

# A Monte Carlo Sampling Scheme for the Ising Model

Roland Häggkvist,<sup>1</sup> Anders Rosengren,<sup>2</sup> Daniel Andrén,<sup>1</sup> Petras Kundrotas,<sup>2,3</sup>  
Per Håkan Lundow,<sup>2</sup> and Klas Markström<sup>1</sup>

*Received August 12, 2002; accepted July 18, 2003*

---

In this paper we describe a Monte Carlo sampling scheme for the Ising model and similar discrete-state models. The scheme does not involve any particular method of state generation but rather focuses on a new way of measuring and using the Monte Carlo data. We show how to reconstruct the entropy  $S$  of the model, from which, e.g., the free energy can be obtained. Furthermore we discuss how this scheme allows us to more or less completely remove the effects of critical fluctuations near the critical temperature and likewise how it reduces critical slowing down. This makes it possible to use simple state generation methods like the Metropolis algorithm also for large lattices.

---

**KEY WORDS:** Monte Carlo methods; density of states; microcanonical.

## 1. INTRODUCTION

The Lenz–Ising model of ferromagnetism has been thoroughly studied since its conception in the 1920's. It was solved in the 1-dimensional case by Ising himself in 1925<sup>(1)</sup> and in the 2-dimensional case without an external field by Onsager in 1944.<sup>(2)</sup> For an introduction to the model, see ref. 3. The partition function of the model is in general notoriously hard to compute. It is defined for a graph  $G$  on  $n$  vertices and  $m$  edges as

$$Z(G; x, y) = \sum_{i,j} a_{i,j} x^i y^j. \quad (1)$$

---

<sup>1</sup> Department of Mathematics, Umeå University, SE-901 87 Umeå, Sweden; e-mail: Roland.Haggkvist@math.umu.se; Daniel.Andren@math.umu.se; Klas.Markstrom@math.umu.se

<sup>2</sup> Department of Physics, AlbaNova University Center, KTH, SE-106 91 Stockholm, Sweden; e-mail: ar@theophys.kth.se and phl@kth.se

<sup>3</sup> Present address: Karolinska institutet, Department of Biosciences at Novum, SE-141 57 Huddinge, Sweden; e-mail: Petras.Kundrotas@csb.ki.se

Here  $a_{i,j}$  is the number of induced subgraphs of  $G$  on  $(n+j)/2$  vertices with  $(m-i)/2$  edges at the boundary. We refer to the index  $i$  as the energy level and  $j$  as the magnetisation. The traditional partition function studied in statistical physics is then obtained by evaluating it at a certain point

$$Z(G; e^K, e^H),$$

where  $K = J/k_B T$ ,  $H = h/k_B T$ , and  $J$  and  $h$  are parameters describing the interaction through edges and with an external magnetic field respectively,  $T$  is the temperature and  $k_B$  is the Boltzmann constant. In order to simplify our presentation we henceforth set  $J = k_B = h = 1$ . From the partition function we obtain various quantities such as the internal energy, magnetisation, susceptibility etc. The main goal in the study of the Ising model on a graph  $G$ , or some family of graphs, is usually to study the model in the vicinity of a *critical temperature*, denoted  $T_c$ , where the model undergoes a phase transition and there determine the behaviour of various critical properties.

As an alternative to actually computing the partition function, which is hard to do, we can use sampling methods to obtain an approximation. Standard methods of sampling include the Metropolis method,<sup>(4)</sup> the Wolff cluster algorithm<sup>(5)</sup> and the Swendsen–Wang process;<sup>(6)</sup> for a textbook treatment of these methods see, e.g., ref. 7.

When working with sampling methods it is natural to focus on finding an approximation of  $Z$  as a function of the temperature  $T$ , or equivalently the coupling  $K$ , although doing so is not necessarily the best way to obtain detailed information about the behaviour of the model and its asymptotics. As noted by many before us, energy and entropy are more fundamental parameters of a thermodynamical system than temperature. See ref. 8 for a modern treatment of the fundamentals of thermodynamics from this point of view.

In this paper we show how to use temperature-based sampling methods to get information about the coefficients  $a_{i,j}$  in Eq. (1) and give some results on the basic behaviour of our sampling scheme. We have been using this method in practice since 1996 in a joint project to explore practical use of graph theoretical methods in statistical physics; our experience with the method is that it is both fast and gives reliable and detailed data for the graphs on which we have used it. See refs. 9–11 for our first publications based on this work. During the same period of time this way of sampling has also been approached in a similar, but not identical, way by others, see, e.g., refs. 12–14. Our approach is in some sense more basic since we avoid involving any specific method of state generation in our scheme, but the methods make use of the same underlying structures. However, we feel that

the use of a graph theoretic language gives a clearer, in fact almost obvious, description of the ideas used, and keeps the amount of terminology needed at a minimum.

## 2. NOTATION AND DEFINITIONS

We will generalise our scheme to that of any simple graph of finite degree. Let  $G$  be a graph on  $n$  vertices  $\{1, 2, \dots, n\}$  and  $m$  edges. If  $u$  and  $v$  are adjacent vertices then  $\{u, v\}$  is an edge. To each vertex  $v$  we assign a spin  $\sigma_v = \pm 1$  and say that the graph is in state  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ . In total a graph can take  $2^n$  different states. Given a state  $\sigma$  its *energy* is  $E(\sigma) = \sum_{\{u,v\}} \sigma_u \sigma_v$ , the sum is taken over all edges  $\{u, v\}$  of  $G$ , and its *magnetisation* is  $M(\sigma) = \sum_v \sigma_v$ , where the sum is taken over all vertices  $v$  in  $G$ . Sometimes we will need their normalised forms  $\nu(\sigma) = E(\sigma)/m$  and  $\mu(\sigma) = M(\sigma)/n$ , both taking values in  $[-1, 1]$ . The *local energy* at a vertex  $v$  is defined as  $\delta(v) = \sum_{u \in N(v)} \sigma_u \sigma_v$ , where  $N(v)$  denotes the set of neighbours of  $v$ . It is easy to see that if we flip a spin  $\sigma_v$  (i.e., change  $\sigma_v$  to  $-\sigma_v$ ) the energy will change by an amount  $-2\delta(v)$  and the magnetisation will change with  $-2\sigma_v$ . The object under study is the partition function, expressed as a formal (generalised) polynomial in variables  $x$  and  $y$

$$Z(G; x, y) = \sum_{\sigma} x^{E(\sigma)} y^{M(\sigma)} = \sum_{i,j} a_{i,j} x^i y^j$$

which defines the coefficients  $a_{i,j}$ . By overloading our notation we also define the somewhat simpler-looking partition function

$$Z(G; x) = Z(G; x, 1) = \sum_{\sigma} x^{E(\sigma)} = \sum_i a_i x^i$$

which defines the coefficients  $a_i$ . The coefficients  $a_i$  are called the density-of-states and constitute the very foundation of this article. Computing these exactly can be frustratingly hard but recently we have managed to obtain them for the  $15 \times 15$  lattice in two variables,<sup>(9)</sup> for the  $256 \times 256$  lattice in one variable,<sup>(15)</sup> and in 3 dimensions for the  $4 \times 4 \times 4$  lattice in two variables and for the  $5 \times 5 \times 5$  lattice in one variable. For larger lattices we have yet to resort to some form of sampling method. For now, we will focus on the one-variable case.

Using  $Z$  we can find the classical thermodynamic quantities for our model. As usual define the *free energy* and the *free energy per vertex* as

$$F = \frac{1}{K} \log Z, \quad f = \frac{1}{n} F.$$

The *internal energy* and *internal energy per vertex* is

$$U = \frac{\partial}{\partial K} \log Z, \quad u = \frac{1}{n} U.$$

The *specific heat* and *specific heat per vertex* is

$$C = K^2 \frac{\partial^2}{\partial K^2} \log Z, \quad c = \frac{1}{n} C.$$

From the density-of-states we define the entropy (or log-density-of-states)

$$S(v) = \frac{\log a_i}{n}$$

where  $v = i/m$ . Since the  $a_i$  are only defined for certain energies, e.g.,  $i$  in the sequence ..., -8, -4, 0, 4, 8, ... for a graph of even degree, we make this function defined on the whole interval  $[-1, 1]$  by using some form of interpolation (usually piecewise linear).

