A study of the multi-canonical Monte Carlo method

1995 J. Phys. A: Math. Gen. 28 6623

(http://iopscience.iop.org/0305-4470/28/23/015)

J. Phys. A: Math. Gen. 28 (1995) 6623–6643. Printed in the UK

A study of the multi-canonical Monte Carlo method

G R Smith and A D Bruce

Department of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3JZ, UK

Received 3 July 1995

Abstract. We present a study of the multi-canonical Monte Carlo method which constructs and exploits Monte Carlo procedures that sample across an extended space of macrostates. We examine the strategies by which the sampling distribution can be constructed, showing, in particular, that a good approximation to this distribution may be generated efficiently by exploiting measurements of the transition rate between macrostates, in simulations launched from sub-dominant macrostates. We explore the utility of the method in the measurement of absolute free energies, and how it compares with traditional methods based on path integration. We present new results revealing the behaviour of the magnetization distribution of a critical finite-sized magnet, for magnetization values extending from the scaling region all the way to saturation.

1. Introduction

The Monte Carlo (MC) method is widely appreciated as an invaluable aid in the exploration of statistical physics. In its most commonly practised form (Boltzmann importance sampling) it allows the microstates of the model system of interest to be visited with the canonical Boltzmann probabilities, and canonical expectation values to be determined as simple averages over the sampled states [1, 2].

The limitations of the method, in its traditional form, are also widely recognized. First, its very faithfulness to the Boltzmann distribution means that it is as susceptible to the problems of metastability as a laboratory experiment: the relative probabilities (the relative free energies) of macroscopically different phases cannot be determined directly because of the intrinsically low probability of paths connecting them. Second, though guaranteed to *sample* with the Boltzmann probabilities, the procedure does not *prescribe* these probabilities: the normalization constant (the partition function, and hence the absolute free energy) for the probabilities is neither required for, nor readily determined by, the standard procedure [2].

The idea of extended sampling, by which we mean sampling from distributions other than the canonical Boltzmann form, also has a long history [3]. The seminal contributions of Torrie and Valleau [4] have recently evolved into the multi-canonical ensemble method of Berg and Neuhaus [5, 6], and the related expanded ensemble method (or 'simulated tempering') of Lyubartsev *et al* [7] and Marinari and Parisi [8].

Our primary concern here is with the multi-canonical Monte Carlo (MCMC) method. In this method, weights are associated with a selected range of values of a nominated macroscopic variable, frequently, but not invariably, the energy; the MC algorithm is then designed so that macrostates (the 'states' of the chosen macroscopic variable) are visited with probabilities that are modified, with respect to the canonical forms, to an extent controlled

0305-4470/95/236623+21$19.50

by the chosen weights. The ideal weights secure a sampling distribution that is flat over an extended range of macrostates.

It is convenient (though not necessarily always wise) to divide the associated issues into two groups: how the sampling distribution should be *constructed* and the ways in which it can be *utilized.*

The construction process poses the core problem, which has formed the principal obstacle [2, 9] to speedier and wider implementation of the ideas of Torrie and Valleau [4]. In most MCMC studies (surveyed recently by Berg [10]) the sampling distribution has been constructed by largely *ad hoc* iterative procedures [11, 12] which use *estimates* of the 'current' sampling distribution (that associated with a particular set of weights), based on observations of the frequency with which macrostates are visited, to refine the choice of weights, which are then used to define the next sampling distribution. Typically the procedure is bootstrapped by the guess of an initial set of weights, based, for example, on an extrapolation of results obtained on a smaller system [5]. The studies reported here (section 2) explore and develop these procedures. We set the the choice of estimator of the current sampling distribution within a Bayesian framework, rationalizing earlier prescriptions [11, 12] . We show that the information provided by observation of the frequency with which *transitions* are made between macrostates provides a potentially more efficient route to an appropriate set of weights than measurements of the frequencies of macrostate *visits.* We also make a little progress (again within a Bayesian framework) towards a weight-update algorithm which reflects confidence levels in the existing weights, and is therefore more robust against sampling error.

The MCMC method has been utilized in a variety of ways, which we may divide broadly into three categories. First, the method can be used to facilitate the equilibration of two phases, by use of a sampling distribution that is extended ('multi-canonical') in an appropriate order parameter. In this way one may locate the line of phase coexistence, and (in effect), measure the difference between the free energies of the two phases. Applications of this kind have, to date, involved extended sampling in the energy (at a thermally driven first-order transition in a lattice model) [5], and the density (at the liquid–vapour transition) in a Lennard–Jones fluid [13]. Second, in conjunction with histogram-reweighting techniques [14, 15], extended sampling in the energy allows one to determine equilibrium properties—including the free energy—at *all* temperatures. Applications here have included studies of spin glasses [11, 16] and tertiary protein structure [17]. Thirdly, one may use extended sampling techniques to access the information inherent in the probabilities of intrinsically unlikely macrostates. The principal application here has been the measurement of interfacial tension (through extended sampling of the magnetization) using the probabilities of inhomogeneous (two-phase) macrostates [18].

The applications presented here contribute to the second and third strands of this programme. We examine the effectiveness of extended energy sampling in determining absolute free-energy values, making an explicit comparison with the MC integration methods traditionally used for such calculations [2, 9]. We also use extended sampling to explore the behaviour of the distribution of the magnetization in a critical magnet beyond the region accessed in conventional Boltzmann sampling [19, 20], right through to saturation. The behaviour in this regime has been the subject of some recent speculation [21, 22], on which the present results cast some light.

Our conclusions are summarized in section 4.

2. Constructing the sampling distribution

2.1. General formulation

We consider a system describable by N local coordinates, which we shall take to have a discrete spectrum. The specific system we shall consider here will be an Ising model on a lattice of N sites. However, the methods we shall discuss are widely applicable; we shall report elsewhere on their application to the study of a structural phase transition [23].

We envisage that the system has microstates, labelled r, with energies E_r, and associated canonical probabilities (at inverse temperature β)

$$p_r^c = \frac{e^{-\beta E_r}}{Z^c(\beta)} \tag{2.1}$$

where $Z^c(\beta)$is the canonical partition function

$$Z^c(\beta) = \sum_r e^{-\beta E_r} \tag{2.2}$$

and the sum extends over the complete, finite, set of microstates, $r = 1, 2 \ldots \Omega_T$, where, for the Ising model, $\Omega_T = 2^N$. Algorithms which sample with the canonical probabilities (2.1) will visit macrostates identified by the values $\mathcal{O}_i$ $(i = 1, 2, \ldots N_m)$ of a chosen macroscopic observable $\mathcal{O}$ with probabilities

$$p_i^c = \sum_{r \in i} p_r^c \equiv \frac{e^{-\beta \mathcal{F}_i}}{Z^c(\beta)} \tag{2.3}$$

where $\mathcal{F}$ is a generalized free-energy function. In particular, for energy macrostates $(\mathcal{O} = E)$,

$$\beta \mathcal{F}_i = \beta E_i - \ln \Omega(E_i) \tag{2.4a}$$

and

$$p_i^c = p^c(\beta, E) \equiv \frac{\Omega(E)\, e^{-\beta E_i}}{Z^c(\beta)} \tag{2.4b}$$

where Ω is the density-of-states function.

For succinctness we shall frequently use a vector notation for macrostate space. Thus, for example, the vector p^c will signify the set of canonical probabilities p_i^c, $i = 1, 2, \ldots N_m$, defined in equation (2.3).

We shall be concerned with algorithms which sample from the extended distribution

$$p_r(\eta) = \frac{e^{-\beta E_r + \eta_i}}{\mathcal{Z}(\beta, \eta)} \qquad r \in i \tag{2.5}$$

which accordingly visit macrostates $\mathcal{O}_i$ with probabilities

$$p_i(\eta) = \frac{e^{-\beta \mathcal{F}_i + \eta_i}}{\mathcal{Z}(\beta, \eta)} \tag{2.6}$$

where $\mathcal{Z}(\beta, \eta)$ is fixed by the normalization condition.

Our aim is to construct, and then to utilize, a set of weights $\eta \equiv \eta_1, \eta_2, \ldots$ approximating to an ideal set η^* with the property

$$\eta_i^* = \beta \mathcal{F}_i + \eta_0^* \tag{2.7}$$

where η_0^* is an arbitrary constant, which we shall fix by the convention that $\min_i(\eta) = 0$. The ideal set of weights (uniquely defined, given this convention) thus secures a sampling distribution that is flat over the space of macrostates. We shall refer to such a *flat* distribution

as *multi-canonical*, whether the macrostates are those of the energy or not. Most work to date has focused on energy macrostates, and the weights have accordingly been parametrized by effective temperatures; we shall keep the notation more general.

In principle, the task of constructing approximations to the ideal multi-canonical set is not independent of the issues involved in their utilization: the 'construction' process involves gathering much of the information needed in the 'utilization' process; and the accuracy with which the multi-canonical weights need to be determined can only be prescribed by consideration of their use. We shall, nevertheless, separate the tasks in this way.

