4 Importance sampling Monte Carlo methods

4.1 INTRODUCTION

In this chapter we want to introduce simple importance sampling Monte Carlo techniques as applied in statistical physics and which can be used for the study of phase transitions at finite temperature. We shall discuss details, algorithms, and potential sources of difficulty using the Ising model as a paradigm. It should be understood, however, that virtually all of the discussion of the application to the Ising model is relevant to other models as well, and a few such examples will also be discussed. Other models as well as sophisticated approaches to the Ising model will be discussed in later chapters. The Ising model is one of the simplest lattice models which one can imagine, and its behavior has been studied for about three-quarters of a century. The simple Ising model consists of spins which are confined to the sites of a lattice and which may have only the values $+1$ or -1. These spins interact with their nearest neighbors on the lattice with interaction constant $\mathcal{J}$; the Hamiltonian for this model was given in Eqn. (2.24) but we repeat it again here for the benefit of the reader:

$$\mathcal{H} = -\mathcal{J} \sum_{i,j} \sigma_i \sigma_j - H \sum_i \sigma_i \tag{4.1}$$

where $\sigma_i = \pm 1$. The Ising model has been solved exactly in one dimension and as a result it is known that there is no phase transition. In two dimensions Onsager obtained exact results (Onsager, 1944) for the thermal properties of $L \times M$ lattices with periodic boundary conditions in zero field which showed that there is a second order phase transition with divergences in the specific heat, susceptibility, and correlation length. In Fig. 4.1 we show configurations for finite $L \times L$ Ising lattices in zero field; these states show the model almost in the groundstate, near the phase transition, and at high temperatures where there are virtually no correlations between spins. Note that in zero field the model has up–down symmetry so that overturning all the spins produces a degenerate state. At high temperature all the clusters of like spins are small, near the transition there is a broad distribution of clusters, and at low temperatures there is a single large cluster of ordered spins and a number of small clusters of oppositely directed spins.

In principle, the Ising model can be simulated using the simple sampling techniques discussed in the previous chapter: spin configurations

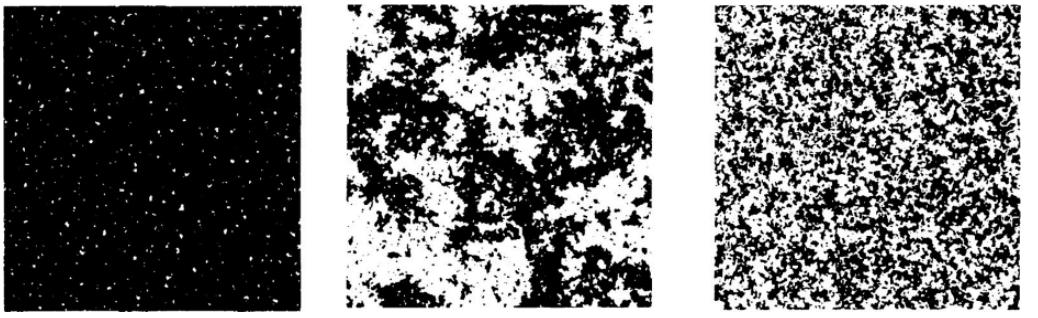

Fig. 4.1 Typical spin configurations for the two-dimensional Ising square lattice: (left) $T \ll T_c$; (center) $T \sim T_c$; (right) $T \gg T_c$.

could be generated completely randomly and their contribution weighted by a Boltzmann factor. Unfortunately most of the configurations which are produced in this fashion will contribute relatively little to the equilibrium averages, and more sophisticated methods are required if we are to obtain results of sufficient accuracy to be useful.

Problem 4.1 Suppose we carry out a simple sampling of the Ising model configurations on an $L \times L$ lattice at $k_B T/J = 1.5$. What is the distribution of the magnetization M of the states that are generated? How large is the probability that a state has a magnetization $M > M_0$, where M_0 is some given value of order unity, e.g. the spontaneous magnetization for $T < T_c$. Use your result to explain why simple sampling is not useful for studying the Ising model.

4.2 THE SIMPLEST CASE: SINGLE SPIN-FLIP SAMPLING FOR THE SIMPLE ISING MODEL

The nearest neighbor Ising model on the square lattice plays a special role in statistical mechanics – its energy, spontaneous magnetization, and correlations in zero magnetic field can be calculated exactly, and this fact implies that the static critical exponents are also known. Critical exponents are known exactly for only a small number of models. The most notable of the exactly soluble models is the two-dimensional Ising square lattice (Onsager, 1944) for which the exact solution shows that the critical exponents which were discussed in Chapter 2 are

$$\alpha = 0, \qquad \beta = 1/8, \qquad \text{and} \qquad \gamma = 7/4. \tag{4.2}$$

We shall first discuss techniques which are suitable for simulating this model so that there are exact results with which the data from the Monte Carlo simulations may be compared.

4.2.1 Algorithm

In the classic, Metropolis method (Metropolis *et al.*, 1953) configurations are generated from a previous state using a transition probability which depends on the energy difference between the initial and final states. The sequence of states produced follows a time ordered path, but the time in this case is referred to as 'Monte Carlo time' and is non-deterministic. (This can be seen from an evaluation of the commutator of the Hamiltonian and an arbitrary spin; the value, which gives the time dependence of the spin, is zero.) For relaxational models, such as the (stochastic) Ising model (Kawasaki, 1972), the time-dependent behavior is described by a master equation (cf. Section 2.2.4)

$$\frac{\partial P_n(t)}{\partial t} = -\sum_{n \neq m} [P_n(t) W_{n \to m} - P_m(t) W_{m \to n}], \tag{4.3}$$

where $P_n(t)$ is the probability of the system being in state n at time t, and $W_{n \to m}$ is the transition rate for $n \to m$. In equilibrium $\partial P_n(t)/\partial t = 0$ and the two terms on the right-hand side of Eqn. (4.3) must be equal. The resultant expression is known as 'detailed balance', as mentioned previously in Eqn. (2.89)

$$P_n(t) W_{n \to m} = P_m(t) W_{m \to n}. \tag{4.4}$$

The probability of the nth state occurring in a classical system is given by

$$P_n(t) = e^{-E_n / k_B T} / Z, \tag{4.5}$$

where Z is the partition function. This probability is usually not exactly known because of the denominator; however, one can avoid this difficulty by generating a Markov chain of states, i.e. generate each new state directly from the preceding state. If we produce the nth state from the mth state, the relative probability is the ratio of the individual probabilites and the denominator cancels. As a result, only the energy difference between the two states is needed, e.g.

$$\Delta E = E_n - E_m. \tag{4.6}$$

Any transition rate which satisfies detailed balance is acceptable. The first choice of rate which was used in statistical physics is the Metropolis form (Metropolis *et al.*, 1953)

$$W_{m \to n} = \tau_o^{-1} \exp(-\Delta E / k_B T) \qquad \Delta E > 0 \tag{4.7a}$$

$$= \tau_o^{-1} \qquad \Delta E < 0, \tag{4.7b}$$

where τ_o is the time required to attempt a spin-flip. (Often this 'time unit' is set equal to unity and hence suppressed in the equations.) The way the Metropolis algorithm is implemented can be described by a simple recipe:

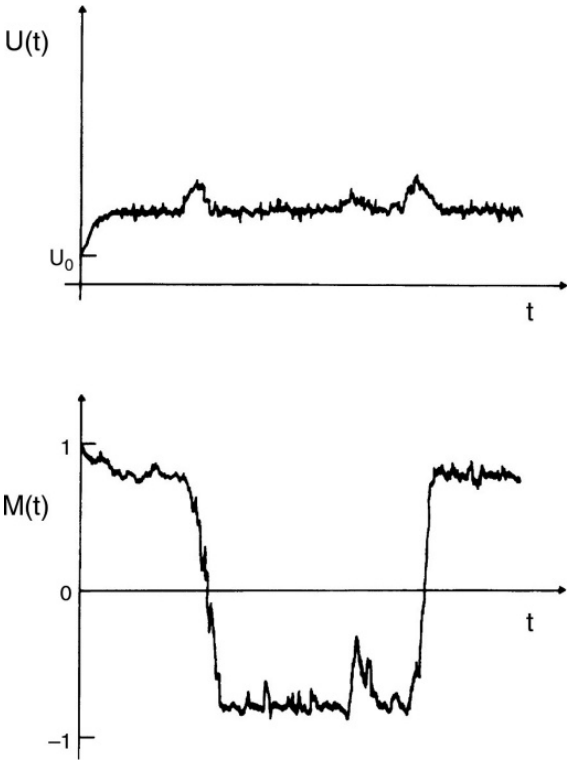

Fig. 4.2 Schematic variation of internal energy and spontaneous magnetization with time for a Monte Carlo simulation of an Ising square lattice in zero field.

Metropolis importance sampling Monte Carlo scheme

(1) Choose an initial state.
(2) Choose a site i.
(3) Calculate the energy change ΔE which results if the spin at site i is overturned.
(4) Generate a random number r such that $0 < r < 1$.
(5) If $r < \exp(-\Delta E/k_{\mathrm{B}}T)$, flip the spin.
(6) Go to the next site and go to (3).

After a set number of spins have been considered, the properties of the system are determined and added to the statistical average which is being kept. Note that the random number r must be chosen *uniformly* in the interval [0,1], and successive random numbers should be uncorrelated. We shall have a great deal more to say about random numbers shortly. The 'standard' measure of Monte Carlo time is the Monte Carlo step/site (MCS/site) which corresponds to the consideration of every spin in the system once. With this algorithm states are generated with a probability proportional to Eqn. (4.5) once the number of states is sufficiently large that the initial transients (see Fig. 4.2) are negligible. Then, the desired averages $\langle A \rangle = \sum_n P_n A_n$ of a variable A simply become arithmetic averages over the entire sample of states which is kept. Note that

if an attempted spin-flip is rejected, the old state is counted again for the averaging.

A typical time development of properties of the system is shown in Fig. 4.2. For early times the system is relaxing towards equilibrium and both the internal energy and order parameter are changing, but with different characteristic time scales. There is a second range of times in which the system is in equilibrium and the properties merely show thermodynamic fluctuations, and at still longer times one can observe global spin inversion; in a finite system this will occur in equilibrium between states of equal energy and spontaneous magnetization which differs only in sign. Of course, the precise results will depend upon many factors including temperature, lattice size, boundary conditions, etc., and all of these considerations will be discussed in forthcoming sections. Figure 4.2 simply provides a starting point for these presentations. In a more complex problem one might not know what the groundstate looks like or what the relevant time scales are. It is thus always wise to take precautions before interpreting the data. Prudent steps to take include repeating a given run with different initial states to see if the same equilibrium distribution is reached and to repeat runs with different random numbers. By working initially with small systems one can generally keep the characteristic times short so that it is easy to make 'long' runs.

A minor variation on the simple Metropolis algorithm described above involves the random selection of sites in the lattice to be considered. If this procedure is used for a system with N sites, 1 MCS/site corresponds to the consideration of N randomly chosen sites. Note that it is likely that some spins will be chosen more than once and some not at all during 1 MCS/site. The time development of the system will look just like that shown in Fig. 4.2, but the explicit variation and time scales will differ from those for the Metropolis method. This random selection of sites must be used if one is not just interested in static equilibrium properties but wishes to record dynamic correlation functions of the corresponding stochastic model.

As shown in the 'principle of detailed balance', Eqn. (4.4), the Metropolis flipping probability is not a unique solution. An alternative method, commonly referred to as 'Glauber dynamics' (Glauber, 1963), uses the single spin-flip transition rate

$$W_{n\to m} = (2\tau_o)^{-1}[1 + \sigma_i \tanh(E_i / k_B T)], \tag{4.8}$$

where $\sigma_i E_i$ is the energy of the ith spin in state n. Unlike the Metropolis method, the Glauber rate is antisymmetric about 0.5 for $E_i \to -E_i$. Müller-Krumbhaar and Binder (1973) showed that both Glauber and Metropolis algorithms are just special cases of a more general transition rate form. In most situations the choice between Glauber and Metropolis dynamics is somewhat arbitrary; but in at least one instance there is a quite important difference. At very high temperatures the Metropolis algorithm will flip a spin on every attempt because the transition probability approaches 1 for $\Delta E > 0$. Thus, in one sweep through the lattice every spin overturns, and in the next sweep every spin overturns again. The process has thus become non-ergodic (see Section 2.1.3) and the system just oscillates between the two states. With the

Glauber algorithm, however, the transition probability approaches 1/2 in this instance and the process remains ergodic.

Simplifications are possible for the Ising model which greatly reduce the amount of computer resources needed. For each spin there are only a small number of different environments which are possible, e.g. for a square lattice with nearest neighbor interactions, each spin may have 4, 3, 2, 1, or 0 nearest neighbors which are parallel to it. Thus, there are only five different energy changes associated with a successful spin-flip and the probability can be computed for each possibility and stored in a table. Since the exponential then need not be computed for each spin-flip trial, a tremendous saving in CPU time results. Although the rapid increase in available computer memory has largely alleviated the problem with storage, large Ising systems may be compressed into a relatively small number of words by packing different spins into a single word. Each bit then describes the state of a spin so that e.g. only a single 32-bit word is needed to describe a 32-spin system. For models with more degrees of freedom available at each site, these simplifications are not possible and the simulations are consequently more resource consumptive.

The Ising model as originally formulated and discussed above may be viewed as a spin-S model with $S = 1/2$, but the definition can be extended to the case of higher spin without difficulty. For spin $S = 1/2$ there are only two states possible at each site, whereas for $S = 1$ there are three possible states, 1, 0, and -1. This means that a nearest neighbor pair can have *three* possible states with different energies and the total space of possible lattice configurations is similarly enlarged. (For higher values of S there will, of course, be still more states.) The spin-S Ising model can be simulated using the method just described with the modification that the 'new' state to which a given spin attempts to flip must be chosen from among multiple choices using another random number. After this is done, one proceeds as before.

One feature of a Monte Carlo algorithm which is important if the method is to be vectorized (these techniques will be discussed in the next chapter) is that the lattice needs to be subdivided into non-interacting, interpenetrating sublattices, i.e. so that the spins on a single sublattice do not interact with each other. This method, known as the 'checkerboard decomposition', can be used without difficulty on scalar computers as well as on vector platforms. If one wishes to proceed through the lattice in order using the checkerboard decomposition, one simply examines each site in turn in a single sublattice before proceeding to the second sublattice. (We mention this approach here simply because the checkerboard decomposition is referred to quite often in the literature.)

4.2.2 Boundary conditions

4.2.2.1 Periodic boundary conditions

Since simulations are performed on finite systems, one important question which arises is how to treat the 'edges' or boundaries of the lattice. These

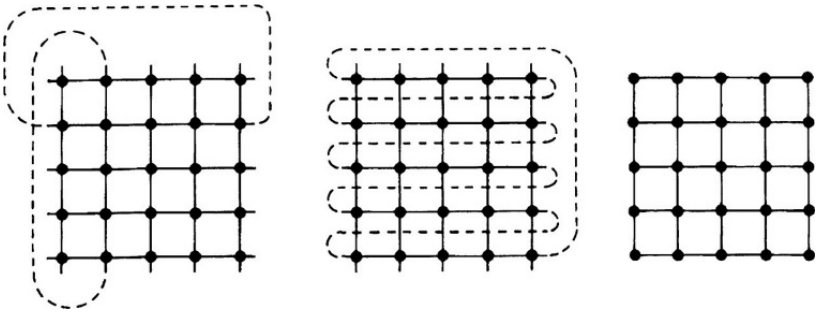

Fig. 4.3 Application of typical boundary conditions for the two-dimensional Ising model: (left) periodic boundary; (center) screw periodic; (right) free edges.

boundaries can be effectively eliminated by wrapping the d-dimensional lattice on a $(d+1)$-dimensional torus. This boundary condition is termed a 'periodic boundary condition' (pbc) so that the first spin in a row 'sees' the last spin in the row as a nearest neighbor and vice versa. The same is true for spins at the top and bottom of a column. Figure 4.3 shows this procedure for a square lattice. This procedure effectively eliminates boundary effects, but the system is still characterized by the finite lattice size L since the maximum value of the correlation length is limited to $L/2$, and the resultant properties of the system differ from those of the corresponding infinite lattice. (These effects will be discussed at length in the next section.) The periodic boundary condition must be used with care, since if the ordered state of the system has spins which alternate in sign from site to site, a 'misfit seam' can be introduced if the edge length is not chosen correctly. Of course, for off-lattice problems periodic boundary conditions are also easily introduced and equally useful for the elimination of edge effects.

4.2.2.2 Screw periodic boundary conditions

The actual implementation of a 'wraparound' boundary condition is easiest by representing the spins on the lattice as entries in a one-dimensional vector which is wrapped around the system. Hence the last spin in a row sees the first spin in the next row as a nearest neighbor (see Fig. 4.3). In addition to limiting the maximum possible correlation length, a result of this form of periodic boundary is that a 'seam' is introduced. This means that the properties of the system will not be completely homogeneous. In the limit of infinite lattice size this effect becomes negligible, but for finite systems there will be a systematic difference with respect to fully periodic boundary conditions which may not be negligible.

4.2.2.3 Antiperiodic boundary conditions

If periodic boundary conditions are imposed with the modification that the sign of the coupling is reversed at the boundary, an interface is introduced into the system. This procedure, known as antiperiodic boundary conditions, is not useful for making the system seem more infinite, but has the salutory effect

of allowing us to work with a single interface in the system. (With periodic boundary conditions interfaces could only exist in pairs.) In this situation the interface is not fixed at one particular location and may wander back and forth across the boundary. By choosing a coordinate frame centered in the local interface center one can nevertheless study the interfacial profile undisturbed by any free edge effects (Schmid and Binder, 1992a, 1992b). Of course, one chooses this antiperiodic boundary condition in only one (lattice) direction, normal to the interface that one wishes to study, and retains periodic boundary conditions in the other direction(s).

In the above example the interface was parallel to one of the surfaces, whereas in a more general situation the interface may be inclined with respect to the surface. This presents no problem for simulations since a tilted interface can be produced by simply taking one of the periodic boundaries and replacing it by a skew boundary. Thus, spins on one side of the lattice see nearest neighbors on the other side which are one or more rows below, depending on the tilt angle of the interface. We then have the interesting situation that the boundary conditions are different in each Cartesian direction and are themselves responsible for the change in the nature of the problem being studied by a simple Monte Carlo algorithm. This is but one example of the clever use of boundary conditions to simplify a particular problem; the reader should consider the choice of the boundary conditions before beginning a new study.

4.2.2.4 Antisymmetric boundary conditions

This type of periodic boundary condition was introduced explicitly for $L \times L$ systems with vortices. (Vortices are topological excitations that occur most notably in the two-dimensional XY-model, see e.g. Section 5.3.9. A vortex looks very much like a whirlpool in two-dimensional space.) By connecting the last spin in row n antiferromagnetically with the first spin in row $(L - n)$, one produces a geometry in which a single vortex can exist; in contrast with pbc only vortex–antivortex pairs can exist (Kawamura and Kikuchi, 1993) on a lattice. This is a quite specialized boundary condition which is only useful for a limited number of cases, but it is an example of how specialized boundaries can be used for the study of unusual excitations.

4.2.2.5 Free edge boundary conditions

Another type of boundary does not involve any kind of connection between the end of a row and any other row on the lattice. Instead the spins at the end of a row see no neighbor in that direction (see Fig. 4.3). This free edge boundary not only introduces finite size smearing but also surface and corner effects due to the 'dangling bonds' at the edges. (Very strong changes may occur near the surfaces and the behavior of the system is not homogeneous.) In some cases, however, the surface and corner behavior themselves become the subjects of study. In some situations free edge boundaries may be more realistic, e.g. in modeling the behavior of superparamagnetic particles or grains,

but the properties of systems with free edge boundaries usually differ from those of the corresponding infinite system by a much greater amount than if some sort of periodic boundary is used. In order to model thin films, one uses pbc in the directions parallel to the film and free edge boundary conditions in the direction normal to the film. In such cases, where the free edge boundary condition is thought to model a physical free surface of a system, it may be appropriate to also include surface fields, modified surface layer interactions, etc. (Landau and Binder, 1990). In this way, one can study phenomena such as wetting, interface localization–delocalization transitions, surface induced ordering and disordering, etc. This free edge boundary condition is also very common for off-lattice problems (Binder, 1983; Landau, 1996).

4.2.2.6 Mean-field boundary conditions

Another way to reduce finite size effects is to introduce an effective field which acts only on the boundary spins and which is adjusted to keep the magnetization at the boundary equal to the mean magnetization in the bulk. The resultant critical behavior is quite sharp, although sufficiently close to T_c the properties are mean-field-like. Such boundary conditions have been applied only sparingly, e.g. for Heisenberg magnets in the bulk (Binder and Müller-Krumbhaar, 1973) and with one free surface (Binder and Hohenberg, 1974).

4.2.2.7 Hyperspherical boundary conditions

In the case of long range interactions, periodic boundary conditions may become cumbersome to apply because each degree of freedom interacts with all its periodic images. In order to sum up the interactions with all periodic images, one has to resort to the Ewald summation method (see Chapter 6). An elegant alternative for off-lattice problems is to put the degrees of freedom on the d-dimensional surface of a $(d + 1)$-dimensional sphere (Caillol, 1993).

Problem 4.2 Perform a Metropolis Monte Carlo simulation for a 10×10 Ising model with periodic boundary conditions. Plot the specific heat (calculated from the fluctuations of the internal energy, see Chapter 2) and the order parameter (estimated as the absolute value of the magnetization) as a function of temperature.

Problem 4.3 Perform a Metropolis Monte Carlo simulation for a 10×10 Ising model with free edge boundary conditions. Plot the specific heat and the order parameter as a function of temperature.

4.2.3 Finite size effects

4.2.3.1 Order of the transition

In the above discussion we have briefly alluded to the fact that the effects of the finiteness of the system could be dramatic. (The reader who has actually worked

out Problems 4.2 and 4.3 will have noted that in a 10×10 lattice the transition is completely smeared out!) Since our primary interest is often in determining the properties of the corresponding infinite system, it is important that we have some sound, theoretically based methods for extracting such behavior for the results obtained on the finite system. One fundamental difficulty which arises in interpreting simulational data, is that the equilibrium, thermodynamic behavior of a finite system is smooth as it passes through a phase transition for *both* first order and second order transitions. The question then becomes, 'How do we distinguish the order of the transition?' In the following sections we shall show how this is possible using finite size scaling.

4.2.3.2 Finite size scaling and critical exponents

At a second order phase change the critical behavior of a system in the thermodynamic limit can be extracted from the size dependence of the singular part of the free energy which, according to finite size scaling theory (Fisher, 1971; Privman, 1990; Binder, 1992), is described by a scaling ansatz similar to the scaling of the free energy with thermodynamic variables T, H (see Chapter 2). Assuming homogeneity and using L and T as variables, we find for the singular part of the free energy that

$$F(L, T) = L^{-(2-\alpha)/\nu} \mathcal{F}(\varepsilon L^{1/\nu}), \tag{4.9}$$

where $\varepsilon = (T - T_c)/T_c$. It is important to note that the critical exponents α and ν assume their infinite lattice values. The choice of the scaling variable $x = \varepsilon L^{1/\nu}$ is motivated by the observation that the correlation length, which diverges as $\varepsilon^{-\nu}$ as the transition is approached, is limited by the lattice size L. (L 'scales' with ξ; but rather than $L/\xi \propto \varepsilon^{\nu} L$, one may also choose $\varepsilon L^{1/\nu}$ as the argument of the function $\mathcal{F}$. This choice has the advantage that F is analytic since $\mathcal{F}$ is analytic in T for finite L.) Appropriate differentiation of the free energy yields the various thermodynamic properties which have corresponding scaling forms, e.g.

$$M = L^{-\beta/\nu} \mathcal{M}^{o}(\varepsilon L^{1/\nu}), \tag{4.10a}$$

$$\chi = L^{\gamma/\nu} \chi^{o}(\varepsilon L^{1/\nu}), \tag{4.10b}$$

$$C = L^{\alpha/\nu} C^{o}(\varepsilon L^{1/\nu}), \tag{4.10c}$$

where $\mathcal{M}^{o}(x)$, $\chi^{o}(x)$, and $C^{o}(x)$ are scaling functions. In deriving these relations, Eqns. (4.10a–c), one actually uses a second argument $HL^{(\gamma+\beta)/\nu}$ in the scaling function $\mathcal{F}$ in Eqn. (4.9), where H is the field conjugate to the order parameter. After the appropriate differentiation has been completed H is then set to zero. Scaling relations such as $2 - \alpha = \gamma + 2\beta$ are also used. Note that the finite size scaling ansatz is valid only for sufficiently large L and temperatures close to T_c. Corrections to scaling and finite size scaling must be taken into account for smaller systems and temperatures away from T_c. Because of the complexity of the origins of these corrections they are not discussed in detail here; readers are directed elsewhere (Liu and Fisher, 1990; Ferrenberg and

Landau, 1991) for a detailed discussion of these corrections and techniques for including them in the analysis of Monte Carlo data. As an example of finite size behavior, in Fig. 4.4 we show data for the spontaneous magnetization of $L \times L$ Ising square lattices with pbc. The raw data are shown in the top portion of the figure, and a finite size scaling plot, made with the exact values of the critical temperature and critical exponents is shown in the bottom portion of the figure. Note that the large scatter of data points in this plot is characteristic of early Monte Carlo work – the computational effort entailed in producing these data from Landau (1976) is easily within the capability of everyone's PC today, and with any moderately fast workstation *today* one can do far better. Exactly at the transition the thermodynamic properties then all exhibit power law behavior, since the scaling functions $\mathcal{M}^o(0), \chi^o(0), C^o(0)$ just reduce to proportionality constants, i.e.

$$M \propto L^{-\beta/\nu}, \tag{4.11a}$$

$$\chi \propto L^{\gamma/\nu}, \tag{4.11b}$$

$$C \propto L^{\alpha/\nu} (C \propto \ln L \text{ if } \alpha = 0), \tag{4.11c}$$

which can be used to extract estimates for the ratio of certain critical exponents. The power law behavior for the order parameter is verified in Fig. 4.4 (bottom) directly noting that for small x all data approach a constant, which is then an estimate of $\mathcal{M}^o(0)$. Note that the scaling functions that appear in Eqn. (4.10) are universal, apart from scale factors for their arguments. The prefactors in Eqn. (4.11) are thus also of interest for the estimation of universal amplitude ratios (Privman *et al.*, 1991).

