juelich.de/nic-series/volume10

Statistical Analysis of Simulations: Data Correlations and Error Estimation

Wolfhard Janke

Institut für Theoretische Physik, Universität Leipzig
Augustusplatz 10/11, 04109 Leipzig, Germany
E-mail: wolfhard.janke@itp.uni-leipzig.de

This lecture gives an introduction to data correlations and error estimation in Monte Carlo simulations. The basic concepts are illustrated for studies of second-order phase transitions in classical spin systems such as the Ising model and for models exhibiting first-order phase transitions such as the q-state Potts model.

1 Introduction

The statistical mechanics of complex systems poses many hard problems which are often difficult to solve by analytical approaches. Numerical simulation techniques will therefore be indispensable tools on our way to a better understanding of such systems. Applications range from physics and chemistry to biological systems and even sociology and economy. Examples include (spin) glasses, disordered magnets, or biologically motivated studies of protein folding, to mention only a few important problems in classical physics. The broad class of quantum statistical problems in condensed matter and elementary particles physics as well as non-perturbative approaches to quantum gravity are further important examples.

Depending on the specific physical problem and the objectives at hand, the simulational approach is either based on molecular dynamics (MD) or Monte Carlo (MC) methods.[1,2] Sometimes even a combination of both methods is used. For the purpose of this lecture I will focus in the following mainly on the MC approach. Thanks to advances in computer technology and significant algorithmic refinements in the past few years, MC studies have reached in many applications an accuracy that allows "quasi-exact" predictions. Since a MC simulation is a stochastic method, it is thus very important to supplement the data with carefully determined, reliable statistical error estimates. But also in other areas, where it is sometimes only feasible to obtain a first qualitative overview of the system behaviour from simulation studies, it is necessary to gain at least a rough idea of the data correlations involved and the order of magnitude of the statistical errors.

The required tools for the data analysis described below are designed for the general case and are perfectly suited for applications to very complex systems. Still, they can be illustrated and tested for very simple examples. In this lecture I will, therefore, mainly concentrate on Ising and Potts models, and sometimes even use synthetically generated stochastic data.

The rest of the paper is organized as follows. In the next section I first recall the definition of the example models and a few standard observables. Then some properties of phase transitions and related aspects of Monte Carlo simulations are briefly summarized. Here emphasis will be placed on those points which are necessary for an understanding of the tools of statistical data analysis which are described in Sec. 3 and illustrated with worked

out examples in Secs. 4 and 5. Section 6 is devoted to error propagation in multicanonical simulations. In Sec. 7, I conclude with a brief summary and a few general comments.

2 Model Systems and Phase Transitions

2.1 Models and Observables

When developing and testing advanced analysis tools for stochastically generated data it is not only convenient to work with very simple models but, in fact, even advantageous since usually at least a few exact analytical results are available for comparison. On the other hand, one should always be prepared that a really complex system may add further complications which are not present in the simple test cases. I nevertheless will follow the "bottom-up" approach, but at several places point out potential difficulties when more complex systems are considered.

The paradigm for a well-controlled second-order phase transition is the well-known Ising model[3] whose partition function is defined as

$$Z_I(\beta) = \sum_{\{\sigma_i\}} \exp(-\beta H_I) \ , \qquad H_I = -\sum_{\langle ij \rangle} \sigma_i \sigma_j \ , \qquad \sigma_i = \pm 1 \ , \tag{1}$$

where $\beta = J/k_B T$ is the inverse temperature in natural units, the spins σ_i live on the sites i of a D-dimensional cubic lattice of volume $V = L^D$, and the symbol $\langle ij \rangle$ indicates that the lattice sum runs over all 2D nearest-neighbour pairs. We always assume periodic boundary conditions. In two dimensions (2D) and zero field this model has been solved exactly, even on finite lattices. For the three-dimensional (3D) model no exact solution is available, but there is an enormous amount of very precise data from MC simulations, high-temperature series expansions and, as far as critical exponents are concerned, also from field theoretical methods. The computer code for MC simulations of this model is easy to program and can be found, for instance, on the accompanying diskette to the article in Ref. 4, where Metropolis, heat-bath, Wolff single-cluster and Swendsen-Wang multiple-cluster update routines are described and implemented. In addition also programs for reweighting analyses and the exact solution of the 2D Ising model are provided in this reference.

Standard observables are the internal energy per site, $e = E/V$, with $E = -d \ln Z_I / d\beta \equiv \langle H_I \rangle$, the specific heat,

$$C/k_B = \frac{de}{d(k_B T)} = \beta^2 \left(\langle H_I^2 \rangle - \langle H_I \rangle^2 \right) / V \ , \tag{2}$$

the magnetization

$$m = M/V = \langle |\mu| \rangle \ , \qquad \mu = \sum_i \sigma_i / V \ , \tag{3}$$

and the susceptibility

$$\chi = \beta V \left(\langle \mu^2 \rangle - \langle |\mu| \rangle^2 \right) \ . \tag{4}$$

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

$$\chi' = \beta V \langle \mu^2 \rangle \ . \tag{5}$$

Similarly, the spin-spin correlation function can then be taken as

$$G(\vec{x}_i - \vec{x}_j) = \langle \sigma_i \sigma_j \rangle \ . \tag{6}$$

At large distances, $G(\vec{x})$ decays exponentially, $G(\vec{x}) \sim \exp(-|\vec{x}|/\xi)$, and the spatial correlation length ξ can be defined as

$$\xi = -\lim_{|\vec{x}|\to\infty} (|\vec{x}|/\ln G(\vec{x})) \ . \tag{7}$$

The standard example exhibiting a first-order phase transition is the q-state Potts model[5] defined by the Hamiltonian

$$H_{\rm P} = -\sum_{\langle ij \rangle} \delta_{\sigma_i \sigma_j} \ , \qquad \sigma_i \in 1, \ldots, q \ , \tag{8}$$

where δ_{ij} is the Kronecker symbol. In 2D, this model is exactly known[6] to exhibit a temperature-driven first-order transition at $\beta_t = \log(1+\sqrt{q})$ for all $q \geq 5$. For $q \leq 4$ the transition is of second order, including the Ising case ($q = 2$) and the special percolation limit ($q = 1$). In 3D, there is strong numerical evidence that for all $q \geq 3$ the transition is of first order.[7]

2.2 Phase Transitions

The theory of phase transitions is a broad subject well covered by many textbooks. Here we shall confine ourselves to those properties that are important for understanding the requirements on the data analysis tools.

The characterising feature of second-order phase transitions is a *divergent* spatial correlation length ξ at the transition point β_c. This causes scale invariance, the "heart" of renormalization group treatments, and is the origin of the important concept of universality. The growth of spatial correlations in the vicinity of the critical point is illustrated in Fig. 1. At β_c one thus expects fluctuations on all length scales, implying power-law singularities in thermodynamic functions such as the correlation length,

$$\xi = \xi_0^{+,-} t^{-\nu} + \ldots \ , \tag{9}$$

where $t \equiv |1 - T/T_c|$ and the $\ldots$ indicate subleading (analytical and confluent) corrections. This defines the (universal) critical exponent ν and the (non-universal) critical amplitudes $\xi_0^{+,-}$ on the high- and low-temperature side of the transition. Similar singularities of the specific heat, magnetization, and susceptibility as sketched in Fig. 2 define the critical exponents α, β, and γ, respectively, which are related with each other through scaling and hyper-scaling relations; only two of them (e.g. ν and γ) may be considered as independent.

When updating the spins with an importance sampling MC process,[2,4,8] the information on the updated state of the spins has to propagate over the correlation volume before one obtains a new, statistically *independent* configuration. The number of update steps this takes is measured by the autocorrelation time τ (a formal definition is given below) which close to β_c scales according to

$$\tau \propto \xi^z \propto t^{-\nu z} \ . \tag{10}$$

Here we have introduced the independent *dynamical* critical exponent z, which depends on the employed update algorithm. For a *local* MC update procedure such as the Metropolis