The construction process clearly has to be iterative in character: it requires a procedure (possibly bootstrapped by a guess at some initial set of weights $\eta^{(1)}$) generating a sequence $\eta^{(2)}, \eta^{(3)} \ldots$ converging to the ideal set η^*. The general structure of such a procedure is clear. We gather information ('data', $D^{(n)}$ say) by MC sampling from the distribution defined by the nth set of weights with microstate probabilities $p_r^{(n)} \equiv p_r(\eta^{(n)})$, visiting macrostates with probabilities $p_i^{(n)} \equiv p_i(\eta^{(n)})$, which define the macrostate probability vector $p^{(n)}$. This information is then used to generate the $(n+1)$th set of weights. It is helpful to consider what is involved within a Bayesian framework [24]. In principle, the $(n+1)$th iterate $\eta^{(n+1)}$ should be identified as an *estimator* of η^* based on the probability distribution for this ideal set, $P(\eta^*|D^{(1)}, D^{(2)}, \ldots D^{(n)})$, constructed using *all* the data $D^{(1)}, D^{(2)}, \ldots D^{(n)}$ gathered up to the nth stage of the procedure. In practice the analytical difficulties in handling the evolution of such a distribution, with n, have led us to implement a rather less ambitious procedure, split into two parts. First, the data gathered at the nth stage (alone) is used to infer a probability distribution for the set $p^{(n)}$ (the true underlying sampling probabilities associated with the nth set of weights). According to Bayes' theorem (with no attempt to fold in the prior information implicit in data from earlier stages of the procedure) this distribution is given, to within a normalization constant, by the *likelihood* function $P(D^{(n)}|p^{(n)})$:

$$P(p^{(n)}|D^{(n)}) \propto P(D^{(n)}|p^{(n)}) . \tag{2.8}$$

This distribution is then used to construct an estimator, $\tilde{p}^{(n)}$, for the sampling probability $p^{(n)}$, and (generally) associated uncertainties. These estimates constitute what we will call *processed data*, $\mathcal{D}^{(n)}$, which define the input to the second part of the procedure. In this second part, the processed data is used to update the distribution for η^* in accord with Bayes' theorem:

$$P(\eta^*|\mathcal{D}^{(1)} \ldots \mathcal{D}^{(n)}) \propto P(\eta^*|\mathcal{D}^{(1)} \ldots \mathcal{D}^{(n-1)}) P(\mathcal{D}^{(n)}|\eta^*) . \tag{2.9}$$

We proceed to explore this structure, giving particular attention to three key issues:

(i) The choice of the *data* $D^{(n)}$ to be gathered in the sampling stage.

(ii) The choice of *estimator* of the sampling probabilities $p^{(n)}$ (given the distribution (2.8)) yielding the processed data $\mathcal{D}^{(n)}$.

(iii) The choice of *algorithm* for weight updating (given the processed data) providing a practical implementation of (2.9).

2.2. *Initial explorations: the choice of probability estimator*

To begin with let us suppose that the *data* $D^{(n)}$ gathered in the course of MC sampling from the distribution $p^{(n)}$ consists simply of a histogram $C^{(n)}$ of the counts of visits to each macrostate, totalling N_c counts in all, say. We shall refer to this strategy as the *visited states* (VS) method. It is, to our knowledge, the basis of all multi-canonical studies to date. If the interval between recorded samplings is sufficiently long in comparison with the

correlation time of the underlying Markov process, the likelihood function of the sampled data $P(D^{(n)}|p^{(n)})$ will be multinomial, so that, appealing to equation (2.8),

$$P(p^{(n)}|D^{(n)}) = \mathcal{N}^{-1} \prod_{i=1}^{N_m} (p_i^{(n)})^{C_i^{(n)}} \qquad \text{with} \qquad \sum_i C_i^{(n)} = N_c \tag{2.10}$$

where $\mathcal{N}$ is determined by the requirement that $P(p^{(n)}|D^{(n)})$ is correctly normalized.

We shall also make the simplest choice as regards the *algorithm* for weight updating. We suppose that the processed data $\mathcal{D}^{(n)}$ consists simply of the *estimate* $\tilde{p}^{(n)}$ we make of $p^{(n)}$ based on the distribution (2.10) (which we shall consider in a moment), together, formally, with the nth weight vector $\eta^{(n)}$. That is,

$$\mathcal{D}^{(n)} \longrightarrow \tilde{p}^{(n)}, \eta^{(n)} . \tag{2.11}$$

We choose, moreover, to make no attempt to fold in prior information gleaned from earlier stages in the process. Then equation (2.9) reduces to the form

$$P(\eta^*|\mathcal{D}^{(n)}) \equiv P(\eta^*|\tilde{p}^{(n)}, \eta^{(n)}) = \delta(\eta^* - \eta^{(n+1)}) \tag{2.12}$$

with

$$\eta_i^{(n+1)} = \eta_i^{(n)} - \ln \tilde{p}_i^{(n)} + k \tag{2.13}$$

where k is a constant (to be fixed by appeal to the convention defined following equation (2.7)).

Now consider the choice of *estimator* $\tilde{p}^{(n)}$. The simplest choice is the *maximum likelihood estimator* (MLE) which locates the maximum in the likelihood function (2.10). The components of this estimator are readily identified as

$$\tilde{p}_i^{(n)} = \frac{C_i^{(n)}}{N_c} \qquad \text{[VS; MLE]} \tag{2.14a}$$

which, in conjunction with (2.13), implies the update scheme

$$\eta_i^{(n+1)} = \eta_i^{(n)} - \ln C_i^{(n)} + k \tag{2.14b}$$

with the understanding that $\eta_i^{(1)} = 0$

The deficiencies of this scheme are clear: it fails where $C_i^{(n)} = 0$, which happens extremely frequently in the early iterations since there are many macrostates that the Boltzmann sampling algorithm does not visit. While there are heuristic ways of avoiding this problem [12, 25–27], it can be skirted more systematically with a revised choice of estimator. Specifically, the *mean value estimator* (MVE), identified as the mean of the likelihood (2.10), has elements

$$\tilde{p}_i^{(n)} = \frac{C_i^{(n)} + 1}{N_c + N_m} \qquad \text{[VS; MVE]} \tag{2.15a}$$

which implies the revised update scheme

$$\eta_i^{(n+1)} = \eta_i^{(n)} - \ln[C_i^{(n)} + 1] + k . \tag{2.15b}$$

In the (troublesome) case $C_i^{(n)} = 0$ this gives the same updating scheme as was introduced, somewhat arbitrarily, by Lee [12]. Its essential effect is to *decrease* the sampling probability (at level $n + 1$) in the region well sampled at level n and thence to *increase* it (uniformly) in the region not sampled at level n.

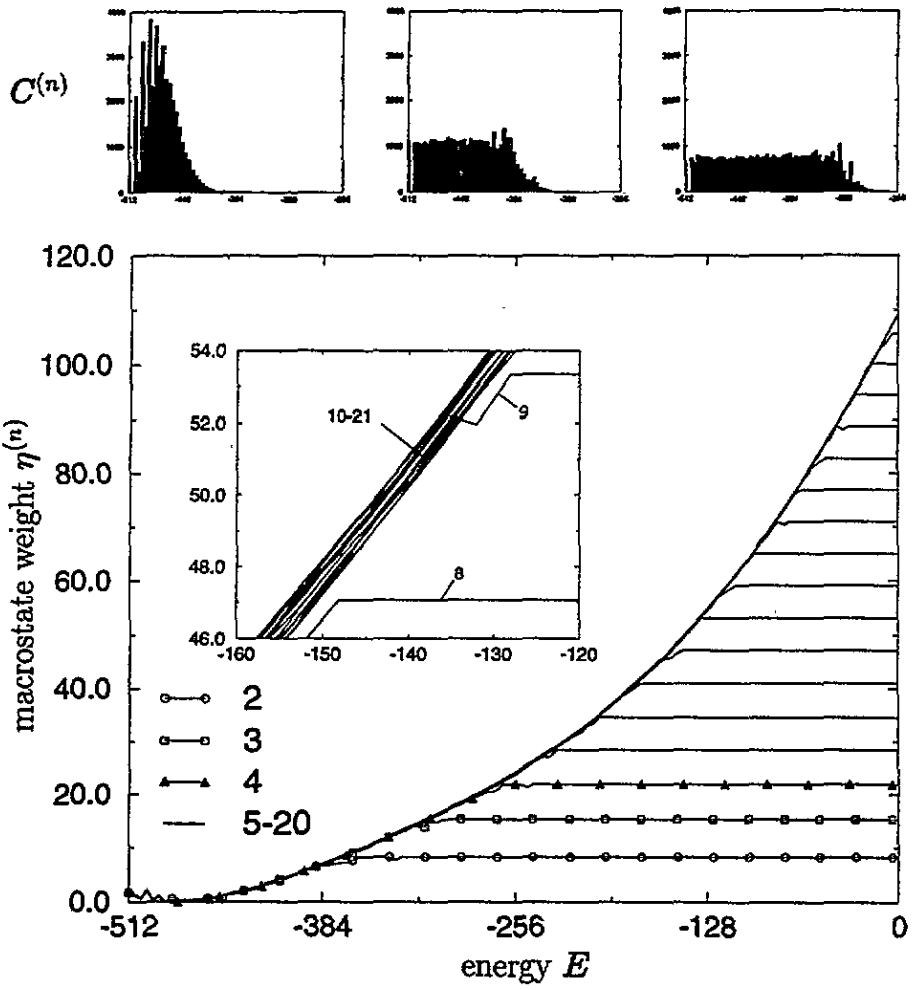

Figure 1. Behaviour of the visited states algorithm with mean value estimator and no prior on weights (equation (2.15*b*)) for a 2D Ising model with $\beta = 0.55$, and $N = 16^2$. The upper figure shows the histograms $C^{(n)}$ of visits to energy macrostates (for $n_c = 300$) at stages $n = 1, 2, 3$; the lower figure shows the resulting set of weights $\eta^{(n)}$ for $n = 2, 3, 4 \ldots 20$. The inset shows detail in a restricted range of energies revealing persistent fluctuations due to the effects of sampling error on the algorithm (2.15*b*).