In addition to these quantities, which are basically just first or second order moments of the probability distribution of order parameter or energy, we may obtain important, additional information by examining higher order moments of the finite size lattice probability distribution. This can be done quite effectively by considering the reduced fourth order cumulant of the order parameter (Binder, 1981). For an Ising model in zero field, for which all odd moments disappear by symmetry, the fourth order cumulant simplifies to

$$U_4 = 1 - \frac{\langle m^4 \rangle}{3 \langle m^2 \rangle^2}. \tag{4.12}$$

As the system size $L \to \infty$, $U_4 \to 0$ for $T > T_c$ and $U_4 \to 2/3$ for $T < T_c$. For large enough lattice size, curves for U_4 cross as a function of temperature at a 'fixed point' value U^* (our terminology here is used in a renormalization group sense, where the rescaling transformation $L' = bL$ with a scale factor $b > 1$ is iterated) and the location of the crossing 'fixed point' is the critical point. Hence, by making such plots for different size lattices one can make a preliminary identification of the universality class from the value of U_4^* and obtain an estimate for T_c from the location of the crossing point. Of course, if the sizes used are too small, there will be correction terms present which prevent all the curves from having a common intersection. Nonetheless there should then be a systematic variation with increasing lattice size towards a

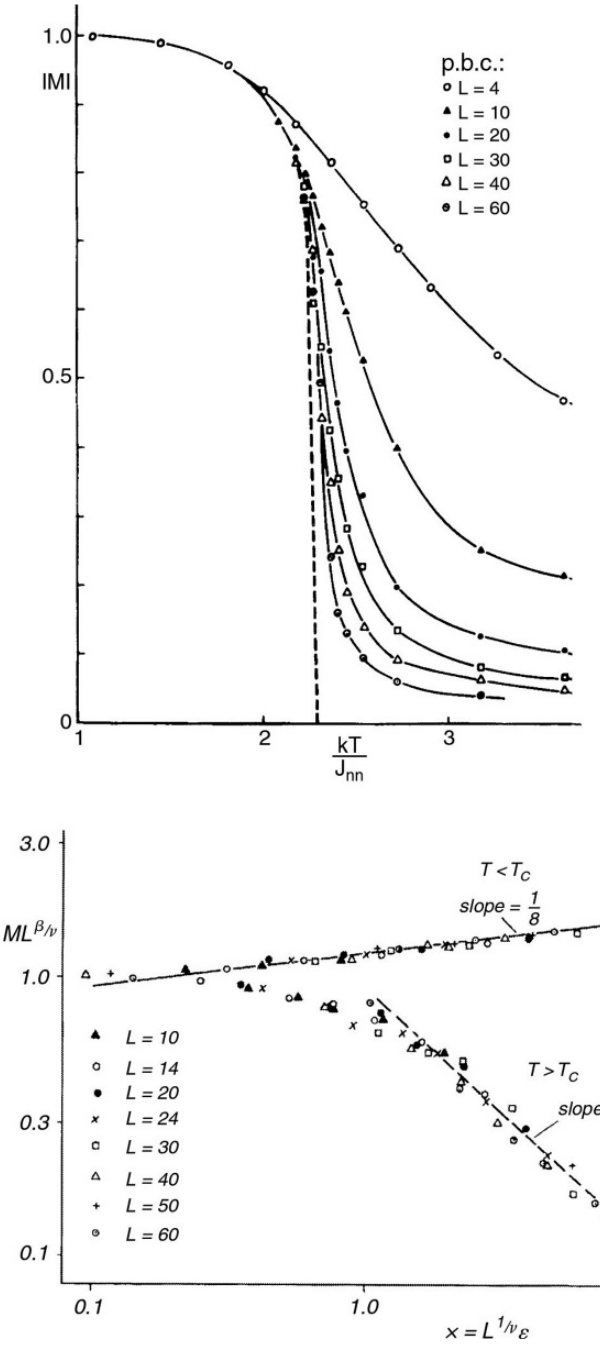

Fig. 4.4 (top) Spontaneous magnetization for $L \times L$ Ising square lattices with periodic boundary conditions; (bottom) finite size scaling plot for the data shown at the top. From Landau (1976).

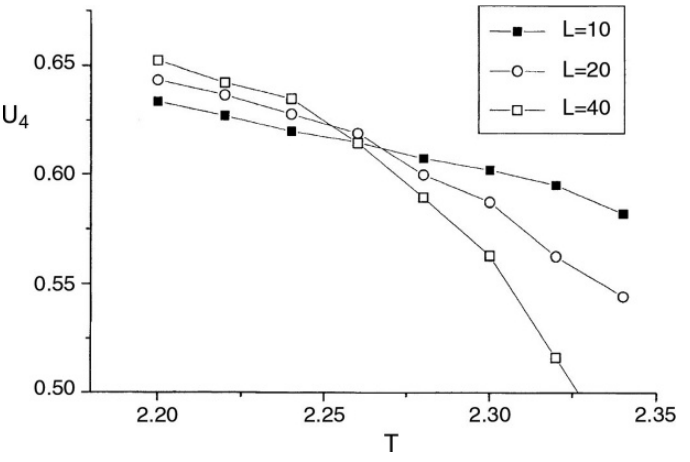

Fig. 4.5 Temperature dependence of the fourth order cumulant for $L \times L$ Ising square lattices with periodic boundary conditions.

common intersection. (The same kind of behavior will also be seen for other models, although the locations of the crossings and values of U_4 will obviously be different.) An example of the behavior which can be expected is shown in Fig. 4.5 for the Ising square lattice in zero field.

Another technique which can be used to determine the transition temperature very accurately relies on the location of peaks in thermodynamic derivatives, for example the specific heat. For many purposes it is easier to deal with inverse temperature so we define the quantity $K = \mathcal{J}/k_B T$ and use K for much of the remainder of this discussion. The location of the peak defines a finite-lattice (or effective) transition temperature $T_c(L)$, or equivalently $K_c(L)$, which, taking into account a correction term of the form L^{-w}, varies with system size like

$$T_c(L) = T_c + \lambda L^{-1/\nu}(1 + bL^{-w}), \tag{4.13a}$$

$$K_c(L) = K_c + \lambda' L^{-1/\nu}(1 + b'L^{-w}), \tag{4.13b}$$

where λ, b, or λ', b' are some (model dependent) constants, and where the exponents will be the same in the two formulations but the prefactors will differ. Because each thermodynamic quantity has its own scaling function, the peaks in different thermodynamic derivatives occur at different temperatures for finite systems, some with positive λ, some with negative λ. To use Eqn. (4.13) to determine the location of the infinite lattice transition it is necessary to have both an accurate estimate for ν and accurate values for finite lattice 'transitions' $K_c(L)$. In a case where neither K_c, ν, nor w are known beforehand, a fit using Eqn. (4.13) involves five adjustable parameters. Hence, a reliable answer is only obtained if data with very good statistical accuracy are used and several quantities are analyzed simultaneously since they must all have the same K_c, ν, and w (see e.g. Ferrenberg and Landau (1991) for an example).

It has been notoriously difficult to determine ν from Monte Carlo simulation data because of the lack of quantities which provide a direct measurement. We

now understand that it is useful to examine several thermodynamic derivatives including that of the fourth order magnetization cumulant U_4 (Binder, 1981). In the finite size scaling region, the derivative varies with L like

$$\frac{\partial U_4}{\partial K} = a L^{1/\nu}(1 + bL^{-w}). \tag{4.14}$$

Additional estimates for ν can be obtained by considering less traditional quantities which should nonetheless possess the same critical properties. For example, the logarithmic derivative of the nth power of the magnetization is

$$\frac{\partial \ln \langle m^n \rangle}{\partial K} = (\langle m^n E \rangle / \langle m^n \rangle) - \langle E \rangle \tag{4.15}$$

and has the same scaling properties as the cumulant slope (Ferrenberg and Landau, 1991). The location of the maxima in these quantities also provides us with estimates for $K_c(L)$ which can be used in Eqn. (4.13) to extrapolate to K_c. For the three-dimensional Ising model consideration of the logarithmic derivatives of $|m|$ and m^2, and the derivative of the cumulant to determine ν proved to be particularly effective.

Estimates for other critical exponents, as well as additional values for $K_c(L)$, can be determined by considering other thermodynamic quantities such as the specific heat C and the finite-lattice susceptibility

$$\chi' = KL^d(\langle m^2 \rangle - \langle |m| \rangle^2). \tag{4.16}$$

Note that the 'true' susceptibility calculated from the variance of m, $\chi = KL^d(\langle m^2 \rangle - \langle m \rangle^2)$, cannot be used to determine $K_c(L)$ because it has no peak. For sufficiently long runs at any temperature $\langle m \rangle = 0$ for $H = 0$ so that any peak in χ is merely due to the finite statistics of the simulation. For runs of modest length, $\langle m \rangle$ may thus have quite different values, depending on whether or not the system overturned completely many times during the course of the run. Thus, repetition of the run with different random number sequences may yield a true susceptibility χ which varies wildly from run to run below T_c. While for $T > T_c$ the 'true' susceptibility must be used if one wishes to estimate not only the critical exponent of χ but also the prefactor, for $T < T_c$ it is χ' and not χ that converges smoothly to the susceptibility of a state that has a spontaneous magnetization in the thermodynamic limit. For $T > T_c$ it is then better to use the result $\langle m \rangle = 0$ for $H = 0$ and estimate χ from $\chi = KL^d \langle m^2 \rangle$.

4.2.3.3 Finite size scaling and first order transitions

If the phase transition is first order, so that the correlation length does not diverge, a different approach to finite size scaling must be used. We first consider what happens if we fix the temperature $T < T_c$ of the Ising square lattice ferromagnet and cross the phase boundary by sweeping the magnetic field H. The subsequent magnetization curves are shown schematically in Fig. 4.6(a). The simplest, intuitive description of the behavior of the probability distribution of states in the system is plotted in Fig. 4.6(b). In the infinite system, in

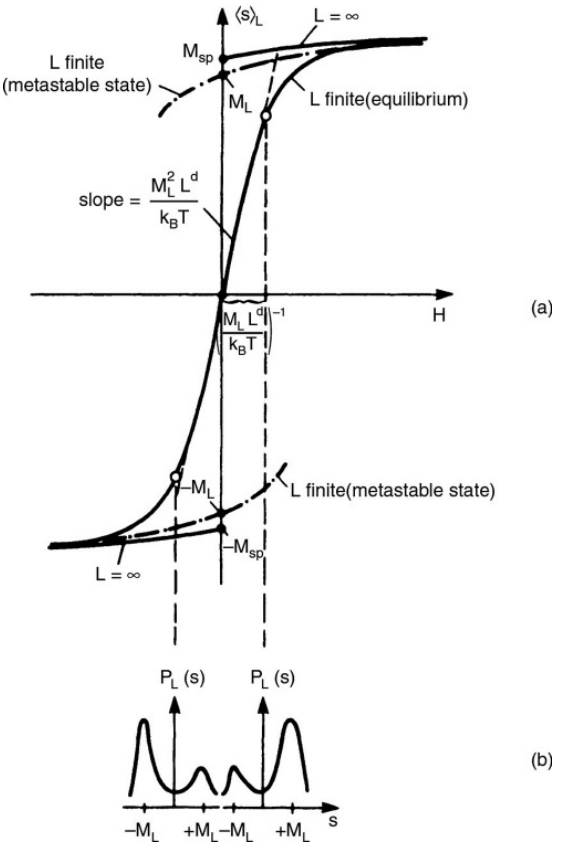

Fig. 4.6 Variation of the magnetization in a finite ferromagnet with magnetic field H. The curves include the infinite lattice behavior, the equilibrium behavior for a finite lattice, and the behavior when the system is only given enough time to relax to a metastable state. From Binder and Landau (1984).

equilibrium, the magnetization changes discontinuously at $H = 0$ from a value $+M_{\rm sp}$ to a value $-M_{\rm sp}$. If, however, L is finite, the system may jump back and forth between two states (see Fig. 4.2) whose most probable values are $\pm M_L$, and the resultant equilibrium behavior is given by the continuous, solid curve. We start the analysis of the finite size behavior by approximating this distribution by two Gaussian curves, one centered on $+M_L$ and one at $-M_L$. In this (symmetric) case, the probability distribution $P_L(s)$ for the magnetization s then becomes

$$P_L(s) = \tfrac{1}{2} L^{d/2} (2\pi k_{\rm B} T \chi_{(L)})^{-1/2} \times \{\exp[-(s - M_L)^2 L^d / (2k_{\rm B} T \chi_{(L)})] + \exp[-(s + M_L)^2 L^d / (2k_{\rm B} T \chi_{(L)})]\}. \quad (4.17)$$

If a magnetic field H is now applied then

$$P_L(s) = A(\exp\{-[(s - M_{\rm sp})^2 - 2\chi s H] L^d / 2k_{\rm B} T \chi\} + \exp\{-[(s + M_{\rm sp})^2 - 2\chi s H] L^d / 2k_{\rm B} T \chi\}), \quad (4.18)$$

where χ is the susceptibility if the system stays in a single phase. The transition is located at the field for which the weights of the two Gaussians are equal; in the Ising square lattice this is, of course, at $H = 0$. It is now straightforward to

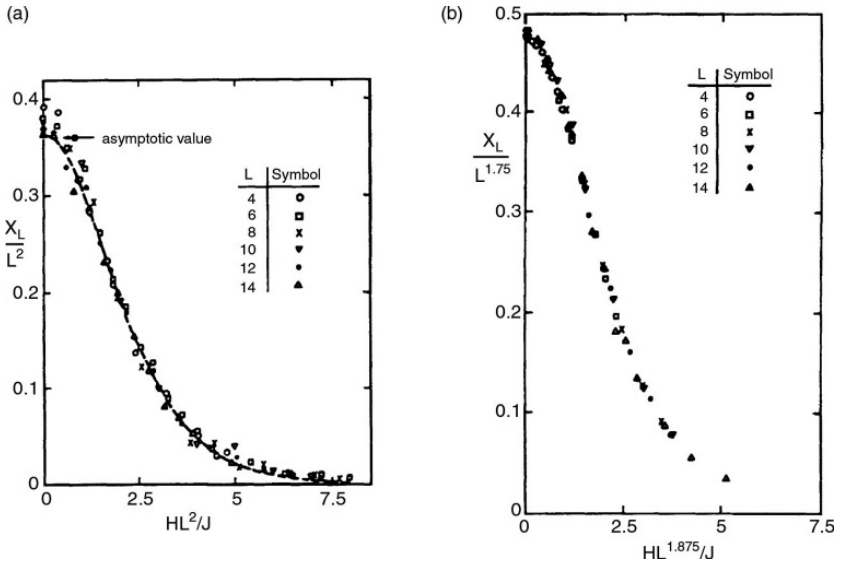

Fig. 4.7 Finite size scaled susceptibility vs. scaled field for the two-dimensional Ising model along paths of constant temperature: (a) $k_B T/\mathcal{J} = 2.1$; (b) $k_B T/\mathcal{J} = 2.269 = T_c$. From Binder and Landau (1984).

calculate the moments of the distribution and thus obtain estimates for various quantities of interest. Thus,

$$\langle s \rangle_L \approx \chi H + M_{sp} \tanh\left[\frac{H M_{sp} L^d}{k_B T}\right] \tag{4.19}$$

and the susceptibility ($\chi_L = K L^d(\langle s^2 \rangle_L - s_L^2)$ is defined in analogy with the 'true' susceptibility) is

$$\chi_L = \chi + M_{sp}^2 L^d \Big/ \left[k_B T \cosh^2\left(\frac{H M_{sp} L^d}{k_B T}\right)\right]. \tag{4.20}$$

This expression shows that length enters only via the lattice volume, and hence it is the dimensionality d which now plays the essential role rather than a (variable) critical exponent as is the case with a second order transition. In Fig. 4.7 we show finite size scaling plots for the susceptibility below T_c and at T_c for comparison. The scaling is quite good for sufficiently large lattices and demonstrates that this 'thermodynamic' approach to finite size scaling for a first order transition works quite well. Note that the approach of χ_L to the thermodynamic limit is quite subtle, because the result depends on the order in which limits are taken: $\lim_{H\to 0} \lim_{L\to\infty} \chi_L = \chi$ (as required for the 'true' susceptibility) but $\lim_{L\to\infty} \lim_{H\to 0} \chi_L / L^d = M_{sp}^2 / k_B T$.

In other cases the first order transition may involve states which are not related by any particular symmetry (Binder, 1987). An example is the two-dimensional q-state Potts model (see Eqn. (2.43)) for $q > 4$ in which there is a temperature-driven first order transition. At the transition the disordered state has the same free energy as the q-fold degenerate ordered state. Again one can describe the distribution of states by the sum of two Gaussians, but these two functions will now typically have rather different parameters (Challa

et al., 1986). The probability distribution function for the internal energy E per lattice site is (E_+, C_+, and E_-, C_- are energy and specific heat in the high temperature phase or low temperature phase right at the transition temperature T_c, respectively, and $\Delta T = T - T_c$)

$$P_L(E) = A\left[\frac{a_+}{\sqrt{C_+}}\exp\left[\frac{-[E-(E_+ + C_+\Delta T)]^2 L^d}{2k_B T^2 C_+}\right] + \frac{a_-}{\sqrt{C_-}}\exp\left[\frac{-[E-(E_- + C_-\Delta T)]^2 L^d}{2k_B T^2 C_-}\right]\right]. \tag{4.21}$$

Here A is a normalizing constant and the weights a_+, a_- are given by

$$a_+ = e^x, \qquad a_- = q\,e^{-x} \tag{4.22}$$

where $x = (T - T_c(\infty))(E_+ - E_-)L^d/(2k_B T T_c)$. Originally, Challa *et al.* (1986) had assumed that at the transition temperature $T_c(\infty)$ of the infinite system each peak of the q ordered domains and the disordered phase has equal *height*, but now we know that they have equal *weight* (Borgs and Kotecký, 1990). From Eqns. (4.21), (4.22) we find that the specific heat maximum occurs at

$$\frac{T_c(L) - T_c}{T_c} = \frac{k_B T_c \ln[q]}{(E_+ - E_-)L^d} \tag{4.23}$$

and the value of the peak is given by

$$C_L\Big|_{\max} \approx \frac{C_+ + C_-}{2} + \frac{(E_+ - E_-)^2 L^d}{4k_B T_c^2}. \tag{4.24}$$

Challa *et al.* (1986) also proposed that a reduced fourth order cumulant of the energy, i.e.

$$V_L = 1 - \frac{\langle E^4\rangle_L}{3\langle E^2\rangle_L^2} \tag{4.25}$$

has a minimum at an effective transition temperature which also approaches the infinite lattice value as the inverse volume of the system. The behavior of V_L for the $q = 10$ Potts model in two dimensions is shown in Fig. 4.8. Thus, even in the asymmetric case it is the volume L^d which is important for finite size scaling. Effective transition temperatures defined by extrema of certain quantities in general differ from the true transition temperature by corrections of order $1/L^d$, and the specific heat maximum scales proportional to L^d (the prefactor being related to the latent heat $E_+ - E_-$ at the transition, see Eqn. (4.24)).

This discussion was included to demonstrate that we understand, in principle, how to analyze finite size effects at first order transitions. In practice, however, this kind of finite size analysis is not always useful, and the use of free energy integrations may be more effective in locating the transition with modest effort. Other methods for studying first order transitions will be presented in later sections.

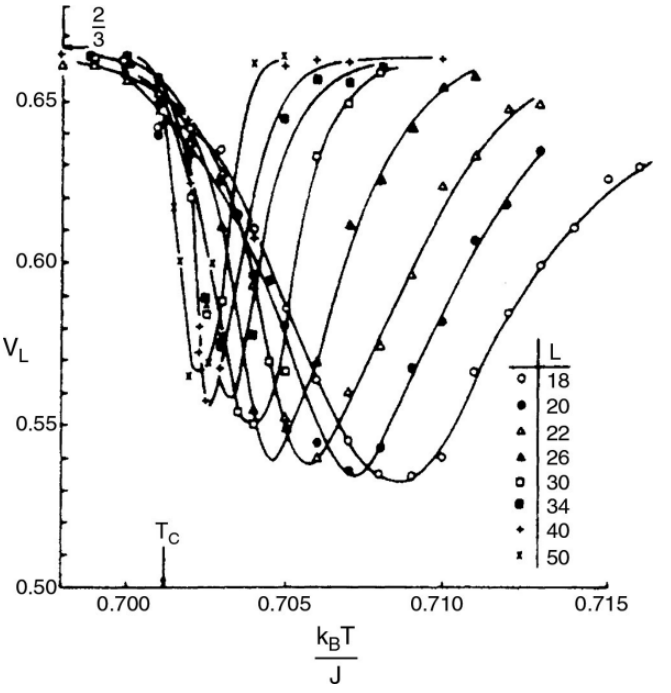

Fig. 4.8 Variation of the 'reduced' fourth order energy cumulant V_L with temperature for the $q = 10$ Potts model in two dimensions. The vertical arrow shows the transition temperature for $L = \infty$. After Challa *et al.* (1986).

4.2.3.4 Finite size subsystem scaling

A theoretical approach to the understanding of the behavior of different systems in statistical physics has been to divide the system into sub-blocks and coarse-grain the free energy to derive scaling laws. We shall see this approach carried out explicitly in Chapter 9 where we discuss Monte Carlo renormalization group methods. The behavior of a sub-block of length scale L'/b in a system of size L' will be different from that of a system of size L'/b because the correlation length can be substantially bigger than the size of the system sub-block. In this case it has been shown (Binder, 1981) that the susceptiblity at the transition actually has an energy-like singularity

$$\langle s^2 \rangle_L \propto L^{2\beta/\nu}[f_2(\infty) - g_2(\xi/L)^{-(1-\alpha)/\nu}]. \tag{4.26}$$

The block distribution function for the two-dimensional Ising model, shown in Fig. 4.9, has a quite different behavior below and above the critical point. For $T < T_c$ the distribution function can be well described in terms of M_{sp}, χ, and the interface tension F_s, while for $T > T_c$ the distribution becomes Gaussian with a width determined by χ. In addition, the advantage of studying subsystems is that in a single run one can obtain information on size effects on many length scales (smaller than the total size of the simulated system, of course).

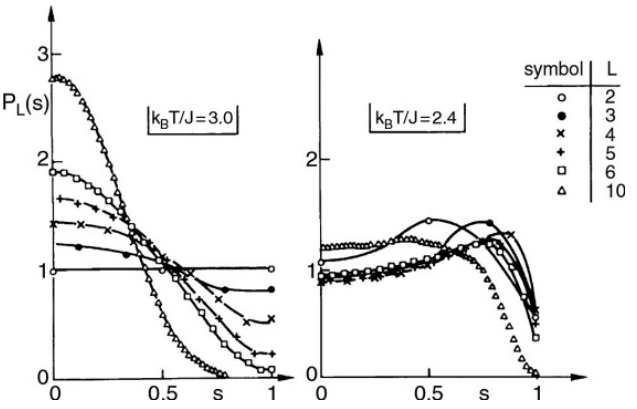

Fig. 4.9 Block distribution function for the two-dimensional Ising model for $L \times L$ sub-blocks: (left) $T > T_c$; (right) $T < T_c$. From Binder (1981).