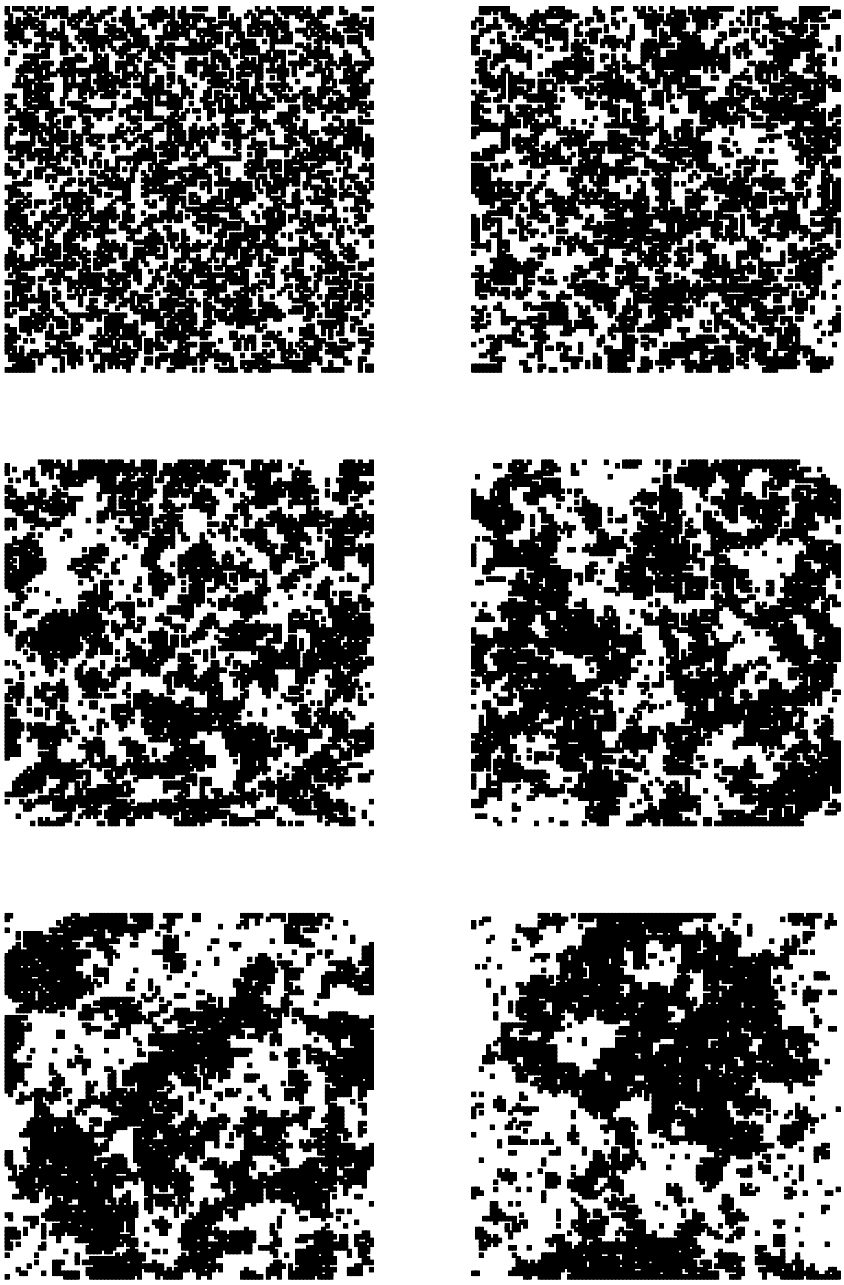

Figure 1. Approach of the critical region (lower right) starting from high temperatures (upper left) illustrating the development of large spatial correlations. Shown are 2D Ising model configurations for a 100×100 lattice with periodic boundary conditions at $\beta/\beta_c = 0.50$, 0.70, 0.85, 0.90, 0.95, and 0.98.

or heat-bath algorithm, one expects that the updated information performs a random walk in configuration space, requiring on the average ξ^2 steps to propagate over a distance proportional to ξ. For local update algorithms one thus expects a dynamical critical exponent of $z \approx 2$. In fact, an exact lower bound is $z \geq \gamma/\nu = 2 - \eta$, and numerical estimates for the Ising model yield $z \approx 2.125$ in 2D and $z \approx 2.03$ in 3D.[9,10] As we will see in the next section, this *critical slowing down* of the dynamics (based on local update rules) is responsible for the dramatic reduction of the statistical accuracy attainable close to a critical point in a given computer time allocation. This is the reason why cluster and multigrid update algorithms have become so important.[8–10] Here the update rules are *non-local*, leading to a significant reduction of critical slowing down. The dynamical critical exponent z varies among the different non-local update schemes and depends on the model class considered. In most cases, however, one finds z smaller than unity, and when cluster algorithms are applied to the 2D Ising model it is even difficult to distinguish z from zero, i.e. a logarithmic divergence.

For systems of finite size, as in any numerical simulation, the correlation length can-

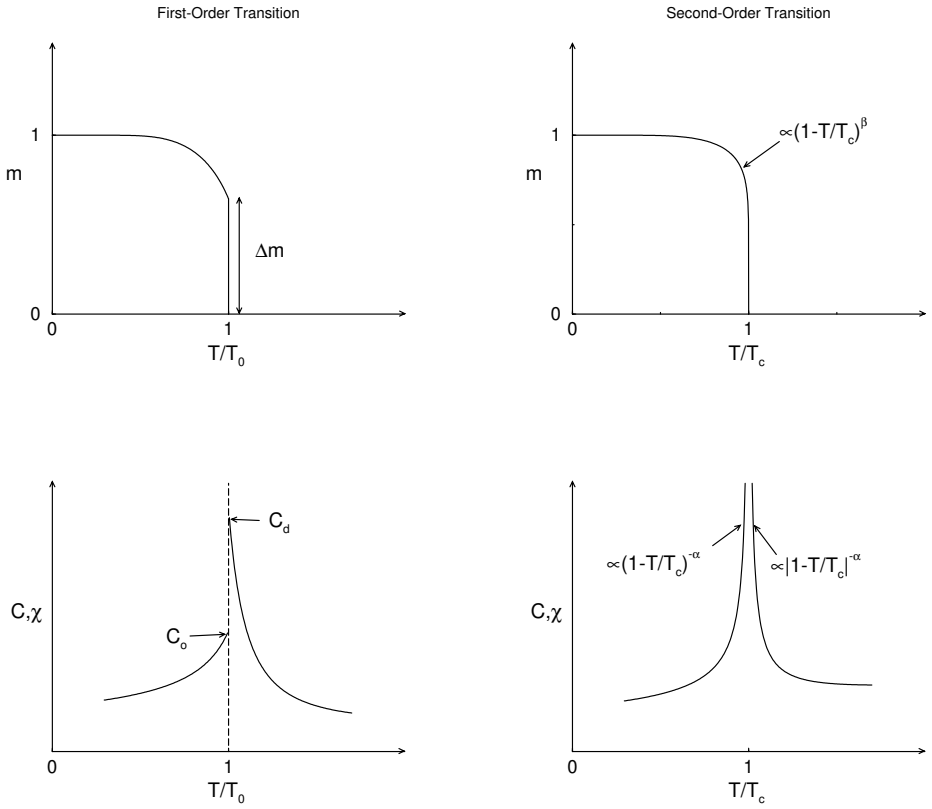

Figure 2. The characteristic behaviour of the magnetization, m, specific heat, C, and susceptibility, χ, at first- and second-order phase transitions.

not diverge, and also the divergences in all other quantities are rounded and shifted. For the specific heat of the 2D Ising model this is illustrated in Fig. 3. In the scaling formulas the role of ξ is then taken over by the linear size of the system, L. By writing $t \propto \xi^{-1/\nu} \longrightarrow L^{-1/\nu}$, it becomes clear how thermodynamic scaling laws, e.g. $\chi \propto t^{-\gamma}$, should be replaced by *finite-size scaling* (FSS) Ansätze, e.g. $\chi \propto L^{\gamma/\nu}$, for finite geometries. In particular, by recalling (10) we obtain for the autocorrelation time a FSS of the form

$$\tau \propto L^z \quad . \tag{11}$$

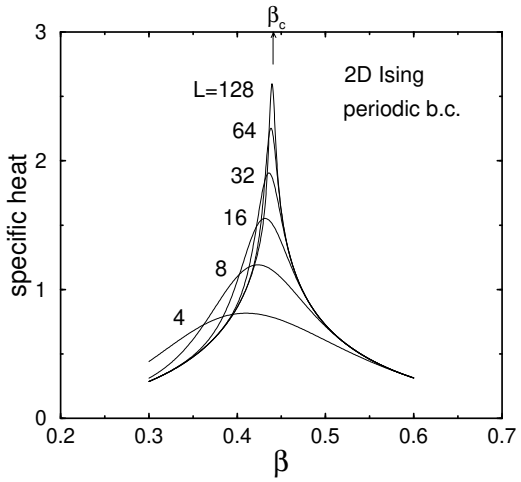

Figure 3. Finite-size scaling behaviour of the specific heat of the 2D Ising model on $L \times L$ lattices close to the infinite-volume critical point $\beta_c = \log(1+\sqrt{2})/2 = 0.440\,686\ldots$.