Figure 1 shows this simple updating scheme in action. Here (and elsewhere in this paper) we have used the 2D Ising model as a testbed. We choose a model of unit coupling constant with the energy

$$E = -\sum_{(ij)} \sigma_i \sigma_j \tag{2.16}$$

where the sum extends over the $2N$ nearest-neighbour bonds in a 2D lattice, and the spins assume values ± 1. In this case we have considered the energy macrostates of an $N = 16^2$ model at inverse temperature $\beta = 0.55$, extending from the ground state $E = -2N$ up to $E = 0$. The total number of recorded macrostate visits N_c is chosen such that $n_c \equiv N_c/N_m = 300$, where $N_m = N/2$ is the number of energy macrostates in the range investigated.

The upper part of figure 1 shows the histogram $C^{(n)}$ of macrostate visits for sampling stages $n = 1, 2, 3$; the lower figure shows the weights $\eta^{(n)}$ for $n = 2, 3, 4$ inferred from this data (and from the following 17 iterations) using (2.15*b*). The general character of the scheme is evident: on each iteration the sampled region widens slightly, becoming roughly multi-canonical (flat) over the part that was sampled before, and extending a little further into the wings. The $(n+1)$th histogram tends to have large fluctuations in the states that were at the edge of the nth, because the poor statistics at the edges of $C^{(n)}$ tend to produce weights $\eta^{(n+1)}$ that are inaccurate here. These fluctuations get smoothed away on subsequent iterations. It is clear that the scheme consistently underestimates the weights to be assigned to regions which have not been sampled. This is to be expected: unless the distribution over macrostates is bimodal, the weight in the unsampled wings of the distribution $p_\cdot^{(n)}$ will typically be many orders of magnitude smaller than that implied by the one count's worth credited by (2.15*a*). Though less than ideal, it is certainly better to err on this side than the other. Thus, for example, we have found that attempts to get faster convergence by extrapolating the weight vector from the sampled into the unsampled region may overestimate the weights there, so that the region well sampled at stage n is missed altogether at stage $n+1$. Convergence is then irregular and awkward, with the latest weights sometimes needing to be discarded and a return made to earlier ones. (We note, however, that *linear* extrapolations are relatively safe in this regard and have been used successfully, in conjunction with various *ad hoc* constraints [10, 11, 25, 28].)

At first glance it would appear from figure 1 that convergence up to $E = -120$ has occurred by the 10th iteration, and convergence over the entire range by the 20th. Closer inspection, however (see the inset to figure 1), shows that this convergence is not complete, and that fluctuations in the weights persist. These fluctuations signal the need for an improved algorithm.

2.3. *Stabilizing the algorithm: a prior on the weights*

The fluctuations in the macrostate weights reflect the fact that the algorithm (2.15*b*) is acutely vulnerable to sampling error. As it stands, this algorithm enjoins us to update the weights by amounts controlled entirely by the histogram of visited states, with no account taken of the confidence levels associated with the existing weights. A recent study [10] offers an *ad hoc* refinement of this strategy which uses the histograms of all previous iterations, each contributing to an extent inversely dependent on the size of local fluctuations. Here we attempt to build on the rather more rigorous framework provided by (2.9). To exploit it we need firstly to extend the scope of the processed data utilized in the weight update procedure, so as to incorporate information about these confidence levels; and secondly to allow for a 'prior' on weights that propagates this information through the update procedure.

To define confidence levels associated with the weights we proceed as follows. The macrostate counts contributing to each histogram $C^{(n)}$ are subdivided to define sub-histograms $C^{(n,m)}$ $(m = 1, 2 \ldots M)$ each comprising N_b counts (so that $M \times N_b = N_c$). Jackknife estimators [29] of revised weights follow by application of (2.15*b*) in the form

$$\eta_i^{(n+1,m)} = \eta_i^{(n)} - \ln[\bar{C}_i^{(n,m)} + 1] + k \tag{2.17}$$

where $\bar{C}_i^{(n,m)}$ comprises the data pooled from all the M histograms except the mth. The mean and standard deviation of the set $\eta_i^{(n+1,m)}$, $(m = 1, 2, \ldots M)$ provide an estimator for the revised weight (of macrostate i), and its uncertainty *given the observations made at stage n alone*. We denote these quantities by $\hat{\eta}_i^{(n+1)}$ and $\hat{\sigma}_i^{(n+1)}$, respectively. Then the processed

data emerging from stage n can be packaged as

$$\mathcal{D}^{(n)} \longrightarrow \hat{\eta}^{(n+1)}, \hat{\sigma}^{(n+1)} \tag{2.18}$$

and the likelihood function featuring in (2.9) assumes the form

$$P(\mathcal{D}^{(n)})|\eta^*) \propto P(\eta^*|\mathcal{D}^{(n)}) = \mathcal{G}\left[\eta^* - \hat{\eta}^{(n+1)}, \hat{\sigma}^{(n+1)}\right] \tag{2.19}$$

where $\mathcal{G}$ is a multi-variate Gaussian distribution (but with co-variance terms suppressed).

This refinement of the likelihood function requires a correspondingly more careful treatment of the other functions appearing in (2.9). Specifically, we shall assume that the prior (on the right-hand side of this equation), which encodes convictions about η^* *before* the observations made at level n, can itself be parametrized in Gaussian form

$$P(\eta^*|\mathcal{D}^{(1)} \cdots \mathcal{D}^{(n-1)}) \propto \mathcal{G}\left[\eta^* - \eta^{(n)}, \sigma^{(n)}\right] \tag{2.20}$$

with a corresponding form (parametrized by $\eta^{(n+1)}$ and $\sigma^{(n+1)}$) for the posterior function (the left-hand side of equation (2.9)).

With these identifications, and appealing to (2.19) we find the update equations

$$\frac{\eta_i^{(n+1)}}{[\sigma_i^{(n+1)}]^2} = \frac{\eta_i^{(n)}}{[\sigma_i^{(n)}]^2} + \frac{\hat{\eta}_i^{(n+1)}}{[\hat{\sigma}_i^{(n+1)}]^2} \tag{2.21a}$$

and

$$\frac{1}{[\sigma_i^{(n+1)}]^2} = \frac{1}{[\sigma_i^{(n)}]^2} + \frac{1}{[\hat{\sigma}_i^{(n+1)}]^2}. \tag{2.21b}$$

The current confidence levels associated with the weights now moderate the extent to which they are altered to reflect the newly gathered data. Thus, if the level-n weights are known only very imprecisely, the algorithm assigns the level-$(n+1)$ weights on the basis of the information gathered at level n alone ($\eta_i^{(n+1)} \simeq \hat{\eta}_i^{(n+1)}$); on the other hand, to the extent that the level-n weights are precisely known, the change at update is small ($\eta_i^{(n+1)} \simeq \eta_i^{(n)}$).

To bootstrap the algorithm it is necessary to make the assignments $\eta_i^{(1)} = 0$ and $\sigma_i^{(1)} = \infty$, and operate the convention that $\sigma_i^{(n)} = \infty$ until n values are reached such that $C_i^{(n-1)} \neq 0$. We also found that it is necessary to locally average the variable $\hat{\sigma}_i$ because, otherwise, the sampling error in $\hat{\sigma}_i^{(n)}$ reintroduces noise in $\eta_i^{(n+1)}$. This problem arises because of the neglect of co-variance terms in (2.19).

Figure 2 shows the algorithm at work, on the same task as that used to test the simple update algorithm (figure 1). The choices $n_c = 50$, $M = 6$ for the refined algorithm ensure the same overall number of observations as that used in the test of the simple algorithm. The key difference between the two schemes is apparent from a comparison of the insets in the two figures: one can see that the persistent weight fluctuations characterizing the simple algorithm are damped out by the refined algorithm. This is a reflection of the additional stability conferred by folding in the confidence levels in existing weights, through the use of the prior. Notwithstanding this advance, comparison of the main portions of the two figures shows that the revised algorithm does nothing to improve the speed at which the sampling progresses into the regime of macrostates with low Boltzmann weight.