4.2.3.5 Field mixing

Up to this point our examples for finite size scaling at critical points have involved the Ising model which is a particularly 'symmetric' model. When viewed as a lattice gas this model has particle–hole symmetry. In more realistic models of fluids, however, this symmetry is lost. Such models consider particles which may move freely in space and which interact via the well known Lennard–Jones form

$$\phi(r) = 4w[(\sigma/r)^{12} - (\sigma/r)^{6}], \tag{4.27}$$

where r is the distance between particles, σ gives the characteristic range of the interaction, and w gives the potential well depth. The critical point of the Lennard–Jones system is described by *two* non-trivial parameter values, the critical chemical potential μ_c and the critical well depth w_c. In general, then, the scaling fields which are appropriate for describing the critical behavior of the system contain linear combinations of the deviations from these critical values:

$$\tau = w_c - w + s\,(\mu - \mu_c)\,, \tag{4.28a}$$
$$h = \mu - \mu_c + r\,(w_c - w)\,, \tag{4.28b}$$

where r and s depend upon the system (Wilding and Bruce, 1992). (For the Ising model $r = s = 0$.) We can now define two relevant densities which are conjugate to these scaling fields

$$\langle \mathcal{E} \rangle = L^{-d}\partial \ln Z_L/\partial\tau = [u - r\rho]/(1 - sr), \tag{4.29a}$$
$$\langle \mathcal{M} \rangle = L^{-d}\partial \ln Z_L/\partial h = [\rho - s\,u]/(1 - sr), \tag{4.29b}$$

which are linear combinations of the usual energy density and particle density. Thus, a generalized finite size scaling hypothesis may be formulated in terms of these generalized quantities, i.e.

$$p_L(\mathcal{M}, \mathcal{E}) \approx \Lambda_{\mathcal{M}}^{+}\Lambda_{\mathcal{E}}^{+}\tilde{p}_{\mathcal{M},\mathcal{E}}(\Lambda_{\mathcal{M}}^{+}\delta\mathcal{M},\, \Lambda_{\mathcal{E}}^{+}\delta\mathcal{E},\, \Lambda_{\mathcal{M}}h,\, \Lambda_{\mathcal{E}}\tau) \tag{4.30}$$

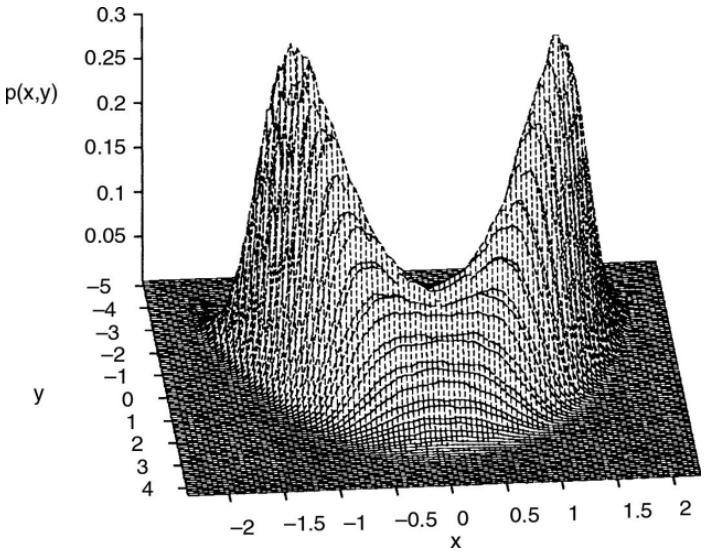

Fig. 4.10 Joint order parameter–energy distribution for an asymmetric lattice gas model as a function of scaling variables $x = a_M^{-1} L^{\beta/\nu}(\mathcal{M} - \mathcal{M}_c)$, $y = a_{\mathcal{E}}^{-1} L^{(1-\alpha)/\nu}(\mathcal{E} - \mathcal{E}_c)$. From Wilding (1995).

where

$$\Lambda_{\mathcal{E}} = a_E L^{1/\nu}, \qquad \Lambda_{\mathcal{M}} = a_M L^{d-\beta/\nu}, \qquad \Lambda_{\mathcal{M}}^{+}\Lambda_{\mathcal{M}} = \Lambda_{\mathcal{E}}^{+}\Lambda_{\mathcal{E}} = L^d \tag{4.31}$$

and

$$\delta\mathcal{M} = \mathcal{M} - \langle\mathcal{M}\rangle_c\,, \qquad \delta\mathcal{E} = \mathcal{E} - \langle\mathcal{E}\rangle_c\,. \tag{4.32}$$

Note that precisely at criticality Eqn. (4.30) simplifies to

$$p_L(\mathcal{M}, \mathcal{E}) \approx \Lambda_{\mathcal{M}}^{+}\Lambda_{\mathcal{E}}^{+}\,\tilde{p}^{*}_{\mathcal{M},\mathcal{E}}(\Lambda_{\mathcal{M}}^{+}\delta\mathcal{M},\, \Lambda_{\mathcal{E}}^{+}\delta\mathcal{E}), \tag{4.33}$$

so that by taking appropriate derivatives one may recapture power law behavior for the size dependence of various quantities. Surprises occur, however, and because of the field mixing contributions one finds that for critical fluids the specific heat

$$C_V = L^d(\langle u^2\rangle - \langle u\rangle^2)/k_B T^2 \sim L^{\gamma/\nu} \tag{4.34}$$

which is quite *different* from that found in the symmetric case. In Fig. 4.10 we show the parameter distribution at criticality as a function of the scaling variables.

4.2.3.6 Finite size effects in simulations of interfaces

As has been discussed in Section 4.2.2, one can deliberately stabilize interfaces in the system by suitable choice of boundary conditions. Such simulations are done with the intention to characterize the interfacial profile between coexisting phases, for instance. Figure 4.11 summarizes some of the standard simulation geometries that have been used for such a purpose, taking the Ising model again as simple example. Since directions parallel and perpendicular to an interface clearly are not equivalent, it also is no longer natural to choose the

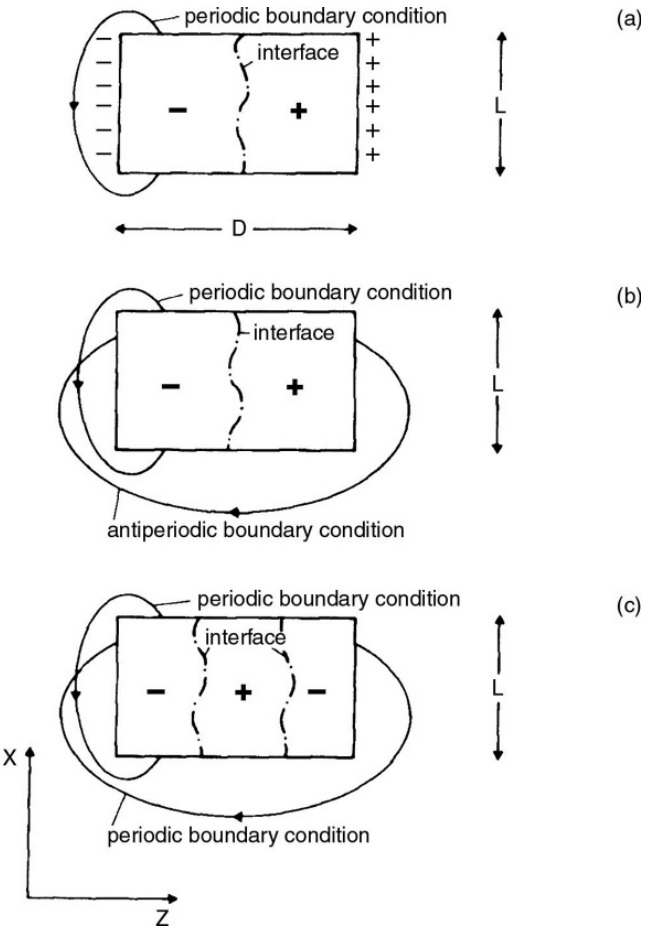

Fig. 4.11 Schematic sketch of three possible simulation geometries to study interfaces in Ising systems: (a) the 'surface field' boundary condition; (b) the antiperiodic boundary condition; and (c) the fully periodic boundary condition. Interfaces between coexisting phases of positive (+) and negative (−) spontaneous magnetization are shown schematically as dash-dotted lines.

same value for the linear dimensions of the simulation box in the parallel and perpendicular directions. Thus, Fig. 4.11 assumes a linear dimension D in the direction across the interface, and another linear dimension L parallel to it. In case (a), the system has periodic boundary conditions in the parallel direction, but free boundaries in the perpendicular directions, with surface magnetic fields (negative ones on the left boundary, positive ones on the right boundary) to stabilize the respective domains, with an interface between them that on average is localized in the center of the film and runs parallel to the boundaries where the surface fields act. We disregard here the possibility that the interface may become 'bound' to one of the walls, and assume high enough temperature so the interface is a 'rough', fluctuating object, not locally localized at a lattice plane (in $d = 3$ dimensions where the interface is two-dimensional). In case (b), an analogous situation with a simple interface is stabilized by an antiperiodic boundary condition, while in case (c), where fully periodic boundary conditions are used, only an even number of interfaces can exist in the system. (In order to avoid the problem that one kind of domain, say the + domain, completely

disappears because the interfaces meet and annihilate each other, we require a simulation at constant magnetization, see Section 4.4.1 below.)

The pictures in Fig. 4.11 are rather schematic, on a coarse-grained level, where both magnetization fluctuations in the bulk of the domains and small-scale roughness of the interface are ignored. But we emphasize the long wavelength fluctuations in the local position of the interface, because these fluctuations give rise to important finite size effects. It turns out that interfaces in a sense are soft objects, with a correlation length of fluctuations parallel to the interface ($\xi_\parallel$) that diverges if L and D tend to infinity: thus the interface is like a system at a critical point.

These fluctuations can be qualitatively accounted for by the concept of 'capillary wave' excitations, i.e. harmonic distortions of the local interface position z away from the average. For $D \to \infty$, one finds that the mean-square width of the interface scales with the parallel linear dimension L (Jasnow, 1984)

$$w^2 \equiv \langle z^2 \rangle - \langle z \rangle^2 \propto \begin{cases} L & d = 2 \quad \text{dimensions} \\ \ln L & d = 3 \quad \text{dimensions,} \end{cases} \tag{4.35}$$

while in the opposite limit where L is infinite and D is varied one finds that $\xi_\parallel$ is finite (Parry and Evans, 1992)

$$\xi_{||} \propto \begin{cases} D^2, & d = 2 \\ \exp(\kappa D), & d = 3, \quad \kappa = \text{constant.} \end{cases} \tag{4.36}$$

Then w^2 also becomes independent of L for large L, but rather depends on the perpendicular linear dimension D,

$$w^2 \propto \begin{cases} D^2, & d = 2, \\ D, & d = 3, \end{cases} \quad D \to \infty. \tag{4.37}$$

Of course, all these relations for the interfacial width make sense only for rather large linear dimensions L and D, such that w in Eqns. (4.35) and (4.37) is much larger than the 'intrinsic width' of the interface. If D is not very large, it is possible that the intrinsic width itself is squeezed down, and one then encounters a regime where $w \propto D$ in $d = 3$ dimensions.

Thus, simulations of interfaces are plagued by various finite size effects. More details and an example (interfaces in binary polymer mixtures) can be found in Werner *et al.* (1997).

4.2.3.7 Final thoughts

In many cases it is possible to perform a direct enumeration of states for a sufficiently small system. Generally this is possible only for systems which are so small that corrections to finite size scaling are important. The results should nonetheless lie on a smooth curve delineating the finite size behavior and can be useful in attempting to extract correction terms. Small lattices play another important role. Since exact results may be obtained for small systems, very useful checks of the correctness of the program may be made. Experience has shown that it is quite easy to make small errors in implementing the different

boundary conditions discussed above, particularly at the corners. For large lattices such errors produce quite small imperfections in the data, but for small lattices the boundary spins make up a substantial fraction of the total system and errors in the data become larger. Thus programming mistakes and other subtle errors are often most visible for small systems.

4.2.4 Finite sampling time effects

When one plans a computer simulation study of a given model using a fixed 'budget' of computer resources, one must make a choice between performing long simulations of small systems or shorter simulations of larger systems. In order to use the available computer time as efficiently as possible, it is important to know the sources of both systematic and statistical errors. One source of systematic errors, finite size effects, was treated in the previous section; here we consider how such errors depend on the number of updates performed, i.e. the length of the run.

4.2.4.1 Statistical error

Suppose $\mathcal{N}$ successive observations A_μ, with $\mu = 1, \ldots, \mathcal{N}$ of a quantity A have been stored in a simulation, with $\mathcal{N} \gg 1$. We consider the expectation value of the square of the statistical error

$$\begin{aligned}\langle(\delta A)^2\rangle &= \left\langle \left[\frac{1}{\mathcal{N}} \sum_{\mu=1}^{\mathcal{N}} (A_\mu - \langle A\rangle) \right]^2 \right\rangle \\ &= \frac{1}{\mathcal{N}^2} \sum_{\mu=1}^{\mathcal{N}} \langle (A_\mu - \langle A\rangle)^2\rangle + \frac{2}{\mathcal{N}^2} \sum_{\mu_1=1}^{\mathcal{N}} \sum_{\mu_2=\mu_1+1}^{\mathcal{N}} (\langle A_{\mu_1} A_{\mu_2}\rangle - \langle A\rangle^2).\end{aligned} \tag{4.38}$$

In order to further explain what this means we now invoke the 'dynamic interpretation' of Monte Carlo sampling in terms of the master equation (Müller-Krumbhaar and Binder, 1973). The index μ which labels each successive configuration then plays the role of a 'time' variable (which may or may not be related to physical time, as discussed elsewhere in this book (see e.g. Sections 2.2.3, 2.3, 4.4, 5.2, etc.). If the states $\{\mathbf{X}_\mu\}$ of the system from which the observations $\{A_\mu\}$ are taken are distributed according to a Boltzmann equilibrium distribution, the origin of this 'time' is indistinguishable from any other instant of this 'time', i.e. there is translational invariance with respect to this 'time' variable so that $\langle A_{\mu_1} A_{\mu_2}\rangle = \langle A_0 A_{\mu_2-\mu_1}\rangle$. Of course, this invariance would not hold in the first part of the Monte Carlo run (see Fig. 4.2), where the system starts from some arbitrary initial state which is not generally characteristic for the desired equilibrium – this early part of the run (describing the 'relaxation towards equilibrium') is hence not considered here and is omitted from the estimation of the average $\langle A\rangle$. The state $\mu = 1$ in Eqn. (4.38) refers to the first state that is actually included in the computation of the average, and not the first state that is generated in the Monte Carlo run.

Using this invariance with respect to the origin of 'time', we can change the summation index μ_2 to $\mu_1 + \mu$ where $\mu \equiv \mu_2 - \mu_1$, and hence

$$\langle(\delta A)^2\rangle = \frac{1}{\mathcal{N}}\left[\langle A^2\rangle - \langle A\rangle^2 + 2\sum_{\mu=1}^{\mathcal{N}}\left(1 - \frac{\mu}{\mathcal{N}}\right)(\langle A_0 A_\mu\rangle - \langle A\rangle^2)\right]. \quad (4.39)$$

Now we explicitly introduce the 'time' $t = \mu\delta t$ associated with the Monte Carlo process where δt is the time interval between two successive observations A_μ, $A_{\mu+1}$. It is possible to take $\delta t = \tau_s/N$, where N is the number of degrees of freedom, and τ_s is a time constant used to convert the transition probability of the Metropolis method to a transition probability per unit time: this would mean that every Monte Carlo 'microstate' is included in the averaging. Since subsequent microstates are often highly correlated with each other (e.g. for a single spin-flip Ising simulation they differ at most by the orientation of one spin in the lattice), it typically is much more efficient to take δt much larger than τ_s/N, i.e. $\delta t = \tau_s$. (This time unit then is called '1 Monte Carlo step/spin (MCS)', which is useful since it has a sensible thermodynamic limit.) But often, in particular near critical points where 'critical slowing down' (Hohenberg and Halperin, 1977) becomes pronounced, even subsequent states $\{\mathbf{X}_\mu\}$ separated by $\delta t = 1$ MCS are highly correlated, and it may then be preferable to take $\delta t = 10$ MCS or $\delta t = 100$ MCS, for instance, to save unnecessary computation. (When we discuss reweighting techniques in Chapter 7 we shall see that this is not always the case.)

Assuming, however, that the 'correlation time' between subsequent states is much larger than δt, we may transform the summation over the discrete 'times' $t = \delta t\ \mu$ to an integration, $t = \delta t\, \mathcal{N}$,

$$\begin{aligned}\langle(\delta A)^2\rangle &= \frac{1}{\mathcal{N}}\left[\langle A^2\rangle - \langle A\rangle^2 + \frac{2}{\delta t}\int_0^t \left(1 - \frac{t'}{t}\right)[\langle A(0)A(t')\rangle - \langle A\rangle^2]dt'\right] \\ &= \frac{1}{\mathcal{N}}(\langle A^2\rangle - \langle A\rangle^2)\left[1 + \frac{2}{\delta t}\int_0^t \left(1 - \frac{t'}{t}\right)\phi_A(t')dt'\right], \end{aligned} \quad (4.40)$$

where we define the normalized time autocorrelation function (also called 'linear relaxation function') $\phi_A(t)$ as

$$\phi_A(t) = \frac{[\langle A(0)A(t)\rangle - \langle A\rangle^2]}{[\langle A^2\rangle - \langle A\rangle^2]}. \quad (4.41)$$

For the magnetization M of an Ising model, this function has already been discussed in Eqns. (2.111) and (2.112). Note that $\phi_A(t=0) = 1$, $\phi_A(t \to \infty) = 0$, and $\phi_A(t)$ decays monotonically with increasing time t. We assume that the time integral of $\phi_A(t)$ exists, i.e.

$$\tau_A \equiv \int_0^\infty \phi_A(t)dt, \quad (4.42)$$

and τ_A then can be interpreted as the 'relaxation time' of the quantity A (cf. Eqn. (2.114)).

Let us now assume that the simulation can be carried out to times $t \gg \tau_A$. Since $\phi_A(t)$ is essentially non-zero only for $t' \leq \tau_A$, the term t'/t in Eqn. (4.40) then can be neglected and the upper integration limit replaced by infinity. This yields (Müller-Krumbhaar and Binder, 1973)

$$\langle(\delta A)^2\rangle = \frac{1}{\mathcal{N}}[\langle A^2\rangle - \langle A\rangle^2]\left(1 + 2\frac{\tau_A}{\delta t}\right). \tag{4.43}$$

We see that $\langle(\delta A)^2\rangle$ is in general *not* given by the simple sampling result $[\langle A^2\rangle - \langle A\rangle^2]/\mathcal{N}$, but is rather enhanced by the factor $(1 + 2\tau_A/\delta t)$. This factor is called the 'statistical inefficiency' of the Monte Carlo method and may become quite large, particularly near a phase transition. Obviously, by calculating $\langle A\rangle$ and $\langle A\rangle^2$, as well as $\langle(\delta A)^2\rangle$ we can estimate the relaxation time τ_{A}. Kikuchi and Ito (1993) demonstrated that for kinetic Ising model simulations such an approach is competitive in accuracy to the standard method where one records $\phi_A(t)$ (Eqn. (4.41)) and obtains τ_A by numerical integration (see Eqn. (4.42)). Of course, if $\tau_A \gg \delta t$, then Eqn. (4.43) may be further simplified by neglecting the unity in the bracket and, using $\mathcal{N}\delta t = t$,

$$\langle(\delta A)^2\rangle = [\langle A^2\rangle - \langle A\rangle^2](2\tau_A/t). \tag{4.44}$$

This means that the statistical error is independent of the choice of the time interval δt, it only depends on the ratio of relaxation time (τ_A) to observation time (t). Conversely, if δt is chosen to be so large that subsequent states are uncorrelated, we may put $\langle A_0 A_\mu\rangle \approx \langle A\rangle^2$ in Eqn. (4.39) to get $\langle(\delta A)^2\rangle = [\langle A^2\rangle - \langle A\rangle^2]/\mathcal{N}$. For many Monte Carlo algorithms τ_A diverges at second order phase transitions ('critical slowing down', see Sections 2.3.3 and 4.2.5), and then it becomes very hard to obtain sufficiently high accuracy, as is obvious from Eqn. (4.44). Therefore the construction of algorithms that reduce (or completely eliminate) critical slowing down by proper choice of global moves (rather than single spin-flips) is of great significance. Such algorithms, which are not effective in all cases, will be discussed in Section 5.1.

Problem 4.4 From a Monte Carlo simulation of an $L = 10$ Ising square lattice, determine the order parameter correlation time at $T = 3.0\, J/k_{\mathrm{B}}$ and at $T = 2.27\, J/k_{\mathrm{B}}$.

Problem 4.5 Perform a Metropolis Monte Carlo simulation for a 10×10 Ising model with periodic boundary conditions. Include the magnetic field H in the simulation and plot both $\langle M\rangle$ and $\langle|M|\rangle$ as a function of field for $k_{\mathrm{B}}T/J = 2.1$. Choose the range from $H = 0$ to $H = 0.05\, J$. Do you observe the behavior which is sketched in Fig. 2.10? Interpret your results.

4.2.4.2 Biased sampling error: Ising criticality as an example

The finite sampling time is not only the source of the statistical error, as described above, but can also lead to systematic errors (Ferrenberg *et al.*,

1991). For example, in the Monte Carlo sampling of response functions the latter are systematically underestimated. This effect comes simply from the basic result of elementary probability theory (see Section 2.2) that in estimating the variance s^2 of a probability distribution using n independent samples, the expectation value $E(s^2)$ of the variance thus obtained is systematically lower than the true variance σ^2 of the distribution, by a factor $(1 - 1/n)$:

$$E(s^2) = \sigma^2(1 - 1/n). \tag{4.45}$$

Since we may conclude from Eqn. (4.43) that for $t \gg t_A$ we have $n = \mathcal{N}/(1 + 2\tau_A/\delta t)$ independent 'measurements', we may relate the calculated susceptibility $\chi_{\mathcal{N}}$ of a spin system to that which we would obtain from a run of infinite length χ_∞ by

$$\chi_{\mathcal{N}} = \chi_\infty \left(1 - \frac{1 + 2\tau_M/\delta t}{\mathcal{N}}\right), \tag{4.46}$$

τ_{M} being the relaxation time of the magnetization, i.e. $A = M$ in Eqns. (4.38)–(4.43).

This effect becomes particularly important at T_{c}, where one uses the values of χ from different system sizes ($N = L^d$ in d dimensions, where L is the linear dimension and the lattice spacing is taken to be unity) to estimate the critical exponent ratio γ/ν (see Section 4.2.3). The systematic error resulting from Eqn. (4.46) will generally vary with L, since the relaxation time τ_{M} may depend on the system size quite dramatically ($\tau_M \propto L^z$, z being the 'dynamic exponent', see Section 2.3.3).

While finite size scaling analyses are now a standard tool, the estimation of errors resulting from Eqn. (4.46) is generally given inadequate attention. To emphasize that neglect of this biased sampling error is not always warranted, we briefly review here some results of Ferrenberg *et al.* (1991) who performed calculations for the nearest neighbor ferromagnetic Ising model. The Monte Carlo simulations were carried out right at the 'best estimate' critical temperature T_{c} of the infinite lattice model ($T_{\mathrm{c}}^{-1} = 0.221\,654 k_{\mathrm{B}}/\mathcal{J}$) for system sizes ranging from $16 \leq L \leq 96$. Well over 10^6 MCS were performed, taking data at intervals $\delta t = 10$ MCS, and dividing the total number of observations $\mathcal{N}_{\mathrm{tot}}$ into g bins of 'bin length' $\mathcal{N}$, $\mathcal{N}_{\mathrm{tot}} = g\mathcal{N}$, and calculating $\chi_{\mathcal{N}}$ from the fluctuation relation. Of course, in order to obtain reasonable statistics they had to average the result over all $g \gg 1$ bins. Figure 4.12 shows the expected strong dependence of $\chi_{\mathcal{N}}$ on both $\mathcal{N}$ and L: while for $L = 16$ the data have settled down to an L-dependent plateau value for $\mathcal{N} \geq 10^3$, for $L = 48$ even the point for $\mathcal{N} = 10^4$ still falls slightly below the plateau, and for $L = 96$ the asymptotic behavior is only reached for $\mathcal{N} \geq 10^5$. (Note that in this calculation a very fast vectorizing multispin coding (Section 5.2.2) single spin-flip algorithm was used.) Thus with a constant number $\mathcal{N}$ as large as $\mathcal{N} = 10^4$ for a finite size scaling analysis, one would *systematically* underestimate the true finite system susceptibility for large L, and an incorrect value of γ/ν in the relation $\ln \chi_{\mathcal{N}}(L) = (\gamma/\nu) \ln L$ would result. However, if we measure τ_M for the different values of L and use Eqn. (4.46), we can correct for this effect. In

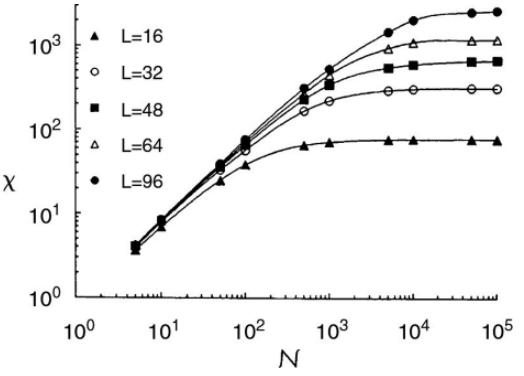

Fig. 4.12 Variation of $\chi_{\mathcal{N}}$ for the susceptibility of $L \times L \times L$ ferromagnetic nearest neighbor Ising lattices at the critical temperature as a function of the 'bin length' $\mathcal{N}$. Different symbols indicate various values of L, as indicated in the figure. From Ferrenberg *et al.* (1991).

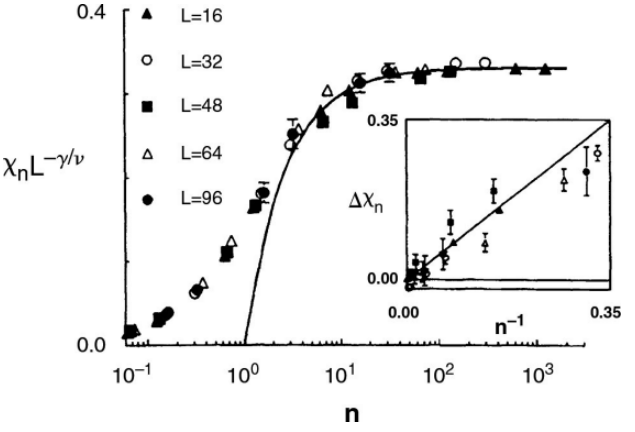

Fig. 4.13 Scaled susceptibility vs. scaled bin length n. The solid line is the function $\chi_\infty L^{-\gamma/\nu}(1 - 1/n)$, using the accepted value $\gamma/\nu = 1.97$. In the insert the reduced systematic error, $\Delta\chi_n = (\chi_\infty - \chi_N)/\chi_N$ is plotted vs. n^{-1} to highlight the large bin length behavior (the solid line, with slope unity, is predicted by Eqn. (4.46)). From Ferrenberg *et al.* (1991).

the present example, the appropriate correlation time τ_M is $\tau_M = 395$, 1640, 3745, 6935, and 15 480, for $L = 16$, 32, 48, 64, and 96, respectively. Using these values we can rewrite Eqn. (4.46) as $\chi_{\mathcal{N}} = \chi_\infty(1 - 1/n)$, computing n as $n = \mathcal{N}/(1 + 2\tau_M/\delta t)$. Figure 4.13 shows that when $\chi_{\mathcal{N}} L^{-\gamma/\nu}$ is plotted vs. n, all the data collapse on a universal function, which for $n \geq 5$ is compatible with the simple result $(1 - 1/n)$.

Another important effect in studies of critical phenomena via importance sampling Monte Carlo is that of cross-correlations between different observables that are measured from the same run (Weigel and Janke, 2008). For example, estimates of the critical exponent α from the size dependence of $d\ln\langle|m|\rangle dK$, $d\ln\langle m^2\rangle/dK$, or dU_4/dK (see Eqns. (4.14), (4.15)), from Monte Carlo 'observations' from the same time series are correlated with each other, and the cross-correlations need to be considered for a reliable error estimation (Weigel and Janke, 2008).