Most phase transitions in nature are of first order.[11–14] An example experienced every day is ordinary melting. The characterising feature of first-order phase transitions are *discontinuities* in the order parameter (the jump Δm of the magnetization m in Fig. 2) or in the energy (the latent heat Δe), or both, at the transition point T_0. This reflects the fact that at T_0, two (or more) phases can coexist. In the example of the melting transition these are the solid (ordered) and liquid (disordered) phases. In contrast to a second-order transition, the correlation length in the coexisting pure phases is finite. Consequently the specific heat, the susceptibility and also the autocorrelation time do not diverge in the *pure* phases. There are, however, superimposed delta function like singularities associated with the jumps of e and m.

For finite systems the singularities are smeared out to narrow peaks with a height proportional to the volume and a width proportional to 1/volume. This signalizes that the system is now capable to flip from one phase into the other via mixed phase configurations. Mixed phase configurations are separated by interfaces which carry an extra free energy σL^{D-1}, where σ is the (reduced) interface tension and L^{D-1} is the projected area

of the interfaces. Since at T_0 the total bulk contribution to the free energy of the two coexisting phases equals that of the pure phases, the probability for the occurrence of mixed phase configurations is suppressed by the interfacial Boltzmann weight $\exp(-2\sigma L^{D-1})$, where the factor two accounts for the topological constraint that with periodic boundary conditions only an even number of interfaces can be present. This argument explains the typical double-peak structure of the energy and magnetization densities at a first-order phase transition. Due to the exponentially suppressed dip between the two peaks, for large system sizes it may take very long before the systems flips from one phase into the other. In fact, the autocorrelation time associated with this "flipping mode" is just the inverse of the suppression factor in the probability density,

$$\tau \propto \exp\left(2\sigma L^{D-1}\right) \quad . \tag{12}$$

Since here the autocorrelations grow exponentially with the system size, this behaviour has been termed "supercritical slowing down" – even though nothing is "critical" at a first-order phase transition.

The standard acceleration methods such as cluster and multigrid algorithms cannot overcome this slowing down problem. The reason is that it is rooted in the shape of the probability density itself. In this situation a completely different strategy is necessary, namely generalized ensemble simulations. One of these techniques are multicanonical simulations[15–17] whose statistical analysis requires special care.

In the multicanonical reweighting approach (for reviews see Refs. 18, 19 and 20) one rewrites the partition function by introducing an auxiliary function $f(E)$ as

$$Z = \sum_{\{s_i\}} e^{-\beta[H-f(E)]} e^{-\beta f(E)} \quad , \tag{13}$$

and adjusts the reweighting factor $\exp(-\beta f(E))$ in such a way that the resulting histogram of E sampled according to the *multicanonical* probability distribution

$$\begin{aligned} p_{\mathrm{muca}}(E) &\propto \exp[-\beta(H - f(E))] \\ &\equiv \exp[-\beta H_{\mathrm{muca}}] \end{aligned} \tag{14}$$

is approximately flat between the two peaks. Here H_{muca} is the central object of a multicanonical simulation, and plays the same role in it as H does in a canonical simulation. Canonical observables $\langle \mathcal{O} \rangle_{\mathrm{can}}$ can be recovered according to

$$\langle \mathcal{O} \rangle_{\mathrm{can}} = \frac{\langle \mathcal{O} w \rangle}{\langle w \rangle} \quad , \tag{15}$$

where $\langle \dots \rangle$ without subscript denote expectation values in the multicanonical distribution and

$$w \equiv \exp(\beta f(E)) \tag{16}$$

is the inverse reweighting factor. The multicanonical probability distribution p_{muca} may be updated using any legitimate MC algorithm, the simplest choice being a local Metropolis update. The fact that all canonical expectation values are expressed as ratios of multicanonical expectation values is the reason for the extra complications in the statistical analysis of multicanonical simulations.

3 Estimators, Autocorrelation Times, Bias and Resampling

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

3.1 Estimators

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

$$\overline{\mathcal{O}} = \frac{1}{N}\sum_{j=1}^{N}\mathcal{O}_j \ , \tag{17}$$

where we assume that the time series contains a total of N measurements. Conceptually it is important to distinguish between the expectation value $\langle\mathcal{O}\rangle$, which is an ordinary number, and the estimator $\overline{\mathcal{O}}$, which is a *random* number fluctuating around the theoretically expected value. Of course, in practice one does not probe the fluctuations of the mean value directly (which would require repeating the whole MC simulation many times), but rather estimates its variance,

$$\sigma^2_{\overline{\mathcal{O}}} = \langle[\overline{\mathcal{O}} - \langle\overline{\mathcal{O}}\rangle]^2\rangle = \langle\overline{\mathcal{O}}^2\rangle - \langle\overline{\mathcal{O}}\rangle^2 \ , \tag{18}$$

from the distribution of the individual measurements $\mathcal{O}_j$. If the N subsequent measurements $\mathcal{O}_j$ were all uncorrelated then the relation would simply be

$$\sigma^2_{\overline{\mathcal{O}}} = \sigma^2_{\mathcal{O}_j}/N \ , \tag{19}$$

where $\sigma^2_{\mathcal{O}_j} = \langle\mathcal{O}_j^2\rangle - \langle\mathcal{O}_j\rangle^2$ is the variance of the individual measurements. Here one assumes, of course, that the simulation is in equilibrium and uses time-translation invariance over the Markov chain. Equation (19) is true for any distribution $\mathcal{P}(\mathcal{O}_j)$ of the $\mathcal{O}_j$. For the energy or magnetization the latter distributions are often plotted as physically directly relevant histograms whose squared width ($= \sigma^2_{\mathcal{O}_j}$) is proportional to the specific heat or susceptibility, respectively.

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

3.2 Autocorrelation Times

Things become more involved for correlated measurements.[21–23] Starting from the second identity in (18) and inserting (17), we obtain

$$\sigma^2_{\overline{\mathcal{O}}} = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \mathcal{O}_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \ . \tag{20}$$

By collecting diagonal and off-diagonal terms we arrive at

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle \right) \ . \tag{21}$$

The first term is identified as the variance of the individual measurements times $1/N$. In the second sum we first use the symmetry $i \leftrightarrow j$ to reduce the summation to $\sum_{i \neq j}^{N} = 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N}$. Then we reorder the summation and use time translation invariance to derive

$$\sigma^2_{\overline{\mathcal{O}}} = \frac{1}{N} \left[\sigma^2_{\mathcal{O}_i} + 2 \sum_{k=1}^{N} \left(\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle \right) \left(1 - \frac{k}{N} \right) \right] \ , \tag{22}$$

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

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \frac{\sigma^2_{\mathcal{O}_i}}{N} 2\tau'_{\mathcal{O},\mathrm{int}} \ . \tag{23}$$

Here we have introduced the so-called (proper) *integrated* autocorrelation time,

$$\tau'_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \left(1 - \frac{k}{N} \right) \ , \tag{24}$$

with

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} \tag{25}$$

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

For large time separations k the autocorrelation function decays exponentially,

$$A(k) \overset{k \to \infty}{\longrightarrow} a e^{-k/\tau_{\mathcal{O},\mathrm{exp}}} \ , \tag{26}$$

where $\tau_{\mathcal{O},\mathrm{exp}}$ is the *exponential* autocorrelation time and a is some constant. Due to the exponential decay of $A(k)$ as $k \to \infty$, in any meaningful simulation with $N \gg \tau_{\mathcal{O},\mathrm{exp}}$,

the correction term in parentheses in (24) can safely be neglected. The usually employed definition of the *integrated* autocorrelation time is thus

$$\tau_{\mathcal{O},\mathrm{int}} = \frac{1}{2} + \sum_{k=1}^{N} A(k) \ . \tag{27}$$

Notice that, in general, $\tau_{\mathcal{O},\mathrm{int}}$ (and also $\tau'_{\mathcal{O},\mathrm{int}}$) is different from $\tau_{\mathcal{O},\mathrm{exp}}$. In fact, one can show[24] that $\tau_{\mathcal{O},\mathrm{int}} \leq \tau_{\mathcal{O},\mathrm{exp}}$ in realistic models. Only if $A(k)$ is a pure exponential, the two autocorrelation times, $\tau_{\mathcal{O},\mathrm{int}}$ and $\tau_{\mathcal{O},\mathrm{exp}}$, coincide (up to minor corrections for small $\tau_{\mathcal{O},\mathrm{int}}$, see Eq. (46) in Sec. 4 below).[23]

The important point of Eq. (23) is that due to temporal correlations of the measurements the statistical error $\epsilon_{\overline{\mathcal{O}}} \equiv \sqrt{\sigma^2_{\overline{\mathcal{O}}}}$ on the MC estimator $\overline{\mathcal{O}}$ is enhanced by a factor of $\sqrt{2\tau_{\mathcal{O},\mathrm{int}}}$. This can be rephrased by writing the statistical error similar to the uncorrelated case as $\epsilon_{\overline{\mathcal{O}}} = \sqrt{\sigma^2_{\mathcal{O}_j}/N_{\mathrm{eff}}}$, but now with a parameter

$$N_{\mathrm{eff}} = N/2\tau_{\mathcal{O},\mathrm{int}} \leq N \ , \tag{28}$$

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. Since some quantities (e.g., the specific heat or susceptibility) can severely be underestimated if the effective statistics is too small,[25] any serious simulation should therefore provide at least a rough order-of-magnitude estimate of autocorrelation times.