2.4. Improving convergence: the transition probability method

It is apparent that the major problem with the visited states method of evolving the multi-canonical distribution is that it is slow to sample from (and thus give information about) regions remote from the dominant, equilibrium, macrostate. It is possible to do better by

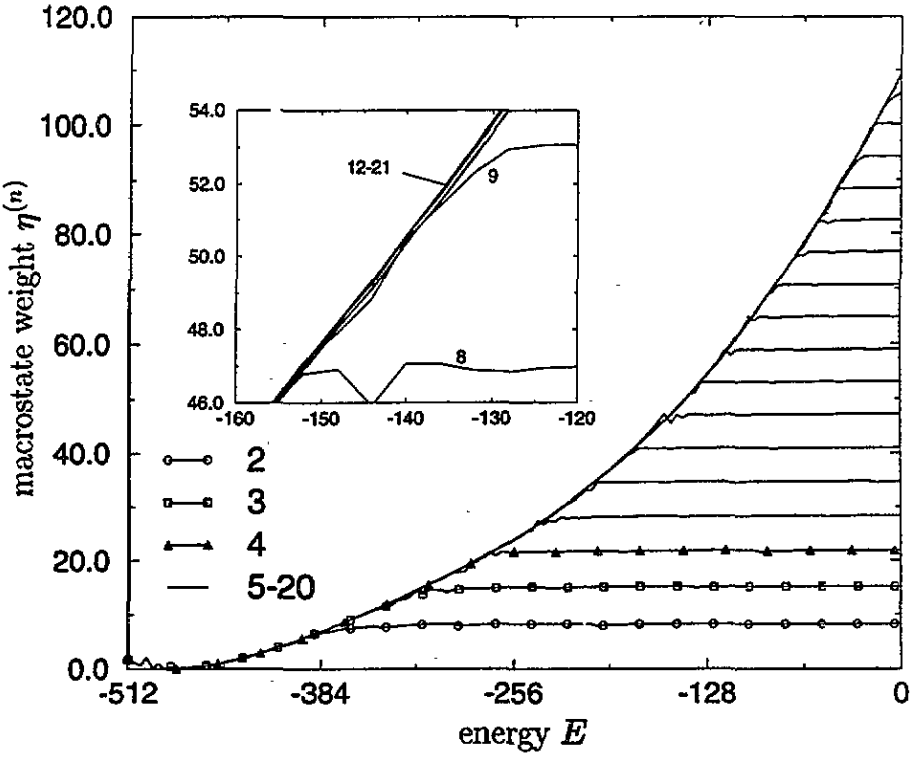

Figure 2. The results of iterations 1–20 for the visited states algorithm with mean value estimator, using a prior on weights (with $n_c = 50$ and $M = 6$). The inset shows detail in a restricted range of energies, revealing the improved convergence of the refined algorithm (2.21*a*), (2.21*b*). (The model parameters are as defined in figure 1.)

extrapolation (cf the discussion in section 2.2) to states as yet unvisited. Here, however, we explore an alternative scheme, which allows us to begin to gather information about all parts of the macrostate space immediately. In this scheme (to be referred to as the *transition probability* (TP) method) the system is prepared in a microstate of intrinsically low probability, such as the ground state. The system is allowed to evolve and the transitions between macrostates are monitored. Inferences about the macrostate probabilities are then made on the basis of the *record of transitions* rather than the *record of visited states*. To our knowledge this method is new (it is not to be confused with Bennett's acceptance ratio method [30]), although very recent work by Kerler *et al* [31] also recognises that TP information forms a fruitful basis for inferences about the sampling distribution.

To explore the basis of the method we must examine the factors controlling the macrostate transition probability matrix. Denote by $\rho_{ij}^{(n)}(t) \equiv P(i \to j|i, p^{(n)})(t)$ the transition probability from macrostate i to macrostate j at time t (measured from the beginning of the nth sampling stage), in a process sampling from an ensemble with macrostate probabilities $p^{(n)}$ (macrostate weights $\eta^{(n)}$). We may write

$$\rho_{ij}^{(n)}(t) = \sum_{r \in i} \sum_{s \in j} P(r|i, p^{(n)})(t) \rho_{rs}^{(n)} \tag{2.22}$$

where $\rho_{rs}^{(n)}$ is the transition matrix for the microstates, at stage n. We consider the implications of this equation in the regime in which the probability distribution of microstates

for a given macrostate can be approximated by its stationary limiting form, which is of the Boltzmann form *independent* of the weights:

$$P(r|i,p^{(n)})(t) \simeq P(r|i,p^c) = \frac{p_r^c}{p_i^c} = \mathrm{e}^{-\beta[E_r - \mathcal{F}_i]} \qquad r \in i\,. \tag{2.23}$$

We shall return to consider the validity of this approximation. Accepting it for the moment, we find that the macrostate transition matrix is itself then stationary, and satisfies

$$\begin{aligned}
\rho_{ij}^{(n)} &= \sum_{r\in i}\sum_{s\in j} P(r|i,p^{(n)})\rho_{rs}^{(n)} \\
&= \sum_{r\in i}\sum_{s\in j} P(r|i,p^{(n)})\rho_{sr}^{(n)}\frac{p_s^{(n)}}{p_r^{(n)}} \\
&= \sum_{r\in i}\sum_{s\in j} P(r|i,p^{(n)})\rho_{sr}^{(n)}\frac{P(s|j,p^{(n)})\sum_{s'\in j}p_{s'}^{(n)}}{P(r|i,p^{(n)})\sum_{r'\in i}p_{r'}^{(n)}} \\
&= \sum_{r\in i}\sum_{s\in j} P(s|j,p^{(n)})\rho_{sr}^{(n)}\frac{p_j^{(n)}}{p_i^{(n)}} \\
&= \frac{p_j^{(n)}}{p_i^{(n)}}\rho_{ji}^{(n)}
\end{aligned} \tag{2.24}$$

where the first step makes use of the detailed balance condition which the *micro*state transition probability matrix satisfies, by construction:

$$\rho_{rs}^{(n)} = \rho_{sr}^{(n)}\frac{p_s^{(n)}}{p_r^{(n)}}\,. \tag{2.25}$$

Equation (2.24) shows that the *macro*state transition probability matrix satisfies its own detailed balance condition, and, to the extent that the approximation (2.23) holds, the eigenvector (of unit eigenvalue) of its transpose gives the macrostate probability vector $p^{(n)}$ associated with the weights $\eta^{(n)}$. An estimate of the transition matrix (based upon a transition count) thus provides an alternative estimator $\tilde{p}^{(n)}$ of the macrostate probabilities $p^{(n)}$.

The utility of this estimator must reflect the reliability of the approximation made in (2.23). The assumption here is, effectively, that the simulation allows a 'local' equilibrium to establish itself within the currently sampled macrostate. To the extent that the macroscopic variables (the macrostate labels) are the slowest to evolve, one may expect this approximation to be reasonable, and to improve with the approach to the multi-canonical limit where the MC dynamics is more diffusive, less directional, and there is more time for relaxation to a local equilibrium. In contrast, the use of visited states estimators (like equation (2.15*a*)) rests on assumptions (multinomial distribution of counts, each bin count independent of the others) which become *less* reliable as the multi-canonical limit is approached.

Let us turn to examine how well the method works in practice, beginning with an outline of some of the implementation details. The system is initialized in a chosen microstate associated with a macrostate remote from equilibrium. At each subsequent MC step we record in an array $C_{ij}^{(n)}$ the transition performed between the macrostate i before the step and the macrostate j after it. Rejected trial moves and accepted moves that do not change the macrostate are recorded, alike, in the diagonal elements $C_{ii}^{(n)}$. Typically (we shall give examples below) the process is repeated, releasing the system from some other location in macrostate space. Then the entire procedure is repeated until the array of recorded

transitions is reasonably *full*. The array $C^{(n)}$ is deemed 'full' when $C^{(n)}_{i,i+1} + C^{(n)}_{i+1,i} > n_t$ for all i (so that n_t plays a role similar to that of n_c in the VS method).

An estimator for the elements of the transition matrix then follows from the recorded transitions, on the basis of a calculation similar to that underlying equation (2.15*a*) [32]:

$$\tilde{\rho}^{(n)}_{ij} = \frac{C^{(n)}_{ij} + 1}{\sum_j (C^{(n)}_{ij} + 1)} \qquad \text{[TP; MVE]}. \tag{2.26}$$

The corresponding estimator for the sampling probability, $\tilde{p}^{(n)}$, follows from the eigenvector of $\tilde{\rho}^{(n)}_{ij}$. Typically the transition matrix is tridiagonal and it is easy to find $\tilde{p}^{(n)}$ using (2.24): we neglect normalization initially and take $\mathrm{V}\tilde{p}^{(n)}_1 = 1$ (where $i = 1$ labels the release state). Then we use $\tilde{p}^{(n)}_{i+1} = \tilde{p}^{(n)}_i \tilde{\rho}^{(n)}_{i,i+1} / \tilde{\rho}^{(n)}_{i+1,i}$ to generate all the other elements successively, and finally impose $\sum_i \tilde{p}^{(n)}_i = 1$. To prevent the buildup of rounding errors we found it essential that the sequence of probabilities thus generated is *increasing*; this aim is realized by iterating from the release state(s), which are chosen because of their low Boltzmann weight. This estimator of $p^{(n)}$ is used to update the macrostate weights, according to the simple updating scheme, equation (2.21*b*). (The algorithm (2.21*a*), (2.21*b*) is of advantage only in the context of weight refinement, near the multi-canonical limit, where, as we shall see the VS method is probably preferable.)