From the derivative of the entropy we obtain the temperature. We will, however, prefer to talk about the coupling, or inverse temperature,  $K$ . As usual, define

$$K(v) = \frac{-n}{m} S'(v).$$

An alternative way to define this quantity is to use the maximum-term method. Suppose  $x$  is a number such that

$$a_{i-k} x^{i-k} \leq a_i x^i \geq a_{i+k} x^{i+k}$$

at some energy  $i$ , where  $k$  is the difference between two consecutive energies. That is,  $a_i x^i$  is the maximum term in the partition function  $Z(G; x)$ . Then

$$\frac{a_{i-k}}{a_i} \leq x^k \leq \frac{a_i}{a_{i+k}},$$

from which it follows, as an aside, that  $a_{i-k} a_{i+k} \leq a_i^2$ , i.e., the sequence is log-concave at energy  $i$ . Setting  $x = e^K$  we get

$$\frac{1}{k} \log \frac{a_{i-k}}{a_i} \leq K \leq \frac{1}{k} \log \frac{a_i}{a_{i+k}}, \quad (2)$$

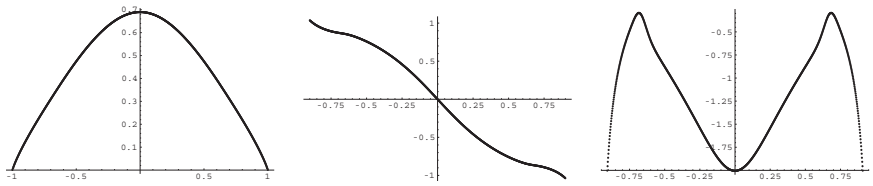


Fig. 1. Plot of  $S(v)$ ,  $S'(v)$ , and  $S''(v)$  for the  $32 \times 32$  lattice.

i.e.,  $K$  is a number in an interval, which is assumed to narrow as our system grows. To make the coupling properly defined we associate the upper bound with energy  $i+k/2$  and use linear interpolation for any other energy. Then the following holds:

$$\begin{aligned}
 S' \left( \frac{i+k/2}{m} \right) &= \frac{S \left( \frac{i+k}{m} \right) - S \left( \frac{i}{m} \right)}{k/m} = \frac{m}{kn} (\log a_{i+k} - \log a_i) \\
 &= -\frac{m}{kn} \log \frac{a_i}{a_{i+k}} = -\frac{m}{n} K \left( \frac{i+k/2}{m} \right)
 \end{aligned}$$

and so our chosen definition follows. Note especially that the coupling is a function of the energy though it is not necessarily one-to-one.

The objective of our sampling technique is to provide data on the first derivative of the entropy.

### 3. SOME THEORY

At this point we will construct a graph  $\mathfrak{G}$  with the states  $\sigma$  of  $G$  as vertices. Let  $\Omega_i$  be the set of states at energy  $i$ . Say that  $\sigma \in \Omega_i$  and  $\tau \in \Omega_j$ . We draw an edge between  $\sigma$  and  $\tau$  if  $\tau$  is the state obtained by flipping a single spin of  $\sigma$ . We are going to count the number of edges between  $\Omega_i$  and  $\Omega_j$ . Given a state  $\sigma$  let  $\eta(\sigma, \ell)$  be the number of vertices having local energy  $\ell$ . Also, let  $\rho(i, \ell)$  be the probability that a vertex has local energy  $\ell$  at energy  $i$ , that is,

$$\rho(i, \ell) = \frac{1}{na_i} \sum_{\sigma \in \Omega_i} \eta(\sigma, \ell).$$

**Lemma 3.1.** The total number of edges in  $\mathfrak{G}$  with one endpoint in  $\Omega_i$  and one endpoint in  $\Omega_{i+2\ell}$  is

$$\sum_{\sigma \in \Omega_i} \eta(\sigma, -\ell) = \sum_{\tau \in \Omega_{i+2\ell}} \eta(\tau, \ell). \tag{3}$$

*Proof.* Flipping a spin having local energy  $- \ell$  changes the energy by  $2\ell$  while reversing the sign of the local energy. The sum over all states at energy  $i$  must be the number of edges between  $\Omega_i$  and  $\Omega_{i+2\ell}$ . An analogous argument holds at energy  $i+2\ell$  so the sums must be equal. Note that the relation is true also for  $\ell = 0$  though the flips will only result in a change of magnetisation while the energy remains unchanged. ■

**Theorem 3.2.** The following relation holds:

$$a_i \rho(i, -\ell) = a_{i+2\ell} \rho(i+2\ell, \ell).$$

*Proof.* We simply manipulate Eq. (3) as follows:

$$a_i \frac{1}{na_i} \sum_{\sigma \in \Omega_i} \eta(\sigma, -\ell) = a_{i+2\ell} \frac{1}{na_{i+2\ell}} \sum_{\tau \in \Omega_{i+2\ell}} \eta(\tau, \ell)$$

and using the definition of  $\rho$  the theorem follows. ■

An immediate consequence is then

**Corollary 3.3.**

$$S' \left( \frac{i+\ell}{m} \right) = -\frac{m}{2\ell n} \log \frac{\rho(i+2\ell, \ell)}{\rho(i, -\ell)}.$$

We thank one of the referees, who pointed out that this was shown also in ref. 13 and used in ref. 12.

## 4. SAMPLING

With the previous section in mind, a sampling process should collect statistics on the probabilities for each local energy, i.e., we want an estimate  $r(i, \ell)$  of  $\rho(i, \ell)$ . We suggest the following scheme:

1. Reset histogram:  $h(i, \ell) \leftarrow 0$  for all  $i$  and  $\ell$ .
2. Generate a state  $\sigma$  using, e.g., the Metropolis or Wolff method.
3. Compute the energy:  $i \leftarrow E(\sigma)$ .
4. For each vertex  $v$  do
5.  $\ell \leftarrow \delta(v)$
6.  $h(i, \ell) \leftarrow h(i, \ell) + 1$
7. End for
8. If more samples are needed then go back to (2), otherwise stop.

When we are satisfied with the quality of our data we note that

$$\rho(i, \ell) \approx r(i, \ell) = \frac{h(i, \ell)}{\sum_{\ell} h(i, \ell)}$$

and from this we can calculate  $S'$  and all quantities derived therefrom. In order to simplify our further discussion we will often rather use the quantity  $R_{\ell}$ , which is just a normalised form of  $r$ , i.e.,  $R_{\ell}(i/m) = r(i, \ell)$ . To make  $R_{\ell}$  defined on the whole interval  $[-1, 1]$ , rather than just points on the form  $i/m$ , we will assume that linear interpolation is used. Note that  $\sum_{\ell} R_{\ell}(v) = 1$  for each energy  $v$ .