3.3 Bias

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

$$\hat{\sigma}^2_{e_i} = \overline{e^2} - \overline{e}^2 = \overline{(e-\overline{e})^2} = \frac{1}{N}\sum_{i=1}^{N} (e_i - \overline{e})^2 \ . \tag{29}$$

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

$$\langle\hat{\sigma}^2_{e_i}\rangle = \langle\overline{e^2} - \overline{e}^2\rangle = \langle\overline{e^2}\rangle - \langle\overline{e}\rangle^2 - \left(\langle\overline{e}^2\rangle - \langle\overline{e}\rangle^2\right) \ , \tag{30}$$

and then use the previously derived result: The first two terms on the r.h.s. of (30) just give $\sigma^2_{e_i}$, and the second two terms in parentheses yield $\sigma^2_{\overline{e}} = \sigma^2_{e_i} 2\tau_{e,\mathrm{int}}/N$, as calculated in (23). Combining these two results we arrive at

$$\langle\hat{\sigma}^2_{e_i}\rangle = \sigma^2_{e_i}\left(1 - \frac{2\tau_{e,\mathrm{int}}}{N}\right) = \sigma^2_{e_i}\left(1 - \frac{1}{N_{\mathrm{eff}}}\right) \neq \sigma^2_{e_i} \ . \tag{31}$$

The estimator $\hat{\sigma}^2_{e_i}$ as defined in (29) thus systematically underestimates the true value by a term of the order of $\tau_{e,\mathrm{int}}/N$. Such an estimator is called *weakly biased* ("weakly" because the statistical error $\propto 1/\sqrt{N}$ is asymptotically larger than the systematic bias; for medium or small N, however, also prefactors need to be carefully considered).

We thus see that for large autocorrelation times or equivalently small effective statistics $N_{\rm eff}$, the bias may be quite large. Since $\tau_{e,\rm int}$ scales quite strongly with the system size for local update algorithms, some care is necessary in choosing the run time N. Otherwise the FSS of the specific heat and thus the determination of the *static* critical exponent α/ν could be completely spoiled by the temporal correlations!

As a side remark we note that even in the completely uncorrelated case the estimator (29) is biased, $\langle\hat\sigma^2_{e_i}\rangle = \sigma^2_{e_i}\,(1-1/N)$, since with our conventions in this case $\tau_{e,\rm int} = 1/2$ (some authors use a different convention in which τ more intuitively vanishes in the uncorrelated case; but this has certain disadvantages in other formulas). In this case one can (and usually does) define a bias-corrected estimator,

$$\hat\sigma^2_{e_i,\rm corr} = \frac{N}{N-1}\hat\sigma^2_{e_i} = \frac{1}{N-1}\sum_{i=1}^N (e_i-\overline{e})^2 \ , \tag{32}$$

which obviously satisfies $\langle\hat\sigma^2_{e_i,\rm corr}\rangle = \sigma^2_{e_i}$. For the squared error on the mean value, this leads to the error formula $\epsilon^2_{\overline{e}} = \hat\sigma^2_{\overline{e},\rm corr} = \hat\sigma^2_{e_i,\rm corr}/N = \frac{1}{N(N-1)}\sum_{i=1}^N (e_i-\overline{e})^2$, i.e., to the celebrated replacement of one of the $1/N$-factors by $1/(N-1)$ "due to one missing degree of freedom".

3.4 Numerical Estimation of Autocorrelation Times

The above considerations show that not only for the error estimation but also for the computation of static quantities themselves it is important to have control over autocorrelations. Unfortunately, it is very difficult to give reliable a priori estimates, and an accurate numerical analysis is often too time consuming. As a rough estimate it is about ten times harder to get precise information on dynamic quantities than on static quantities like critical exponents. A (biased) estimator $\hat A(k)$ for the autocorrelation function is obtained by replacing in (25) the expectation values (ordinary numbers) by mean values (random variables), e.g., $\langle\mathcal{O}_i\mathcal{O}_{i+k}\rangle$ by $\overline{\mathcal{O}_i\mathcal{O}_{i+k}}$. With increasing k the relative variance of $\hat A(k)$ diverges rapidly. To get at least an idea of the order of magnitude of $\tau_{\mathcal{O},\rm int}$ and thus the correct error estimate (23), it is useful to record the "running" autocorrelation time estimator

$$\hat\tau_{\mathcal{O},\rm int}(k_{\rm max}) = \frac{1}{2} + \sum_{k=1}^{k_{\rm max}} \hat A(k) \ , \tag{33}$$

which approaches $\tau_{\mathcal{O},\rm int}$ in the limit of large $k_{\rm max}$ where, however, its statistical error increases rapidly. As a compromise between systematic and statistical errors, an often employed procedure is to determine the upper limit $k_{\rm max}$ self-consistently by cutting off the summation once $k_{\rm max} \geq 6\hat\tau_{\mathcal{O},\rm int}(k_{\rm max})$. In this case an a priori error estimate is available,[9, 10, 23]

$$\epsilon_{\tau_{\mathcal{O},\rm int}} = \tau_{\mathcal{O},\rm int}\sqrt{\frac{2(2k_{\rm max}+1)}{N}} \approx \tau_{\mathcal{O},\rm int}\sqrt{\frac{12}{N_{\rm eff}}} \ . \tag{34}$$

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

which depends on the considered quantity), thus implying $N \approx 10^8$. Since in each sweep L^2 spins have to be updated and assuming that each spin update takes about a μsec, we end up with a total time estimate of about 10^6 seconds $\approx$ 1 CPU-day to achieve this accuracy.

Another possibility is to approximate the tail end of $A(k)$ by a single exponential as in (26). Summing up the small k part exactly, one finds[26]

$$\tau_{\mathcal{O},\mathrm{int}}(k_{\max}) = \tau_{\mathcal{O},\mathrm{int}} - ce^{-k_{\max}/\tau_{\mathcal{O},\exp}} \quad , \tag{35}$$

where c is a constant. The latter expression may be used for a numerical estimate of both the exponential and integrated autocorrelation times.[26]

3.5 Binning Analysis

As the preceding discussions have shown, ignoring autocorrelations can lead to a severe underestimation of statistical errors. Invoking the full machinery of autocorrelation analysis, 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 original data into bins or blocks, one forms a new, shorter time series which is almost uncorrelated and can thus be analyzed by standard means. But even if the data are completely uncorrelated in time, one still has to handle the problem of error estimation for quantities that are not directly measured in the simulation but are computed as a non-linear combination of "basic" observables. This problem can either be solved by error propagation or by using the Jackknife method described in the next subsection.

Let us assume that the time series consists of N correlated measurements $\mathcal{O}_i$. One then forms N_B non-overlapping blocks of length k such that $N = N_B k$ (assuming that N was chosen cleverly; otherwise one has to discard some of the data and redefine N) and computes the block average $\mathcal{O}_{B,n}$ of the n-th block,

$$\mathcal{O}_{B,n} \equiv \frac{1}{k}\sum_{i=1}^{k} \mathcal{O}_{(n-1)k+i} \quad , \qquad n = 1, \ldots, N_B \quad . \tag{36}$$

The mean value over all block variables obviously satisfies $\overline{\mathcal{O}}_B = \overline{\mathcal{O}}$. If the block length k is large enough ($k \gg \tau$), the blocks are basically uncorrelated in time and their variance can be computed according to the unbiased estimator (32), leading to the squared statistical error of the mean value,

$$\epsilon^2_{\overline{\mathcal{O}}} \equiv \sigma^2_{\overline{\mathcal{O}}} = \sigma^2_B/N_B = \frac{1}{N_B(N_B-1)}\sum_{n=1}^{N_B}(\mathcal{O}_{B,n} - \overline{\mathcal{O}}_B)^2 \quad . \tag{37}$$

By comparing with (23) we see that $\sigma^2_B/N_B = 2\tau_{\mathcal{O},\mathrm{int}}\sigma^2_{\mathcal{O}_i}/N$, showing that one may also use

$$2\tau_{\mathcal{O},\mathrm{int}} = k\sigma^2_B/\sigma^2_{\mathcal{O}_i} \tag{38}$$

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

3.6 Jackknife Analysis

Instead of considering rather small blocks of lengths k and their fluctuations as in the binning method, in a Jackknife analysis[27,28] one forms N_B large Jackknife blocks $\mathcal{O}_{J,n}$ containing all data but one of the previous binning blocks,

$$\mathcal{O}_{J,n} = \frac{N\overline{\mathcal{O}} - k\mathcal{O}_{B,n}}{N-k} \ , \qquad n = 1, \ldots, N_B \ . \tag{39}$$