Problem 4.6 Carry out Monte Carlo simulations for an $L = 10$ Ising square lattice with different run lengths for $T = 2.8\, J/k_B$. Calculate the susceptibility and plot it vs. run length. Extract an estimate for the infinite lattice susceptibility.

4.2.4.3 Relaxation effects

When one starts a simulation run, typically equilibrium states for the system are not yet known. The Metropolis algorithm requires some initial state of the system, however, this choice will probably not be characteristic for the equilibrium that one wishes to study. For example, one may intend to study the critical region of an Ising ferromagnet but one starts the system for example in a state where all spins are perfectly aligned, or in a random spin configuration. Then it is necessary to omit the first $\mathcal{N}_0$ configurations from the averages, since they are not yet characteristic for equilibrium states of the system (see Fig. 4.2). Therefore any Monte Carlo estimate $\overline{A}$ of an average $\langle A \rangle$ actually reads

$$\overline{A} = \frac{1}{\mathcal{N} - \mathcal{N}_0} \sum_{\mu=\mathcal{N}_0+1}^{\mathcal{N}} A(\mathbf{X}_\mu) = \frac{1}{t - t_0} \int_{t_0}^{t} A(t')dt', \tag{4.47}$$

where $t_0 = \mathcal{N}_0 \delta t$. Time-displaced correlation functions $\langle A(t)\, B(0) \rangle$ as they appear in Eqn. (4.40) are actually estimated as

$$\overline{A(t')B(0)} = \frac{1}{t - t' - t_0} \int_{t_0}^{t-t'} A(t' + t'')B(t'')dt'', \qquad t - t' > t_0. \tag{4.48}$$

As emphasized above, times t_0 must be chosen which are large enough that thermal equilibrium has been achieved, and therefore time averages along the Monte Carlo 'trajectory' in phase space, as defined in Eqns. (4.47) and (4.48), make sense.

However, it is also interesting to study the non-equilibrium relaxation process by which equilibrium is approached, starting from a non-equilibrium initial state. In this process, $A(t') - \overline{A}$ depends on the observation time t' systematically, and an ensemble average $\langle A(t') \rangle - \langle A(\infty) \rangle$ ($\lim_{t\to\infty} \overline{A} = \langle A \rangle = \langle A(\infty) \rangle$ if the system is ergodic) is non-zero. Hence we define

$$\langle A(t) \rangle = \sum_{\{\mathbf{X}\}} P(\mathbf{X}, t) A(\mathbf{X}) = \sum_{\{\mathbf{X}\}} P(\mathbf{X}, 0) A\{\mathbf{X}(t)\}. \tag{4.49}$$

In the second step of this equation we have used the fact that the ensemble average involved is actually an average weighted by the probability distribution $P(\mathbf{X}, 0)$ of an ensemble of initial states $\{\mathbf{X}(t = 0)\}$ which then evolve as described by the master equation of the associate Monte Carlo process. In practice, Eqn. (4.49) means an average over $m \gg 1$ independent runs

$$\overline{A(t)} = \frac{1}{m} \sum_{l=1}^{m} A^{(l)}(t), \tag{4.50}$$

with $A^{(l)}(t)$ being the observable A observed at time t in the lth run of this non-equilibrium Monte Carlo averaging (these runs also differ in practice by use of different random numbers for each realization (l) of the time evolution).

Using Eqn. (4.49) we can define a non-linear relaxation function which was already considered in Eqn. (2.115)

$$\phi_A^{nl}(t) = [\langle A(t)\rangle - \langle A(\infty)\rangle] / [\langle A(0)\rangle - \langle A(\infty)\rangle] \tag{4.51}$$

and its associated relaxation time

$$\tau_A^{(nl)} = \int_0^\infty \phi_A^{(nl)}(t)dt. \tag{4.52}$$

The condition that the system is well equilibrated then simply reads

$$t_0 \gg \tau_A^{(nl)}. \tag{4.53}$$

This inequality must hold for all physical observables A, and hence it is important to focus on the slowest relaxing quantity (for which $\tau_A^{(nl)}$ is largest) in order to estimate a suitable choice of t_0. Near second order phase transitions, the slowest relaxing quantity is usually the order parameter M of the transition, and not the internal energy. Hence the 'rule-of-thumb' published in some Monte Carlo investigations that the equilibration of the system is established by monitoring the time evolution of the internal energy is clearly not a reliable procedure. This effect can readily be realized by examining the finite size behavior of the times $\tau_M^{(nl)}$, $\tau_E^{(nl)}$, at criticality, cf. Eqns. (2.116) and (2.117)

$$\tau_M^{(nl)} \propto L^{z-\beta/\nu}, \qquad \tau_E^{(nl)} \propto L^{z-(1-\alpha)/\nu}, \tag{4.54}$$

where the exponent of the order parameter is β, of the critical part of the energy is $1-\alpha$, and of the correlation length is ν. Typically β/ν is much less than $(1-\alpha)/\nu$ and the correlation time associated with the magnetization diverges much faster than that of the internal energy.

We also wish to emphasize that starting the system in an arbitrary state, switching on the full interaction parameters instantly, and then waiting for the system to relax to equilibrium is not always a very useful procedure. Often this approach would actually mean an unnecessary waste of computing time. For example, in systems where one wishes to study ordered phases at low temperature, it may be hard to use fully disordered states as initial configurations since one may freeze in long-lived multidomain configurations before the system relaxes to the final monodomain sample. In glass-like systems (spin glass models, etc.) it is advisable to produce low temperature states by procedures resembling slow cooling rather than fast quenching. Sometimes it may be preferable to relax some constraints (e.g. self-avoiding walk condition for polymers) first and then to switch them on gradually. There are many 'tricks of the trade' for overcoming barriers in phase space by suitably relaxing the system by gradual biased changes in its state, gradually switching on certain terms in the Hamiltonian, etc., which will be mentioned from time to time later.

4.2.4.4 Back to finite size effects again: self-averaging

Suppose we observe a quantity A in n statistically independent observations made while the system is in equilibrium, and calculate its error

$$\Delta_A(n, L) = \sqrt{(\langle A^2\rangle - \langle A\rangle^2)/n}, \qquad n \gg 1. \tag{4.55}$$

We now ask, does this error go to zero if $L \to \infty$? If it does, A is called 'self-averaging', while if it yields an L-independent non-zero limit, we say A exhibits 'lack of self-averaging'. In pure phases away from phase boundaries, extensive quantities (energy per site E, magnetization per site M, etc.) have a Gaussian distribution whose variance scales inversely with the volume L^d,

$$P_L(A) = L^{d/2}(2\pi C_A)^{-1/2} \exp[-(A - \langle A\rangle)^2 L^d/2C_A]. \tag{4.56}$$

If, for example, $A = M$ then $C_A = k_B T\chi$, and if $A = E$, then $C_E = k_B T^2 C$ with C being the specific heat, etc. For these quantities, we hence see that errors scale as $\Delta_A(n, L) \propto (nL^d)^{-1/2}$. This property is called 'strong self-averaging' (Milchev *et al.*, 1986), in contrast to the behavior at critical points where the exponent governing the power law for the size dependence is smaller, $\Delta_A(n, L) \propto (nL^{x_1^A})^{-1/2}$ ($x_1^M = 2\beta/\nu$, $x_1^E = 2(1-\alpha)/\nu$; this situation is termed 'weak self-averaging').

The situation differs drastically if we consider quantities that are sampled from fluctuation relations (such as C, $\chi, \ldots$), rather than quantities that are spatial averages of a simple density (such as E, $M, \ldots$). We still can formally use Eqn. (4.55), but we have to replace A by $C_A = (\delta A)^2 L^d$ in this case,

$$\Delta_{C_A}(n, L) = L^d n^{-1/2} \sqrt{\langle(\delta A)^4\rangle - \langle(\delta A)^2\rangle^2}, \qquad \delta A = A - \langle A\rangle. \tag{4.57}$$

Since for the Gaussian distribution, Eqn. (4.55), $\langle(\delta A)^4\rangle = 3\langle(\delta A)^2\rangle^2$, Eqn. (4.56) reduces to ($C_A = L^d\langle(\delta A)^2\rangle$)

$$\Delta_{C_A}(n, L) = L^d n^{-1/2}\langle(\delta A)^2\rangle\sqrt{2} = C_A\sqrt{2/n}. \tag{4.58}$$

Consequently, the size L^d cancels out precisely, and the relative error $\Delta_{C_A}(n, L)/C_A = \sqrt{2/n}$ is completely universal. It only depends on the number n of statistically independent observations. Thus, increasing L at fixed n will strongly improve the accuracy of quantities such as E and M, but nothing is gained with respect to the accuracy of χ, C, etc. Thus, it is more economical to choose the smallest size which is still consistent with the condition $L \gg x$ and to increase n rather than L to improve the accuracy. For those researchers who feel that the best approach is to study the largest system size possible, we believe that an analysis of fluctuation relations in subsystems (Section 4.2.3.4) is mandatory.

4.2.5 Critical relaxation

The study of critical slowing down in spin models has formed an extremely active area of research, and Monte Carlo simulations have played an important

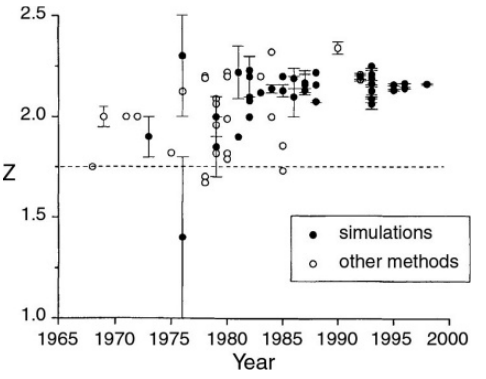

Fig. 4.14 Variation of the estimate of the dynamic exponent z for the two-dimensional Ising model as a function of time. The horizontal dashed line shows the value $\gamma/\nu = 1.75$ which is a lower bound.

role in developing an understanding of critical relaxation. The basic features of the underlying theory were presented in Section 2.3.3 and we now wish to examine the implementation of these ideas within the context of computer simulation. As we shall see below, the interest in this problem extends well beyond the determination of the dynamic critical exponent for a particular sampling algorithm since in any simulation there are multiple time scales for different quantities which must be understood even if the topic of interest is the static behavior.

Critical relaxation has been studied for many years for a number of different spin models but with uncertain results. Thus, in spite of the fact that the static behavior of the two-dimensional Ising model is known exactly, the determination of the critical relaxation has remained a rather elusive goal. As shown in Fig. 4.14, there have been estimates made for the dynamic critical exponent z over a period of more than 30 years using a number of different theoretical and numerical methods, and we may only just be coming to an accurate knowledge of the exponent for a few models (Landau *et al.*, 1988; Wansleben and Landau, 1991; Ito, 1993; Ozeki and Ito, 2007). In the following sub-sections we shall briefly examine the different features associated with critical relaxation and the different ways that Monte Carlo data can be used to extract an estimate for z.

4.2.5.1 Non-linear relaxation

As we have already seen, the approach of a thermodynamic property A to its equilibrium behavior occurs in a characteristic fashion and is described by a simple, non-linear relaxation function, $\phi_A(t)$, given by Eqn. (2.115). The accurate determination of this relaxation function is non-trivial since knowledge of the equilibrium value of the quantity being studied is needed. This necessitates performing simulations which are long compared to the non-linear relaxation time to insure that an equilibrium value can be measured; however, to guarantee that the statistical errors are small for the non-linear relaxation function it is also necessary to make many equivalent runs with different random number sequences and average the data together. As a result some balance between the number and length of the runs must be achieved.

Finally, the long time behavior of the non-linear relaxation function can be fitted by an exponential function to determine the asymptotic relaxation time $\tau \propto \xi^z$ (Eqn. (2.113)) while the integral of ϕ_A can be used to estimate the non-linear relaxation time τ_{nl}. The variation of τ_{nl} with temperature as the critical point is approached may then be used to estimate the dynamic exponent, although finite size effects will become important quite close to T_c. From Eqns. (2.116) and (2.117) we recall that $\tau_{nl} \propto \xi^{z_{nl}^A}$ with an exponent that is always smaller than z but is related to z by a scaling law, $z_{nl}^A = z - \beta^A/\nu$, β^A being the exponent of the 'critical part' of the quantity A. ($\beta^A = \beta$ if A is the order parameter and $\beta^A = 1 - \alpha$ if A is the energy, etc.) How is it possible that $z_{nl}^A < z$ although the asymptotic decay of $\phi_A(t)$ occurs with the 'linear' relaxation time τ which is governed by the exponent z? The solution to this puzzle is that the asymptotic decay sets in only when $\phi_A(t)$ has decayed down to values of the order of the static critical part of A, i.e. is of the order of $\xi^{-\beta_A/\nu} \sim \varepsilon^{\beta_A}$. Near T_c these values are small and accuracy is hard to obtain. Alternatively, if the critical temperature is well known, the critical exponent can be determined from the finite size behavior at T_c.

For an infinite system at T_c the magnetization will decay to zero (since this is the equilibrium value) as a power law

$$m(t) \propto t^{-\beta/z\nu}, \tag{4.59}$$

where β and ν are the static critical exponents which are known exactly for the two-dimensional Ising model. Eventually, for a finite lattice the decay will become exponential, but for sufficiently large lattices and sufficiently short times, a good estimate for z can be determined straightforwardly using Eqn. (4.59). (The study of multiple lattice sizes to insure that finite size effects are not becoming a problem is essential.) Several different studies have been successfully carried out using this technique. For example, Ito (1993) used multilattice sampling and carefully analyzed his Monte Carlo data, using Eqn. (4.59), for systems as large as $L = 1500$ to insure that finite size effects were not beginning to appear. (A skew periodic boundary was used in one direction and this could also complicate the finite size effects.) From this study he estimated that $z = 2.165(10)$. Stauffer (1997) examined substantially larger lattices, $L = 496\,640$, for times up to $t = 140$ MCS/site using this same method and concluded that $z = 2.18$. Although these more recent values appear to be well converged, earlier estimates varied considerably. For a review of the problems of non-linear relaxation in the Ising model see Wang and Gan (1998). We also note that there exists yet another exponent which appears in non-equilibrium relaxation at criticality when we start the system at T_c in a random configuration. The magnetization then has a value of $\sim N^{-1/2}$, and increases initially like $M(t) \propto t^\theta$ with a new exponent θ (Janssen *et al.*, 1989; Li *et al.*, 1994).

More recent studies of non-linear, short time relaxation have produced rather impressive results for both dynamic and static exponents of several well-known models (see, e.g., Li *et al.*, 1996; Zheng, *et al.*, 2003). They used

the dynamic finite size scaling of the moments of the magnetization, $M^{(k)}$, to extract exponent estimates. For zero initial magnetization

$$M^{(k)}(t, \varepsilon, L) = b^{k\beta/\nu}\mathcal{M}^{(k)}(b^z t, b^{-1/\nu}\varepsilon, bL) \tag{4.60}$$

where $\varepsilon = (T - T_c)/T_c$, and b gives the scale factor between two different lattice sizes. As the lattice size approaches the thermodynamic limit, we recover Eqn. (4.59) for long time decay at T_c. The values of z that were obtained (Li *et al.*, 1996), however, were slightly below other estimates with this same method. Recent large scale Monte Carlo simulations that examined the short time non-linear relaxation (Zheng *et al.*, 2003) were even able to extract corrections to scaling for the two-dimensional XY-model and the two-dimensional fully frustrated XY-model. They found different behavior depending upon whether or not they began with an ordered or disordered state. (For the initially ordered state, the scaling form in Eqn. (4.60) requires modification.) Data were averaged over more than 10^4 runs so statistical error bars were small. An appraisal of the so-called non-equilibrium relaxation (NER) method to estimate critical exponents on the basis of Eqn. (4.60), together with a review of various applications, has been given by Ozeki and Ito (2007).

4.2.5.2 Linear relaxation

Once a system is in equilibrium the decay of the time-displaced correlation function is described by a linear relaxation function (cf. Eqn. (2.111)). The generation of the data for studying the linear relaxation can be carried out quite differently than for the non-linear relaxation since it is possible to make a single long run, first discarding the initial approach to equilibrium, and then treating many different points in the time sequence as the starting point for the calculation of the time-displaced correlation function. Therefore, for a Monte Carlo run with N successive configurations the linear correlation function at time t can be computed from

$$\phi_{AA}(t) = \Gamma\left(\frac{1}{N-t}\sum_{t'}^{N-t} A(t')A(t'+t) - \frac{1}{(N-t)^2}\sum_{t'}^{N-t} A(t')\sum_{t''}^{N-t} A(t'')\right), \tag{4.61}$$

where $\Gamma = (\langle A^2\rangle - \langle A\rangle^2)^{-1}$. From this expression we see that there will indeed be many different estimates for short time displacements, but the number of values decreases with increasing time displacement until there is only a single value for the longest time displacement. The characteristic behavior of the time-displaced correlation function shown in Fig. 4.15 indicates that there are three basic regions of different behavior. In the early stages of the decay (Region I) the behavior is the sum of a series of exponential decays. Actually it is possible to show that the initial slope of $\phi_{AA}(t)$, $(d\phi_{AA}(t)/dt)_{t=0} = \tau_I^{-1}$, defines a time τ_I which scales as the static fluctuation, $\tau_I \propto (\langle A^2\rangle - \langle A\rangle^2)$. Since $\phi_{AA}(t)$ is non-negative, this result implies that $\tau > \tau_I$, and hence the inequality $z > \gamma/\nu$ results when we choose $A = M$, i.e. the order parameter.

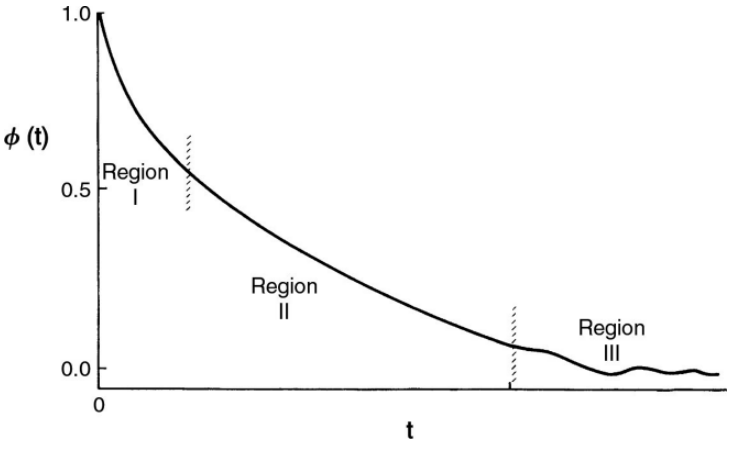

Fig. 4.15 Schematic behavior of the time-displaced correlation function as defined by Eqn. (4.61).

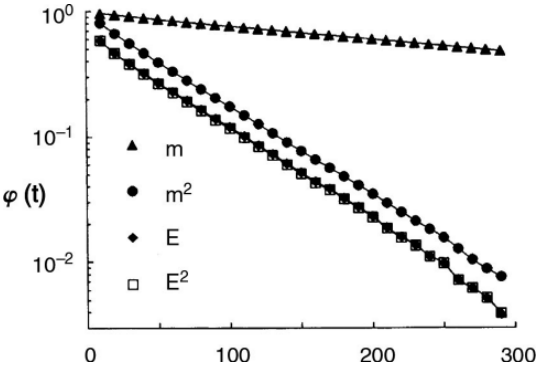

Fig. 4.16 Linear relaxation function for different quantities for the three-dimensional Ising model at the critical temperature with $L = 16$ and periodic boundary conditions. From Ferrenberg *et al.* (1991).

If instead $A = E$, i.e. the energy, the initial decay is rather rapid since $\tau_I \propto C \propto \varepsilon^{-\alpha}$, where α is the specific heat exponent. Nevertheless, the asymptotic decay of $\phi_{EE}(t)$ is governed by an exponential relaxation $e^{-t/\tau}$ where τ diverges with the same exponent z as the order parameter relaxation time. For a more detailed discussion see Stoll *et al.* (1973). In Region II the time dependence of the relaxation function can be fitted by a single exponential described by a correlation time τ which diverges as the critical point is approached. Finally, in Region III the statistical errors become so large that it becomes impossible to perform a meaningful fit. The difficulty, of course, is that it is never completely obvious when the data have entered the regime where they are described by a single exponential, so any analysis must be performed carefully. Generally speaking, the early time regime is much more pronounced for the internal energy as compared to the order parameter. This is clear from the above remark that the initial relaxation time for the energy scales like the specific heat. In order to compare the decay of different quantities, in Fig. 4.16 we show a semi-logarithmic plot for the three-dimensional Ising model. From this figure we can see that the magnetization decay is quite slow and is almost perfectly linear over the entire range. In contrast, the internal energy shows quite pronounced contributions from multiple decay modes at short times and

has a much shorter relaxation time in the asymptotic regime. Note that both m^2 and E^2 have time reversal symmetry (but m does not) and have the same asymptotic relaxation time as does E. For time displacements greater than about 250 MCS/site the statistical fluctuations begin to grow quite quickly and it becomes difficult to analyze the data in the asymptotic regime.

Although the general approach is straightforward, there are nonetheless considerable subtleties in this kind of analysis. The use of skew periodic boundary conditions simplifies the computer code but introduces a 'seam' into the model which provides a correction for small lattices. The relaxation function is a biased estimator so the length of the individual runs must also be quite long to eliminate another source of small corrections. (In fact, for finite length runs the relaxation function will oscillate about a small *negative* value at very long times.) Lastly, it is often necessary to perform least squares fits over different ranges of time to ascertain where noise is becoming a problem at long times.

A completely different approach to the analysis of the correlations in equilibrium, which does not require the computation of the relaxation function, is through the determination of the 'statistical inefficiency' described for example by Eqn. (4.43). A 'statistical dependence time' τ_{dep} is calculated by binning the measurements in time and calculating the variance of the mean of the binned values; as the size of the bins diverges, the estimate τ_{dep} approaches the correlation time. Kikuchi and Ito (1993) used this approach to study the three-dimensional Ising model and found that $z = 2.03$ (4).

The analysis of non-linear relaxation is still a popular tool for the study of critical phenomena in various models, but judging the accuracy of the resulting estimates is difficult. For instance, da Silva *et al.* (2013) used Eqn. (4.59) to analyze the critical dynamics at the tricritical point of the two-dimensional Blume–Capel model. Despite an impressive statistical effort (10 000 independent runs) and a careful data analysis, their final estimate for the correlation length exponent, $\nu = 0.537$ (6), differs from the exactly known value $\nu = 5/9 = 0.5555$ by three times the quoted error. As has been discussed earlier (Eqns. (4.13) and (4.14)), it is advisable to include the effects of correction terms to the leading critical behavior in a finite size scaling analysis. Presumably, then, this problem of correction terms still is a challenge in the context of non-linear relaxation studies.

4.2.5.3 Integrated vs. asymptotic relaxation time

As we saw earlier in this chapter, an integrated correlation time may be extracted by integrating the relaxation function; and it is this correlation time, given in Eqn. (4.42), which enters into the calculation of the true statistical error. The resulting integrated correlation time also diverges as the critical point is approached, but the numerical value may be different in magnitude from the asymptotic correlation time if there is more than one exponential that contributes significantly to the relaxation function. This is relatively easy to see if we look at the behavior of the internal energy E with time shown in

Fig. 4.16: from this figure we can see that both E and m^2 have the same asymptotic relaxation time, but m^2 will have a much larger integrated relaxation time. When one examines all of the response functions it becomes clear that there are a number of different correlation times in the system, and the practice of only measuring quantities at well separated intervals to avoid wasting time on correlated data may actually be harmful to the statistical quality of the results for some quantities.

4.2.5.4 Dynamic finite size scaling

The presence of finite size effects on the dynamic (relaxational) behavior can be used to estimate the dynamic critical exponent. Dynamic finite size scaling for the correlation time τ can be written

$$\tau = L^z \mathcal{F}(\varepsilon L^{1/\nu}), \tag{4.62}$$

so at the critical point the correlation time diverges with increasing lattice size as

$$\tau \propto L^z. \tag{4.63}$$

As in the case of statics, this finite size scaling relation is valid only as long as the lattice size L is sufficiently large that corrections to finite size scaling do not become important. The behavior of the correlation time for the order parameter and the internal energy may be quite different. For example, in Fig. 4.17 we show the finite size behavior of both correlation times for the three-dimensional Ising model. As a result we see that the asymptotic dynamic exponents for both quantities are consistent, but the amplitudes of the divergencies are almost an order of magnitude different.

Of course, it is also possible to extract an estimate for z using finite size data and Eqn. (4.62). In this approach a finite size scaling plot is made in the same manner as for static quantities with the same requirement that data for different sizes and temperatures fall upon a single curve. Here too, when the data are too far from T_c, scaling breaks down and the data no longer fall upon the same curve. In addition, when one tries to apply dynamic finite size scaling, it is important to be aware of the fact that $\phi_{\mathrm{MM}}(t)$ does not decay with a single relaxation time but rather with an entire spectrum, i.e.

$$\phi_{\mathrm{MM}}(t \to \infty) = c_1 e^{-t/\tau_1} + c_3 e^{-t/\tau_3} + \cdots, \qquad \tau_1 > \tau_3 > \cdots \tag{4.64}$$

where $c_1, c_3, \ldots$ are amplitudes, and all times $\tau_1 \propto \tau_3 \propto \cdots L^z$. Only the amplitudes $\hat{\tau}_n (\tau_n = \hat{\tau}_n L^z)$ decrease with increasing n. Note that we have used odd indices here because M is an 'odd operator', i.e. it changes sign under spin reversal. The second largest relaxation time τ_2 actually appears for the leading asymptotic decay of the 'even operators' such as E or M^2 which are invariant under spin reversal.