Figure 3 shows the results of this procedure applied to the energy macrostates of an Ising model. In this case, the simulations were initiated (successively) from the ground

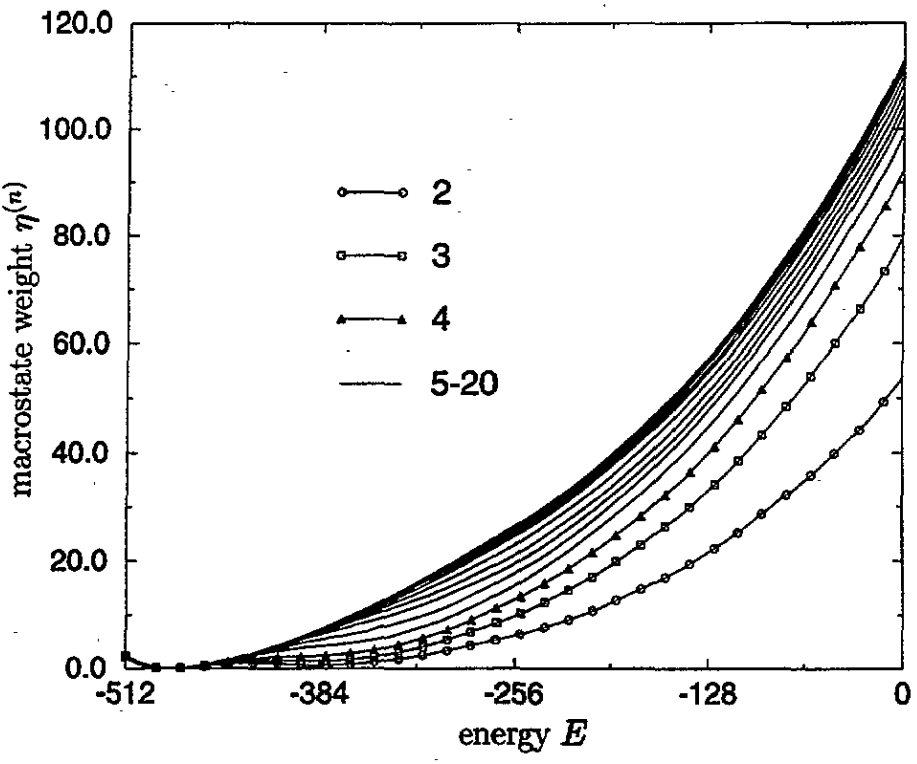

Figure 3. Evolution of the weights of energy macrostates using the transition probability method (equations (2.26), (2.24), and (2.21*a*), (2.21*b*)), for iterations 1–20. (The model parameters are as defined in figure 1.)

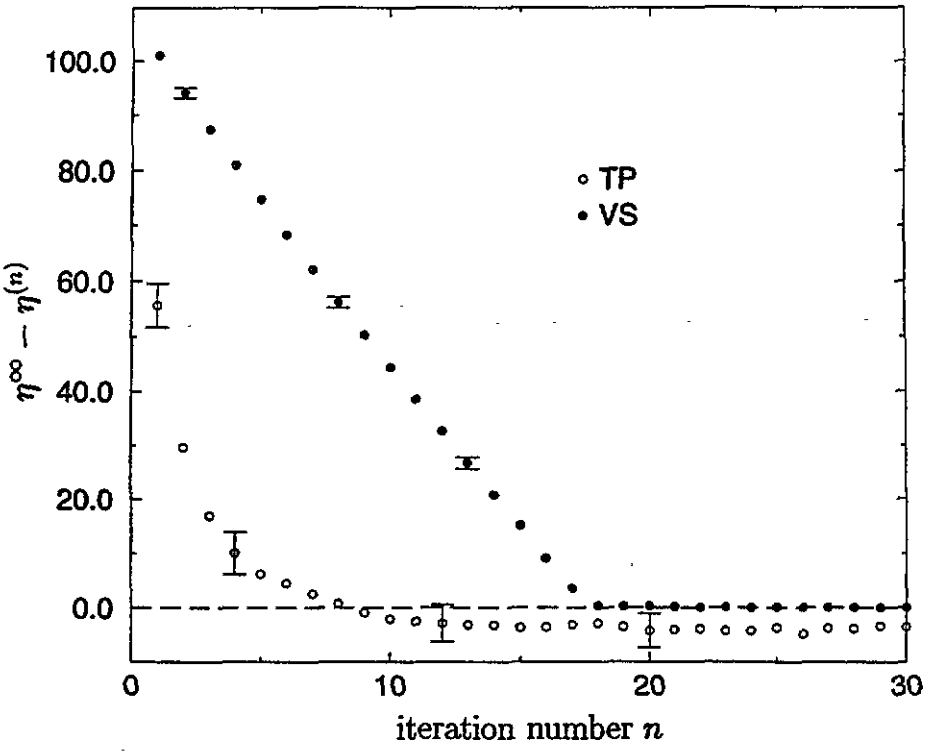

Figure 4. The evolution of the weight $\eta^{(n)}$ for the $E \simeq 0$ energy macrostate as a function of iteration number, for the visited states method (VS) and the transition probability method (TP). In each case the ordinate shows the *difference* $\eta^{\infty} - \eta^{(n)}$ where η^{∞} is defined by the limiting behaviour of the refined VS algorithm (2.21*a*). (The model parameters are as defined in figure 1.)

state (with energy $E = -2N$) and from a microstate chosen randomly from the infinite-temperature ensemble (with $E \simeq 0$). The parameter n_t was set to 600, so that each iteration took about as long as its equivalent in our tests of the visited states method. To keep the matrix $\tilde{p}_{ij}^{(n)}$ tridiagonal, we binned the energy macrostates so that the width of each bin was $\Delta E = 8$. Figure 3 shows that, in contrast to the visited states method (cf figure 2), the transition probability scheme yields significant information about the whole space of macrostates already from the earliest iterations. This comparison is made more explicitly in figure 4 which shows the weights assigned by the two schemes to the $E = 0$ macrostate, as a function of iteration number. The TP method clearly wins over the course of the *early* iterations, converging rapidly to a good approximation to the multicanonical limit. (We shall return to discuss the behaviour over the later iterations.)

We have also applied the TP method to the magnetization macrostates of an $N = 32^2$ Ising model at and below its critical temperature. This problem provides a simple example of the important case of a canonically distributed macroscopic variable with *two* distinct peaks. This is the situation that has to be confronted whenever we consider a system with a phase boundary, separating phases distinguished by different values of that (or some other) macroscopic variable. To deal with this kind of problem we need to extend the spectrum of the initial states used to launch a simulation sequence, to ensure that the entire space of macrostates is sampled effectively. In this case we have used *three* launch states. The ordered microstates (with $M = \pm M_s$ where $M_s \equiv N$ is the saturation magnetization) define

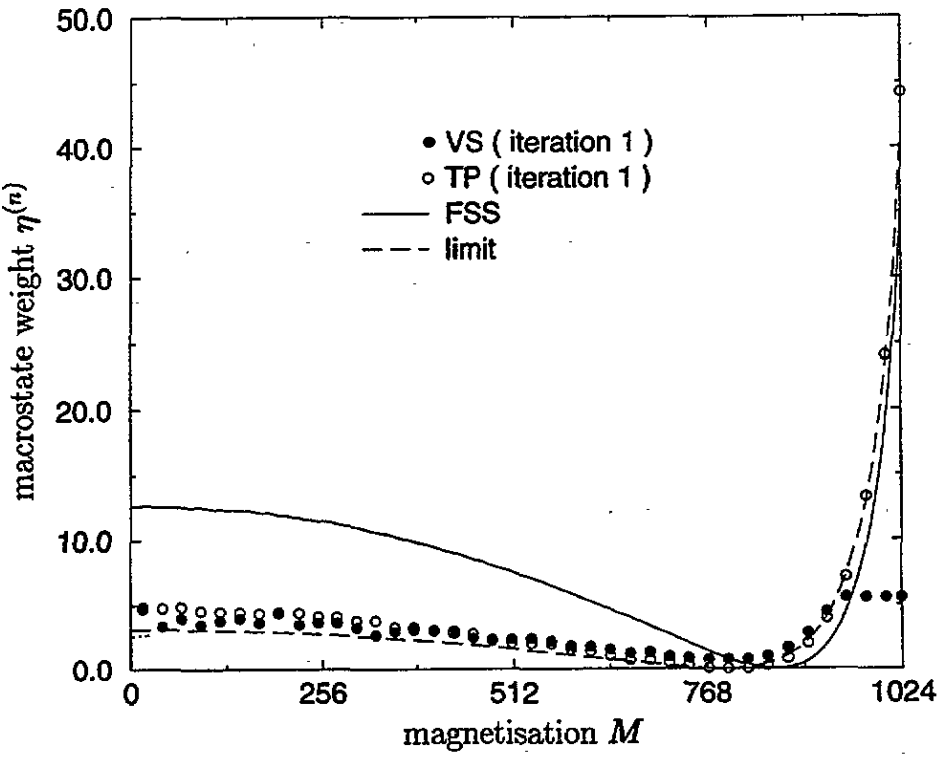

Figure 5. The weight of magnetization macrostates for a 2D Ising model with $\beta = \beta_c = 0.4406$, and $N = 32^2$, inferred from one iteration of TP, one iteration of VS and naive finite size scaling of the limiting weights for the $N = 16^2$ system. The full curve shows the multi-canonical limit established from an extended VS procedure.

two of these states; the third is a microstate chosen randomly from the set associated with the $M = 0$ macrostate. Using these three initial states the TP algorithm yields a sequence of weights which cover the entire macrostate space and which converge to a practically usable approximation to the multi-canonical limit *on the first iteration* (figure 5), in contrast to the result of a VS run of the same length (also shown in figure 5). The faster convergence compared with the application to energy macrostates reflects differences in the macrostate TP matrix in the two cases. In the case of magnetization macrostates the matrix ρ_{ij} is naturally tridiagonal. For energy macrostates it is tridiagonal only if (as we have chosen to do) the energy is blocked (coarse-grained). While preserving the simplicity of a tridiagonal matrix, the blocking compromises the approximation underlying equation (2.23), which presupposes that the degrees of freedom *within* each 'macrostate' (which now include the energy) relax on a faster time-scale than that characterizing the transitions *between* macrostates. In fact, it can be shown that *fewer* transitions occur in the direction (through macrostate space) in which $p^{(n)}$ is increasing (and *more* occur in the opposite direction) than would occur if *local equilibrium were established within each bin. As a result, the eigenvector estimator* continually underestimates the change in weights required.