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

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \frac{N_B - 1}{N_B} \sum_{n=1}^{N_B} (\mathcal{O}_{J,n} - \overline{\mathcal{O}}_J)^2 \ . \tag{40}$$

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

4 A Simplified Model

It is always useful to have a few exact results available against which the numerical techniques can be checked. Of course, to continue analytically, we have to make some simplifying assumptions about the distribution $P(e)$ from which the e_i are drawn and about the temporal correlations of these measurements. In the following we shall hence assume that the e_i are Gaussian random variables. Furthermore, without loss of generality we simplify the notation and normalize the energies to have zero expectation, $\langle e_i \rangle = 0$, and unit variance, $\langle e_i^2 \rangle = 1$. This is convenient but inessential. Finally, the temporal correlations are modelled by a bivariate time series with correlation coefficient ρ ($0 \le \rho < 1$),

$$\begin{aligned} e_0 &= e_0' \ , \\ e_i &= \rho e_{i-1} + \sqrt{1-\rho^2}\, e_i' \ , \qquad i \ge 1 \ , \end{aligned} \tag{41}$$

where the e_i' are *independent* Gaussian random variables satisfying $\langle e_i' \rangle = 0$ and $\langle e_i' e_j' \rangle = \delta_{ij}$. By iterating the recursion (41) it is then easy to see that

$$e_k = \rho e_{k-1} + \sqrt{1-\rho^2}\, e_k' = \rho^k e_0 + \sqrt{1-\rho^2} \sum_{l=1}^{k} \rho^{k-l} e_l' \ , \tag{42}$$

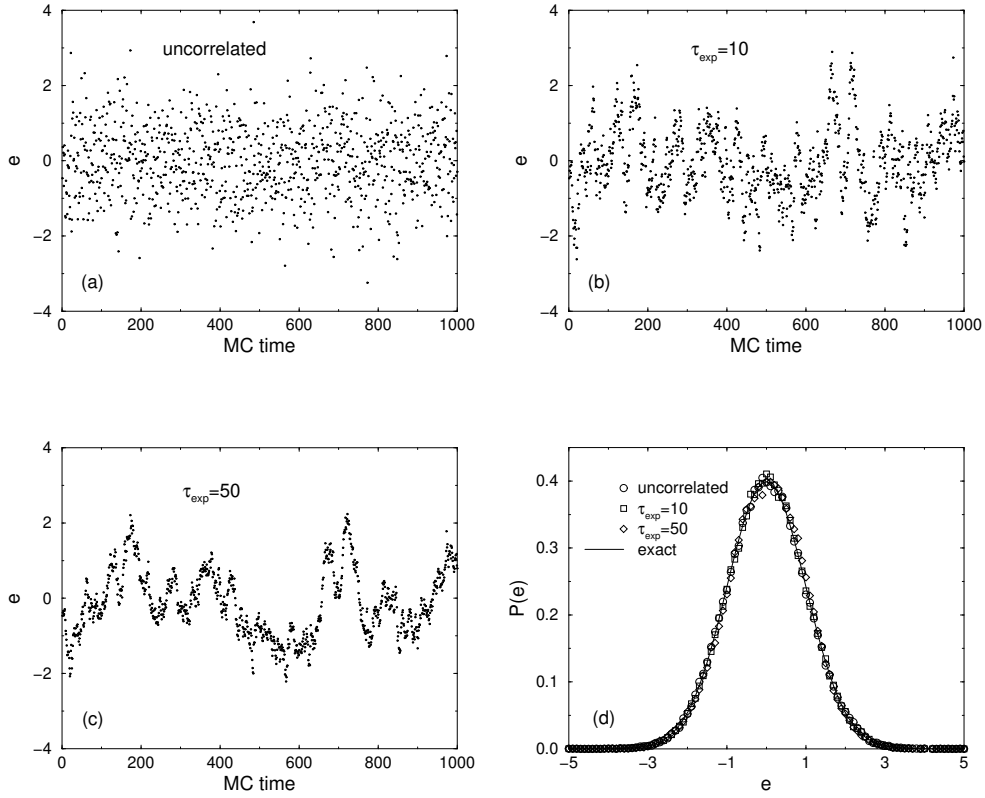

Figure 4. One percent of the "MC time" evolution according to the bivariate Gaussian process (41) in (a) the uncorrelated case respectively with (b) $\tau_{\rm exp} = 10$ and (c) $\tau_{\rm exp} = 50$. All three time evolutions with a total of 100 000 consecutive "measurements" lead to the same Gaussian histogram shown in (d).

and consequently

$$A(k) = \langle e_0 e_k \rangle = \rho^k \equiv e^{-k/\tau_{\rm exp}} \quad . \tag{43}$$

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

$$\tau_{\rm exp} = -1/\ln \rho \quad . \tag{44}$$

It should be stressed that in realistic situations a purely exponential decay can only be expected asymptotically for large k where the slowest mode dominates. For smaller time separations usually also many other modes contribute whose correlation time is smaller. The visual appearance of uncorrelated and correlated data with $\tau_{\rm exp} = 10$ and 50 is depicted in Figs. 4(a)-(c) where in each case one percent of the total "MC time" evolution consisting of 100 000 consecutive "measurements" is shown. Despite the quite distinct temporal evolutions, histogramming the time series leads to the same Gaussian distribution

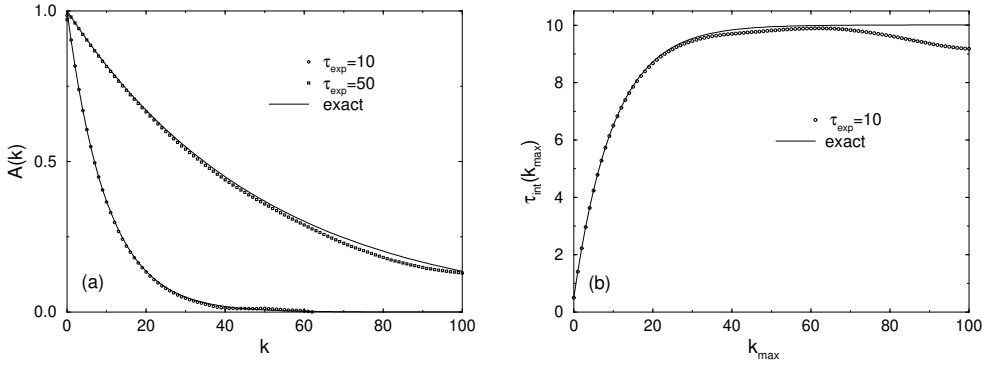

Figure 5. (a) Autocorrelation functions and (b) integrated autocorrelation time for $\tau_{\rm exp} = 10$ on the basis of 100 000 "measurements" in comparison with exact results for the bivariate Gaussian model shown as the solid lines.

within error bars, as it should, cf. Fig. 4(d). The corresponding autocorrelation functions $A(k)$ are shown in Fig. 5(a).

The integrated autocorrelation time can also be calculated exactly,

$$\tau_{\rm int} = \frac{1}{2} + \sum_{k=1}^{\infty} A(k) = \frac{1}{2}\frac{1+\rho}{1-\rho} = \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp}) \ . \tag{45}$$

For $\tau_{\rm exp} \gg 1$ this can be approximated by

$$\tau_{\rm int} = \tau_{\rm exp}\left[1 + \frac{1}{12\tau_{\rm exp}^2} + \mathcal{O}(1/\tau_{\rm exp}^4)\right] \ , \tag{46}$$

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

As explained in the last section, one usually truncates the summation in (45) self-consistently at about $k_{\rm max} = 6\tau_{\rm int}$ ($\approx 6\tau_{\rm exp}$) since $A(k)$ becomes very noisy for large time separations. Observing that (45) is nothing but a geometric series, also the resulting correction can be calculated exactly,

$$\tau_{\rm int}(k_{\rm max}) \equiv \frac{1}{2} + \sum_{k=1}^{k_{\rm max}} A(k) = \frac{1}{2}{\rm cth}(1/2\tau_{\rm exp})\left[1 - \frac{2e^{-(k_{\rm max}+1)/\tau_{\rm exp}}}{1+e^{-1/\tau_{\rm exp}}}\right] \ , \tag{47}$$

which simplifies in the case of large $\tau_{\rm exp} \gg 1$ to

$$\tau_{\rm int}(k_{\rm max}) = \tau_{\rm int}\left[1 - \frac{2\tau_{\rm exp}}{2\tau_{\rm exp}+1}e^{-k_{\rm max}/\tau_{\rm exp}}\right] \ , \tag{48}$$

showing that with increasing $k_{\rm max}$ the asymptotic value of $\tau_{\rm int} \equiv \tau_{\rm int}(\infty)$ is approached exponentially fast. This is illustrated in Fig. 5(b) for the bivariate Gaussian time series with $\tau_{\rm exp} = 10$. Here we also see that for too large $k_{\rm max}$ the estimate for $\tau_{\rm int}(k_{\rm max})$ can deviate quite substantially from the exact value due to its divergent variance. The usually employed self-consistent cutoff would be around $6\tau_{\rm exp} = 60$ where $\tau_{\rm int}(k_{\rm max}) \approx 9.89$.