All of these relaxation times, τ_n, have a scaling behavior as written in Eqn. (4.62); however, it is important to note that τ_1 is distinct from all

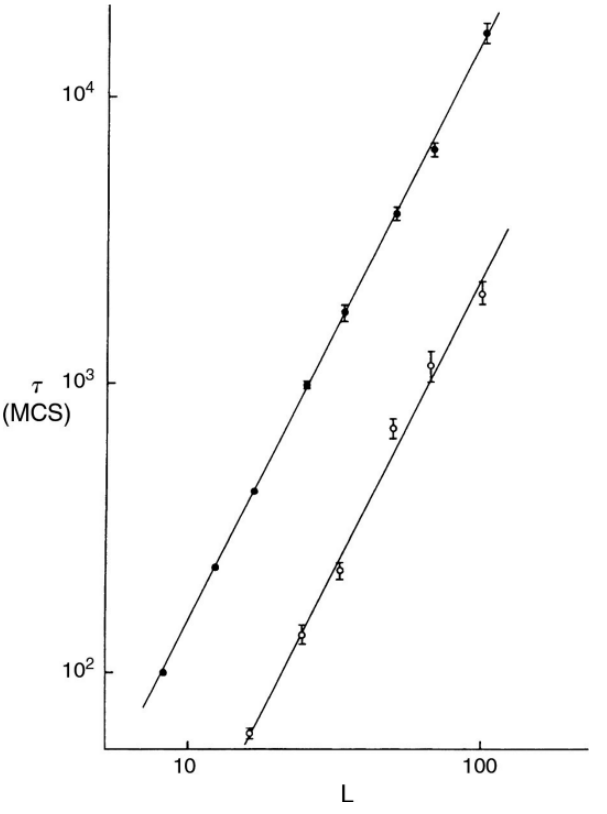

Fig. 4.17 Dynamic finite size scaling analysis for the three-dimensional Ising model at T_c. Closed circles are for the order parameter, open circles are for the internal energy. From Wansleben and Landau (1991).

other relaxation times because it increases monotonically as the temperature is lowered through T_c, while all other τ_n have their maximum somewhere in the critical region (Koch *et al.*, 1996; Koch and Dohm, 1998). The reason for this uninterupted increase of τ_1, is that below T_c it develops into the ergodic time τ_e which describes how long it takes for the system to tunnel between regions of phase space with positive and negative magnetizations. This process must occur through a high energy barrier ΔF between the two regions and $\tau_e \propto L^z \exp(\Delta F/k_B T)$. Actually, ΔF can be estimated for an Ising system (for a simulation geometry of an L^d system with periodic boundary conditions) as $2\sigma L^{d-1}$, where σ is the interfacial free energy of the system. This corresponds to the creation of a domain with two walls running through the entire simulation box to reverse the sign of the spontaneous magnetization. Thus, we obtain the estimate

$$\tau_1 = \tau_e \propto L^z \exp(2L^{d-1}\sigma/k_B T). \tag{4.65}$$

This monotonic increase of τ_1 with decreasing T corresponds to the increase in the fluctuation $\langle M^2\rangle - \langle M\rangle^2$ ($= \langle M^2\rangle$ for $H = 0$). Remember, however, that below T_c we have to use $\langle M^2\rangle - \langle |M| \rangle^2$ to take into account the symmetry

breaking; and in the same vein, below T_c it is the next relaxation time τ_3 which characterizes the decay of magnetization fluctuations in a state with non-zero spontaneous magnetization.

4.2.5.5 Final remarks

In spite of the extensive simulational work done on critical relaxation, the quality of the estimates of the dynamic exponent z is not nearly as high as that of the estimates for static exponents. The diverse techniques described above are simple in concept but complicated in their implementation. Nonetheless a reasonably good consensus has emerged for the two-dimensional Ising model between the 'best' estimates from Monte Carlo simulation, series expansion, and a clever analysis based on variational approximations of the eigenstates of the Markov matrix describing heat-bath single spin-flip dynamics (Nightingale and Blöte, 1998).

4.3 OTHER DISCRETE VARIABLE MODELS

4.3.1 Ising models with competing interactions

The Ising model with nearest neighbor interactions has already been discussed several times in this book; it has long served as a testing ground for both new theoretical methods as well as new simulational techniques. When additional couplings are added the Ising model exhibits a rich variety of behavior which depends on the nature of the added interactions as well as the specific lattice structure. Perhaps the simplest complexity can be introduced by the addition of next-nearest neighbor interactions, $\mathcal{J}_{nnn}$, which are of variable strength and sign so that the Hamiltonian becomes

$$\mathcal{H} = -\mathcal{J}_{nn} \sum_{i,j} \sigma_i \sigma_j - \mathcal{J}_{nnn} \sum_{i,k} \sigma_i \sigma_k - H \sum_i \sigma_i , \tag{4.66}$$

where the first sum is over nearest neighbor pairs and the second sum over next-nearest pairs. It is straightforward to extend the single spin-flip Metropolis method to include $\mathcal{J}_{nnn}$: the table of flipping probabilities becomes a two-dimensional array and one must sum separately over nearest and next-nearest neighbor sites in determining the flipping energy. In specialized cases where the magnitudes of the couplings are the same, one can continue to use a one-dimensional flipping probability array and simply include the contribution of the next-nearest neighbor site to the 'sum' of neighbors using the appropriate sign. If the checkerboard algorithm is being used, the next-nearest neighbor interaction will generally connect the sub-lattices; in this situation the system need merely be decomposed into a greater number of sublattices so that the spins on these new sublattices do not interact. An example is given below for the Ising square lattice.

Example

For the Ising square lattice with nearest neighbor coupling the simplest checkerboard decomposition is shown on the left. If next-nearest neighbor coupling is added the simplest possible checkerboard decomposition is shown on the right.

1	2	1	2	1	2	1	2
2	1	2	1	3	4	3	4
1	2	1	2	1	2	1	2
2	1	2	1	3	4	3	4

If both nearest and next-nearest neighbor interactions are ferromagnetic, the system will only undergo a transition to a ferromagnetic state and there are seldom complexities. One simple case which may lead to difficulties is when there are only nearest neighbor interactions which are quite different in magnitude in different directions. This may then lead to a situation in which well-ordered chains form at some relatively high temperature, and long range order sets in only at a much lower temperature. In this case it becomes very difficult for chains to overturn to reach the groundstate because each individual spin in the chain is effectively 'held in place' by its neighbors (Graim and Landau, 1981). If, however, the couplings are both antiferromagnetic, or of opposite sign, there may be multiple configurations of quite similar free energy which are separated from each other by a significant free energy barrier. The resultant sequence of states may then also have a complicated time dependence. For the simple case of nearest neighbor, antiferromagnetic interactions only, below the transition temperature the system may alternate between two different states, one in which sublattice 1 is up and sublattice 2 is down, and one in which all spins are reversed. If a strong, antiferromagnetic next-nearest neighbor interaction is added it will be necessary to decompose the system into four interpenetrating, next-nearest neighbor sublattices (s.l.), and there will be *four* different ordered states as shown below:

state	s.l.1	s.l.2	s.l.3	s.l.4
1	+	+	−	−
2	−	−	+	+
3	+	−	+	−
4	−	+	−	+

One important consequence of this behavior is that the relevant order parameter changes. For some ranges of couplings it is not immediately clear which kind of order will actually result and multiple order parameters (and their finite size behavior) must then be determined. Even if the simple antiferromagnetic states are lowest in free energy, the states shown above may be close in free energy and may appear due to fluctuations. The net result is that one must pay close attention to the symmetry of the states which are produced and to the resultant time dependence.

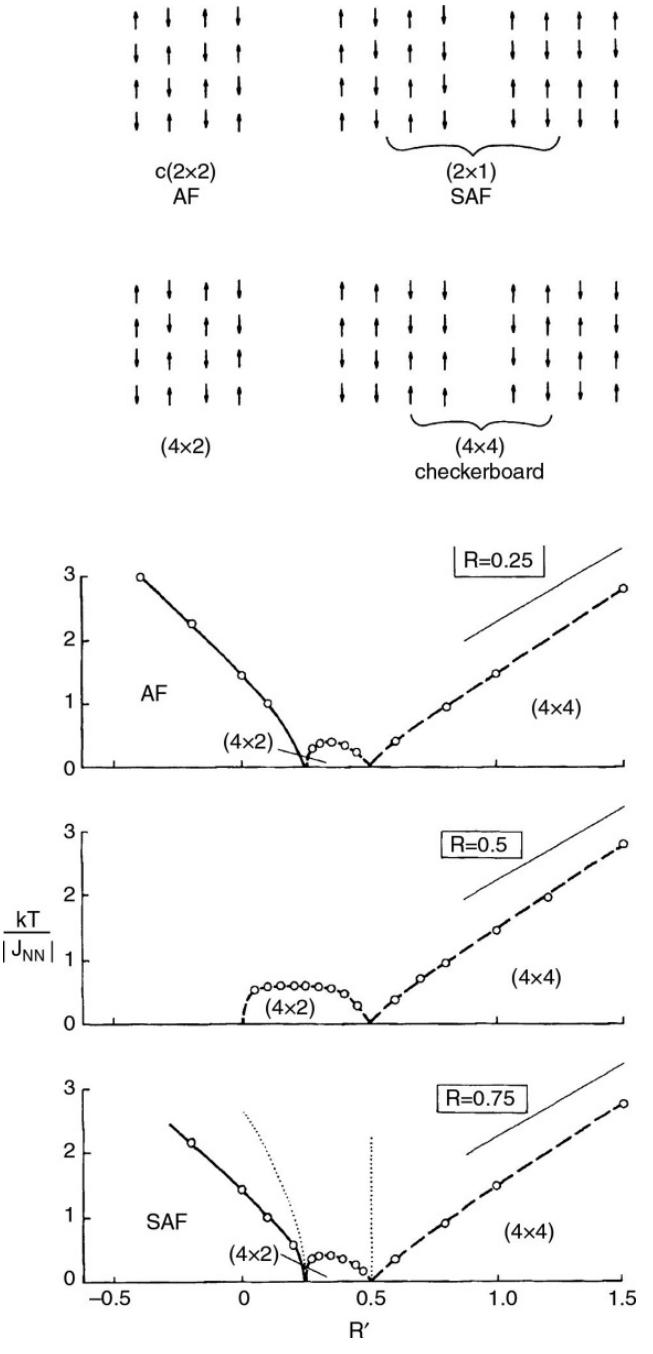

Fig. 4.18 Ising square lattice with nn-, nnn-, and third nnn-couplings: (top) possible spin configurations; (bottom) phase diagrams for different ratios of $R = \mathcal{J}_{\text{nnn}} / \mathcal{J}_{\text{nn}}$ and $R' = \mathcal{J}_{3\text{nn}} / \mathcal{J}_{\text{nn}}$. From Landau and Binder (1985).

With the inclusion of third nearest neighbor interactions the number of different states which appear becomes larger still. Other, metastable domain states also become prevalent. In Fig. 4.18 we show a number of different possible spin configurations for the Ising square lattice with competing interactions. In the bottom part of this figure we then show phase diagrams, deduced from Monte Carlo studies, for three different values of nnn-neighbor coupling as the

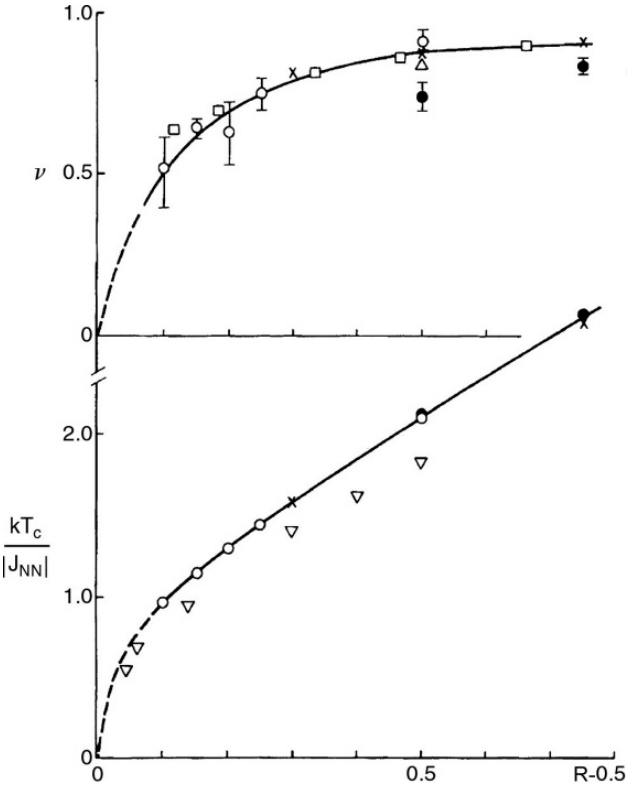

Fig. 4.19 Critical behavior for the superantiferromagnetic state in the Ising square lattice. (○) Results of the Monte Carlo block distribution analysis; (△) Monte Carlo results using finite size scaling; (×) MCRG results; (•) series expansion estimates; (□) finite strip width RG; (▽) real space RG results. From Landau and Binder (1985).

3nn-interaction is varied. For different regions of couplings, different states become lowest in free energy, and unit cells as large as 4×4 are needed to index them. When the interactions become complex, it may well be possible that entropic effects play a substantial role in determining which states actually appear. It may then be helpful to calculate multiple order parameters in order to determine which states are actually realized.

Interesting new physics may arise from competing interactions. In one of the simplest such examples, the addition of antiferromagnetic nnn-coupling to an nn-Ising square lattice antiferromagnet produces the degenerate 'superantiferromagnetic' state described earlier with non-universal critical exponents (those of the XY-model with fourth order anisotropy). The order parameter must then be redefined to take into account the degeneracy of the ordered state, but the finite size analyses which were described in Section 4.2.3 of this chapter can still be applied. For example, the crossing of the fourth order cumulant still occurs but at a different value than for the simple Ising model. Monte Carlo data were used to determine the variation of the critical temperature as well as the change in critical exponents with coupling. In Fig. 4.19 we show the comparison between the Monte Carlo estimates for T_c, as well as for ν, obtained from an analysis of the fourth order cumulant. For comparison, results obtained from a number of other methods are shown. Finite size scaling of the fourth order cumulant (block spin scaling) data showed quite clearly that the critical behavior was non-universal. This study is now rather old and

higher resolution could be easily obtained with modern computing equipment; but even these data suffice to show the variation with coupling and to test other theoretical predictions. For a detailed study of the critical behavior of this model, see Landau and Binder (1985). But this is not the whole story. Later theoretical work and simulations have suggested that the transition became first order as $R \to 0.5$ (theoretical aspects and latest simulation results are given in Jin *et al.* (2012)). The observed behavior of this system is subtle, and careful finite size and finite sampling time analysis, as described earlier in this chapter, is needed to unambiguously determine the nature of the phase boundary.

A very interesting case occurs when a competing antiferromagnetic interaction is added in only one lattice direction to an Ising ferromagnet to produce the so-called ANNNI model (Selke, 1992). For sufficiently strong antiferromagnetic interaction, the model exhibits a phase transition from the disordered phase to a 'modulated' phase in which the wavelength of the ordering is incommensurate with the lattice spacing. In $d = 2$ dimensions this phase is a 'floating phase' with zero order parameter and a power law decay of the correlation function; in $d = 3$ the ordered region contains a multitude of transitions to high order commensurate phases, i.e. phases with order which has periods that are much larger than the lattice spacing. The detailed behavior of this model to date is still incompletely understood.

4.3.2 *q*-state Potts models

Another very important lattice model in statistical mechanics in which there are a discrete number of states at each site is the q-state Potts model (Potts, 1952) with Hamiltonian

$$\mathcal{H} = -\mathcal{J} \sum_{i,j} \delta_{\sigma_i \sigma_j}, \tag{4.67}$$

where $\sigma_i = 1, 2, \ldots, q$. Thus a bond is formed between nearest neighbors *only* if they are in the same state. From the simulations perspective this model is also quite easy to simulate; the only complication is that now there are multiple choices for the new orientation to which the spin may 'flip'. The easiest way to proceed with a Monte Carlo simulation is to randomly choose one of the $q - 1$ other states using a random number generator and then to continue just as one did for the Ising model. Once again one can build a table of flipping probabilities, so the algorithm can be made quite efficient. Simple q-state Potts models on periodic lattices are known to have first order transitions for $q > 4$ in two dimensions and for $q > 2$ in three dimensions. For q close to the 'critical' values, however, the transitions become very weakly first order and it becomes quite difficult to distinguish the order of the transition without prior knowledge of the correct result. These difficulties are typical of those which arise at other weakly first order transitions; hence, Potts models serve as very useful testing grounds for new techniques, in particular in $d = 2$ where many exact results are available for comparison (Baxter, 1982).

Problem 4.7 Perform a Monte Carlo simulation of a $q = 3$ Potts model on a square lattice. Plot the internal energy as a function of temperature. Estimate the transition temperature.

Problem 4.8 Perform a Monte Carlo simulation of a $q = 10$ Potts model on a square lattice. Plot the internal energy as a function of temperature. Estimate the transition temperature. How do these results compare with those in Problem 4.7?

4.3.3 Baxter and Baxter–Wu models

Another class of simple lattice models with discrete states at each site involves multispin couplings between neighbors. One of the simplest examples is the Baxter model (1972) which involves Ising spins on two interpenetrating (next-nearest neighbor) sublattices on a square lattice; the two sublattices are coupled by a (nearest neighbor) four spin interaction so that the total Hamiltonian reads:

$$\mathcal{H} = -\mathcal{J}_{\mathrm{nnn}} \sum_{i,k} \sigma_i \sigma_k - \mathcal{J}_{\mathrm{nnn}} \sum_{j,l} \sigma_j \sigma_l - \mathcal{J}_{\mathrm{nn}} \sum_{i,j,k,l} \sigma_i \sigma_j \sigma_k \sigma_l, \qquad (4.68)$$

where the first two sums are over nnn-pairs and the last sum is over nn-plaquettes. Once again, there are only a discrete number of possible states involving each site, i.e. the number of 'satisfied' next-nearest neighbor pairs and the number of four spin plaquettes, so that tables of flipping probabilities can be constructed. There are obviously multiple degenerate states because of the different possible orientations of each of the sublattices, so the order parameter must be carefully constructed. The critical behavior of the Baxter model is non-universal, i.e. it depends explicitly on the values of the coupling constants.

Another simple, discrete state lattice model with somewhat subtle microscopic behavior considers Ising spins on a triangular lattice with nearest neighbor three-spin coupling; the model, first proposed by Baxter and Wu (1973), has the Hamiltonian

$$\mathcal{H} = -\mathcal{J}_{\mathrm{nn}} \sum_{i,j,k} \sigma_i \sigma_j \sigma_k. \qquad (4.69)$$

Even though the model is extremely simple, in a Monte Carlo simulation it has surprisingly complex behavior because different fluctuations occur at different time scales. The groundstate for this system is four-fold degenerate as shown in Fig. 4.20. This also means that the order parameter is complicated and that regions of the system may be in states which look quite different. If clusters of different ordered states 'touch' each other, a domain wall-like structure may be created with the result that the energy of the system is increased by an amount which depends upon the size of the overlap. The energy fluctuations then contain multiple kinds of excitations with different time scales, and care must be taken to insure that all characteristic fluctuations are sampled. The correlation between the time dependence of the energy and the microscopic behavior is shown in Fig. 4.21 which clearly underscores the utility of even simple scientific visualization techniques to guide our understanding of numerical results. (These data are also rather old and using modern computers it is easy to make much longer runs; they nonetheless represent an example of complexity

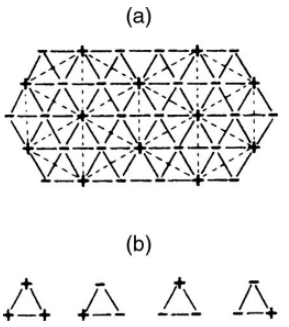

Fig. 4.20. Degenerate groundstates for the Baxter–Wu model: (a) ordered ferrimagnetic groundstate (solid lines connect nearest neighbors, dashed lines are between next-nearest neighbors); (b) elementary (nearest neighbor) plaquettes showing the four different degenerate groundstates.

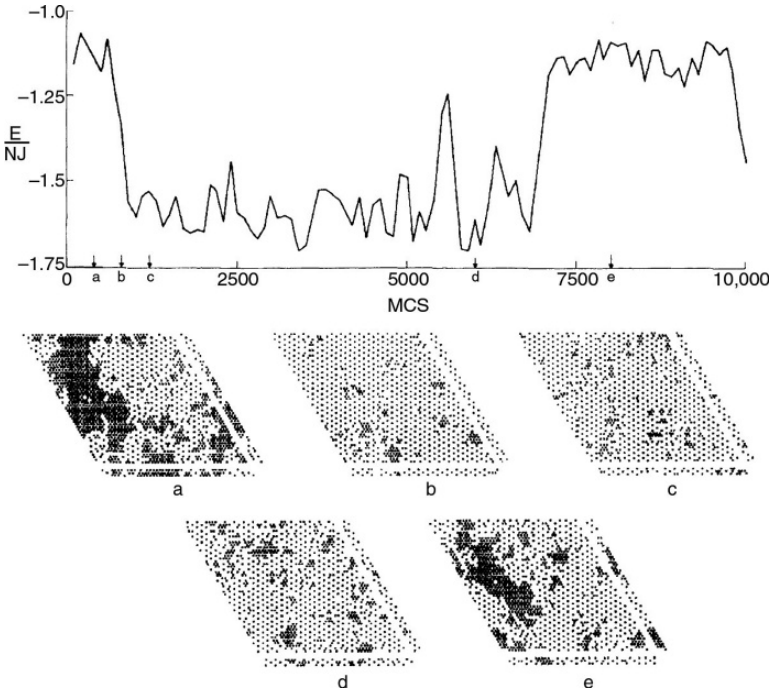

Fig. 4.21 Time dependence of the internal energy of the Baxter–Wu model and the development of 'domain-like states'. Periodic boundaries are copied from one side to another as shown in the lower portion of the figure. From Novotny and Landau (1981).

which may also occur in other systems.) This behavior also demonstrates the advantages of making occasional very long runs to test for unexpected behavior.

4.3.4 Clock models

Models with spins which may assume a continuous range of directions will be discussed in the next chapter, but a set of models which may be thought of as limiting cases of such continuous spin models with anisotropy in two

dimensions are the so-called 'clock' models. In the q-state clock model the spins can only point in one of the q possible directions on a clock with q hours on it. The Hamiltonian then looks very much like that of a continuous spin model, but we must remember that the spins may only point in a discrete number of positions:

$$\mathcal{H} = -\mathcal{J} \sum_{i,j} S_i \cdot S_j. \tag{4.70}$$

As $q \to \infty$ the model becomes a continuous spin model. Just as in the case of a high spin Ising model, the number of possible nearest neighbor states can become quite large and the flip probability table can become big. Nonetheless the Monte Carlo algorithm proceeds as before, first using a random number to select a possible new state and then calculating the energy change which a 'flip' would produce. It can also be shown that for $q = 4$, the clock model becomes exactly identical to an Ising model with interaction $\mathcal{J}/2$, so the program can be tested by comparing with the known behavior for finite Ising models. For $q > 4$, the clock model becomes a limiting case for the XY-model with q-fold anisotropy. This model has two Kosterlitz–Thouless transitions and the interpretation of the data, and location of the transitions, becomes a quite subtle matter (Challa and Landau, 1986). It is possible to use a very large value of q to approximate a continuous spin XY-model and thus take advantage of the tricks that one can employ when dealing with a model with discrete states. One must not forget, however, that asymptotically near to the transition the difference between the two models becomes evident.

4.3.5 Ising spin glass models

The field of spin glasses has a voluminous literature and the reader is directed elsewhere for in-depth coverage (see e.g. Binder and Young, 1986; Marinari *et al.*, 2000; Crisanti and Ritort, 2003). Spin glasses are magnetic systems with competing interactions which result in frozen-in disorder reminiscent of that which occurs in ordinary glass. Thus, although there is no long range order, there will be short range order with a resultant cusp in the magnetic susceptibility. Below the spin glass temperature T_f there is hysteresis and a pronounced frequency dependence when a small oscillating field is applied. These effects arise because the geometry and/or interactions give rise to 'frustration', i.e. the inability of the system to find an ordered state which satisfies all interacting neighbors. One of the simplest spin glass models (with short range interactions) employs Ising spins σ_i with Hamiltonian

$$\mathcal{H} = -\sum_{i,j} \mathcal{J}_{ij} \sigma_i \sigma_j - H \sum_i \sigma_i, \tag{4.71}$$

where the distribution $P(\mathcal{J}_{ij})$ of 'exchange constants' $\mathcal{J}_{ij}$ is of the Edwards–Anderson form

$$P(\mathcal{J}_{ij}) = [2\pi((\Delta\mathcal{J}_{ij})^2)]^{-1/2} \exp[-(\mathcal{J}_{ij} - \overline{\mathcal{J}}_{ij})^2/2(\Delta\mathcal{J}_{ij})^2] \tag{4.72}$$

or the $\pm\mathcal{J}$ form

$$P(\mathcal{J}_{ij}) = p_1\delta(\mathcal{J}_{ij} - \mathcal{J}) + p_2\delta(\mathcal{J}_{ij} + \mathcal{J}). \tag{4.73}$$

Explicit distributions of bonds are placed on the system and Monte Carlo simulations can be performed using techniques outlined earlier; however, near the spin glass freezing temperature T_f and below the time scales become very long since there is a very complicated energy landscape and the process of moving between different 'local' minima becomes difficult. Of course, the final properties of the system must be computed as an average over multiple distributions of bonds. One complication which arises from spin glass behavior is that the spontaneous magnetization of the system is no longer a good order parameter. One alternative choice is the Edwards–Anderson parameter

$$q = \overline{\langle\sigma_i\rangle^2} \tag{4.74}$$

where $\langle\cdots\rangle$ denotes the expectation value for a single distribution of bonds and the $\overline{\cdots}$ indicates an average over all bond distributions. Another choice is the local parameter

$$q = \frac{1}{N}\sum_i \sigma_i\phi_i^l, \tag{4.75}$$

where ϕ_i^l represents the spin state of site i in the lth groundstate. The Monte Carlo simulations reveal extremely long relaxation times, and the data are often difficult to interpret. (For more recent developments in this field see, e.g., Young and Kawashima, 1996; Katzgraber *et al.*, 2001, 2004; Young and Katzgraber, 2004.) In the next chapter we shall discuss improved methods for the study of spin glasses.