#### 4.1. Some Details

We pause here for a moment to clarify some details. While the reader may have the  $L \times L \times L$  lattice at the back of his head when reading this article, the arguments should hold for any family of graphs having finite bounded degree. The 2D square lattices are 4-regular so the local energy can take the values  $0, \pm 2, \pm 4$  while for the 6-regular 3D cubic lattices the local energy can also take the value  $\pm 6$ . In general, if the graph has maximum degree  $\Delta$  then the local energy can only take the values  $\Delta, \Delta-2, \Delta-4, \dots, -\Delta$ . To simplify our argument we will assume that we are working with graphs of even degree. Flipping a spin can change the energy by at most  $\pm 2\Delta$ . In connection with Eq. (2) above we defined the coupling as

$$K\left(\frac{i+k}{m}\right) = \frac{1}{2k} \log \frac{a_i}{a_{i+2k}},$$

where  $2k$  was considered the difference between two consecutive energies, giving us a sequence of points where the coupling is defined. We extend this definition by removing the ‘‘consecutive energies’’ condition. Doing this should be reasonable given that we have limited ourselves to graphs of finite degree; this ensures that the relative energy change when flipping a spin goes to zero as  $n$  goes to infinity. Not only have we one sequence of couplings, but  $\Delta/2$  such. That all these sequences approach an asymptotic function  $K(v)$  is here taken as an axiom. From Theorem 3.2 it then follows that

$$K\left(\frac{i+\ell}{m}\right) = \frac{1}{2\ell} \log \frac{\rho(i+2\ell, \ell)}{\rho(i, -\ell)}$$

for  $\ell = 2, 4, \dots, \Delta$ . We will refer to them individually by subscribing them with an  $\ell$ , as in  $K_{\ell}$ .

## 4.2. The Sequence

After having acquired the sequences  $K_\ell$ , the next problem is to mold them into one final sequence denoted  $K^*$ . There are several approaches to this and our suggestion should be thought of as just this: a suggestion. For a different solution, see ref. 12.

First we define the function  $K_\ell^*$  as

$$K_\ell^*(v) = \frac{1}{2\ell} \log \frac{R_\ell(v + \ell/m)}{R_{-\ell}(v - \ell/m)}$$

for  $\ell > 0$ , and finally we define

$$K^*(v) = \frac{1}{1 - R_0(v)} \sum_{\ell > 0} (R_\ell(v) + R_{-\ell}(v)) K_\ell^*(v),$$

which we will use as the estimate of  $K(v)$ .

## 5. CONSISTENCY

How do we know if our sampled data is good? The neat thing with this scheme is that there are built-in consistency checks. If our data is bad then we can see it! The samples have returned estimates of the various  $\rho(i, k)$ . By plotting certain quotients of them we can judge their quality. As stated in Theorem 3.2 we have

$$\frac{a_i}{a_{i+2\ell}} = \frac{\rho(i+2\ell, \ell)}{\rho(i, -\ell)}.$$

Let us now suppose that our graph is 4-regular, that is, the possible values of  $\ell$  are 0,  $\pm 2$ ,  $\pm 4$ . We then have

$$\frac{a_i}{a_{i+4}} = \frac{\rho(i+4, 2)}{\rho(i, -2)},$$

$$\frac{a_i}{a_{i+8}} = \frac{\rho(i+8, 4)}{\rho(i, -4)}.$$

On the other hand we have

$$\frac{\rho(i+8, 4)}{\rho(i, -4)} = \frac{a_i}{a_{i+8}} = \frac{a_i}{a_{i+4}} \frac{a_{i+4}}{a_{i+8}} = \frac{\rho(i+4, 2)}{\rho(i, -2)} \frac{\rho(i+8, 2)}{\rho(i+4, -2)}. \quad (4)$$



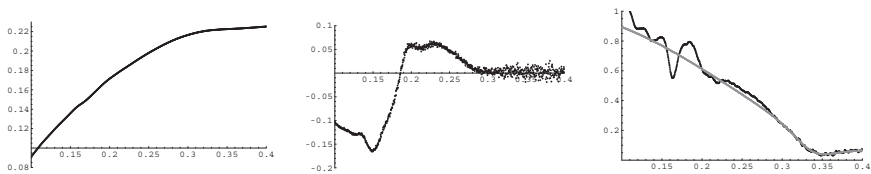


Fig. 2. Plot of  $\frac{1}{4} \log \frac{r(i+4,2)}{r(i,-2)}$ ,  $\psi(i)$  and the derivative of  $\frac{1}{4} \log \frac{r(i+4,2)}{r(i,-2)}$  together with a correctly sampled derivative for the  $16 \times 16 \times 16$  cube.

If we let  $\psi(i)$  be the difference between the left- and right-hand side of this equation for our estimates  $r$ , then plotting  $\psi$  will reveal many kinds of inconsistencies in the sampled data.

**Example 5.1.** Let us look at an example. In order to produce some bad data we ran a Metropolis program on the 16-cube in the following way. We swept through the cube in typewriter fashion using the usual Metropolis rule to update each spin. Each sweep was continued until we had managed to actually flip  $16^3/2$  spins. Then we made a measurement and started a new sweep. Any experienced sampler will probably be wincing by now.

In the leftmost part of Fig. 2 we see our approximation of  $K(v)$  from these data. Basically the curve looks good apart from the slight dip at  $v \approx 0.17$ , but this could very well be explained by sampling noise. In the middle part of the figure we see our indicator  $\psi(i)$ , which is clearly not concentrated around zero for low values of  $v$ , thus indicating that our samples are bad. In the right part of the figure we see a plot of the derivative of our approximation together with a correct sampling. Here we clearly see a systematic deviation from the correct curve in the interval indicated as bad by our middle plot.

This method of sampling is of course rather obviously flawed, but in the same manner more subtle algorithmic errors will be caught, thus giving us further confidence that we are actually sampling according to the proper Boltzmann distribution.

## 6. COMMENTS AND RELATIONS TO OTHER METHODS

Normally we require  $R_\ell$  for a range of consecutive energies. Using, e.g., the Metropolis method as the engine for data production, we may need to sample at several different temperatures to obtain data for the desired window of energies. This is of course not a problem since we may simply add the histograms, i.e., the  $h(i, \ell)$ , from the different temperatures.

Observe that it does not matter how the data are generated as long as the distribution of  $R_\ell$  at each particular energy is correct. This was also realised by de Oliveira<sup>(12)</sup> in the so-called Broad Histogram Method.

What we propose is therefore not so much a sampling method as a book-keeping scheme for obtaining more information from our hard-earned data. This is at the relatively small cost of adding that extra loop in our algorithm above, and as will be seen later we can expect this extra complexity to be compensated for by a reduction in critical slowing down. We are thus free to choose among the most efficient state generation methods available to us. For smaller systems this will typically be a Metropolis algorithm and for large systems we may turn to one of the cluster algorithms.

Though the approach is very different from that of, e.g., the single histogram method<sup>(16,17)</sup> there is not necessarily any conflict in combining both methods while we are at it. We just have to keep in mind that the single histogram method relies on sampled data from one well-chosen temperature only and yields data at other temperatures by a transformation procedure that zooms in on the corresponding spot in the histogram. A good comparison between different methods can be found in ref. 18 along with numerous references.

A method which is clearly related to ours is the Transition Matrix Monte Carlo (TMMC) method of Wang and Swendsen.<sup>(18)</sup> One major difference between the methods is that in the TMMC method one does not separate the data analysis from the state generation as we do. Indeed, given our sampled data one can reconstruct the results of a run of the TMMC method for any choice of flip rate dependent on only the local and total energies which correctly samples the Boltzmann distribution. For a generalised TMMC based on a non-local flip rate this will not be the case. Our method can in a sense be seen as the underlying rationale for the spectrum of local flip rate TMMC methods. Another example of the close relation between the methods is that Eq. (4) already appears in ref. 18, in the form of an identity for Markov chain transition probabilities.

## 7. CRITICAL SLOWING DOWN AND CRITICAL FLUCTUATIONS

When sampling close to the critical temperature, or energy, one typically encounters two phenomena which severely hamper sampling of different interesting quantities, *critical fluctuations* and *critical slowing down*. We now wish to examine these two phenomena in the context of our sampling scheme and demonstrate that even though the method cannot remove them it can reduce their effect significantly.

A third aspect, the increase in equilibration time, is not alleviated by our scheme. Thus we must at first obtain our equilibrated system by the usual long equilibration run and whatever model-specific methods that might be available to speed this part up.

## 7.1. Critical Fluctuations

When sampling the Ising model at a fixed temperature at, or close to,  $T_c$  one finds that many properties, such as the magnetisation and energy, gain enormously in variance, thus increasing the number of samples needed in order to achieve a desired accuracy. Let us examine the structure of this variance.