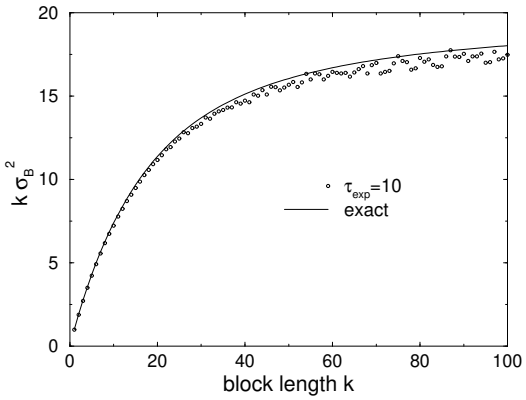

Figure 6. Binning analysis of 100 000 "measurements" in the bivariate Gaussian model. The solid line shows the exact result.

Let us now turn to the binning analysis, assuming that we decompose the total number of measurements N into N_B non-overlapping blocks of length k such that $N = N_B k$. The nth block of measurements then yields a block average of

$$e_{B,n} \equiv \frac{1}{k} \sum_{i=1}^{k} e_{(n-1)k+i} \quad . \tag{49}$$

In our simple example the expected value is, of course, zero, $\langle e_{B,n} \rangle = \frac{1}{k} \sum_{i=1}^{k} \langle e_{(n-1)k+i} \rangle = 0$. Therefore, the variance of the block variables is just the expectation value of $e_{B,n}^2$,

$$\begin{aligned} \sigma_B^2 = \langle e_{B,n}^2 \rangle &= \frac{1}{k^2} \sum_{i,j=1}^{k} \rho^{|i-j|} \\ &= \frac{1}{k^2} \left[k + 2 \sum_{i=1}^{k} \sum_{j=1}^{i-1} \rho^{i-j} \right] \\ &= \frac{1}{k} \left[1 + \frac{2\rho}{1-\rho} - \frac{2\rho}{k} \frac{1-\rho^k}{(1-\rho)^2} \right] \quad . \end{aligned} \tag{50}$$

Recalling (45) this can be rewritten as

$$k\sigma_B^2 = 2\tau_{\rm int} \left[1 - \frac{\tau_{\rm int}}{k} \left(1 - e^{-k/\tau_{\rm exp}} \right) / \cosh^2(1/2\tau_{\rm exp}) \right] \quad , \tag{51}$$

and for $\tau_{\rm exp} \gg 1$ to a very good approximation as

$$\begin{aligned} k\sigma_B^2 &\approx 2\tau_{\rm int} \left[1 - \frac{\tau_{\rm int}}{k} \left(1 - e^{-k/\tau_{\rm exp}} \right) \right] \\ &\approx 2\tau_{\rm exp} \left[1 - \frac{\tau_{\rm exp}}{k} \left(1 - e^{-k/\tau_{\rm exp}} \right) \right] \quad , \end{aligned} \tag{52}$$

showing that with increasing block length k the asymptotic value $2\tau_{\mathrm{int}}$ is approached according to a power law. For an illustration see Fig. 6.

5 A Realistic Example

In this section the autocorrelation and error analysis is illustrated for a realistic but still very simple model: The two-dimensional (2D) Ising model, simulated with the Metropolis algorithm at the infinite-volume critical point $\beta_c = \ln(1+\sqrt{2})/2 \approx 0.440\,686\,793\,4\ldots$ on a 16×16 square lattice with periodic boundary conditions. The raw data are the time series with $1\,000\,000$ measurements of the energy and magnetization taken after each sweep over the lattice, after discarding the first $200\,000$ sweeps to equilibrate the system.

A small part of the time evolution of the energy and magnetization is shown in Fig. 7. Notice the very different time scales for the e and m plots. The energy plot should be compared with the Gaussian model time series in Figs. 4(b) and (c).

Next, using the complete time series the autocorrelation functions were computed according to (25). The only difference to the analysis of the simplified model is that instead of using the Gaussian data one now reads in the Ising model time series. The result for the energy autocorrelations is shown in Fig. 8. On the log-scale of Fig. 8(b) we clearly see the asymptotic linear behaviour of $\ln A(k)$. Apart from the noise for large k, which is also present in the simplified model for finite statistics, the main difference to the artificial data of the simplified model lies in the small k behaviour. For the Ising model we clearly notice an initial fast drop, corresponding to faster relaxing modes, before the asymptotic behaviour sets in. This is, in fact, the generic behaviour of autocorrelation functions in realistic models.

Once the autocorrelation function is known, it is straightforward to sum up the integrated autocorrelation time. The result for the energy is depicted in Fig. 9, yielding an estimate of $\tau_{e,\mathrm{int}} \approx 27$. Also shown is the binning analysis which yields consistent results as it should (the horizontal line shows $2\tau_{e,\mathrm{int}} \approx 54$).

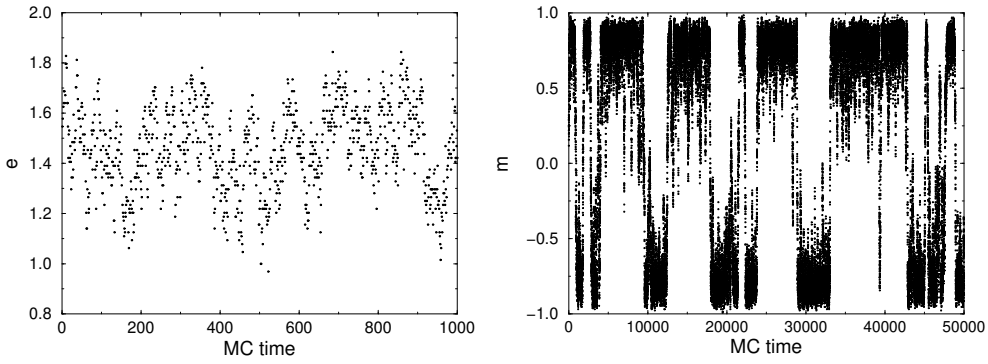

Figure 7. Part of the time evolution of the energy e and magnetization m for the 2D Ising model on a 16×16 lattice at β_c.

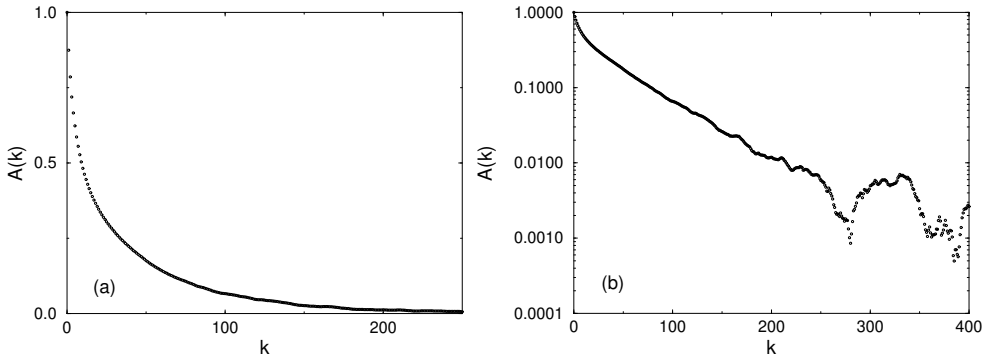

Figure 8. (a) Autocorrelation function of the energy for the 2D Ising model on a 16×16 lattice at β_c. (b) The same data as in (a) on a logarithmic scale, revealing the fast initial drop for small k and the noisy behaviour for large k.