4.3.6 Complex fluid models

In this section we discuss briefly the application of Monte Carlo techniques to the study of microemulsions, which are examples of complex fluids. Microemulsions consist of mixtures of water, oil, and amphiphilic molecules and for varying concentrations of the constituents can form a large number of structures. These structures result because the amphiphilic molecules tend to spontaneous formation of water–oil interfaces (the hydrophilic part of the molecule being on the water-rich side and the hydrophobic part on the oil-rich side of the interface). These interfaces may then be arranged regularly (lamellar phases) or randomly (sponge phases), and other structures (e.g. vesicles) may form as well. Although real complex fluids are best treated using sophisticated off-lattice models, simplified, discrete state lattice models have been used quite successfully to study oil–water–amphiphilic systems (see, e.g., Gompper and Goos, 1995). Models studied include the Ising model with nn- and nnn-interaction and multispin interactions and the Blume–Emery–Griffiths (BEG) model with three spin coupling. These models can be easily studied using the methods described earlier in this chapter, although because of the

complicated structures which form, relaxation may be slow and the system may remain in metastable states. These systems have also been studied using a Ginzburg–Landau functional and spatial discretization. Thus the free energy functional

$$F\{\Phi\} = \int d^3r (c(\nabla^2\Phi)^2 + g(\Phi)(\nabla\Phi)^2 + f(\Phi) - \mu\Phi) \tag{4.76}$$

for a scalar order parameter Φ becomes

$$F(\Phi(r_{ij})) = c \sum_i \left(\sum_{k=1}^{3} \frac{\phi(\overline{X}_i + \hat{e}_k) - 2\phi(\overline{X}_i) + \phi(\overline{X}_i - \hat{e}_k)}{a_o^2} \right)^2 + \sum_{ij} g \left\{ \tfrac{1}{2}[\phi(\overline{X}_i) + \phi(\overline{X}_j)] \right\} \left[\frac{\phi(\overline{X}_i) - \phi(\overline{X}_j)}{a_o} \right]^2 + f(\phi) - \mu\phi \tag{4.77}$$

where a_o is the lattice constant and the $\hat{e}_k$s are the lattice vectors. Monte Carlo moves are made by considering changes in the local order parameter, i.e.

$$\Phi \rightarrow \Phi + \Delta\Phi \tag{4.78}$$

with the usual Metropolis criterion applied to determine if the move is accepted or not. Monte Carlo simulations have been used to determine phase diagrams for this model as well as to calculate scattering intensities for neutron scattering experiments.

4.4 SPIN-EXCHANGE SAMPLING

4.4.1 Constant magnetization simulations

For the single spin-flipping simulations described above, there were no conserved quantities since both energy and order parameter could change at each flip. A modification of this approach in which the magnetization of the system remains constant may be easily implemented in the following fashion. Instead of considering a single spin which may change its orientation, one chooses a pair of spins and allows them to attempt to exchange positions. This 'spin-exchange' or Kawasaki method (Kawasaki, 1972) is almost as easy to implement as is spin-flipping. In its simplest form, spin-exchange involves nearest neighbor pairs, but this constraint is not compulsory. (If one is not interested in simulating the time dependence of a model for a physical system, it may even be advantageous to allow more distant neighbor interchanges.) For instance, such an algorithm was already implemented by Binder and Stauffer (1972) for the simulation of the surface area of 'liquid droplets' of down spins surrounded by a 'gas' of up spins, with the additional constraint that the number of down spins in the 'droplet' remains constant. One examines the interacting near neighbors of both spins in the pair and determines the change in energy if the spins are interchanged. This energy difference is then

used in the acceptance procedure described above. Obviously, a pair of spins has a greater number of near neighbors than does a single spin, and even with nearest-neighbor coupling only a checkerboard decomposition requires more than two sublattices. Nonetheless, spin-exchange is straightforward to implement using table building and other tricks which can be used for spin-flip Monte Carlo. The behavior which results when this method is used is quite different from that which results using spin-flipping and will be discussed in the next several sections.

Problem 4.9 Simulate an $L = 10$ Ising square lattice using Kawasaki dynamics. Choose an initially random state and quench the system to $T = 2.0\, J/k_B$. Plot the internal energy as a function of time. Make a 'snapshot' of the initial configuration and of the last configuration generated.

4.4.2 Phase separation

At a first order transition the system separates into two distinct regions, each of which is typical of one of the two coexisting phases. (The basic ideas have been introduced in Section 2.3.) If, for example, a disordered system is quenched from some high temperature to below the critical temperature, the disordered state becomes unstable. If this is done in an AB binary alloy in which the number of each kind of atom is fixed, phase separation will occur (Gunton *et al.*, 1983). Because of the Ising-lattice gas-binary alloy equivalence, a Monte Carlo simulation can be carried out on an Ising model at fixed magnetization using spin-exchange dynamics. The structure factor $S(k, t)$ can be extracted from the Fourier transform of the resultant spin configurations and used to extract information about the nature of the phase separation. As a specific example we consider the physical situation described by Fig. 2.9 in which a binary alloy containing vacancies may evolve in time by the diffusion of atoms and vacancies. A vacancy site is chosen at random and it attempts to exchange position with one of its nearest neighbors. The probability of a jump which involves an energy change $\delta\mathcal{H}$ in which the vacancy exchanges site with an A-atom (B-atom) is denoted $W_A(W_B)$ and is given by

$$W_A = \begin{cases} \Gamma_A & \text{if } \delta\mathcal{H} < 0 \\ \Gamma_A \exp(-\delta\mathcal{H}/k_B T) & \text{if } \delta\mathcal{H} > 0 \end{cases} \tag{4.79}$$

$$W_B = \begin{cases} \Gamma_B & \text{if } \delta\mathcal{H} < 0 \\ \Gamma_B \exp(-\delta\mathcal{H}/k_B T) & \text{if } \delta\mathcal{H} > 0. \end{cases} \tag{4.80}$$

The ratio of the jump rates is then given by $\Gamma = \Gamma_B/\Gamma_A$. As an example of the results which are obtained from this Monte Carlo procedure we show characteristic results which are obtained for the structure factor for four different jump rates in Fig. 4.22. Data are shown for five different times following the quench and show the evolution of the system. For wave vectors that are small enough ($k < k_c$) the equal-time structure factor grows with time: this is

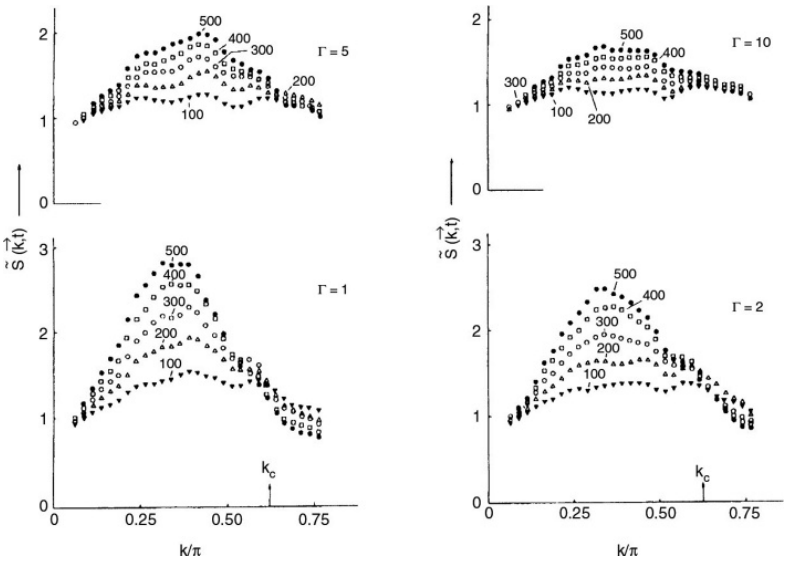

Fig. 4.22 Smoothed structure factor of an AB binary alloy with vacancies: $c_A = c_B = 0.48$, $c_V = 0.04$. From Yaldram and Binder (1991).

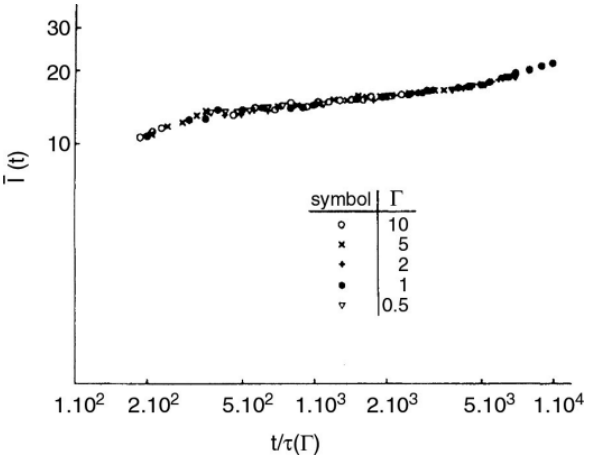

Fig. 4.23 Log–log plot of the mean cluster size vs. scaled time for phase separation in the AB binary alloy. From Yaldram and Binder (1991).

the hallmark of spinodal decomposition (see Section 2.3.2). Another important property of the developing system which needs to be understood is the development of the mean cluster size $\bar{l}$ as a function of time where

$$\bar{l}(t) = \sum_{l \geq 10} l n_l(t) \Big/ \sum_{l \geq 10} n_l(t) \tag{4.81}$$

and n_l is the number of clusters of size l. In Fig. 4.23 we show the mean cluster size against the scaled time for five different values of the jump rate. The scaling time $\tau(\Gamma)$ not only describes the behavior of the mean cluster size but is also appropriate to describe the scaling of the internal energy.

Of course, the example discussed above only refers to a simple model in order to illustrate the type of questions that can be asked. It is possible to

combine kinetic Monte Carlo methods to model the vacancy mechanism of atomic hopping processes in alloys with a quantitatively accurate description of effective interactions appropriate for real materials (Müller *et al.*, 2000, 2001, 2002). Extracting these effective interactions from 'first principles' electronic structure calculations, one derives the appropriate transition probabilities to be used in the Monte Carlo simulation.

4.4.3 Diffusion

In this section we consider lattice gas models which contain two species A and B, as well as vacancies which we denote by the symbol V. The sum of the concentrations of each species c_A, c_B, c_V is held fixed and the total of all the components is unity, i.e. $c_A + c_B + c_V = 1$. In the simulations particles are allowed to change positions under various conditions and several different types of behavior result. (See Fig. 2.9 for a schematic representation of interdiffusion in this model.)

First we consider non-interacting systems. In the simplest case there is only one kind of particle in addition to vacancies, and the particles undergo random exchanges with the vacancies. Some particles are tagged, i.e. they are followed explicitly, and the resultant diffusion constant is given by

$$D_t = f_c V D_{sp}, \tag{4.82}$$

where D_{sp} is the single particle diffusion constant in an empty lattice, V is the probability that a site adjacent to an occupied site is vacant, and f_c is the (backwards) correlation factor which describes the tendency of a particle which has exchanged with a vacancy to exchange again and return to its original position. This correlation can, of course, be measured directly by simulation. The process of interdiffusion of two species is a very common process and has been studied in both alloys and polymer mixtures. By expressing the free energy density f of the system in terms of three non-trivial chemical potentials μ_A, μ_B, μ_V, i.e.

$$f = \mu_A c_A + \mu_B c_B + \mu_V c_V, \tag{4.83}$$

we can write a Gibbs–Duhem relation, valid for an isothermal process:

$$c_A d\mu_A + c_B d\mu_B + c_V d\mu_V = 0. \tag{4.84}$$

The conservation of species leads to continuity equations

$$\partial c_A/\partial t + \vec{\nabla} \cdot \vec{j}_A = 0; \quad \partial c_B/\partial t + \vec{\nabla} \cdot \vec{j}_B = 0; \quad \frac{\partial c_V}{\partial t} + \vec{\nabla} \cdot \vec{j}_V = 0. \tag{4.85}$$

The constitutive linear equations relating the current densities to the gradients of the chemical potentials are ($\beta = 1/k_B T$)

$$\begin{aligned} \vec{j}_A &= -\beta\lambda_{AA}\nabla\mu_A - \beta\lambda_{AB}\nabla\mu_B - \beta\lambda_{AV}\nabla\mu_V, \\ \vec{j}_B &= -\beta\lambda_{BA}\nabla\mu_A - \beta\lambda_{BB}\nabla\mu_B - \beta\lambda_{BV}\nabla\mu_V, \\ \vec{j}_V &= -\beta\lambda_{VA}\nabla\mu_A - \beta\lambda_{VB}\nabla\mu_B - \beta\lambda_{VV}\nabla\mu_V, \end{aligned} \tag{4.86}$$

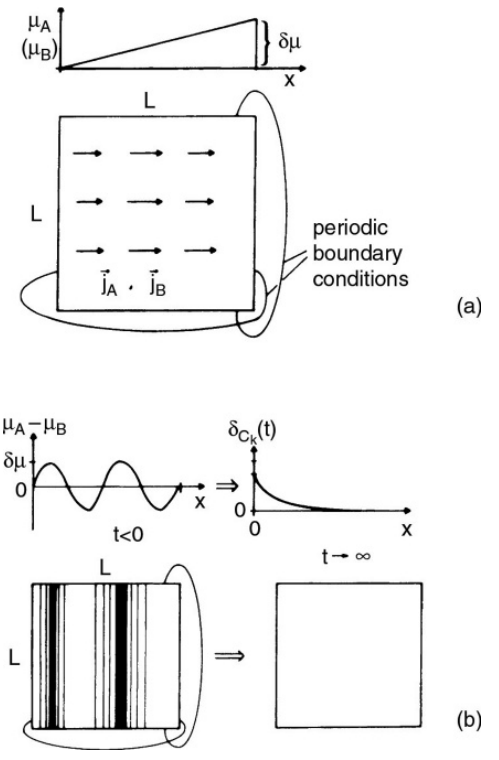

Fig. 4.24 An AB binary alloy model for the study of Onsager coefficients. (a) A linear gradient of the chemical potential μ_A (or μ_B, respectively) leading to steady-state current. (b) Periodic variation of the chemical potential difference, commensurate with the linear dimension L and leading to a concentration wave $\delta(x) = \delta c_k \times \exp(2\pi ix/\lambda)$.

where the λ_{ij} are known as Onsager coefficients. The Onsager symmetry relations reduce the number of independent parameters since $\lambda_{AB} = \lambda_{BA}, \ldots$ and the conservation of the total number of 'particles' allows us to eliminate the Onsager coefficients connected to the vacancies. The remaining Onsager coefficients can be estimated from Monte Carlo simulations of their mobilities when forces act on one of the species. In Fig. 4.24 we show a schematic view of how to set up a model. A combination of a chemical potential gradient and judicious choice of boundary conditions allows us to measure currents and thus extract estimates for Onsager coefficients. (Note that a linear increase in the chemical potential with position is inconsistent with a static equilibrium in a box, because of the periodic boundary condition: particles leaving the box through the right wall re-enter through the left wall.) For small enough $\delta\mu$ there is a linear relationship between chemical potential and the currents. Using the continuity equations together with the constitutive current expressions, we can extract coupled diffusion equations whose solutions yield decays which are governed by the Onsager coefficients. All three Onsager coefficients were successfully estimated for the non-interacting alloy (Kehr *et al.*, 1989). While the phenomenological description of diffusion in alloys as outlined above involves many unknown parameters, the obvious advantage of the simulation is that these parameters can be 'measured' in the simulation from their definition. Other scenarios may be studied by simulation. If a periodic variation of the chemical potential is created instead (see Fig. 4.24b), a concentration

wave develops. Following the ideas of linear response theory, we 'shut off' this perturbation at $t = 0$, and simply watch the decay of the concentration with time. A decay proportional to exp $(-D_{\text{int}}k^2t)$ where $k = 2\pi/\lambda$ allows us to determine the interdiffusion constant D_{int}.

Monte Carlo simulations were also used to study tracer diffusion in the binary alloy and no simple relationship was found to interdiffusion.

Diffusion can also be considered in interacting systems. Within the context of the Ising lattice gas model a particle can jump to a nn-vacancy site with probability

$$P(i \to l_i) = \exp(-\Delta E/k_{\text{B}}T), \tag{4.87}$$

where

$$\Delta E = \begin{cases} \varepsilon(l - z + 1) & \text{for repulsion } (\varepsilon < 0) \\ \varepsilon l & \text{for attraction } (\varepsilon > 0) \end{cases}, \tag{4.88}$$

where z is the coordination number and l is the number of nn-particles in the initial state. Monte Carlo simulations were used to study both self-diffusion and collective diffusion as a function of the concentration of vacancies and of the state of order in the alloy (Kehr and Binder, 1984). Similarly, two-dimensional models of adsorbed monolayers can be considered and the self-diffusion and collective diffusion can be studied (Sadiq and Binder, 1983; Ala-Nissila *et al.*, 2002). Again, it is possible to combine such modeling (see also Kang and Weinberg, 1989; Fichthorn and Weinberg, 1991) of adatom hopping processes with an atomistically realistic description of the energy minima of the adsorption sites and the energy barriers separating them, using 'first principles' electronic structure calculations to predict the corresponding hopping rates and transition probabilities for the resulting 'kinetic Monte Carlo' modeling.

This approach (also sometimes termed '*ab initio* atomistic thermodynamics', e.g. Reuter and Scheffler (2002, 2003) can also be extended to model kinetic processes far from thermal equilibrium, such as the kinetics of heterogeneous catalysis (Reuter *et al.*, 2004a, 2004b).

More comments on related Monte Carlo simulations for non-equilibrium processes and 'kinetic Monte Carlo' methods will be given later in this book (Chapter 10).

4.4.4 Hydrodynamic slowing down

The conservation of the concentration (or magnetization) during a simulation also has important consequences for the kinetics of fluctuations involving long length scales. If we consider some quantity A which has density ρ_{A} the appropriate continuity equation is

$$\frac{\partial \rho_{\text{A}}(x,t)}{\partial t} + \nabla \cdot \vec{\text{j}}_{\text{A}}(x,t) = 0 \tag{4.89}$$

where $\vec{j}_A$ is a current density. Near equilibrium and for local changes of A, we may approximate the current by

$$\vec{j}_A(x,t) = D_{AA}\nabla a(x,t). \tag{4.90}$$

Taking the Fourier transform of Eqn. (4.89) and integrating we find

$$A(k,t) = A(k,\infty) + [A(k,0) - A(k,\infty)]e^{-D_{AA}k^2 t}. \tag{4.91}$$

This equation exhibits 'hydrodynamic slowing down' with characteristic time $\tau_{AA}(k) = (D_{AA}k^2)^{-1}$. This argument justifies the result already discussed in Section 2.3.4. Thus, equilibrium will be approached quite slowly for all properties which describe long wavelength (i.e. small k) properties of the system.

4.5 MICROCANONICAL METHODS

4.5.1 Demon algorithm

In principle, a microcanonical method must work at perfectly constant energy. The demon algorithm first proposed by Creutz (1983) is not strictly microcanonical, but for large systems the difference becomes quite small. The procedure is quite simple. One begins by choosing some initial state. A 'demon' then proceeds through the lattice, attempting to flip each spin in turn and either collecting energy given off by a spin-flip or providing the energy needed to enable a spin-flip. The demon has a bag which can contain a maximum amount of energy, so that if the capacity of the bag is reached no spin-flip is allowed which gives off energy. On the other hand, if the bag is empty, no flip is possible that requires energy input. Thus, the energy in the bag E_D will vary with time, and the mean value of the energy stored in the bag can be used to estimate the mean value of the inverse temperature $K = \mathcal{J}/k_B T$ during the course of the simulation for a square lattice,

$$K = \tfrac{1}{4}\ln(1 + 4\mathcal{J}/\langle E_D\rangle). \tag{4.92}$$

If the bag is too big, the simulation deviates substantially from the microcanonical condition; if the bag is too small, it becomes unduly difficult to produce spin-flips. Note that once the initial state is chosen, the method becomes deterministic.

Problem 4.10 Simulate an $L = 10$ Ising square lattice using the microcanonical 'demon' method at two different values of energy E and estimate the temperatures. Carry out canonical ensemble simulations at these temperatures and compare the values of energy with your initial choices of E.

4.5.2 Dynamic ensemble

This method uses a standard Monte Carlo method for a system coupled to a suitably chosen finite bath (Hüller, 1993). We consider an N-particle system

with energy E coupled to a finite reservoir which is an ideal gas with M degrees of freedom and kinetic energy k. One then studies the micro-canonical ensemble of the total, coupled system with fixed total energy G. An analysis of detailed balance shows that the ratio of the transition probabilities between two states is then

$$\frac{W_{b\to a}}{W_{a\to b}} = (G - E_a)^{\frac{N-2}{N}} \Big/ (G - E_b)^{\frac{N-2}{N}} \approx e^{-\zeta(E_b - E_a)} \tag{4.93}$$

where $\zeta = (N-2)/2Nk_b$ and where $k_b = (G - E_b)/N$ is the mean kinetic energy per particle in the bath. The only difference in the Monte Carlo method is that the effective inverse temperature ζ is adjusted dynamically during the course of the simulation. Data are then obtained by computing the mean value of the energy on the spin system $\langle E \rangle$ and the mean value of the temperature from $\langle k_b \rangle$. This method becomes accurate in the limit of large system size. Plots of E vs. T then trace out the complete 'van der Waals loop' at a first order phase transition.

4.5.3 Q2R

The Q2R cellular automaton has been proposed as an alternative, microcanonical method for studying the Ising model. In a cellular automaton model the state of each spin in the system at each time step is determined completely by consideration of its nearest neighbors at the previous time step. The Q2R rule states that a spin is flipped if, and only if, half of its nearest neighbors are up and half down. Thus, the local (and global) energy change is zero. A starting spin configuration of a given energy must first be chosen and then the Q2R rule applied to all spins; this method is thus also deterministic after the initial state is chosen. Thermodynamic properties are generally well reproduced, although the susceptibility below T_c is too low. (Other cellular automata will be discussed in Chapter 8.)

Problem 4.11 Simulate an $L = 10$, $q = 10$ Potts model square lattice using a microcanonical method and estimate the transition temperature. How does your answer compare with that obtained in Problem 4.8?

4.6 GENERAL REMARKS, CHOICE OF ENSEMBLE

We have already indicated how models may be studied in different ensembles by different methods. There are sometimes advantages in using one ensemble over the other. In some cases there may be computational advantages to choosing a particular ensemble, in other situations there may be a symmetry which can be exploited in one ensemble as opposed to the other. One of the simplest cases is the study of a phase diagram of a system with a tricritical point. Here there are both first order and second order transitions. As shown in Fig. 4.25 the phase boundaries look quite different when shown in the canonical and grand-canonical ensembles. Thus, for low 'density' (or magnetization in

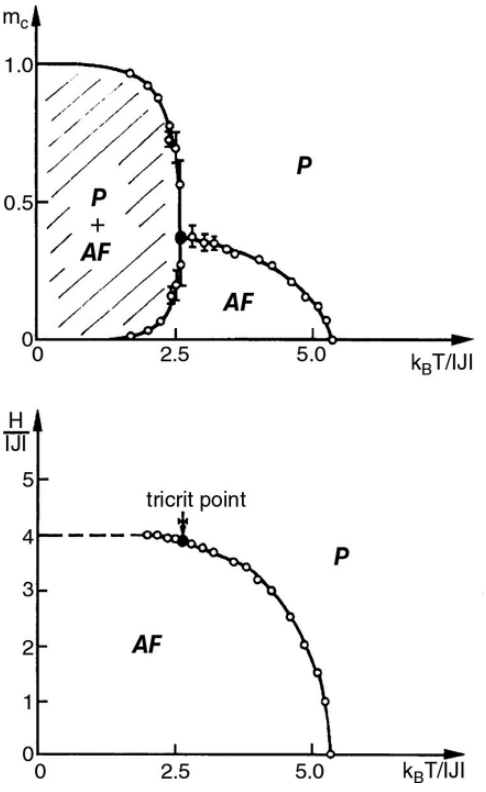

Fig. 4.25 Phase diagram for an Ising antiferromagnet with nearest and next-nearest neighbor couplings with a tricritical point: (top) canonical ensemble, the shaded area is a region of two-phase coexistence; (bottom) grand canonical ensemble.

magnetic language) two phase transitions are encountered as the temperature is increased whereas if the 'field' is kept fixed as the temperature is swept only a single transition is found. Of course, to trace out the energy–field relation in the region where it is double valued, it is preferable to use a microcanonical ensemble (as was described in the previous section) or even other ensembles, e.g. a Gaussian ensemble (Challa and Hetherington, 1988). The use of a microcanonical ensemble for the study of protein folding will be given in a later chapter.

A situation in which it is preferable to use an ensemble which differs from those available to the experimentalist is the case of fluid or solid binary (A, B) mixtures. In the laboratory, for a given volume V and temperature T, the particle numbers N_A and N_B will be fixed (i.e. the relevant ensemble is the canonical ensemble). In a simulation it is often preferable to work in the 'semi-grand canonical ensemble' in which only the total number of particles $N = N_A + N_B$ is fixed and an additional intensive variable, the chemical potential difference μ enters the Boltzmann factor in the transition probability. Due to the difference in chemical potential $\Delta\mu$, 'identity switches', $A \to B$ or $B \to A$ may occur as attempted Monte Carlo moves. This is not 'alchemy' (like medieval chemists trying to transform lead into gold) but a valid method in statistical mechanics that is preferred because of its faster equilibration (in particular for solid alloys where ordered superstructures like in β-brass or copper–gold alloys occur). An example for the use of this ensemble will be given in Chapter 6 of this book.

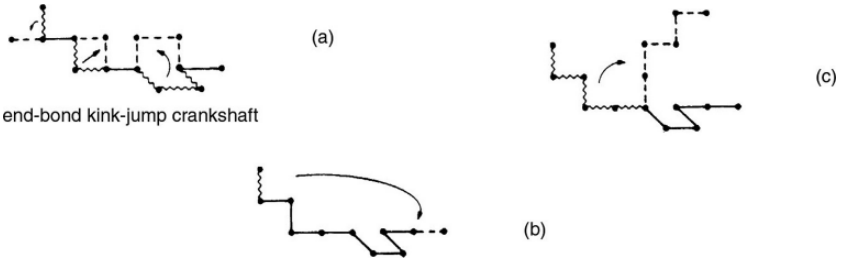

Fig. 4.26 Dynamic Monte Carlo algorithms for SAWs on a simple cubic lattice: (a) generalized Verdier–Stockmayer algorithm; (b) slithering snake algorithm; (c) pivot algorithm.