Assume that we can somehow generate a series of completely independent samples of some property  $M$  at a fixed temperature  $T$ . Having the magnetisation and energy of our graph in mind is a good guide for our thoughts here.

If we consider  $M$  as a random variable we can think of it as produced by a two-step procedure: first we choose an energy  $i$  at random from the energy distribution  $\Pr(i|T)$  at temperature  $T$  and then we choose a random value of  $M$  from its distribution  $\Pr(M|i)$  at energy  $i$ .

Let us calculate the mean and variance of  $M$ .

$$\begin{aligned} E(M|T) &= \int M \Pr(M|T) dM = \int M \left( \int \Pr(M|i) \Pr(i|T) di \right) dM \\ &= \int \Pr(i|T) \left( \int M \Pr(M|i) dM \right) di = \int E(M|i) \Pr(i|T) di \\ &= E(E(M|i)|T)). \end{aligned}$$

So we find that the mean of  $M$  given  $T$  is just the mean of  $M$  at each energy weighted by the probability for the energies, the “mean of the means.” Nothing unexpected here.

Next we look at the variance.

$$\begin{aligned} V(M|T) &= E(M^2|T) - E^2(M|T) = \int M^2 \Pr(M|T) dM - E^2(M|T) \\ &= \int M^2 \left( \int \Pr(M|i) \Pr(i|T) di \right) dM - E^2(M|T) \\ &= \int \Pr(i|T) \left( \int M^2 \Pr(M|i) dM \right) di - E^2(M|T) \end{aligned}$$

$$\begin{aligned}
&= \int \Pr(i | T) E(M^2 | i) di - E^2(M | T) \\
&= \int \Pr(i | T) E(M^2 | i) di - \int \Pr(i | T) E^2(M | i) di \\
&\quad + \int \Pr(i | T) E^2(M | i) di - E^2(M | T) \\
&= \int \Pr(i | T) (E(M^2 | i) - E^2(M | i)) di \\
&\quad + \int \Pr(i | T) E^2(M | i) di - E^2(M | T) \\
&= \int \Pr(i | T) (E(M^2 | i) - E^2(M | i)) di + E(E^2(M | i) | T) - E^2(M | T) \\
&= \int \Pr(i | T) (V(M | i)) di + E(E^2(M | i) | T) - E^2(E(M | i) | T) \\
&= \underbrace{E(V(M | i) | T)}_{E_g} + \underbrace{V(E(M | i) | T)}_{V_g}.
\end{aligned}$$

So we see that the variance has two parts,  $E_g$ —the mean of the variances and  $V_g$ —the variance of the means, and this is where it gets interesting.

Let us first examine  $V_g$ . If our system undergoes a phase transition at a temperature  $T_c$  we usually expect some observable  $M$  to have one distribution when  $T < T_c$  and another quite different distribution when  $T > T_c$ . This means that for a temperature close to  $T_c$  our system will tend to spend part of the time near one value of  $M$  and part of the time at a completely different value, depending on whether the system is at a high or low energy. This will make  $V_g$  large since the expected value of  $M$  varies a lot at such  $T$  and no matter how large the system is the variance at  $T_c$  will tend to remain large, perhaps even grow. However this is completely due to the fact that one has combined values of  $M$  from several different energies, and so our way of bookkeeping will not suffer from this problem.

The other part of the variance,  $E_g$ , comes from the variance of  $M$  within each energy level. If we are at an energy different from  $E_c$ , the mean energy at  $T_c$ , this variance will actually decrease as the system grows larger, typically as some power of the system size, and even for  $E_c$  it will be bounded.

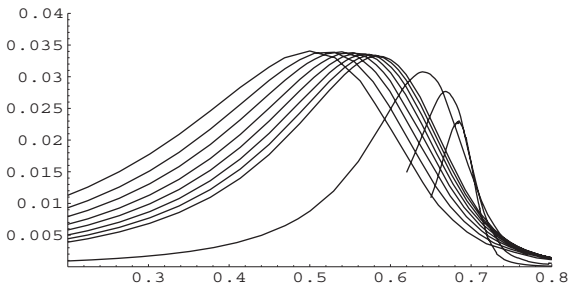


Fig. 3. The variance of the magnetisation as a function of energy for the  $L$  by  $L$  square lattice.  $L=8, \dots, 15, 32, 64, 128$ .

**Example 7.1.** In Fig. 3, taken from ref. 9, we can see how the variance within a fixed energy level develops for the modulus of the magnetisation of square lattices of increasing side. The smaller lattices have been computed exactly and the three larger lattices are from Monte Carlo simulations. We can clearly see here that the variance decreases with increasing  $n$  for all energies, even at  $E_c$ . The exact results for the magnetic susceptibility of the square lattice would lead us to believe that the maximum variance decays as  $L^{-4}$ . In Fig. 4 we see the corresponding curves for the simple cubic lattice.

So we see that of the two components of the critical fluctuations that classical fixed-temperature Monte Carlo suffers from, only one,  $E_g$ , affects our way of doing things and in fact this component will diminish when the size of the system grows.

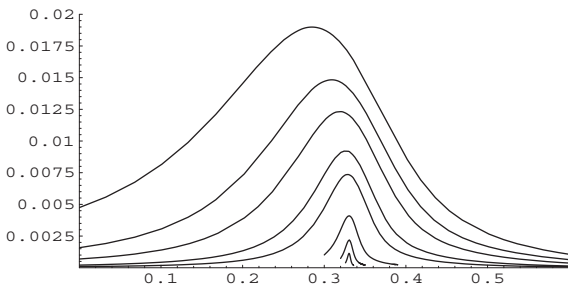


Fig. 4. The variance of the magnetisation as a function of energy for the simple cubic lattice of side  $L$ .  $L = 4, 6, 8, 12, 16, 32, 64, 128$ .

## 7.2. Critical Slowing Down

When running a simulation at a fixed temperature we start with a system in some state  $S_0$  and use some procedure  $P(S)$  (Metropolis algorithm, Wolff cluster algorithm or whatever our favourite might be) to generate a new state  $S_1 = P(S_0)$  from  $S_0$ , then a state  $S_2 = P(S_1)$  and so on. For a given simulation method  $P$  there is a number  $\tau(T)$  such that  $S_\tau$  is considered sufficiently independent from the initial state  $S_0$  in order to allow  $S_0$  and  $S_\tau$  to be used as independent samples (for a fuller discussion of this see, for example, ref. 7). This number  $\tau(T)$  is called the *correlation time* of our simulation and is the victim of our second crippling phenomenon, critical slowing down.

Many properties will gain in variance as the temperature gets closer to  $T_c$  and so the natural thing to do would of course be to collect a larger number of samples when we are close to  $T_c$ , in order to reduce the error in our estimates. However in order to get independent samples we must always run our simulation for a time period which is at least the correlation time before we make a new sample, and right at  $T_c$  the correlation time of many sampling methods tends to grow dramatically. This sudden increase in correlation time is known as *critical slowing down*.

In order to reduce the effect of critical slowing down a number of methods have been developed to replace the basic Metropolis algorithm near the critical temperature. Most well known are probably the Wolff cluster algorithm and the Swendsen–Wang algorithm, both of which reduce the correlation time compared to the Metropolis significantly. However they are also significantly more complex than the Metropolis algorithm and so are only of practical advantage for a large enough system, but for really large systems they can once more be at a disadvantage due to their higher demand for memory. The good performance of the Swendsen–Wang algorithm is also restricted by the number of spin values. As shown in ref. 19 there is a  $q_0(d)$  such that for the  $q$ -state Potts model on  $\mathbb{Z}^d$  with  $q \geq q_0(d)$  the Swendsen–Wang algorithm has a mixing (correlation) time which is exponential in the number of vertices. The exact value of  $q_0$  is, to the authors's knowledge, not yet known, but is conjectured to be equal to the smallest  $q$  such that the  $d$ -dimensional Potts model has a first-order phase transition.