These two examples show that a few iterations of the TP method can provide a good approximation to the ideal multi-canonical weights, over any desired range of macrostate space. For many purposes the approximation may be sufficiently good that no further

refinement is necessary. However, our studies suggest that the limiting weights emerging from TP may be slightly biased with respect to the multi-canonical ideal. The bias reflects the fact that (2.23) is not fully satisfied. Elsewhere [23, 33] we have shown that this problem disappears (the local equilibrium condition (2.23) is automatically satisfied) in the multi-canonical limit, *provided* the launch macrostates are chosen *randomly* from the macrostate space under investigation (and that the Markov chain generated from each launch state has the same length). An alternative procedure, applicable in the present context, would be to *enforce* (2.23) by following each macrostate transition with a sequence of spin-updates *constrained* to preserve the macrostate, thus establishing local equilibrium before the next macrostate transition is attempted. Of course, doing this reduces the TP method's speed advantage over the VS method and for the final stages of weight-refinement, where the weights change only a little between iterations, it may still be more efficient to use VS.

3. Utilizing the sampling distribution

3.1. Energy weighting: free-energy estimation

As we have already noted (and shall shortly explore) standard Boltzmann sampling methods do not provide immediate access to the free energy. The task of devising appropriate extensions of the standard techniques, to deal with this problem, has attracted continuing attention over the years: many different schemes have emerged. The most widely practiced involve integration of the canonically averaged energy along some path connecting the system of interest to some reference system. Reviews are to be found in [2, 9, 33] (which includes an extended bibliography). Here we shall focus on the utility of the multi-canonical distribution, and compare it with integration methods (IM).

The task of determining the free energy defined by

$$F(\beta) \equiv -\frac{1}{\beta} \ln\left[Z^c(\beta)\right] \tag{3.1}$$

entails finding a satisfactory way of estimating, by MC sampling, the partition function sum (equation (2.2))

$$Z^c(\beta) = \sum_r \phi_r \qquad \text{with} \quad \phi_r \equiv \mathrm{e}^{-\beta E_r}. \tag{3.2}$$

We estimate the sum by realizing a set of N_c microstates indexed by $t = 1, 2 \ldots N_c$ drawn from the general distribution (equation (2.5))

$$p_r(\eta, \hat{\beta}) = \frac{\mu_r}{\mathcal{Z}(\hat{\beta}, \eta)} \qquad \text{with} \quad \mu_r \equiv \mathrm{e}^{-\hat{\beta}E_r + \eta_i} \qquad r \in i \tag{3.3}$$

where $\hat{\beta}$ may be different from β, and the weights refer to *energy* macrostates. Then

$$\left\langle \frac{1}{N_c} \sum_{t=1}^{N_c} \frac{\phi_t}{\mu_t} \right\rangle_{\hat{\beta},\eta} = \frac{1}{\mathcal{Z}(\hat{\beta}, \eta)} \sum_r \phi_r = \frac{Z^c(\beta)}{\mathcal{Z}(\hat{\beta}, \eta)} \tag{3.4a}$$

and

$$\left\langle \frac{1}{N_c} \sum_{t=1}^{N_c} \frac{1}{\mu_t} \right\rangle_{\hat{\beta},\eta} = \frac{\Omega_T}{\mathcal{Z}(\hat{\beta}, \eta)} \tag{3.4b}$$

where we adopt the notation

$$\langle\ [\cdot]\ \rangle_{\hat{\beta},\eta} \equiv \sum_r [\ \cdot\ ] p_r(\hat{\beta}, \eta) \tag{3.5}$$

and $\Omega_T = \sum_r$ is the total number of microstates. Combining equations (3.4*a*) and (3.4*b*) to eliminate the unknown normalization constant associated with the sampling distribution (3.3), we find that the partition function may be written as

$$Z^c(\beta) = \left[\left\langle \frac{\Omega_T}{N_c} \sum_{t=1}^{N_c} \frac{\phi_t}{\mu_t} \right\rangle_{\hat{\beta},\eta}\right] \Big/ \left[\left\langle \frac{1}{N_c} \sum_{t=1}^{N_c} \frac{1}{\mu_t} \right\rangle_{\hat{\beta},\eta}\right]. \tag{3.6}$$

It follows that we may construct a ratio estimator for the partition function, $Z^c(\beta) \stackrel{\text{eb}}{=} \tilde{Z}^c(\beta)$ with

$$\begin{aligned} \tilde{Z}^c(\beta) &= \left[\frac{\Omega_T}{N_c} \sum_{t=1}^{N_c} \frac{\phi_t}{\mu_t}\right] \Big/ \left[\frac{1}{N_c} \sum_{t=1}^{N_c} \frac{1}{\mu_t}\right] \\ &= \left[\Omega_T \sum_{i}^{N_m} C_i(\hat{\beta}, \eta)\, \mathrm{e}^{(\hat{\beta}-\beta)E_i - \eta_i}\right] \Big/ \left[\sum_{i}^{N_m} C_i(\hat{\beta}, \eta)\, \mathrm{e}^{\hat{\beta}E_i - \eta_i}\right] \end{aligned} \tag{3.7}$$

where $C_i(\hat{\beta}, \eta)$ denotes the number of occurrences of energy macrostate i amongst the N_c observations drawn from the sampling distribution (3.3). (We note that the right-hand side of (3.7) does not provide an *unbiased* estimator of $Z^c(\beta)$ [34]. However, the bias is $O(1/N_c)$ and therefore small on the scale of the sampling error; it can be suppressed by the use of 'double-jackknife bias-corrected estimators' [29].) Now, irrespective of the sampling process (that is, irrespective of the values assigned to the weights η),

$$C_i(\hat{\beta}, \eta)\, \mathrm{e}^{\hat{\beta}E_i - \eta_i} \sim \Omega(E_i) \tag{3.8a}$$

and

$$C_i(\hat{\beta}, \eta)\, \mathrm{e}^{(\hat{\beta}-\beta)E_i - \eta_i} \sim \mathrm{e}^{-\beta E_i} \Omega(E_i) \sim p^c(\beta, E_i)\,. \tag{3.8b}$$

Thus, the sum featuring in the denominator of (3.7) is dominated by energies in the vicinity of the maximum E^∞ of the density of states function $\Omega(E)$ (equation (2.4*a*)), while the sum in the *numerator* is dominated by energies close to $\bar{E}^c(\beta)$, the most probable energy in the canonical distribution (2.4*b*). While Boltzmann sampling can provide a reliable estimate of the latter sum (it samples most effectively in the region $E \simeq \bar{E}^c$) it will provide an unreliable estimate of the former (it will sample hardly at all the region $E \simeq E^\infty$). This is why Boltzmann sampling methods are inappropriate for this problem. In contrast, the multi-canonical distribution extends over *all* ranges of energy, allowing *both* sums to be estimated reliably, for any β. The multi-canonical distribution evolved for one temperature may thus be used to determine the free-energy (and, indeed all other properties that can be written as averages of an operator over energy macrostates) at any other temperature, constituting a powerful extension of the histogram reweighting technique [14].

This is not to say that the multi-canonical distribution (of whatever temperature) is *optimal* for the free-energy importance sampling process. We shall return to this point briefly in section 4; here we need to note only that the results of the estimation process are rather insensitive to the *details* of the sampling distribution, provided only that its samples effectively in the two regions dominating the sums featuring in (3.7); thus, in practice, we need only a set of weights η which approximate the multi-canonical ideal η^* to within terms of order unity.

To benchmark the utility of the multi-canonical approach we again appeal to the zero-field 2D Ising model, where exact results for the free energy exist not only in the thermodynamic limit [35] but also for systems of finite size [36]. The MCMC process is divided into two parts: an initial stage which determines an acceptable approximation to

the multi-canonical set; and a production stage in which the distribution is sampled to yield the two sums featuring in the estimator $\tilde{Z}^c(\beta)$ (equation (3.7)) and thence the free-energy density estimator

$$f(\beta) \stackrel{\text{eb}}{=} \tilde{f} \equiv -\frac{1}{N\beta} \ln\left[\tilde{Z}^c(\beta)\right]. \tag{3.9}$$

We applied this method to an $N = 32^2$ Ising system. The multi-canonical weights were determined by using the the transition probability method of section 2.4, on a system of inverse temperature $\hat{\beta} = 0.55$. The production runs entailed some 10^7 lattice sweeps, generated in ten 'blocks' with jackknife blocking used to estimate errors for all results. The results are shown in figure 6,where they are compared with exact results (for the $N = 32^2$ system) constructed from [36]. To make the accuracy obtained clearly visible, we have plotted the *difference* Δf between the simulation and exact results. Over the entire range, the (1σ) error bars are smaller than 0.0002; the largest fractional error is 0.01%.