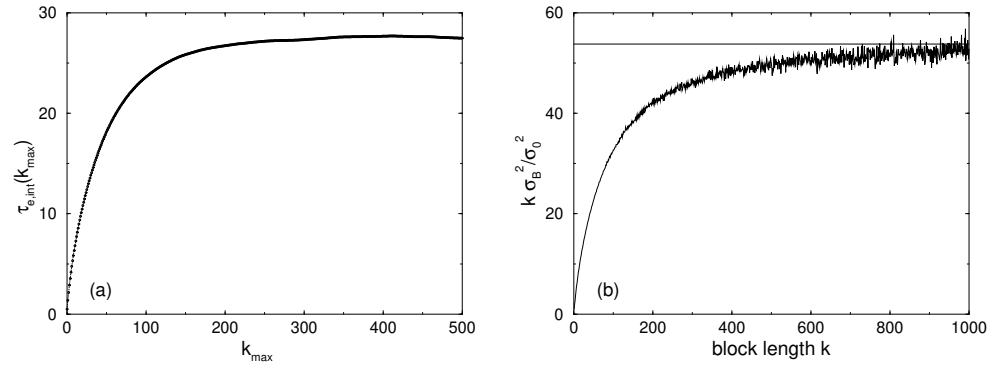

Figure 9. (a) Integrated autocorrelation time and (b) binning analysis for the energy of the 2D Ising model on a 16×16 lattice at β_c. The horizontal line in (b) shows $2\tau_{e,\mathrm{int}} \approx 54$.

6 Error Propagation in Multicanonical Simulations

As a rather extreme example for dealing with non-linear combinations of basic observables we consider now the error analysis of multicanonical simulations.[26] As shown in (15), for any observable $\mathcal{O}$ expectation values in the *canonical* ensemble, $\langle \mathcal{O} \rangle_{\mathrm{can}}$, can be calculated as

$$\langle \mathcal{O} \rangle_{\mathrm{can}} = \frac{\langle \mathcal{O} w \rangle}{\langle w \rangle} \quad , \tag{53}$$

where $\langle \dots \rangle$ (without subscript) denote expectation values with respect to the *multicanonical* distribution and $w = \exp(\beta f)$ is the inverse reweighting factor. In a MC simulation with a total number of N measurements these values are, as usual, estimated by the mean

values

$$\langle \mathcal{O} w \rangle \approx \overline{\mathcal{O} w} \equiv \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}_i w_i \ , \tag{54}$$

$$\langle w \rangle \approx \overline{w} \equiv \frac{1}{N} \sum_{i=1}^{N} w_i \ , \tag{55}$$

where $\mathcal{O}_i$ and w_i denote the measurements for the i-th configuration. Hence $\langle \mathcal{O} \rangle_{\rm can}$ is estimated by

$$\langle \mathcal{O} \rangle_{\rm can} \approx \hat{\mathcal{O}} \equiv \frac{\overline{\mathcal{O} w}}{\overline{w}} \ . \tag{56}$$

The estimator $\hat{\mathcal{O}}$ is biased,

$$\langle \hat{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle_{\rm can} \left[1 - \frac{\langle \overline{\mathcal{O} w}; \overline{w} \rangle}{\langle \overline{\mathcal{O} w} \rangle \langle \overline{w} \rangle} + \frac{\langle \overline{w}; \overline{w} \rangle}{\langle \overline{w} \rangle \langle \overline{w} \rangle} + \cdots \right] \ , \tag{57}$$

and fluctuates around $\langle \hat{\mathcal{O}} \rangle$ with variance, i.e. squared statistical error

$$\epsilon^2_{\hat{\mathcal{O}}} = \langle \mathcal{O} \rangle^2_{\rm can} \left[\frac{\langle \overline{\mathcal{O} w}; \overline{\mathcal{O} w} \rangle}{\langle \overline{\mathcal{O} w} \rangle^2} + \frac{\langle \overline{w}; \overline{w} \rangle}{\langle \overline{w} \rangle^2} - 2 \frac{\langle \overline{\mathcal{O} w}; \overline{w} \rangle}{\langle \overline{\mathcal{O} w} \rangle \langle \overline{w} \rangle} + \cdots \right] \ . \tag{58}$$

Here we used the abbreviation

$$\langle \overline{\mathcal{O} w}; \overline{w} \rangle \equiv \langle \overline{\mathcal{O} w}\, \overline{w} \rangle - \langle \overline{\mathcal{O} w} \rangle \langle \overline{w} \rangle \ , \tag{59}$$

etc. to denote (connected) correlations of the mean values, which can be computed as

$$\langle \overline{\mathcal{O} w}; \overline{w} \rangle = \langle \mathcal{O}_i w_i; w_i \rangle \frac{2 \tau^{\rm int}_{\mathcal{O} w; w}}{N} \ , \tag{60}$$

where

$$\tau^{\rm int}_{\mathcal{O} w; w} = \tau^{\rm int}_{w; \mathcal{O} w} = \frac{1}{2} + \sum_{k=1}^{N} \frac{\langle \mathcal{O}_0 w_0; w_k \rangle}{\langle \mathcal{O}_0 w_0; w_0 \rangle} \left(1 - \frac{k}{N} \right) \tag{61}$$

is the associated integrated autocorrelation time of measurements in the multicanonical distribution. Hence the statistical error is given by

$$\begin{aligned} \epsilon^2_{\hat{\mathcal{O}}} = \langle \mathcal{O} \rangle^2_{\rm can} \Bigg[& \frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle \mathcal{O}_i w_i \rangle^2} \frac{2 \tau^{\rm int}_{\mathcal{O} w; \mathcal{O} w}}{N} + \frac{\langle w_i; w_i \rangle}{\langle w_i \rangle^2} \frac{2 \tau^{\rm int}_{w; w}}{N} \\ & - 2 \frac{\langle \mathcal{O}_i w_i; w_i \rangle}{\langle \mathcal{O}_i w_i \rangle \langle w_i \rangle} \frac{2 \tau^{\rm int}_{\mathcal{O} w; w}}{N} \Bigg] \ . \end{aligned} \tag{62}$$

Since for uncorrelated measurements $\tau^{\rm int}_{\mathcal{O} w; \mathcal{O} w} = \tau^{\rm int}_{\mathcal{O} w; w} = \tau^{\rm int}_{w; w} = 1/2$, it is useful to define an *effective* multicanonical variance[a]

$$\sigma^2_{\hat{\mathcal{O}}} = \langle \mathcal{O} \rangle^2_{\rm can} \left[\frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle \mathcal{O}_i w_i \rangle^2} + \frac{\langle w_i; w_i \rangle}{\langle w_i \rangle^2} - 2 \frac{\langle \mathcal{O}_i w_i; w_i \rangle}{\langle \mathcal{O}_i w_i \rangle \langle w_i \rangle} \right] \ , \tag{63}$$

[a] In the multicanonical distribution this is nothing but an *abbreviation* of the expression on the r.h.s. but *not* the variance in the multicanonical distribution.

3D 4-state Potts model, multibondic simulations							
L	$\tau^{\rm int}_{E;E}$	$\tau^{\rm int}_{Ew;Ew}$	$\tau^{\rm int}_{w;w}$	$\tau^{\rm int}_{Ew;w}$	$\tau^{\rm eff}_{e}$	$\tau^{\rm jack}_{e}$	$\tau^{\rm flip}_{e}$
8	71(5)	39(2)	9(1)	18(1)	77(5)	63	119(2)
10	95(10)	62(3)	38(2)	50(3)	107(8)	103	181(3)
12	148(12)	81(3)	53(2)	66(3)	189(34)	205	298(4)
14	229(24)	105(3)	74(2)	87(3)	316(21)	326	468(6)
16	303(27)	131(3)	100(4)	113(3)	488(65)	498	655(10)
20	584(41)	206(5)	166(4)	183(4)	940(58)	1009	1298(18)
24	1008(427)	280(10)	239(8)	256(11)	1607(473)	1471	2434(82)
30	2730(1293)	340(15)	334(13)	324(13)	3085(173)	4972	5429(238)
L	$\tau^{\rm int}_{B;B}$	$\tau^{\rm int}_{Bw;Bw}$	$\tau^{\rm int}_{w;w}$	$\tau^{\rm int}_{Bw;w}$	$\tau^{\rm eff}_{b}$	$\tau^{\rm jack}_{b}$	$\tau^{\rm flip}_{b}$
8	69(5)	40(2)	9(1)	18(1)	75(5)	62	106(2)
10	93(10)	62(3)	38(2)	50(3)	107(7)	102	177(3)
12	146(12)	81(3)	53(2)	66(3)	189(34)	204	286(4)
14	225(23)	104(3)	74(2)	87(3)	317(21)	327	451(6)
16	299(27)	131(3)	100(4)	113(3)	490(64)	498	649(10)
20	578(41)	206(5)	166(4)	183(4)	941(58)	1009	1281(18)
24	1005(427)	280(10)	239(8)	256(10)	1611(449)	1479	2418(83)
30	2723(1364)	340(15)	334(13)	324(13)	3081(172)	4978	5429(237)

Table 1. Autocorrelation times in multibondic simulations of the three-dimensional 4-state Potts model. Error estimates are obtained with the Jackknife method on the basis of 100 blocks for $L = 8 - 20$, 50 blocks for $L = 24$, and 40 blocks for $L = 30$.