So far this discussion of the correlation time has been focused on temperature. How does it behave in our setting? It is of course equally important for us to get independent samples of the properties that we wish to study; however what we need is not independent samples at a given temperature but rather at a given energy, and this is a fundamental difference. Even though we might have generated two states  $S_1$  and  $S_2$  which are

clearly correlated and thus unsuitable for sampling in the temperature setting, this will not cause us trouble as long as they are at sufficiently different energies. In the Ising model on a graph with maximum degree  $\Delta$  an energy difference  $2\Delta + 1$  will suffice when we are sampling local energies, and for magnetisation we only require that we end up at different energies.

Let us take a more classical example of this idea. Let us assume that we are studying the Ising model on some lattice and that we are interested in three quantities, the energy, the magnitude of the magnetisation, and the cluster size. Further assume that we are working at a temperature at which the correlation time corresponds to three Metropolis sweeps of our lattice. Now rather than doing three sweeps between each measurement we can do one sweep and then measure the energy, one further sweep and then measure the magnetisation, one further sweep and measure the cluster size, one sweep and then measure energy, and so on. In this way we always keep an amount of time corresponding to the correlation time between successive measurements of a given quantity, even though consecutive measurements are made on correlated states.

Returning to the present setting, what we need to ensure is that within a time  $\tau$  we do not expect to take two samples at the same energy. Thus we want to find a modified correlation time  $\tau'(T)$  such that if we sample at time intervals  $\tau'(T)$  the probability of finding two samples at the same energy in a time less than  $\tau(T)$  is sufficiently small.

First of all we need to find out how many samples we can make at a given temperature before we expect to find two samples at the same energy. The relevant quantity here is the so-called return time of the system. The *return time*  $r(E, T)$  of an energy  $E$  is the expected time before a system starting in a state with energy  $E$  and temperature  $T$  returns to a state with energy  $E$ , and the minimum return time,  $r_{\min}(T)$ , is the minimum value of the return time given  $T$ , that is,

$$r_{\min}(T) = \min_E r(E, T).$$

If we make sure that we do not take more than  $r_{\min}(T)$  samples during a time period of length  $\tau(T)$  we can expect our samples to be sufficiently independent. This gives  $\tau'(T)$  as

$$\tau'(T) = \frac{\tau(T)}{r_{\min}(T)}.$$

So we are led to consider the behaviour of  $r_{\min}(T)$  for  $T$  close to  $T_c$ . Finding  $r_{\min}(T)$  exactly is probably as hard as, or harder than, solving our

underlying model, but under rather weak assumptions about energy distribution for our model we can get a useful approximation. From the theory of Markov chains we know that we can express  $r_{\min}(T)$  as

$$r_{\min}(T) = \min_E (\Pr(E|T))^{-1},$$

and so we want to estimate the highest probability for any energy given at a given  $T$ . If the energy distribution  $\Pr(E|T)$  is reasonably close in shape to a normal distribution we find that  $r_{\min}$  is proportional to the standard deviation of the energy distribution. For the particular case of the Ising model we can thus express  $r_{\min}$  in terms of the specific heat as

$$r_{\min}(T) \propto (T^2 C(T))^{\frac{1}{2}} = TC(T)^{\frac{1}{2}}.$$

Observe that we are using the specific heat, not the specific heat per vertex.

So for a system where the specific heat diverges at  $T_c$  we expect  $\tau'$  to either diverge significantly slower than  $\tau$  or even converge to a finite value.

**Example 7.2.** Let us look at the Ising model on the 3-dimensional cubic lattice. First we look at some Monte Carlo data from a Metropolis simulation of a cubic lattice with side 64 and cyclic boundary conditions. In Fig. 5 we have plotted the inverse minimum return time as a function of  $K$ .

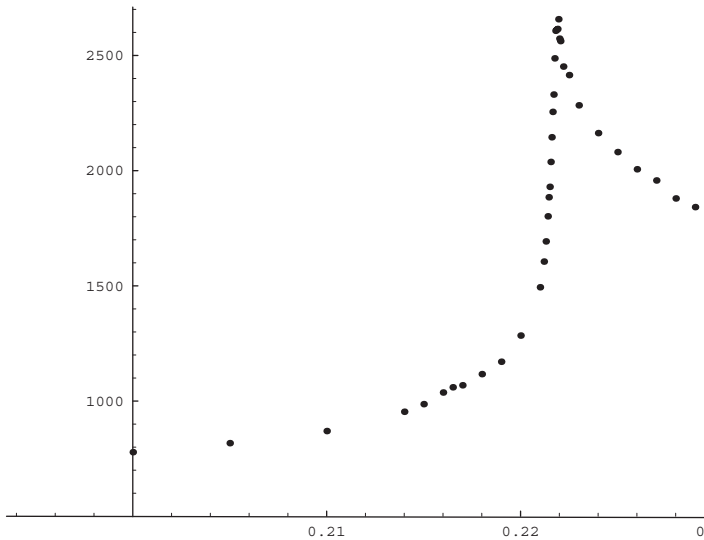


Fig. 5. The inverse minimum return time as a function of  $K$  for the Ising model on the 64-cube.



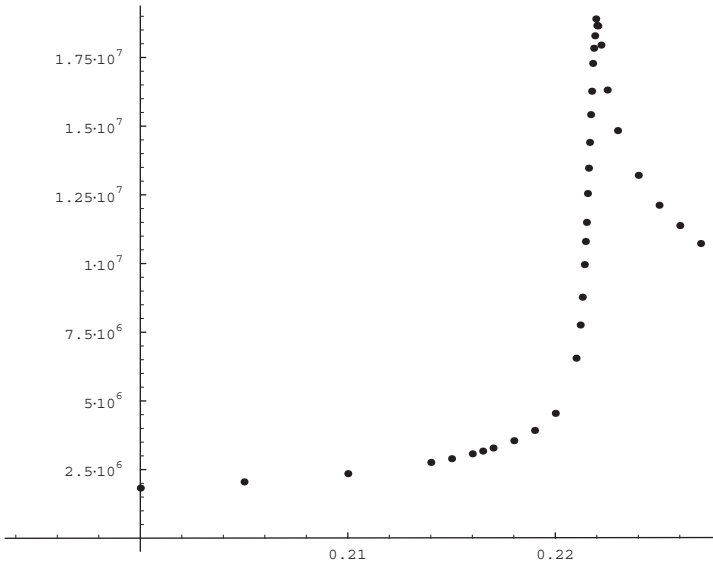


Fig. 6. The variance of the energy as a function of  $K$ .

We can see a strong increase in inverse return time as we get closer to  $K_c$ , and also that the return time overall is larger for  $K > K_c$ .

In Fig. 6 we have the variance of  $E$  and in Fig. 7 we see the rescaled quotient between  $r_{\min}(K)^2$  and  $V(E|K)$ . The quotient stays bounded around 3 and is consistent with  $r_{\min}$  being proportional to  $C^{\frac{1}{2}}$ . In Figs. 8 and 9 we see the corresponding quotients with  $r_{\min}(K)^{1.5}$  and  $r_{\min}(K)^{2.5}$ , both indicating that these exponents do not match the observed behaviour. For this short Monte Carlo run, around 20 CPU-hours<sup>4</sup> per value of  $K$  after equilibration, the curve is quite noisy.

According to the standard simulation literature we have that  $\tau \propto \xi^z$ , where  $\xi$  is the correlation length and  $z$  the so-called dynamic exponent of the algorithm. For the Metropolis algorithm in 3 dimensions we have that  $z = 2.08\dots$  (see ref. 20). For a finite lattice with side  $L$  the correlation length is bounded by  $L$  and so we find that  $\tau \propto L^z$ . Similarly one finds that

$$c(T) \propto \xi^{\frac{\alpha}{\nu}},$$

where  $c$  is the specific heat per vertex,  $\alpha$  is the critical exponent for  $c$  and  $\nu$  is the critical exponent for the correlation length. As above,  $\xi$  is bounded by  $L$  and so

$$c \propto L^{\frac{\alpha}{\nu}}.$$

<sup>4</sup> A 160 MHz pwr2 processor was used.