For comparison, we have also determined the free energy, for the same system, using one of the standard integration methods (IM) which exploit the fact that free-energy *derivatives* are related to canonical averages that *are* accessible to conventional Boltzmann sampling. Thus, in particular, the free-energy can be determined by integrating the measured *energy*

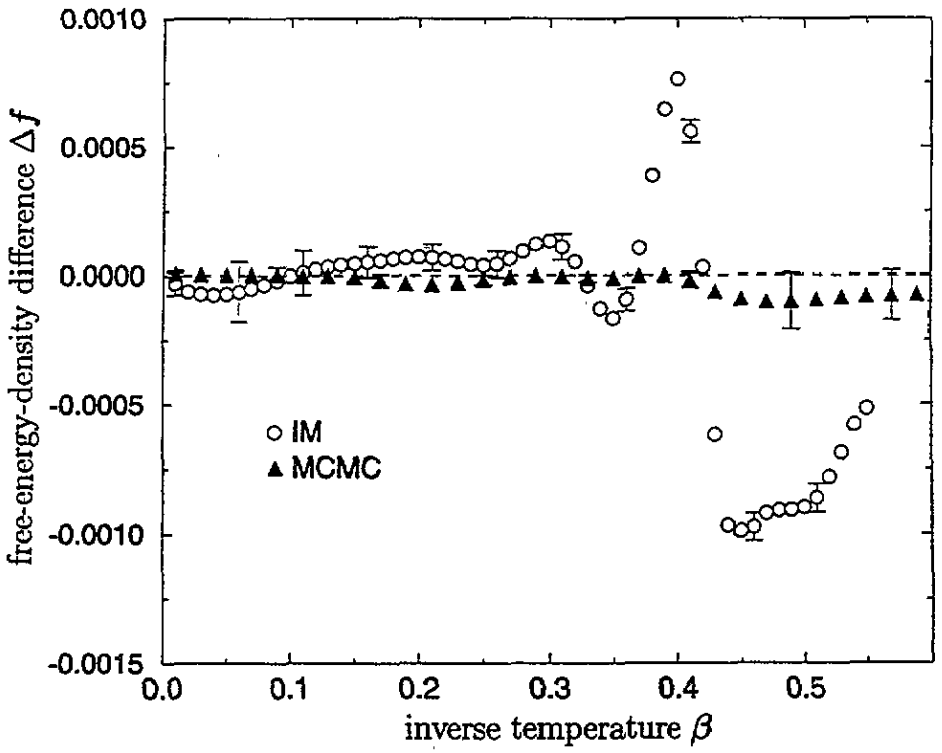

Figure 6. Comparison of thermodynamic integration (IM) and multi-canonical (MCMC) calculations of the free energy density of the $N = 32^2$ Ising model, displayed as *differences* from the exact finite-size results [36].

with respect to inverse-temperature:

$$\beta f(\beta) = \beta_r f(\beta_r) + \frac{1}{N}\int_{\beta_r}^{\beta} \bar{E}^c(\beta')\,\mathrm{d}\beta' \tag{3.10}$$

with an appropriate choice of reference temperature β_r. In this case we chose $\beta_r = 0$, so that $\beta_r f(\beta_r) = -\ln 2$. We made measurements of $\bar{E}^c$ using Boltzmann sampling simulations for 11 evenly spaced values of inverse temperature between 0.05 and 0.55, with 10^6 lattice sweeps at each temperature, entailing a total compute time comparable with that of the MCMC calculation. An interpolating spline was fitted to the data points and the integral in equation (3.10) evaluated numerically.

The results are also shown in figure 6. At small β, MCMC and IM methods yield comparable accuracy. However, at larger β the IM results deviate significantly from the exactly established values. This is not a random error: the error bars, which represent the measured spread of the estimator, obtained by jackknife blocking, are approximately the same size as those on the multi-canonical data; rather, it is a systematic error which can be traced to the phase transition that lies on the path of integration (at $\beta_c = 0.44\ldots$, in the infinite-N limit). In this region the variation of $\bar{E}^c$ with β is too rapid (the heat capacity is logarithmically divergent at β_c in the thermodynamic limit) to be reflected adequately in the data points. To reduce this error we would have had to space the integration points differently, clustering them around the phase transition point. By contrast no such special care is required in the multi-canonical method. In the vicinity of the phase transition the error bars get larger because of critical fluctuations, but they still contain the line $\Delta f = 0$ and are thus trustworthy confidence measures, in contrast to those associated with the thermodynamic integration data.

3.2. *Magnetization weighting: from scaling to saturation*

As our second application of multi-canonical methods we turn to consider the canonical distribution of magnetization macrostates in a critical Ising model.

It is well established [37, 38] that, for a critical system whose linear dimension L is large compared to microscopic lengths, the canonical probability density function (PDF) of the magnetization $m \equiv L^{-d}M$ has the universal scaling form

$$p(m)\,\mathrm{d}m \simeq p^*(x)\,\mathrm{d}x \qquad \text{with} \quad x \equiv m/m_\sigma \tag{3.11}$$

where

$$m_\sigma \equiv \langle m^2 \rangle^{1/2} \sim L^{-d/(1+\delta)} \tag{3.12}$$

with δ the equation of state exponent. The function $p^*(x)$ is believed to be universal, in the sense that it describes the distribution of the order parameter in all systems of the same universality class (and subject to the same boundary conditions). Thus, for example, there is good reason to believe that the distribution of the magnetization in a critical Ising model has the same form as the distribution of the density in a critical fluid of the same spatial dimension [20]. This correspondence has provided a fruitful basis for the accurate determination of critical points in simulations of fluids [13].

Although renormalization-group calculations [38–40] yield some understanding of the structure of the universal scaling function $p^*(x)$, most of what is known to date rests on simulations probing the region accessible to conventional Boltzmann sampling. Here we use multi-canonical methods to explore the form of $p^*(x)$ *beyond* this region. The motivation for this study is a recent conjecture [21, 22] for the large-x behaviour of $p^*(x)$:

$$p^*(x) \simeq p_\infty x^\psi \mathrm{e}^{-a_\infty x^{\delta+1}} \tag{3.13a}$$

with

$$\psi = \frac{\delta - 1}{2} \tag{3.13b}$$

while p_∞ and a_∞ are universal constants (implicit in the form of p^*). The structure of the exponential (equation (3.13*a*)) is suggested by rigorous results for the 2D Ising model [41] and is consistent with MC studies of the Ising universality class [19]. The value assigned to the exponent of the power-law prefactor (equation (3.13*b*)) has been shown [22] to lead to a successful account of the value of the Privman–Fisher universal amplitude [42], characterizing the free energy of a finite-sized system at its critical point. This form of prefactor also emerges from a recently developed theory [21] which argues that the critical order parameter distribution may be related to the stable distributions of probability theory. However, this theory also suggests the existence of *additional* non-universal contributions to the order parameter distribution, falling off as a *power* at large x and thus asymptotically dominant.

To explore these issues we have performed a multi-canonical study of magnetization macrostates in the critical 2D Ising model (of sizes $L = 32$ and $L = 64$), extending the region sampled all the way to the saturation magnetization $|M| = N = L^2$. To evolve an appropriate set of weights we used the TP method initially, with final refinements using VS. The final set of weights extend over a range 0 to 191 (for $L = 64$), manifesting a variation of some 83 *decades* in the magnetization PDF itself.

The probability p_0 that the system will be found in (a particular) ground state, with the saturation magnetization, is directly measurable in this ensemble. It is easy to see from equation (2.3) (with $\mathcal{F}_0 = E_0 = -2N$) that this probability provides another route to the determination of the free energy through

$$f(\beta_c) = -2 + \frac{\ln p_0}{N\beta_c}. \tag{3.14}$$

The results which follow from the application of these equations are shown in table 1. The remarkable precision gives a measure of the accuracy with which the ground state probabilities have been measured.

Table 1. The critical free-energy density of the 2D Ising model obtained by direct measurement of the ground-state probability, using the multi-canonical ensemble in conjunction with equation (3.14), compared with exact results (quoted to five decimal places).

	Critical free energy density	
L	MCMC	Exact [36]
32	−2.111 15(4)	−2.111 07
64	−2.109 99(2)	−2.110 01

To test the ansatz (3.13*a*), (3.13*b*) we plot, in figure 7(*a*), the function

$$q^*(x) \equiv -\ln\left[x^{-7} p^*(x)\right] \tag{3.15}$$

with $\delta = 15$ (and thus $\psi = 7$), appropriate in $d = 2$. In a region described by equations (3.13*a*) and (3.13*b*) this function should be *linear* in the scaling variable x. The figure shows that such a region does exist, at intermediate values of x, and that it extends to larger values of x for the larger system. This is indeed what one must expect: universal scaling behaviour requires the limit $L \to \infty$ at fixed x, since only then are corrections to

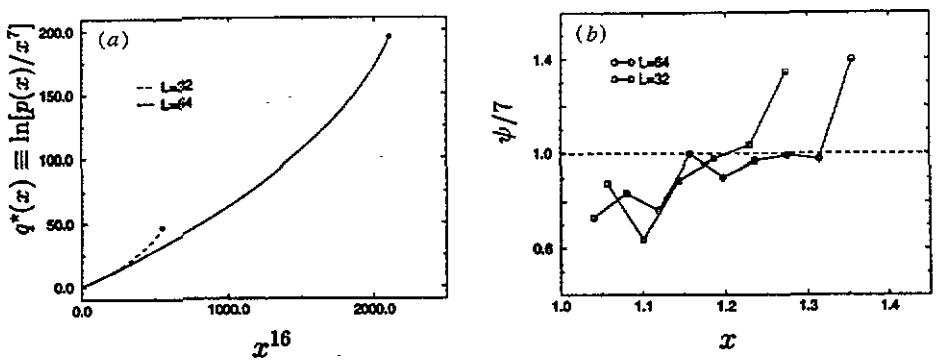

Figure 7. (*a*) The function $q^*(x) \equiv -\ln[x^{-7}p^*(x)]$ plotted against x^{16} obtained by sampling from the multi-canonical distribution established for $L = 32$ and $L = 64$. The data points in which the lines terminate are exact results deduced from [36]. (*b*) The results of fitting the ansatz (3.13*a*) to the critical magnetization PDF, over a series of windows of the scaling variable x for $L = 32$ and $L = 64$. The ordinate shows the ratio of the best fit value of the exponent ψ to the prediction (3.13*b*). The abscissa shows the value of the scaling variable locating the centre of the window used in each fit.

scaling guaranteed to be negligible; the limit $x \to \infty$ at fixed L must eventually take us outside this regime. While we have no detailed understanding of the factors controlling the behaviour in this region, beyond the scaling limit, the results in figure 7(*a*) do not support the suggestion [21] that the asymptotic behaviour of the magnetization PDF is a power-law decay. If it were, then at large enough x we would expect that $q^*(x) \sim \ln x$, which would be concave (i.e. have a negative second derivative); the observed behaviour is convex.