4.7 STATICS AND DYNAMICS OF POLYMER MODELS ON LATTICES

4.7.1 Background

Real polymers are quite complex and their simulation is a daunting task (Binder, 1995). There are a number of physically realistic approximations which can be made, however, and these enable us to construct far simpler models which (hopefully) have fundamentally the same behavior. First we recognize that the bond lengths of polymers tend to be rather fixed as do bond angles. Thus, as a more computationally friendly model we may construct a 'polymer' which is made up of bonds which connect nearest neighbor sites (monomers) on a lattice and which obey an excluded volume constraint. The sites and bonds on the lattice do not represent individual atoms and molecular bonds but are rather the building blocks for a coarse-grained model. Even within this simplified view of the physical situation simulations can become quite complicated since the chains may wind up in very entangled states in which further movement is almost impossible.

4.7.2 Fixed bond length methods

The polymer model just described may be viewed as basically a form of self-avoiding-walk (SAW) which can be treated using Monte Carlo growth algorithms which have already been discussed (see Section 3.8.3). Another class of algorithms are dynamic in nature and allow random moves of parts of the polymer which do not allow any change in the length of a bond connecting two monomers. The range of possible configurations for a given polymer model can be explored using a variety of different 'dynamic' Monte Carlo algorithms which involve different kinds of move, three examples of which are shown in Fig. 4.26. In the generalized 'kink-jump' method single sites may be moved, obeying the restriction that no bond length changes. In the 'slithering snake' (reptation) method, a bond is removed from one end and then glued to the other end of the polymer in a randomly chosen orientation. In Fig. 4.26c we show the pivot ('wiggle') move, in which a large part of the chain is rotated about a single site in the chain. (Obviously, not all moves reflect real, physical

time development.) Different kinds of moves are useful for avoiding different kinds of 'trapped' configurations, and an intelligent choice of trial moves is essential in many cases. There are a large number of off-lattice models which are useful for studying more complex behavior, but these are beyond the scope of consideration here. More details about the methods shown in Fig. 4.26 can be found in Kremer and Binder (1988) and additional methods are discussed by Sokal (1995) and Attig *et al*. (2004). For dense melts a new kind of non-local move shows great promise. Termed the 'double pivot', this trial move breaks bonds in two neighboring chains and attempts to reconnect the monomers such that the chains remain monodisperse. A more detailed description is given by Baschnagel *et al*. in Attig *et al*. (2004).

Of course, for dense systems of long polymers, simple methods of simulation become quite inefficient. One very successful innovative algorithm builds upon old ideas from the early days of Monte Carlo simulations (Rosenbluth and Rosenbluth, 1955; Wall and Erpenbeck, 1959) by combining the biasing of the weights of new configurations with enrichment. The resulting algorithm (Grassberger, 1997), known as PERM ('pruned and enriched Rosenbluth method', sometimes also termed the 'go with the winners' algorithm), has greatly extended the size of systems that may be studied with a reasonable amount of effort. In the application to the simplest case of self-avoiding walks, chains do not die when an attempt is made to form a bond to an already occupied site. Instead, such attempts are avoided completely, but a bias is introduced by giving different weights to the chains that are actually produced by the addition of 'acceptable' bonds. In a systematic fashion, chains with too low a weight are eliminated, i.e. 'pruned', and chains whose weight exceeds a certain value are copied, i.e. 'enriched'. As a result, all chains contributed with approximately the same weight and the exponential attrition of the simple methods is avoided. PERM has been used to simulate chains of lengths up to 10^6 in the investigation of three-dimensional θ-polymers (see Section 4.7.6). More details on this PERM algorithm and hints to the recent literature will be given in Section 4.7.8.

Bond fluctuation Monte Carlo method

(1) Choose an initial state.
(2) Randomly choose a monomer.
(3) Randomly choose a 'plaquette' (from among the allowed possibilities) to which a move will be attempted.
(4) Check the excluded volume and bond length restrictions; if these are violated return to step (2).
(5) Calculate the energy change ΔE which results if the move is accepted.
(6) Generate a random number r such that $0 < r < 1$.
(7) If $r < \exp(-\Delta E/k_B T)$, accept the move.
(8) Choose another monomer and go to (3).

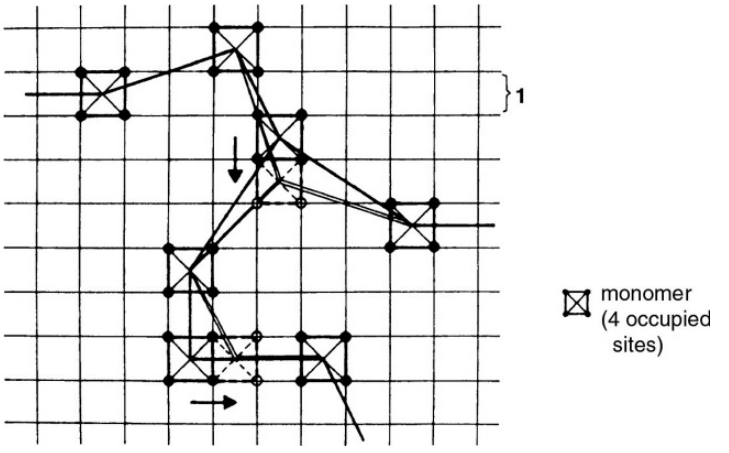

Fig. 4.27 Sample moves for the Bond fluctuation algorithm on a square lattice.

4.7.3 Bond fluctuation method

A very powerful 'dynamic' method which relaxes the rigid bond constraint slightly employs the 'bond fluctuation' model (Carmesin and Kremer, 1988). In this approach a monomer now occupies a nearest neighbor plaquette and attempts to move randomly by an amount which does not stretch or compress the bonds to its neighbors too much, and in the process to expand the range of configuration space which can be explored. Note that these moves may also allow some change in the bond angle as well as bond length. The excluded volume constraint is obeyed by not allowing overlap of monomer plaquettes. Examples of possible moves are shown in Fig. 4.27. At each step a randomly chosen monomer moves to a randomly chosen plaquette subject to excluded volume constraints as well as the limitations on bond length mentioned above. The bond fluctuation method can be effective in getting the system out of 'blocking' configurations and, as shown in Fig. 4.27, can also be applied to lattice model branched polymers.

The PERM algorithm described in Section 4.7.2 has also been successfully applied to the bond fluctuation model (Grassberger, 1997) using a stochastic version of the algorithm described in the preceding section.

4.7.4 Enhanced sampling using a fourth dimension

For densely packed systems, such as collapsed polymers, the relaxation times can become exceedingly long. This problem arises because the combination of the high density and excluded volume requires cooperative rearrangements of atoms in order for substantial changes to occur. A novel and general approach to the reduction of the characteristic time scales in dense systems (Paul and Müller, 2001) allows the particles of a three-dimensional system to move in four spatial dimensions. Every state of the system with all particles having the same coordinate in the fourth direction is then a valid configuration of

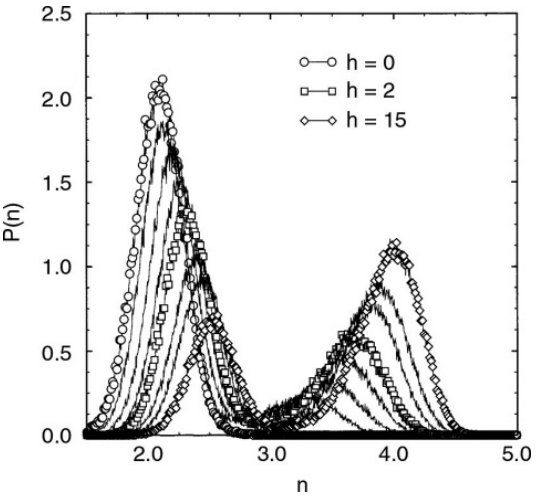

Fig. 4.28 Distribution of the number of contacts per monomer (internal energy) for the four-dimensional expanded ensemble algorithm for an $N = 256$ homopolymer. For $h = 15$ a three-dimensional configuration is produced; for $h = 0$ there is equal occupation of the replicated lattices. After Paul and Müller (2001).

the three-dimensional system of interest. The Hamiltonian of this expanded system is given by

$$\mathcal{H} = \mathcal{H}_o + \sum_{i=1}^{N} h\, x_4(i) \tag{4.94}$$

where $\mathcal{H}_o$ is the Hamiltonian of the physical (i.e. three-dimensional) system and $x_4(i)$ is the coordinate of the ith particle in the fourth dimension. The effective applied field h determines how the particles are distributed in the fourth dimension. The partition function of the expanded ensemble is then

$$\mathcal{Z} = \sum_{h} \frac{1}{W(h)} \sum_{\{c\}} \exp\left\{ -\left(\mathcal{H}_o + \sum_{i=1}^{N} h\, x_4(i) \right) \Big/ k_B T \right\}. \tag{4.95}$$

This approach has been implemented for simulations using the bond fluctuation method to study the coil-globule transition of a system of homopolymers on a simple cubic lattice. The linear dimension in the fourth dimension was only $L_4 = 2$, so the system is composed of two three-dimensional lattices. Figure 4.28 shows data for the distribution of the number of contacts for different values of h when the system is deep in the collapsed state. For $h = 0$, both three-dimensional sublattices are equally occupied and only a single peak appears in the distribution. For $h = 15$, however, a double peak is clearly evident with the two maxima corresponding to a liquid state and to a solid state. The collapse of polymer chains and the subsequent crystallization has also been studied within the context of the bond fluctuation model using the Wang–Landau algorithm (see Section 7.8) by Rampf *et al.* (2006).

4.7.5 The 'wormhole algorithm' – another method to equilibrate dense polymeric systems

For dilute polymer solutions simulations sampling difficulties arise only when the chain lengths are very large, and then methods such as the 'pivot algorithm' (Madras and Sokal, 1988; Sokal, 1995) or the PERM method (Grassberger 1997) described in Section 4.7.2, are very useful. In fact, with the latter method one can reach chain lengths of $N = 10^6$ monomers, at least for favorable cases such as self-avoiding walks with attractive nearest neighbor interaction on the simple cubic lattice near the θ-point. Unfortunately, none of these powerful methods works for very dense polymeric systems.

A new algorithm that is suitable for dense systems of both homopolymers and heteropolymers was recently invented by Houdayer (2002) and applied by Houdayer and Müller (2002) to elucidate the phase behavior of random copolymer melts. This so-called 'wormhole algorithm' is a generalization of the reptation algorithm (also known as the 'slithering snake' algorithm, see Fig. 4.26b), and is able to completely displace a polymer in a time that scales proportional to N^2. The algorithm consists of the following steps. (i) 'Wormhole drilling step' – one attempts to move one randomly chosen end-monomer to a new, random position. The old bond is broken and a virtual one appended to the other end of the polymer. (This may be a bond of arbitrarily large length.) (ii) Standard reptation step: randomly choose one end-monomer and try to move it to the other end of the polymer by connecting it with a randomly chosen bond (drawn from the standard set of bonds of the particular model that is being simulated). For this move one considers the virtual bond as if it were a normal one so that the polymer has only two ends. (iii) End test: if the polymer is in two pieces, proceed to step (ii). Otherwise, the trial move is complete and is accepted with probability one, while steps (i) and (ii) are accepted only according to the standard Metropolis probability $P_{\mathrm{M}}(\Delta E) = \min[1, \exp(-\Delta E/k_{\mathrm{B}}T)]$, where ΔE is the energy difference generated by the trial move.

For a proof that this algorithm satisfies detailed balance see Houdayer (2002). Obviously, the nature of the monomers and their order along the chain are preserved so this algorithm can be used for heteropolymers.

4.7.6 Polymers in solutions of variable quality: θ-point, collapse transition, unmixing

So far the only interaction between monomers that are not nearest neighbors along the chain, is the (infinitely strong) repulsive excluded volume interaction. Obviously, this is an extremely simplified view of the actual interactions between the effective monomers that form a real macromolecule. Physically, this corresponds to the 'athermal' limit of a polymer chain in a good solvent: the solvent molecules do not show up explicitly in the simulation, they are just represented by the vacant sites of the lattice.

Given the fact that interactions between real molecules or atoms in fluids can be modeled rather well by the Lennard–Jones interaction, which is strongly repulsive at short distances and weakly attractive at somewhat longer distances, it is tempting to associate the above excluded volume interaction (incorporated both in the SAW and the bond fluctuation model) with the repulsive part of the Lennard–Jones interaction, and add an attractive energy which acts at somewhat longer distances, to represent the attractive part of the Lennard–Jones interaction. The simplest choice for the SAW model is to allow for an energy, ε, if a pair of monomers (which are not nearest neighbors along the chain) occupy nearest neighbor sites on the lattice. In fact, such models can be (and have been) studied by simple sampling Monte Carlo methods as described in Chapter 3. To do this one simply has to weigh each generated SAW configuration with a weight proportional to the Boltzmann factor $\exp(n\varepsilon / k_{\mathrm{B}}T)$, n being the number of such nearest neighbor contacts in each configuration. However, the problem of generating a sufficiently large statistical sample for long chains is now even worse than in the athermal case: we have seen that the success rate to construct a SAW from unbiased growth scales as $\exp(-\text{const.}\, N)$, for chains of N steps, and actually a very small fraction of these successfully generated walks will have a large Boltzmann weight. Therefore, for such problems, the 'dynamic' Monte Carlo methods treated in the present chapter are clearly preferred.

While in the case of the pure excluded volume interaction the acceptance probability is either one (if the excluded volume constraint is satisfied for the trial move) or zero (if it is not), we now have to compute for every trial move the change in energy $\Delta E = \Delta n\varepsilon$ due to the change Δn in the number of nearest neighbor contacts due to the move. This energy change has to be used in the acceptance probability according to the Metropolis method in the usual way, for all trial moves that satisfy the excluded volume constraint. This is completely analogous to the Monte Carlo simulation of the Ising model or other lattice models discussed in this book.

Of course, it is possible to choose interaction energies that are more complicated than just nearest neighbor. In fact, for the bond fluctuation model discussed above it is quite natural to choose an attractive interaction of somewhat longer range, since the length of an effective bond (remember that this length is in between 2 and $\sqrt{10}$ lattice spacings in $d = 3$ dimensions) already creates an intermediate length scale. One then wishes to define the range of the attractive interaction such that in a dense melt (where 50% or more of the available lattice sites are taken by the corners of the cubes representing the effective monomers) an effective monomer interacts with all nearest neighbor effective monomers that surround it. This consideration leads to the choice (e.g. Wilding *et al.*, 1996) that effective monomers experience an energy ε if their distance r is in the range $2 \leq r \leq \sqrt{6}$ and zero else. In the bond fluctuation algorithm quoted above, the presence of some energy parameters such as ε was already assumed.

What physical problems can one describe with these models? Remember that one typically does not have in mind to simulate a macromolecule in

vacuum but rather in dilute solution, so the vacant sites of the lattice represent the small solvent molecules, and hence ε really represents a difference in interactions $(\varepsilon_{mm} + \varepsilon_{ss})/2 - \varepsilon_{ms}$ where ε_{mm}, ε_{ss}, ε_{ms} stand for interactions between pairs of monomers (mm), solvent (ss), and monomer–solvent (ms), respectively. In this sense, the model is really a generalization of the ordinary lattice model for binary alloys (A, B), where one species (A) is now a much more complicated object, taking many lattice sites and exhibiting internal configurational degrees of freedom. Thus already the dilute limit is non-trivial, unlike the atomic binary mixture where both species (A, B) take a lattice site and only the concentrated mixture is of interest. Changing the parameter $\varepsilon/k_B T$ then amounts to changing the quality of the solvent: the larger $\varepsilon/k_B T$ the more the polymer coil contracts, and thus the mean square radius of gyration $\langle R^2_{gyr}\rangle_{N,T}$ is a monotonically decreasing function when $\varepsilon/k_B T$ increases. Although this function is smooth and non-singular for any finite N, a singularity develops when the chain length N diverges: for all temperatures T exceeding the so-called 'theta temperature', θ, we then have the same scaling law as for the SAW, $\langle R^2_{gyr}\rangle_{N,T} = A(T)N^{2\nu}$ with $\nu \approx 0.588$, only the amplitude factor $A(T)$ depends on temperature, while the exponent does not. However, for $T = \theta$ the macromolecule behaves like a simple random walk, $\langle R^2_{gyr}\rangle_{N,T} = A'(\theta)N$ (ignoring logarithmic corrections), and for $T < \theta$ the chain configurations are compact, $\langle R^2_{gyr}\rangle_{N,T} = A''(T)N^{2/3}$. This singular behavior of a single chain is called the 'collapse transition'. (Generalizations of this simple model also are devised for biopolymers, where one typically has a sequence formed from different kinds of monomers, such as proteins where the sequence carries the information about the genetic code. Simple lattice models for proteins will be discussed in Chapter 13, and more sophisticated models for protein folding will follow in Chapter 14.)

Now we have to add a warning for the reader: just as power laws near a critical point are only observed sufficiently close, also the power laws quoted above are only seen for $N \to \infty$; in particularly close to θ one has to deal with 'crossover' problems: for a wide range of N for T slightly above θ the chain already behaves classically, $\langle R^2_{gyr}\rangle \propto N$, and only for very large N does one have a chance to detect the correct asymptotic exponent. In fact, the θ-point can be related to tricritical points in ferromagnetic systems (de Gennes, 1979). Thus the Monte Carlo study of this problem is quite difficult and has a long history. Now it is possible to simulate chains typically for N of the order of 10^4, or even longer, and the behavior quoted above has been nicely verified, both for linear polymers and for star polymers (Zifferer, 1999). A combination of all three algorithms shown in Fig. 4.26 is used there.

The simulation of single chains is appropriate for polymer solutions only when the solution is so dilute that the probability that different chains interact is negligible. However, a very interesting problem results when only the concentration of monomers is very small (so most lattice sites are still vacant) but typically the different polymer coils already strongly penetrate each other. This case is called the 'semidilute' concentration regime (de Gennes, 1979). For

good solvent conditions, excluded volume interactions are screened at large distances, and the gyration radius again scales classically, $\langle R^2_{\rm gyr}\rangle_{N,T} = A(T,\phi)N$, where ϕ is the volume fraction of occupied lattice sites. While the moves of types (a) and (b) in Fig. 4.26 are still applicable, the acceptance probability of pivot moves (type c) is extremely small, and hence this algorithm is no longer useful. In fact, the study of this problem is far less well developed than that of single polymer chains, and the development of better algorithms is still an active area of research (see e.g. the discussion of the configurational bias Monte Carlo algorithm in Chapter 6 below). Thus, only chain lengths up to a few hundred are accessible in such many-chain simulations.

When the solvent quality deteriorates, one encounters a critical point $T_c(N)$ such that for $T < T_c(N)$ the polymer solution separates into two phases: a very dilute phase ($\phi_I(T) \to 0$) of collapsed chains, and a semidilute phase ($\phi_{II}(T) \to 1$ as $T \to 0$) of chains that obey Gaussian statistics at larger distances. It has been a longstanding problem to understand how the critical concentration $\phi_c(N)$ ($= \phi_I(T_c) = \phi_{II}(T_c)$) scales with chain length N, as well as how $T_c(N)$ merges with θ as $N \to \infty$, $\phi_c(N) \propto N^{-x}$, $\theta - T_c(N) \propto N^{-y}$, where x, y are some exponents (Wilding *et al.*, 1996). A study of this problem is carried out best in the grand-canonical ensemble (see Chapter 6), and near $T_c(N)$ one has to deal with finite size rounding of the transition, very similar to the finite size effects that we have encountered for the Ising model.

This problem of phase separation in polymer solutions is just one problem out of a whole class of many-chain problems, where the 'technology' of an efficient simulation of configurations of lattice models for polymer chains and the finite size scaling 'technology' to analyze critical phenomena and phase coexistence need to be combined in order to obtain most useful results. One other example, the phase diagram of 'equilibrium polymers', will now be described in more detail below.

4.7.7 Equilibrium polymers: a case study

Systems in which polymerization is believed to take place under conditions of chemical equilibrium between the polymers and their respective monomers are termed 'living polymers'. These are long linear-chain macromolecules that can break and recombine, e.g. liquid and polymer-like micelles. (In fact, in the chemistry community the phrase 'living polymers' is applied to radical initiated growth, or scission, that can occur only at one end of the polymer. In the model presented here, these processes can occur any place along the polymer chain. These systems are sometimes now referred to as 'equilibrium polymers'.) In order to study living polymers in solutions, one should model the system using the dilute $n \to 0$ magnet model (Wheeler and Pfeuty, 1981); however, theoretical solution presently exists only within the mean field approximation (Flory, 1953). For semiflexible chains Flory's model predicts a first order phase transition between a low temperature ordered state of stiff parallel rods and a high temperature disordered state due to disorientation of the chains.

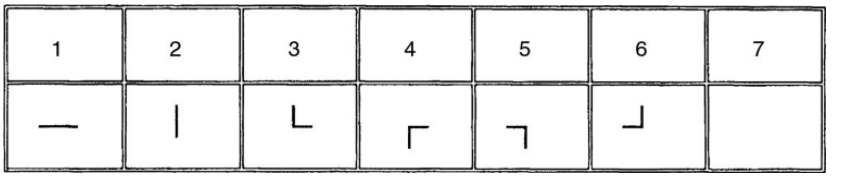

Fig. 4.29 Different allowed monomer bond states and their Potts representation.

Simulating the behavior of a system of living polymers is extremely difficult using a description which retains the integrity of chains as they move because the dynamics becomes quite slow except in very dilute solutions. An alternative model for living polymers, which is described in more detail elsewhere (Milchev, 1993), maps the system onto a model which can be treated more easily. Consider regular L^d hypercubic lattices with periodic boundary conditions and lattice sites which may either be empty or occupied by a (bifunctional) monomer with two strong (covalent) 'dangling' bonds, pointing along separate lattice directions. Monomers fuse when dangling bonds of nearest-neighbor monomers point toward one another, releasing energy $v > 0$ and forming the backbone of self-avoiding polymer chains (no crossing at vertices). Right-angle bends, which ensure the semiflexibility of such chains, are assigned an additional activation energy $\sigma > 0$ in order to include the inequivalence between rotational isomeric states (e.g. *trans* and *gauche*) found in real polymers. The third energetic parameter, w, from weak (van der Waals) *inter* chain interactions, is responsible for the phase separation of the system into dense and sparse phases when T and/or μ are changed. w is thus the work for creation of empty lattice sites (holes) in the system. One can define $q = 7$ possible states, S_i, of a monomer i on a two-dimensional lattice (two straight 'stiff' junctions, $S_i = 1, 2$, four bends, $S_i = 3, \ldots, 6$, and a hole $S_i = 7$), and $q = 16$ monomer states in a simple cubic lattice. The advantage of this model is that it can be mapped onto an unusual q-state Potts model and the simulation can then be carried out using standard single spin-flip methods in this representation. The Hamiltonian for the model can be written:

$$\mathcal{H} = \sum_{i<j} \mathcal{F}_{ij} n(S_i) n(S_j) - \sum_i (\mu + \varepsilon) n(S_i), \tag{4.96}$$

where $n(S_i) = 1$ for $i = 1, 2, \ldots, 6$, and $n(S_i) = 0$ (a hole) for $i = 7$ in two dimensions. Note that the interaction constant depends on the mutual position of the nearest neighbor monomer states, $\mathcal{F}_{ij} \neq \mathcal{F}_{ji}$. Thus, for example, $\mathcal{F}_{13} = -w$ whereas $\mathcal{F}_{31} = -v$. The local energies $\varepsilon_i = \sigma$ for the bends, and $\varepsilon_i = 0$ for the *trans* segments. The mapping to the different Potts states is shown in Fig. 4.29. The groundstates of this model depend on the relative strengths of v, w and σ; long chains at low temperature are energetically favored only if $v/w > 1$. This model may then be simulated using single spin-flip methods which have already been discussed; thus the polymers may break apart or combine quite easily. (The resultant behavior will also give the correct static properties of a polydisperse solution of 'normal' polymers, but the time development will obviously be incorrect.) Even using the Potts model mapping,

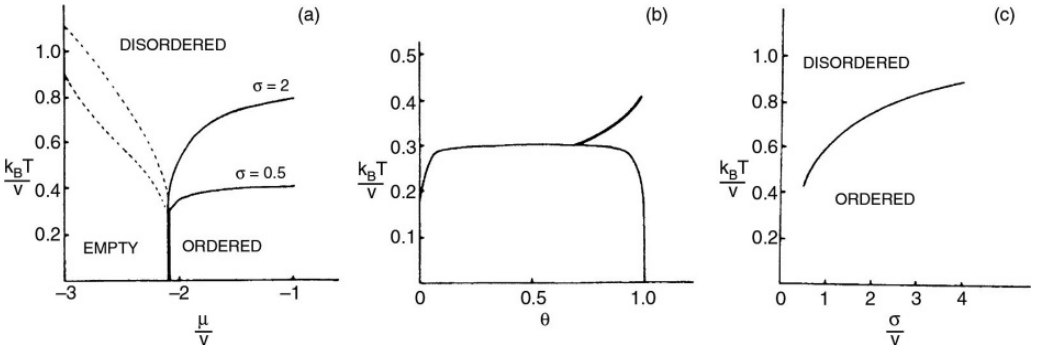

Fig. 4.30 Phase diagram of the two-dimensional system of living polymers for $v = 2.0$, $w = 0.1$: (a) T_c vs. chemical potential μ for two values of the rigidity parameter σ. The single line indicates a second order phase transition, the double line denotes a first order transition, and dots mark the Lifshitz line. (b) T_c as a function of coverage θ for $\sigma = 0.5$. (c) Variation of T_c with σ for $\mu = -1.4$. From Milchev and Landau (1995).

equilibration can be a problem for large systems so studies have been restricted to modest size lattices. An orientational order parameter must be computed: in two dimensions $\psi = \langle c_1 - c_2 \rangle$ (c_i is the concentration of segments in the ith state) where c_1 and c_2 are the fractions of stiff *trans* segments pointing horizontally and vertically on the square lattice. In $d = 3$ there are many more states than are shown in the figure, which is only for $d = 2$, and we do not list these explicitly here. In $d = 3$ then, the order parameter is defined as $\psi = \sqrt{(c_1 - c_2)^2 + (c_1 - c_8)^2 + (c_2 - c_8)^2}$, and c_1, c_2, c_8 are the fractions of *trans* bonds pointing in the x, y and z directions.