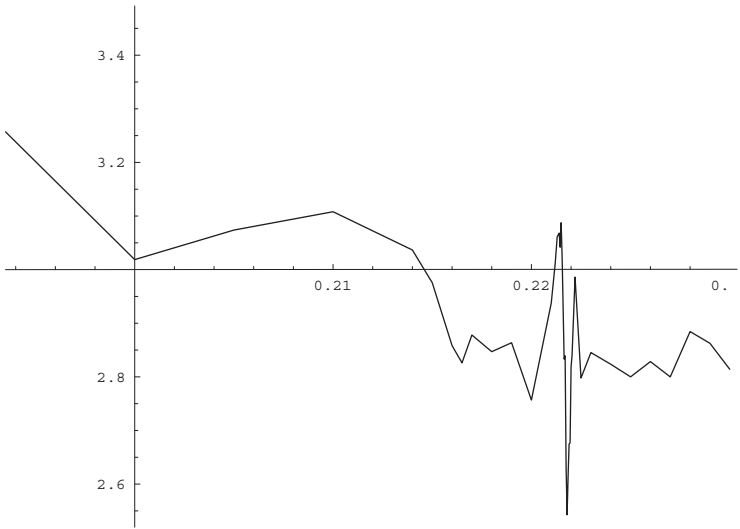


Fig. 7. The quotient of  $r_{\min}(K)^2$  and  $V(E|K)$ . The curve stays bounded in the vicinity of 3 as we get closer to  $K_c$ , but sampling noise is quite strong near  $K_c$  for this short Monte Carlo run.

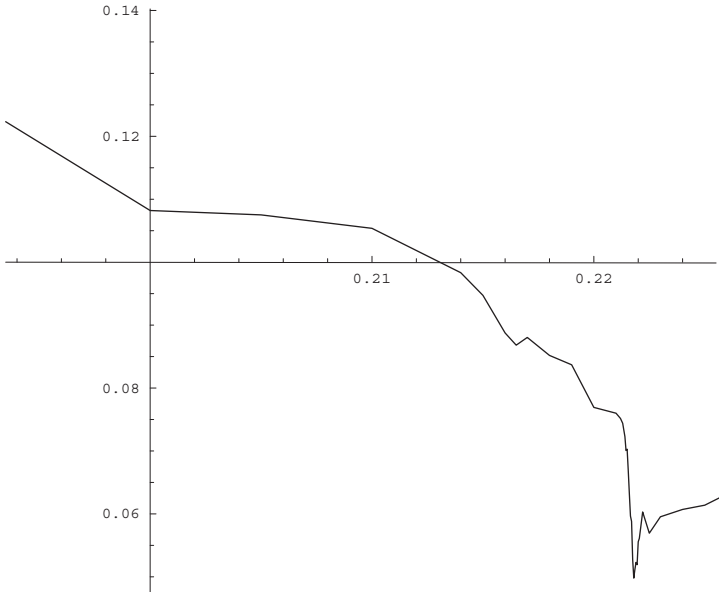


Fig. 8. The quotient of  $r_{\min}(K)^{1.5}$  and  $V(E|K)$ . The curve has a clear trend towards 0 as we get close to  $K_c$ .

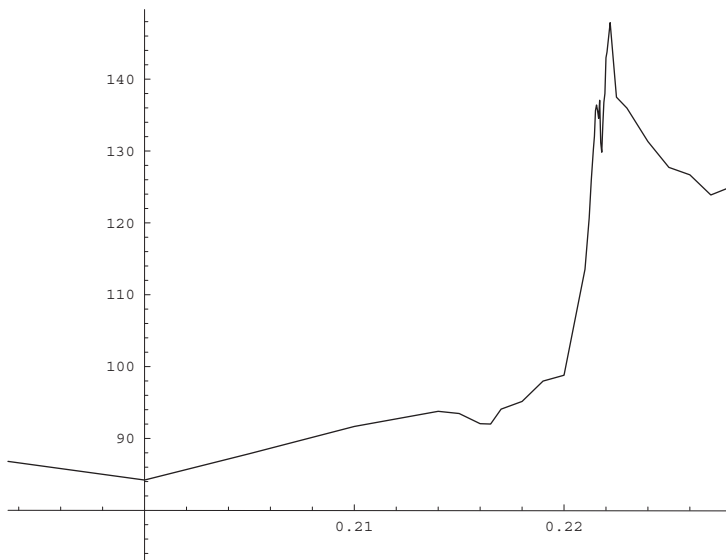


Fig. 9. The quotient of  $r_{\min}(K)^{2.5}$  and  $V(E|K)$ . The quotient begins to diverge as we get closer to  $K_c$ .

The specific heat is just  $C(T) = L^d c(T)$ , where  $d$  is the dimension of the lattice; in our case  $d = 3$ . This gives us

$$\tau' \propto \frac{\tau}{C^{\frac{1}{2}}} \propto \frac{L^z}{L^{\frac{d+\frac{\alpha}{\nu}}{2}}} = L^{\frac{2z-d-\frac{\alpha}{\nu}}{2}}.$$

Given that  $\alpha \approx 0.110$ ,  $\nu \approx 0.63$ , and  $d = 3$ , we find that  $\tau' \propto L^{0.49}$  and so we still expect a slowing-down effect for the Metropolis algorithm, however significantly less so than before. Given that we have ignored many constants here and the simplicity of the Metropolis algorithm we expect it to be faster in practice than many of the cluster-based algorithms with lower values of  $z$ , unless the lattices are extremely large.

Another fact worth noting is that as the dimension increases the exponent on  $L$  will tend to decrease, and for the Ising model we actually expect it to become negative in high enough dimensions.

So we see that focusing on energy allows one to worry less about the increasing correlation time near  $T_c$  and use a fast and simple algorithm for all but the largest lattices. However this does not come without a cost. Due to the increased variance in  $E$  we will get fewer hits per time unit on any

given energy when we get closer to  $T_c$  and we will thus have to sample for a longer time before we have a useful number of samples at each energy. But since the variance of many properties is significantly smaller within an energy level than at the corresponding temperature, as seen in the previous discussion, this is not necessarily as troublesome as it may seem at first. This point is exemplified in ref. 11, where as few as 10(!) samples per energy level on a large number of levels turn out to give very nice results.

## 8. DISCUSSION AND EXAMPLES

In this section we will look at two quick examples where the sampling scheme may prove itself. The first example, the 2D Ising model, will reveal little beyond what is already well-known, apart from the usefulness of the algorithm. The second example, however, the 5-state 2D Potts model, should demonstrate clearly what we may achieve. Before we proceed, a short discussion of what we *cannot* achieve is necessary.

A quantity commonly studied at, or near,  $K_c$  is the specific heat per vertex  $c(K)$ . For example, it is well-known that  $c(K_c) \propto \log L$  for the 2D square Ising model and, most likely, grows as  $L^z$  for some unknown exponent  $z \geq 0$  in the 3D case. To obtain this quantity we need estimates of the complete sequence of coefficients  $a_i$ .

How can we obtain the specific heat  $c(K)$  from our data? Well, we cannot necessarily do that. Note that what we have is

$$K^*(v) \approx K(v) = \frac{1}{k} \log \frac{a_i}{a_{i+k}}, \quad v = \frac{i+k/2}{m},$$

where  $i$  and  $i+k$  are non-empty consecutive energy levels. That is, we have only information on the quotients  $a_i/a_{i+k}$ . Of course, if our sampled data include very high energy levels where we know the exact value of  $a_i$ , we may recover the entire sequence of  $a_i$ . However, if our target system is, e.g., a cube of linear order 256 (for which we have used the method<sup>(10)</sup>) this will indeed be an arduous task since it requires collecting data from some  $10^7$  energy levels. We should mention that the method described in ref. 21 suffers from the same problem and the authors offer the same conclusions.