such that the squared error (62) can be written in the usual form

$$\epsilon^2_{\hat{\mathcal{O}}} \equiv \sigma^2_{\hat{\mathcal{O}}} \frac{2\tau_{\mathcal{O}}}{N} \ , \tag{64}$$

with $\tau_{\mathcal{O}}$ collecting the various autocorrelation times in an averaged sense. For a comparison with canonical simulations we need one further step since

$$(\epsilon^2_{\hat{\mathcal{O}}})^{\rm can} = \langle \overline{\mathcal{O}}; \overline{\mathcal{O}} \rangle_{\rm can} = (\sigma^2_{\mathcal{O}_i})^{\rm can} \frac{2\tau^{\rm can}_{\mathcal{O}}}{N} \tag{65}$$

but $\sigma^2_{\hat{\mathcal{O}}} \neq (\sigma^2_{\mathcal{O}_i})^{\rm can} = \langle \mathcal{O}_i; \mathcal{O}_i \rangle$. Hence we define an effective autocorrelation time $\tau^{\rm eff}_{\mathcal{O}}$ through

$$\epsilon^2_{\hat{\mathcal{O}}} = (\sigma^2_{\mathcal{O}_i})^{\rm can} \frac{2\tau^{\rm eff}_{\mathcal{O}}}{N} = (\epsilon^2_{\hat{\mathcal{O}}})^{\rm can} \frac{\tau^{\rm eff}_{\mathcal{O}}}{\tau^{\rm can}_{\mathcal{O}}} \ , \tag{66}$$

i.e.,

$$\tau^{\rm eff}_{\mathcal{O}} = \frac{\sigma^2_{\hat{\mathcal{O}}}}{(\sigma^2_{\mathcal{O}_i})^{\rm can}} \tau_{\mathcal{O}} \ . \tag{67}$$

For symmetric distributions and *odd* observables we have $\langle \mathcal{O}_i w_i \rangle \equiv 0$ and this simplifies to

$$\epsilon^2_{\hat{\mathcal{O}}} = \frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle w_i \rangle^2} \, 2\tau^{\rm int}_{\mathcal{O}w;\mathcal{O}w} \ , \tag{68}$$

3D 4-state Potts model, multibondic simulations						
L	$\frac{<Ew;Ew>}{<Ew>^2}$	$\frac{<w;w>}{<w>^2}$	$\frac{<Ew;w>}{<Ew><w>}$	$\sigma_e^2 \times 10^2$	$\varepsilon_e \times 10^2$	$\varepsilon_e^{\rm jack} \times 10^2$
8	0.11208(70)	0.05993(28)	0.07032(46)	9.751(67)	0.462(13)	0.417
10	0.2429(25)	0.15050(89)	0.1799(16)	8.363(70)	0.423(15)	0.415
12	0.4471(52)	0.3418(26)	0.3759(37)	7.731(71)	0.383(45)	0.398
14	0.6555(76)	0.5422(42)	0.5790(57)	7.331(83)	0.393(14)	0.399
16	1.047(14)	0.9210(87)	0.961(11)	6.988(92)	0.369(25)	0.373
20	1.561(22)	1.433(15)	1.468(18)	6.992(94)	0.363(13)	0.376
24	2.175(71)	2.038(56)	2.077(62)	6.15(22)	0.70(11)	0.673
30	2.555(91)	2.328(65)	2.381(73)	8.11(18)	1.119(32)	1.420

Table 2. Variances, covariances and expectation values of multibondic simulations of the three-dimensional 4-state Potts model, which enter the effective error estimate ε_e. σ_e^2 is the canonical variance of the energy. For the number of Jackknife blocks used in the computation of $\varepsilon_e^{\rm jack}$ see Table 1.

such that

$$\tau_{\mathcal{O}} = \tau^{\rm int}_{\mathcal{O}w;\mathcal{O}w} \ , \tag{69}$$

and

$$\tau^{\rm eff}_{\mathcal{O}} = \frac{\sigma^2_{\hat{\mathcal{O}}}}{(\sigma^2_{\hat{\mathcal{O}}_i})^{\rm can}} \tau^{\rm int}_{\mathcal{O}w;\mathcal{O}w} \ . \tag{70}$$

These formulas have been tested and compared with a standard Jackknife analysis in Refs. 26 and 29. As an example the various quantities computed are shown in Tables 1 and 2 for the case of the 3D 4-state Potts model simulated with the multibondic algorithm,[30,31] a cluster algorithm variant of the multicanonical method. Here E denotes as usual the energy and B is the number of active bonds.

7 Summary

Thanks to the great improvements in Monte Carlo simulation methodology over the last decade, the results for at least certain model classes have reached such a high degree of precision that careful and reliable statistical error analyses become more and more important. The interplay of autocorrelations and correlations between basic observables in non-linear combined quantities requires rather involved analyses which are well understood in principle but still a bit cumbersome to implement in practice. Still, after spending months or even years of computer time for the generation of the raw data, it is certainly worth spending this comparatively little extra effort; since a Monte Carlo simulation is a stochastic method, trustworthy error estimates are an important part of the final results.

Acknowledgments

I would like to thank Bernd Berg, Stefan Kappler, Tilman Sauer and Martin Weigel for enjoyable collaborations and many useful discussions on various aspects of this lecture.

References

1. D. Frenkel and B. Smit, *Understanding Molecular Simulation – From Algorithms to Applications* (Academic Press, San Diego, 1996).
2. D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
3. There is by now a huge amount of Ising model material available on the World Wide Web, including animations. For a list of useful links, see e.g. `http://oscar.cacr.caltech.edu/Hrothgar/Ising/references.html`.
4. W. Janke, *Monte Carlo Simulations of Spin Systems*, in: *Computational Physics: Selected Methods – Simple Exercises – Serious Applications*, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996), p. 10.
5. R.B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).
6. F.Y. Wu, Rev. Mod. Phys. **54**, 235 (1982); *ibid.* **55**, 315(E) (1983).
7. W. Janke and R. Villanova, Nucl. Phys. **B489**, 679 (1997).
8. W. Janke, *Nonlocal Monte Carlo Algorithms for Statistical Physics Applications*, Mathematics and Computers in Simulations **47**, 329 (1998).
9. 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, 1989.
10. A.D. Sokal, *Bosonic Algorithms*, in: *Quantum Fields on the Computer*, ed. M. Creutz (World Scientific, Singapore, 1992), p. 211.
11. 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), p. 269.
12. K. Binder, Rep. Prog. Phys. **50**, 783 (1987).
13. H.J. Herrmann, W. Janke, and F, Karsch (eds.), *Dynamics of First Order Phase Transitions* (World Scientific, Singapore, 1992).
14. 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.
15. B.A. Berg and T. Neuhaus, Phys. Lett. **B267**, 249 (1991).
16. B.A. Berg and T. Neuhaus, Phys. Rev. Lett. **69**, 9 (1992).
17. W. Janke, in Ref. 13, p. 365.
18. B.A. Berg, Fields Inst. Commun. **26**, 1 (2000).
19. W. Janke, *Monte Carlo Methods for Sampling of Rare Event States*, in: *Classical and Quantum Dynamics in Condensed Phase Simulations*, Proceedings of the International School of Physics "Computer Simulation of Rare Events and the Dynamics of Classical and Quantum Condensed-Phase Systems" and Euroconference on "Technical advances in Particle-based Computational Material Sciences", Lerici, Italy, July 7 – 18, 1997, eds. B. Berne, G. Ciccotti, and D. Coker (World Scientific, Singapore, 1998), p. 195.
20. W. Janke, *Multicanonical Monte Carlo Simulations*, Physica **A254**, 164 (1998).
21. M.B. Priestley, *Spectral Analysis and Time Series*, 2 vols. (Academic, London, 1981), Chapters 5-7.
22. T.W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).

23. N. Madras and A.D. Sokal, J. Stat. Phys. **50**, 109 (1988).
24. A.D. Sokal and L.E. Thomas, J. Stat. Phys. **54**, 797 (1989).
25. A.M. Ferrenberg, D.P. Landau, and K. Binder, J. Stat. Phys. **63**, 867 (1991).
26. W. Janke and T. Sauer, J. Stat. Phys. **78**, 759 (1995).
27. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics [SIAM], Philadelphia, 1982).
28. R.G. Miller, Biometrika **61**, 1 (1974).
29. M.S. Carroll, W. Janke, and S. Kappler, J. Stat. Phys. **90**, 1277 (1998).
30. W. Janke and S. Kappler, Phys. Rev. Lett. **74**, 212 (1995).
31. W. Janke, *Multicanonical Multigrid and Multibondic Algorithms*, in: *Multiscale Phenomena and Their Simulation*, Proceedings of the International Conference, Bielefeld, September 30 – October 4, 1996, eds. F. Karsch, B. Monien, and H. Satz (World Scientific, Singapore, 1997), p. 147.