Figure 7(*b*) provides a more stringent test of the prefactor structure suggested in equations (3.13*a*), (3.13*b*). It shows the result of fitting the ansatz (3.13*a*) to the measured magnetization PDF over a series of windows of x-values. The results provide substantial support for the assignment (3.13*b*) for the prefactor exponent ψ.

4. Conclusions

In this paper we have explored both the techniques for generating multi-canonical distributions, and some of the applications of multi-canonical sampling. We divide our conclusions accordingly.

The studies reported in section 2.3 take us only a little way towards realizing the full Bayesian formulation of the MCMC programme set out in section 2.1. The algorithm defined by equations (2.21*a*) and (2.21*b*) allows one to propagate only a limited amount of the information gathered in earlier iterations; its pay-off—the suppression of noise in the multicanonical distribution it delivers—has only marginal impact on the quality of averages formed by sampling from this distribution. However, we anticipate that the procedure will prove rather more valuable when the MCMC strategy is developed so as to allow physical quantities of interest to be computed *directly* from the multicanonical weights, rather than by further simulation within the multicanonical ensemble.

The transition probability method described in section 2.4 is, we believe, of immediate practical utility. It provides an efficient way of obtaining a good approximation to a multi-canonical set of weights, covering any targeted range of macrostate space. In some circumstances the weights that TP delivers may be quite adequate (cf section 3.1). In others

(cf section 3.2) one may wish only to use the TP weights to bootstrap some further weight-refinement process (such as the refined VS algorithm of section 2.3). In this respect TP is complementary to the finite-size-scaling (FSS) strategy in common use [5, 12, 18], where an initial set of weights is inferred on the basis of Boltzmann sampling measurements on a system small enough to allow the macrostate space to be fully explored. In some cases (homogeneous systems, such as the Ising model in the one phase region) an FSS bootstrap followed by visited-states refinement is a simple and effective procedure. But FSS of the weights is not always straightforward to implement (the ideal weights are not simply extensive either in a two-phase region, or at a critical point: cf figure 5); and sometimes not applicable at all (in the context of a spin glass [11]). We have not yet addressed the spin-glass problem. But here we have demonstrated the capacity of TP to handle the critical point, and in further explorations of the method, to be described elsewhere [23], have applied it successfully to the determination of a multi-canonical sampling distribution bridging between two different solid phases. We note that the method may also be applied to determine the weights associated with the sub-ensembles that feature in the expanded ensemble techniques developed by Lyubartsev *et al* [7] and Marinari and Parisi [8].

As regards the *applications* of the multi-canonical distribution, we observe, first, that the multi-canonical distribution is surely not *optimal* as regards the estimation of the canonical averages that determine the free energy. This is intuitively clear from the fact that the macrostates that lie between the two energies that locate the dominant regions of numerator and denominator in equation (3.7) contribute hardly anything to either sum; their weight is of importance only because it controls the probability of the system tunnelling between these two regions. Thus there is no reason *a priori* to expect that a flat sampling distribution will be optimal for this problem or, indeed, for the estimation of any canonical average. The issue of what sampling distribution *is* optimal, for the estimation of a given average, was addressed in the earliest days of computer simulation [3], and has recently attracted renewed attention [27]. On the basis of our own investigations [33], we observe that although the multi-canonical distribution is not optimal for any ensemble average, the gains associated with the optimization process are marginal. The key point is that multi-canonical sampling is never bad in the way that Boltzmann sampling can be bad. Boltzmann sampling is bad for some observables (those which contain the averages of exponentials) because the fraction of the sampling weight that it puts in the region of macrostate space which dominates the ensemble average may be exponentially small. Multi-canonical sampling can never have this problem, because it puts an equal amount of weight in every region of macrostate space.

Secondly, our study of the multi-canonical distribution for the magnetization (section 3.2) has demonstrated the remarkable precision with which the method may be used to determine probabilities with extremely small absolute values. This capacity, already exploited in studies of interfaces between coexisting phases [18], has many interesting potential applications.

However, it is in the context of the phase behaviour of condensed matter systems that the multi-canonical method seems set to have the most striking impact. The studies reported in section 3.1 show that multi-canonical sampling is at least as accurate as traditional integration methods in the measurement of absolute free energy values, and is much better able to deal with the problems arising from the occurrence of a phase transition in the space sampled. The advantages of the multi-canonical approach are sharpened when it comes to the task of determining the relative free energies of two phases—in effect, the generic problem of locating a phase boundary. Integration methods need to resort to double-tangent constructions; as shown in the seminal paper [5] multi-canonical methods allow the two phases to be sampled in the same simulation. Although further ingenuity will surely be

required to evolve the multi-canonical distributions appropriate to different types of phase boundary, the number of potential applications here is enormous.

Acknowledgment

GRS acknowledges the support of an SERC research studentship.

References

[1] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[2] Binder K 1985 *J. Comput. Phys.* **59** 1
[3] Fosdick L D 1963 *Methods Comput. Phys.* **1** 245
[4] Torrie G M and Valleau J P 1974 *Chem. Phys. Lett.* **28** 578
[5] Berg B A and Neuhaus T 1991 *Phys. Lett.* **267B** 249; 1992 *Phys. Rev. Lett.* **68** 9
[6] Berg B A 1993 *Int. J. Mod. Phys.* C **4** 249
[7] Lyubartsev A P, Martsinovski A A, Shevkunov S V and Vorontsov-Velyaminov P N 1992 *J. Chem. Phys.* **96** 1776
[8] Marinari E and Parisi G 1992 *Europhys. Lett.* **19** 451
[9] Frenkel D 1986 *Molecular-Dynamics Simulation of Statistical-Mechanical Systems* ed G Ciccotti and W G Hoover (Amsterdam: North-Holland)
[10] Berg B A 1995 Multicanonical recursions *Preprint*
[11] Berg B A and Celik T 1992 *Phys. Rev. Lett.* **69** 2292
[12] Lee J 1993 *Phys. Rev. Lett.* **71** 211, 2353 (erratum)
[13] Wilding N B 1995 *Phys. Rev.* E **52** 602
[14] Ferrenberg A M and Swendsen R H 1988 *Phys. Rev. Lett.* **61** 2635, 1658 (erratum)
[15] Baumann B and Berg B A 1985 *Phys. Lett.* **164B** 131
[16] Berg B A, Hansmann U H E and Celik T 1993 *Europhys. Lett.* **22** 63; 1994 *Phys. Rev.* B **50** 16444
[17] Hansmann U H E and Okamoto Y 1993 *J. Comput. Chem.* **14** 1333
[18] Berg B A, Hansmann U H E and Neuhaus T 1993 *Z. Phys.* B **90** 229
[19] Nicolaides D and Bruce A D 1988 *J. Phys. A: Math. Gen.* **21** 233
[20] Bruce A D and Wilding N B 1992 *Phys. Rev. Lett.* **68** 193
[21] Hilfer R 1994 *Z. Phys.* B **96** 63
Hilfer R and Wilding N B 1995 *J. Phys. A: Math. Gen.* **28** L281
[22] Bruce A D 1995 *J. Phys. A: Math. Gen.* **28** 3345
[23] Smith G R and Bruce A D 1995 Multi-canonical Monte-Carlo study of solid–solid phase coexistence in a model colloid *Preprint*
[24] Jeffreys H 1939 *The Theory of Probability* (Oxford: Clarendon)
[25] Mezei M 1987 *J. Comput. Phys.* **68** 237
[26] Geyer C J and Thompson E A 1992 *J. R. Stat. Soc.* B **54** 657
[27] Hesselbo B and Stinchcombe R B 1995 *Phys. Rev. Lett.* **74** 2151
[28] Wilding N B, Müller M and Binder K 1994 *J. Chem. Phys.* **101** 4324
[29] Berg B A 1992 *Comput. Phys. Commun.* **69** 7
[30] Bennett C H 1976 *J. Comput. Phys.* **22** 245
[31] Kerler W, Rebbi C and Weber A 1995 *Nucl. Phys.* B **S42** 675
[32] Martin J J 1967 *Bayesian Decision Problems and Markov Chains* (New York: Wiley)
[33] Smith G R 1995 The measurement of free energy by Monte Carlo computer simulation *PhD Thesis* University of Edinburgh
[34] Kennedy A D 1993 *Nucl. Phys.* B **S30** 96
[35] Onsager L 1944 *Phys. Rev.* **65** 117
[36] Ferdinand A E and Fisher M E 1969 *Phys. Rev.* **185** 833
[37] Binder K 1981 *Z. Phys.* B **43** 119
[38] Bruce A D 1981 *J. Phys. C: Solid State Phys.* **14** 3667
[39] Eisenriegler E and Tomaschitz R 1987 *Phys. Rev.* B **35** 4876
[40] Burkhardt T W and Derrida B 1985 *Phys. Rev.* B **32** 7273
[41] McCoy B M and Wu T T 1973 *The Two-Dimensional Ising Model* (Cambridge, MA: Harvard University Press)
[42] Privman V and Fisher M E 1984 *Phys. Rev.* B **30** 322