On the other hand, the specific heat may not be as holy a grail as it at first may seem. Consider instead the following derivation

$$c(K) = K^2 \frac{\partial u}{\partial K} = \frac{K^2}{\partial K / \partial u} = \frac{m K^2(v)}{n K'(v)} = \frac{-(S'(v))^2}{S''(v)},$$

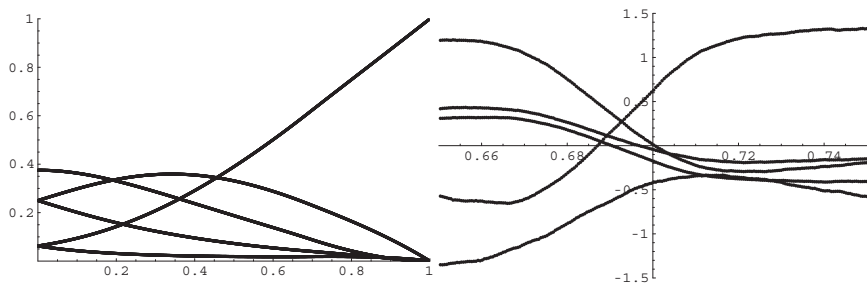


Fig. 10. The  $64 \times 64$ -lattice. Plots of  $R_\ell$  and  $R_\ell''$  versus  $\nu$ , for  $\ell = 0, \pm 2, \pm 4$ .

where  $K = K(\nu)$ . This holds only asymptotically and when  $K(\nu)$  is one-to-one. Since  $K(\nu)$ , and consequently  $S'(\nu)$ , is well-behaved, even in the critical region, we could then rather focus on  $K'(\nu)$  or  $S''(\nu)$ . The order with which  $K'$  goes to zero is thus directly related to how fast the specific heat grows to infinity. Working in this way we can achieve asymptotic results for quantities parameterised by  $K$ , or  $T$ , as well by combining the asymptotics for  $K(\nu)$  and  $S'(\nu)$ .

## 8.1. The Ising Model

Our first example will be the Ising model on the simple square  $64 \times 64$ -lattice. The model is solved exactly for the infinite case and we also have exact data to rely upon for this particular case, i.e., we have the actual coefficients  $a_i$ , see ref. 15.

Recall that for the infinite case  $K_c = \log(\sqrt{2} + 1)/2 \approx 0.440687$  and  $\nu_c = 1/\sqrt{2} \approx 0.707107$ . About  $3 \cdot 10^6$  samples per energy level were collected in the critical region. Three couplings were used, with the main run on  $K_c = 0.4406\dots$ , and two short runs at one value of  $K$  on each side of  $K_c$ . The total runtime was 10 CPU days on a 333 MHz PowerPC-604.

Figure 10 shows a plot of the  $R_\ell$  functions and their second derivatives. Had we compared  $K$  and  $K^*$  there would be no discernible difference, even without any smoothing of data. In Fig. 11 we plot instead the first and second derivative of  $K^*$ . Data are smoothed only slightly to let some noise remain. The thin line is based upon exact data for the  $64 \times 64$ -lattice.

## 8.2. The Potts Model

In our second example we look at the 5-state Potts model on the  $L \times L$  square lattice. Here the spins can take the values 1, 2, 3, 4, 5 and the energy of a state is defined as  $E(\sigma) = \sum_{uv} \delta(\sigma_u, \sigma_v)$ , where  $\delta$  is the Kronecker delta

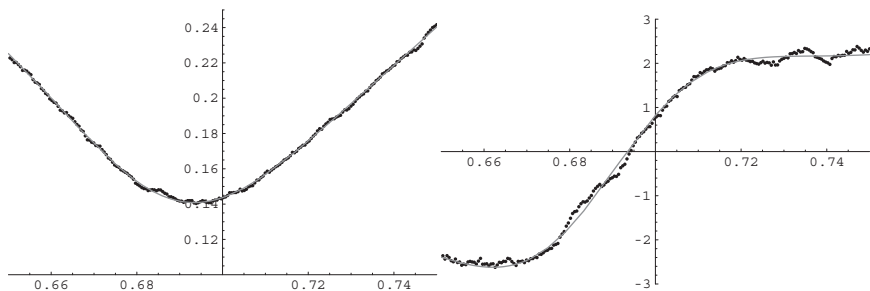


Fig. 11. The Ising model on the  $64 \times 64$ -lattice. First and second derivative of  $K$  (thin line) and  $K^*$  versus  $\nu$ .

and the sum is, as usual, taken over all edges. The energy can take any integer value in  $[0, m]$ , rather than just multiples of 4. Also, the energy can change by  $0, \pm 1, \pm 2, \pm 3, \pm 4$  when flipping a spin to a different state. We define  $\rho(i, \ell)$  as the probability for changing the energy by  $\ell$  when flipping a spin to a different value. Thus we have

$$\frac{a_i}{a_{i+\ell}} = \frac{\rho(i+\ell, -\ell)}{\rho(i, \ell)}.$$

With  $R_\ell(i/m)$  defined as before, i.e., an estimate of  $\rho(i, \ell)$ , we then have

$$K_\ell^*(\nu) = \frac{1}{\ell} \log \frac{R_{-\ell}(\nu + \ell/2m)}{R_\ell(\nu - \ell/2m)},$$

which we, as before, mold into a single  $K^*$ . The critical coupling for the  $q$ -state model is, quoting from ref. 22,

$$K_c = \log(1 + \sqrt{q}).$$

At this point a jump in the energy takes place between  $\nu^-$  and  $\nu^+$ . The average of these we will refer to as the critical energy

$$\nu_c = \frac{\nu^- + \nu^+}{2} = \frac{1}{2} \left( 1 + \frac{1}{\sqrt{q}} \right)$$

and according to ref. 23 the latent heat for  $q \geq 5$ , i.e., their difference, is

$$\nu^\pm = \nu^+ - \nu^- = \left( 1 + \frac{1}{\sqrt{q}} \right) \tanh \frac{\theta}{2} \prod_{k=1}^{\infty} \tanh^2 k\theta,$$

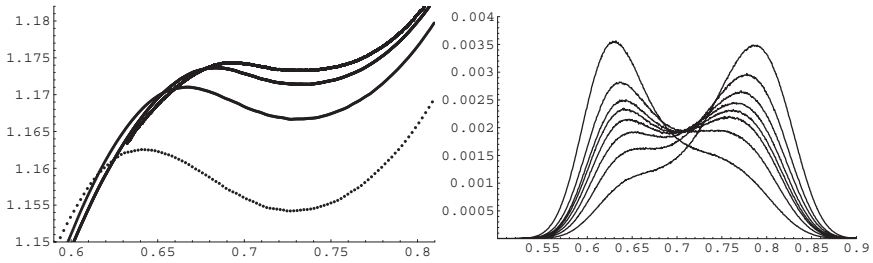


Fig. 12. Left:  $K(v)$  for  $L = 16, 32, 64, 128$ . Right: energy distribution for  $L = 32$  at  $K = 1.165, 1.167, 1.168, 1.1685, 1.169, 1.170, 1.171, 1.173$ .

where  $2 \cosh \theta = \sqrt{q}$ . For  $q = 5$  this gives  $K_c \approx 1.1744$ ,  $v_c \approx 0.7236$  and  $v^\pm \approx 0.02646$ . Also we have  $v^- \approx 0.7104$  and  $v^+ \approx 0.7368$ . Let us see if the sampling scheme at hand can provide some verification of this.

A short computer run was made for demonstrative, rather than analytical, purposes. In the critical region we collected roughly  $3 \cdot 10^6$  measurements per energy level for  $L = 8$ ,  $10^6$  for  $L = 16$ ,  $3 \cdot 10^5$  for  $L = 32$ ,  $10^5$  for  $L = 64$  and only  $10^4$  for  $L = 128$ . The runtime was about 30 CPU days for  $L = 128$ , about 15 CPU days for  $L = 64$ , using a 333 Mhz PowerPC-604. Three temperatures were used and in this case it is really necessary to use more than one temperature in order to get enough samples over the entire energy interval of interest.