For two dimensions at $T = 0$; the lattice is completely empty below $\mu_c = -(v + w)$. Finite temperature phase transitions were found from the simulation data and, as an example, the resultant phase diagram for $v = 2.0$, $w = 0.1$ is shown in Fig. 4.30 for two different values of σ. In both cases the transition is first order at low temperatures, but above a tricritical point $T_t = 0.3$, it becomes second order. While for $\mu > \mu_c$ the density is quite high in both the ordered phase as well as the high temperature disordered phase, for $\mu < \mu_c$ the lattice is virtually empty below a temperature (the Lifshitz line) at which a rather steep (but finite) increase in θ is accompanied by pronounced maxima in the second derivatives of the thermodynamic potentials. A finite size scaling analysis along the second order portion of the boundary indicates critical behavior consistent with that of the two-dimensional Ising model. Figure 4.30 shows the phase diagram in $\theta - T$ space; the first order portion of the phase boundary has opened up into a large coexistence region leaving only a relatively small area of the pure ordered phase. Figure 4.30c shows that as the chains become stiffer, T_c rises monotonically.

On a simple cubic lattice the groundstate is triply degenerate with parallel rods pointing along any of the three Cartesian axes. Moreover, a sort of a smectic ordered state with planes of differently oriented parallel rigid chains will be formed at low temperature if the *inter* chain interaction, w, between nearest neighbor monomers does not differentiate between pairs of rods which

are parallel (in plane) or which cross at right angles when they belong to neighboring planes. Viewing these bonds as rough substitutes for the integral effect of longer range interactions, one could assume that the ws in both cases would differ so that in the former case (parallel rods) $\boldsymbol{w}_{\parallel}$ is somewhat stronger than the latter one, $\boldsymbol{w}_{\perp}$. Such an assumption leads to a groundstate consisting only of stiff chains, parallel to one of the three axes, whereby the order parameter in three dimensions attains a value of unity in the ordered state. A finite size scaling analysis of data for both $\boldsymbol{w}_{\perp} \neq \boldsymbol{w}_{\parallel}$ and $\boldsymbol{w}_{\perp} = \boldsymbol{w}_{\parallel}$ showed that the transition was first order.

4.7.8 The pruned enriched Rosenbluth method (PERM): a biased sampling approach to simulate very long isolated chains

At the beginning of this chapter we have seen that the importance sampling algorithm estimates averages in the canonical ensemble

$$\langle A \rangle = \frac{1}{Z} \sum_{\alpha} A(\alpha) \exp[(-\beta E(\alpha)],$$
$$Z = \sum_{\alpha} Q(\alpha) = \sum_{\alpha} \exp[-\beta E(\alpha)], \tag{4.97}$$

by choosing a subset of M microstates $\{\alpha\}$ of the system such that the probability $p(\alpha)$ of choosing a state α is proportional to $1/Q(\alpha)$, and hence $\langle A \rangle$ reduces to a simple arithmetic average $\overline{A}$ over the sample of M states,

$$\overline{A} = (1/M) \sum_{\alpha=1}^{M} A(\alpha). \tag{4.98}$$

Obviously, in this way direct information about both the partition function Z and the free energy $F = -(1/\beta)\ln Z$ is lost. (We shall come back to this problem in Section 5.8.) Another problem with importance sampling that we discussed earlier, is that the microstates that are generated are highly correlated with each other (unlike the simple sampling methods described in Chapter 3). With biased sampling methods one introduces a probability $p(\alpha)$ with which states are selected, so that $(W(\alpha) = Q(\alpha)/p(\alpha))$

$$Z = \frac{1}{M} \sum_{\alpha=1}^{M} W(\alpha), \qquad \overline{A} = \frac{1}{M} \sum_{\alpha=1}^{M} A(\alpha) W(\alpha) / \overline{Z}. \tag{4.99}$$

Obviously, if we chose $p(\alpha) = 1/M$, we would have simple sampling; for $p(\alpha) = Q(\alpha)$ we are back to importance sampling. Clever intermediate choices will hopefully yield a compromise and independence of configurations. This is the strategy of PERM (Grassberger, 1997; Hsu *et al.*, 2003; Hsu and Grassberger, 2011).

We explain the approach for the case of a self-avoiding walk (SAW) of N steps on a square (dimensionality $d = 2$) or simple cubic ($d = 3$) lattice. Every

lattice site can be visited only once, the bond length is the lattice spacing, and an energy $\varepsilon < 0$ is won if two non-bonded monomers occupy neighboring lattice sites. The partition function is then $Z = \sum q^m$ with the sum extending over all SAWs of N steps, $q \equiv \exp(-\beta\varepsilon)$, where m denotes the total number of non-bonded nearest neighbor pairs.

In the methods due to Rosenbluth and Rosenbluth (1955), SAWs are constructed step by step and their weight W_N calculated recursively: the first monomer is placed on an arbitrary lattice site and its weight is defined as $W_0 = 1$; for the first step one has $2d$ possibilities, so $W_1 = 2d$. For subsequent steps one scans the neighborhood of the chain end to identify the number n_{free} of free sites where in the step a monomer can be added to the previous chain end (the walk is abandoned if $n_{\text{free}} = 0$). After this step the weight is updated according to $W_N = w_N W_{N-1}$ with $w_n = q^{m_n} \times n_{\text{free}}$, m_n being the number of neighbors of the new site that are already occupied by non-bonded sites. This procedure yields

$$W_N = \prod_{n=0}^{N} w_n. \tag{4.100}$$

However, for long polymers this method fails because: (i) at some step we encounter the 'attrition problem', and (ii) the full weight W_N will show enormously large fluctuations.

PERM overcomes these limitations (to a large extent) using the idea of 'population control', i.e. by pruning some low-weight configurations and cloning (enriching) all those configurations with high weight, as the chain grows. Two thresholds W_n^+ and W_n^- define what is meant by 'low' or 'high' weights. If at a step n the weight W_n according to Eqn. (4.100) is larger than W_n^*, k copies of the current configuration are made, each copy getting a weight $W_n = w_n W_{n-1}/(k+1)$. If W_n is less than W_n^-, a random number $r \in [0, 1]$ is drawn: if $r < 1/2$ the configuration is 'killed', otherwise it is kept and its weight doubled. Pruning and cloning then leaves all averages unchanged and improves importance sampling very much. The price that must be paid is that the configurations become more and more correlated the larger N: thus the method cannot be continued indefinitely. Of course, the choice of the thresholds W_n^+, W_n^- is crucial: one finds, in practice, that often $W_n^+ = C_+ Z_n$ and $W_n^- = C_- Z_n$, with C_+, C_- being constants of order unity and $C_+/C_- = 10$ works well.

The copies made in the enrichments are placed on a stack, and a depth-first implementation is used. At each time one handles a single configuration only until the chain has grown to the desired maximum length N (if it was killed due to attrition or if the stack is empty, a new trial is started). Otherwise, one returns to the configuration at the top of the stack and the simulation continues. Since only a single configuration has to be remembered during the run, much less memory is required rather than when one used an explicit 'population' of many configurations of chains growing in parallel ('breadth-first' implementation).

Of course, configurations obtained from different clones of the same ancestor are correlated. One calls the set of all such configurations a 'tour', and one

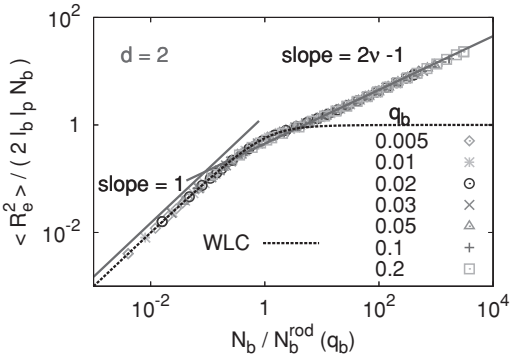

Fig. 4.31 Log–log plot of the rescaled mean-square end-to-end distance of semiflexible self-avoiding walks on the square lattice versus the rescaled chain length, for a broad range of values $q_b = \exp(-\beta\varepsilon_b)$, as indicated. The theoretical prediction of the wormlike chain (WLC) model is shown as a full curve. The power laws for rigid rods (slope = 1) and SAWs in $d = 2$ (slope $= 2\nu - 1 = 1/2$) are also shown. After Hsu and Binder (2012).

needs to consider the distribution $P(\ln W_t)$, where W_t is the tour weight. Only when the weighted distribution $W_t P(\ln W_t)$ has its maximum in a region where $P(\ln W_t)$ is still well sampled can the results safely be trusted.

In reality, this is not the whole story: often one also needs to use a bias in the growth process. For example, if one simulates a chain where one end is fixed at the origin and the other exposed to some force $\vec{F}$, it is advantageous to use a bias for the next step of the walk according to the direction of $\vec{F}$. A less trivial bias is based on the idea that a walk 'tries' to avoid the region of space where it just came from, so one may use the knowledge of the last k steps to bias the walk accordingly. For details on this so-called 'k-step Markovian anticipation' we refer the reader to the literature (Hsu and Grassberger, 2011).

We also note that this algorithm can be used for many related models such as branched objects like star polymers, lattice animals, bottle-brush polymers with rigid backbones, etc. (Hsu and Grassberger, 2011). Here we show only one example to illustrate the power of this algorithm, semiflexible polymers: rather than the energy ε introduced above (which describes the variation of the quality of the solvent the polymer is in) we use an energy cost ε_b for bond bending, whenever the walk makes a kink by $\pm 90^\circ$. So the partition function becomes ($q_b = \exp(-\beta\varepsilon_B)$)

$$Z_N(q_B) = \sum_{configurations} C(N, N_{bend}) q_b^{N_{bend}}, \tag{4.101}$$

where N_{bend} such kinks occur in a configuration. Here it is advantageous to use a bias which gives less weight to steps making a turn as ε_b increases. As an example, Fig. 4.31 shows typical results obtained for this model on the square lattice (Hsu and Binder, 2012), choosing results for chains of length $N = 25\,600$ and a wide range of parameters q_b. The mean square end-to-end distance $\langle R^2 \rangle$ of the chains is scaled by the product of the contour length $N\ell_b$ (where ℓ_b is the bond length) and the Kuhn length $\ell_K = 2\ell_p$, with the persistence length

ℓ_p being extracted from the initial decay of bond-autocorrelation functions with the 'chemical distance' along the chain. The chain length is scaled by the Kuhn length. The reason for this scaling is that the standard model for semiflexible polymers, the 'wormlike chain (WLC)' (Kratky and Porod, 1949) would predict that all data should superimpose on a master curve that describes the crossover from rod-like polymer to simple random walks. The data do fall on a master curve, but a different one, describing the crossover from rods to $d = 2$ SAWs, which follows the scaling relation $\langle R^2 \rangle \propto N^{2\nu}$ with $\nu = 3/4$. These results show that the WLC model, often used to discuss the configurations of semiflexible biopolymers adsorbed on substrates, is actually completely invalid (apart from the trivial regime of very short stiff chains which are stretched out as linear rods, for $N/\ell_b \ll \ell_K$ in $d = 2$ dimensions).

As a caveat, we mention that cases are known for which the PERM algorithm needs particular care, because the distribution of configurations that one needs to sample has several rather well separated peaks. This happens e.g. for the case of a polymer that is partially confined in a tube, but for smaller length L and diameter D, fixed with one end to the bottom of the tube. If L is large enough, the SAW will form linear string of 'blobs' (and the blobs forming the stem inside the tube are stretched due to an entropic force and are, therefore, elongated). Only when the biasing of the grown walks is carefully adjusted to such a situation can one sample such a bimodal distribution correctly (Hsu *et al.*, 2008). Thus, when one uses PERM it pays off to carefully consider the physics of the problem that one is dealing with, rather than using the method like a 'blackbox'.

4.8 SOME ADVICE

We end this chapter by summarizing a few procedures which in our experience can be useful for reducing errors and making simulations studies more effective. These thoughts are quite general and widely applicable. While these 'rules' provide no 'money-back' guarantee that the results will be correct, they do provide a prudent guideline of steps to follow.

(1) In the very beginning, *think*.
What problem do you really want to solve and what method and strategy is best suited to the study. You may not always choose the best approach to begin with, but a little thought may reduce the number of false starts.

(2) In the beginning think small.
Work with small lattices and short runs. This is useful for obtaining rapid turnaround of results and for checking the correctness of a program. This also allows us to search rather rapidly through a wide range of parameter space to determine ranges with physically interesting behavior.

(3) Test the random number generator.
Find some limiting cases where accurate, or exact values of certain properties can be calculated, and compare your results of your

algorithm with different random number sequences and/or different random number generators.

(4) Look at systematic variations with system size and run length.
Use a wide range of sizes and run lengths and then use scaling forms to analyze data.

(5) Calculate error bars.
Search for and estimate both statistical and systematic errors. This enables both you and other researchers to evaluate the correctness of the conclusions which are drawn from the data.

(6) Make a few very long runs.
Do this to ensure that there is not some hidden time scale which is much longer than anticipated.

REFERENCES

Ala-Nissila, T., Ferrando, R., and Ying, S. C. (2002), *Adv. Phys.* **51**, 949.

Attig, N., Binder, K., Grubmüller, H., and Kremer, K. (2004), *Computational Soft Matter : From Synthetic Polymers to Proteins* (NIC Directors, Jülich).

Baxter, R. J. (1972), *Ann. Phys. (N.Y.)* **70**, 193.

Baxter, R. J. (1982), *Exactly Solved Models in Statistical Mechanics* (Academic Press, London).

Baxter, R. J. and Wu, F. Y. (1973), *Phys. Rev. Lett.* **31**, 1294.

Binder, K. (1981), *Z. Phys. B* **43**, 119.

Binder, K. (1983), in *Phase Transitions and Critical Phenomena*, vol. 8, eds. C. Domb and J. L. Lebowitz (Academic Press, London), p. 1.

Binder, K. (1987), *Rep. Prog. Phys.* **50**, 783.

Binder, K. (1992), in *Computational Methods in Field Theory*, eds. C. B. Lang and H. Gausterer (Springer, Berlin).

Binder, K. (ed.) (1995), *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford University Press, New York).

Binder, K. and Hohenberg, P. C. (1974), *Phys. Rev. B* **9**, 2194.

Binder, K. and Landau, D. P. (1984), *Phys. Rev. B* **30**, 1477.

Binder, K. and Müller-Krumbhaar, H. (1973), *Phys. Rev. B* **7**, 3297.

Binder, K. and Stauffer, D. (1972), *J. Stat. Phys.* **6**, 49.

Binder, K. and Young, A. P. (1986), *Rev. Mod. Phys.* **58**, 801.

Borgs, C. and Kotecký, R. (1990), *J. Stat. Phys.* **61**, 79.

Caillol, J. M. (1993), *J. Chem. Phys.* **99**, 8953.

Carmesin, I. and Kremer, K. (1988), *Macromolecules* **21**, 2878.

Challa, M. S. S. and Hetherington, J. H. (1988), in *Computer Simulation Studies in Condensed Matter Physics I*, eds. D. P. Landau, K. K. Mon, and H.-B. Schüttler (Springer, Heidelberg).

Challa, M. S. S. and Landau, D. P. (1986), *Phys. Rev. B* **33**, 437.

Challa, M. S. S., Landau, D. P., and Binder, K. (1986), *Phys. Rev. B* **34**, 1841.

Creutz, M. (1983), *Phys. Rev. Lett.* **50**, 1411.

Crisanti, A. and Ritort, F. (2003), *J. Phys. A* **36**, R181.

Da Silva, R., Fernandes, H. A., de Felicio, J. R. D., and Figueiredo, W. (2013), *Comput. Phys. Commun.* **184**, 2371.

de Gennes, P. G. (1979), in *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca), Chapter 1.

Ferrenberg, A. M. and Landau, D. P. (1991), *Phys. Rev. B* **44**, 5081.
Ferrenberg, A. M., Landau, D. P., and Binder, K. (1991), *J. Stat. Phys.* **63**, 867.
Fichthorn, K. A. and Weinberg, W. H. (1991), *J. Chem. Phys.* **95**, 1090.
Fisher, M. E. (1971), in *Critical Phenomena*, ed. M. S. Green (Academic Press, London), p. 1.
Flory, P. J. (1953), *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York).
Glauber, R. J. (1963), *J. Math. Phys.* **4**, 294.
Gompper, G. and Goos, J. (1995), in *Annual Reviews of Computational Physics II*, ed. D. Stauffer (World Scientific, Singapore), p. 101.
Graim, T. and Landau, D. P. (1981), *Phys. Rev. B* **24**, 5156.
Grassberger, P. (1997), *Phys. Rev. E* **56**, 3682.
Gunton, J. D., San Miguel, M., and Sahni, P. S. (1983), in *Phase Transitions and Critical Phenomena*, eds. C. Domb and J. L. Lebowitz (Academic Press, London), vol. 8, p. 267.
Hohenberg, P. C. and Halperin, B. I. (1977), *Rev. Mod. Phys.* **49**, 435.
Houdayer, J. (2002), *J. Chem. Phys.* **116**, 1783.
Houdayer, J. and Müller, M. (2002), *Europhys. Lett.* **58**, 660.
Hsu, H.-P. and Binder, K. (2012), *J. Chem. Phys.* **136**, 024901.
Hsu, H.-P. and Grassberger, P. (2011), *J. Stat. Phys.* **144**, 597.
Hsu, H.-P., Binder, K., Klushin, L. I., and Skvortsov, A. M. (2008), *Phys. Rev. E* **78**, 041803.
Hsu, H.-P., Mehra, V., Nadler, W., and Grassberger, P. (2003), *J. Chem. Phys.* **118**, 444.
Hüller, A. (1993), *Z. Phys. B* **90**, 207.
Ito, N. (1993), *Physica A* **196**, 591.
Janssen, H. K., Schaub, B., and Schmittmann, B. (1989), *Z. Phys. B* **73**, 539.
Jasnow, D. (1984), *Rep. Prog. Phys.* **47**, 1059.
Jin, S., Sen, A., and Sandvik, A. (2012), *Phys. Rev. Lett.* **108**, 045702.
Kang, H. C. and Weinberg, W. H. (1989), *J. Chem. Phys.* **90**, 2824.
Katzgraber, H. G., Lee, L. W., and Young, A. P. (2004), *Phys. Rev. B* **70**, 014417.
Katzgraber, H. G., Palassini, M., and Young, A. P. (2001), *Phys. Rev. B* **64**, 184422.
Kawamura, H. and Kikuchi, M. (1993), *Phys. Rev. B* **47**, 1134.
Kawasaki, K. (1972), in *Phase Transitions and Critical Phenomena*, vol. 2, eds. C. Domb and M. S. Green (Academic Press, London).
Kehr, K. W. and Binder, K. (1984), in *Applications of the Monte Carlo Method in Statistical Physics*, ed. K. Binder (Springer, Heidelberg).
Kehr, K. W., Reulein, S., and Binder, K. (1989), *Phys. Rev. B* **39**, 4891.
Kikuchi, M. and Ito, N. (1993), *J. Phys. Soc. Japan* **62**, 3052.
Koch, W. and Dohm, V. (1998), *Phys. Rev. E* **58**, 1179.
Koch, W., Dohm, V., and Stauffer, D. (1996), *Phys. Rev. Lett.* **77**, 1789.
Kratky, O. and Porod, G. (1949), *J. Colloid Sci.* **4**, 35.
Kremer, K. and Binder, K. (1988), *Computer Phys. Rep.* **7**, 261.
Landau, D. P. (1976), *Phys. Rev. B* **13**, 2997.
Landau, D. P. (1996), in *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*, eds. K. Binder and G. Ciccotti (Societa Italiana di Fisica, Bologna).
Landau, D. P. and Binder, K. (1985), *Phys. Rev. B* **31**, 5946.
Landau, D. P. and Binder, K. (1990), *Phys. Rev. B* **41**, 4633.
Landau, D. P., Tang, S., and Wansleben, S. (1988), *J. de Physique* **49**, C8–1525.

Li, Z. B., Ritschel, U., and Zhang, B. (1994), *J. Phys. A: Math. Gen.* **27**, L837.
Li, Z., Schülke, L., and Zheng, B. (1996), *Phys. Rev. E* **53**, 2940.
Liu, A. J. and Fisher, M. E. (1990), *J. Stat. Phys.* **58**, 431.
Madras, N. and Sokal, A. (1988), *J. Stat. Phys.* **50**, 109.
Marinari, E., Parisi, G., Ricci-Tersenghi, F., Ruiz-Lorenzo, J. J., and Zuliani, F. (2000), *J. Stat. Phys.* **98**, 973.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), *J. Chem Phys.* **21**, 1087.
Milchev, A. (1993), *Polymer* **34**, 362.
Milchev, A. and Landau, D. P. (1995), *Phys. Rev. E* **52**, 6431.
Milchev, A., Binder, K., and Heermann, D. W. (1986), *Z. Phys. B* **63**, 527.
Müller, S., Wolveston, C., Wang, L.-W., and Zunger, A. (2000), *Acta Mater.* **48**, 4007.
Müller, S., Wang, L.-W., Zunger, A., and Wolveston, C. (2001), *Europhys. Lett.* **55**, 33.
Müller, S., Wang, L.-W., and Zunger, A. (2002), *Model. Sim. Mater. Sci. Eng.* **10**, 131.
Müller-Krumbhaar, H. and Binder, K. (1973), *J. Stat. Phys.* **8**, 1.
Nightingale, M. P. and Blöte, H. W. J. (1998), *Phys. Rev. Lett.* **80**, 1007.
Novotny, M. A. and Landau, D. P. (1981), *Phys. Rev. B* **24**, 1468.
Onsager, L. (1944), *Phys. Rev.* **65**, 117.
Ozeki, Y. and Ito, N. (2007), *J. Phys. A: Math Theor.* **40**, R149.
Parry, A. D. and Evans, R. (1992), *Physica A* **181**, 250.
Paul, W. and Müller, M. (2001), *J. Chem. Phys.* **115**, 630.
Potts, R. B. (1952), *Proc. Cambridge Philos. Soc.* **48**, 106.
Privman, V. (1990) (ed.), *Finite Size Scaling and Numerical Simulation of Statistical Systems* (World Scientific, Singapore).
Privman, V., Hohenberg, C., and Aharony, A. (1991), in *Phase Transitions and Critical Phenomena*, vol. 14, eds. C. Domb and J. L. Lebowitz (Academic Press, London).
Rampf, F., Binder, K., and Paul, W. (2006), *J. Polymer Sci. B: Polym. Phys.* **44**, 2542.
Reuter, K. and Scheffler, M. (2002), *Phys. Rev. B* **65**, 035406.
Reuter, K. and Scheffler, M. (2003), *Phys. Rev. Lett.* **90**, 046103.
Reuter, K., Frenkel, D., and Scheffler, M. (2004a), *Phys. Rev. Lett.* **93**, 116105.
Reuter, K., Stampfl, C., and Scheffler, M. (2004b), in *Handbook of Materials Modeling, Vol. 1, Fundamental Models and Methods*, ed. S. Yip (Springer, Dordrecht).
Rosenbluth, M. N. and Rosenbluth, A. W. (1955), *J. Chem. Phys.* **23**, 356.
Sadiq, A. and Binder, K. (1983), *Surface Sci.* **128**, 350.
Schmid, F. and Binder, K. (1992a), *Phys. Rev. B* **46**, 13553.
Schmid, F. and Binder, K. (1992b), *Phys. Rev. B* **46**, 13565.
Selke, W. (1992), in *Phase Transitions and Critical Phenomena*, Vol. 15, eds. C. Domb and J. L. Lebowitz (Academic Press, London), p. 1.
Sokal, A. D. (1995), in *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, ed. K. Binder (Oxford University Press, New York), Chapter 2.
Stauffer, D. (1997), *Physica A* **244**, 344.
Stoll, E., Binder, K., and Schneider, T. (1973), *Phys. Rev. B* **8**, 3266.
Wall, F. T. and Erpenbeck, J. J. (1959), *J. Chem. Phys.* **30**, 634.
Wang, J.-S. and Gan, C. K. (1998), *Phys. Rev. E* **57**, 6548.
Wansleben, S. and Landau, D. P. (1991), *Phys. Rev. B* **43**, 6006.
Weigel, M. and Janke, W. (2008), *Phys. Rev. E* **81**, 066701.

Werner, A., Schmid, F., Müller, M., and Binder, K. (1997), *J. Chem. Phys.* **107**, 8175.

Wheeler, J. C. and Pfeuty, P. (1981), *Phys. Rev. A* **24**, 1050.

Wilding, N. B. (1995), in *Computer Simulation Studies in Condensed Matter Physics VIII*, eds. D. P. Landau, K. K. Mon, and H.-B. Schüttler (Springer, Heidelberg).

Wilding, N. B. and Bruce, A. D. (1992), *J. Phys. Condens. Matter* **4**, 3087.

Wilding, N. B., Müller, M., and Binder, K. (1996), *J. Chem. Phys.* **105**, 802.

Yaldram, K. and Binder, K. (1991), *J. Stat. Phys.* **62**, 161.

Young, A. P. and Katzgraber, H. G. (2004), *Phys. Rev. Lett.* **93**, 207203.

Young, A. P. and Kawashima, N. (1996), *Int. J. Mod. Phys. C* **7**, 327.

Zheng, B., Ren, F., and Ren, H. (2003), *Phys. Rev. E* **68**, 046120.

Zifferer, G. (1999), *Macromol. Theory Simul.* **8**, 433.