Note the dip in the curve  $K(v)$  in the critical region in the left plot of Fig. 12. This is a clear indication that we are looking at latent heat. Indeed, as the right plot shows, we see a bimodal distribution of energies for some values of  $K$  near  $K_c$ , and this is the true culprit. Before we continue, we state the following definitions:

$$\begin{aligned}
 v^- &= \min \{v : K'(v) = 0\}, \\
 v^+ &= \max \{v : K'(v) = 0\}, \\
 v^\pm &= v^+ - v^-,
 \end{aligned}$$

where  $v^\pm$  will be our definition of latent heat. Our samples suggested the data in Table I, where the error should be in the last decimal.

Let us return to the subject of energy distributions. Looking, for example, at the distribution for  $K = 1.1685$ , where the peaks are about equal in height, we have a maximum at  $v = 0.641$  and  $v = 0.761$  and in between lies a minimum at  $v = 0.700$ . Looking at the left plot again we see that the three energies where  $K(v) = 1.1685$  are 0.642, 0.700, and 0.761 respectively. Considering the fairly low precision in our data this is as close

**Table I. Critical Data for 5-State 2D Potts Model**

$L$	$v^-$	$v^+$	$K(v^-)$	$K(v^+)$
8	0.606	0.730	1.1374	1.1225
16	0.642	0.728	1.1625	1.1543
32	0.667	0.732	1.1710	1.1667
64	0.683	0.734	1.1736	1.1714
128	0.693	0.733	1.1743	1.1733

to a perfect match as makes no difference. Compare this with the quite irrelevant average energy 0.709.

The distributions are clearly bimodal for  $K = 1.168, 1.1685, 1.169$  and  $1.170$ . They are clearly *not* bimodal for  $1.165$  and  $1.173$ . This is less clear for  $1.167$  and  $1.171$ . A closer scrutiny reveals that the distribution is actually slightly bimodal for  $K = 1.167$  but it is not clear which category  $1.1710$  falls into. Ideally, the distributions should be bimodal for  $K(v^+) < K < K(v^-)$  and since  $K(v^-) = 1.1710$  we are obviously looking at a case in point.

In the left plot of Fig. 13 we plot  $v^+$  and  $v^-$  versus  $L^{-2/3}$ . The straight lines were fitted to the data for the three largest lattices. These are  $0.710 - 0.435x$  for  $v^-$  and  $0.734 - 0.020x$  for  $v^+$ , giving a spot-on candidate for  $v^-$  but a deviation from the correct  $v^+$  in the third decimal. Due to this our latent heat will be asymptotically  $v^\pm = 0.024$  and  $v_c = 0.722$ , not entirely off their correct values.

The right plot of Fig. 13 shows the couplings at the energies  $v^-$  and  $v^+$  versus  $L^{-1}$ . Unfortunately they seem to scale with different exponents though if we choose the exponents  $5/3$  and  $4/3$  respectively, then we get very nice fits. The curves are given by  $1.1746 - 1.193x^{5/3}$  and  $1.1748 - 0.834x^{4/3}$  and fit very well to the data points. The asymptotic values differ

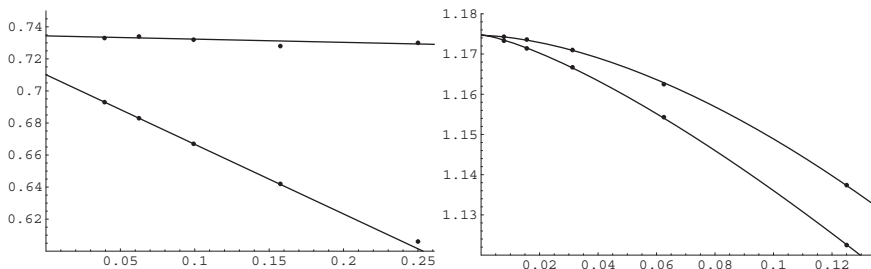


Fig. 13. Left: energies  $v^-$  and  $v^+$  vs  $L^{-2/3}$ . Right: couplings  $K(v^-)$  and  $K(v^+)$  vs  $L^{-1}$ .



slightly in the fourth decimal from each other and from the correct value  $K_c \approx 1.1744$ , but what is to be expected from three decimals in the input?

## ACKNOWLEDGMENTS

This work was supported by the Swedish Natural Sciences Research Council, The Göran Gustafsson Foundation, and the National Graduate School in Scientific Computing (NGSSC). We have used the computing facilities at HPC2N Umeå, and PDC Stockholm.

## REFERENCES

1. E. Ising, Beitrag zur Theorie des Ferromagnetismus, *Z. Physik* **31**:253–258 (1925).
2. L. Onsager, Crystal statistics I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**:117–149 (1944).
3. B. A. Cipra, An introduction to the Ising model, *Amer. Math. Monthly* **94**:937–959 (1987).
4. N. Metropolis, A. Rosenblut, M. Rosenbluth, A. Teller, and E. Teller, Equation of state calculation by fast computing machines, *J. Chem. Phys.* **21**:1087–1092 (1953).
5. U. Wolff, Collective Monte Carlo updating for spin systems, *Phys. Rev. Lett* **62**:361 (1989).
6. R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
7. M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (The Clarendon Press Oxford University Press, New York, 1999).
8. E. H. Lieb and Jakob Yngvason, The physics and mathematics of the second law of thermodynamics, *Phys. Rep.* **310**:1–96 (1999).
9. R. Häggkvist and P. H. Lundow, The Ising partition function for 2D grids with cyclic boundary: Computation and analysis, *J. Stat. Phys.* **108**:429–457 (2002). See also <http://www.math.umu.se/~phl>.
10. Roland Häggkvist, Anders Rosengren, Daniel Andrén, Petras Kundrotas, Per Håkan Lundow, and Klas Markström, *On the 3-Dimensional Ising Model*, manuscript.
11. P. H. Lundow and Klas Markström, *A New Look at the Phase Transition of the 5-Dimensional Ising Model*, manuscript.
12. P. M. C. de Oliveira, Broad histogram simulation; Microcanonical Ising dynamics, *Int. J. Mod. Phys. C* **3**:497–503 (1998).
13. B. A. Berg and U. H. E. Hansmann, Configuration space for random walk dynamics, *Eur. Phys. J. B* **6**:395–398 (1998).
14. J.-S. Wang and R. H. Swendsen, Transition matrix Monte Carlo method, *J. Stat. Phys.* **106**:245–285 (2002).
15. Roland Häggkvist, Anders Rosengren, Daniel Andrén, Petras Kundrotas, Per Håkan Lundow, and Klas Markström, *Exact Calculation of the Ising Partition Function for Large Square Grids*, manuscript.
16. A. M. Ferrenberg and D. P. Landau, Critical behavior of the three-dimensional Ising model: A high-resolution Monte Carlo study, *Phys. Rev. B* **44**:5081–5091 (1991).
17. D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, 2000).

18. J.-S. Wang and R. H. Swendsen, *Transition Matrix Monte Carlo*, arXiv:cond.mat/0104418 v1, 23 Apr 2001 (2001).
19. C. Borgs, J. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. Vu, *Torpid Mixing of Some mcmc Algorithms in Statistical Physics*, Proceedings of FOCS '99 (1999), pp. 218–229. Preprint available at <http://www.math.cmu.edu/~af1p/papers.html>.
20. F. Wang and M. Suzuki, Finite-size scaling properties of the damage distance and dynamical critical exponent for Ising model, *Physica A* **220**:534–541 (1995).
21. F. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, *Phys. Rev. Let.* **86**:2050–2053 (2001).
22. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, 1982).
23. R. J. Baxter, Potts model at the critical temperature, *J. Phys. C* **6**:L445–L448 (1973).