Exact sampling from nonattractive distributions using summary states

Andrew M. Childs,[1] Ryan B. Patterson,[1] and David J. C. MacKay[2]
[1]*Physics Department, California Institute of Technology, Pasadena, California 91125*
[2]*Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
(Received 16 May 2000; published 22 February 2001)

Propp and Wilson's method of coupling from the past allows one to efficiently generate *exact* samples from attractive statistical distributions (e.g., the ferromagnetic Ising model). This method may be generalized to nonattractive distributions by the use of *summary states*, as first described by Huber. Using this method, we present exact samples from a frustrated antiferromagnetic triangular Ising model and the antiferromagnetic $q=3$ Potts model. We discuss the advantages and limitations of the method of summary states for practical sampling, paying particular attention to the slowing down of the algorithm at low temperature. In particular, we show that such slowing down can occur in the absence of a physical phase transition.

DOI: 10.1103/PhysRevE.63.036113 PACS number(s): 05.50.+q, 02.70.Rr, 02.50.Ga

I. INTRODUCTION

In many statistical problems, physical and otherwise, it is useful to be able to draw samples from a complex distribution. For example, in statistical physics one is interested in the Boltzmann distribution

$$P(\sigma)=\frac{e^{-\beta E(\sigma)}}{Z}, \tag{1}$$

where $E(\sigma)$ describes the energy of a system in configuration σ, β is the inverse temperature (we set $k_B=1$), and Z is a normalizing constant (the partition function). In general, $E(\sigma)$ may be easy to evaluate for a particular configuration, but the number of possible configurations makes it impractical to draw directly from the distribution. Yet some efficient method of sampling is desirable, as this would allow one to calculate properties of the system that might not be easily computed by analytical means.

In traditional Monte Carlo sampling methods [1], such as the Metropolis-Hastings method [2] and Gibbs sampling [3] (also known as the heat bath algorithm), one constructs an ergodic Markov chain whose stationary distribution is the desired distribution. By starting in some state and evolving the chain for a sufficiently long time, one can approximate a sample from the desired distribution. Unfortunately, such a sample is exact only in the limit of infinite time. In practice, it is often difficult to determine how long to wait to achieve sufficiently good samples, and one inevitably either produces poor samples or wastes time by running the Markov chain for longer than necessary.

However, in 1996, Propp and Wilson demonstrated the possibility of *exact sampling* by the method of coupling from the past, allowing one to produce perfect samples in a finite number of steps [4]. In the most general case, their method requires the infeasible task of running a Markov chain for every possible initial state of the system. But for certain distributions, termed *attractive* (such as a ferromagnetic Ising model), Propp and Wilson showed that the task may be greatly simplified by tracking only extremal states, permitting the practical calculation of exact samples. This method was generalized to antiattractive distributions by Häggström and Nelander [5]. More recently, Huber showed that one can instead track just a single state that summarizes one's knowledge of the system [6]. Because it does not require that the states be partially ordered, this last method—which we call the *summary state method*—is applicable to nonattractive distributions.

Using the summary state method, we have drawn exact samples from the antiferromagnetic triangular Ising model and from the three-state Potts antiferromagnet on a square lattice. Figure 1 shows one such sample. In Sec. II, we describe the methods that make this possible. In Sec. III, we briefly discuss the Ising and Potts models. We present results from the exact sampling of these models in Sec. IV. Finally, we discuss the convergence properties of the summary state method, and we suggest practical generalizations.

II. COUPLING FROM THE PAST AND THE SUMMARY STATE METHOD

Propp and Wilson's method of coupling from the past is based on the observation that, for a fixed choice of the random numbers used to propagate a Markov chain, its possible paths in state space from different initial states may ulti-

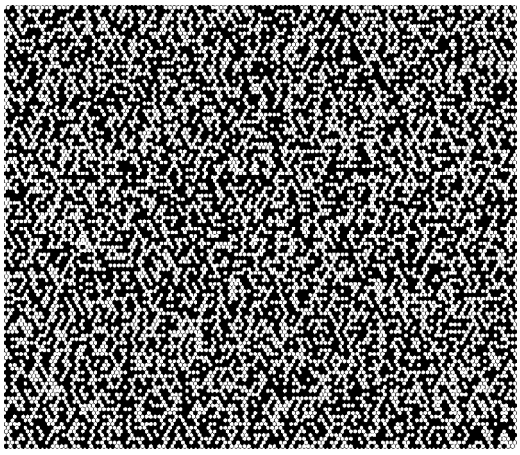

FIG. 1. An exact sample from a triangular Ising antiferromagnet with 14 400 spins at $\beta=4.9^{-1}\approx 0.2041$ and zero applied magnetic field.

mately coalesce into a single trajectory. Once two initial states lead to the same state, they will remain in the same state thereafter.

Consider simulating a Markov chain from every possible initial state at some fixed time $t=-T$, with the goal of taking a sample at $t=0$. If all the chains coalesce before $t=0$, then this finite procedure yields the same results as a Monte Carlo simulation started at an infinite time in the past, so the result is an exact sample. If the chains fail to coalesce, one can simply double the starting time to $-2T$, reusing the random numbers for the interval $[-t,0]$ (i.e., treating the random numbers as a function of simulation time), and repeat until coalescence is achieved.

Having to follow every possible state would make this method exponentially costly. But for problems that admit a partial ordering of the states and which are ''attractive''—that is, which preserve the ordering under evolution of the Markov chain—the computation can be vastly simplified by tracking only the extremal states. An example of an attractive system is the ferromagnetic Ising model under single-spin-flip (Glauber) dynamics.

Huber [6] and Harvey and Neal [7] have shown that the method of Propp and Wilson may be generalized using a single summary state instead of a pair of extremal states. This single state summarizes one's knowledge of the possible states of the system, allowing the state of some subsystems to be uncertain.

For example, suppose the system is a collection of variables σ_i taking on the values $\{\pm 1\}$. Conventional single-site heat bath updating sets

$$\sigma_i \mapsto \begin{cases} +1 & \text{if } u \leqslant P(\sigma_i=+1|\overline{\sigma_i}) \\ -1 & \text{if } u > P(\sigma_i=+1|\overline{\sigma_i}), \end{cases} \tag{2}$$

where u is uniformly distributed on $[0,1]$ and $\overline{\sigma_i}$ denotes the set of all variables but the ith. To implement summary states, we allow each variable to take on the additional value ? which indicates uncertainty. We then run a modified Markov chain on this system: σ_i is updated according to Eq. (2) if the result is the same for any possible assignment of ± 1 to the ?'s in $\overline{\sigma_i}$; otherwise, $\sigma_i \mapsto ?$. As in the Propp and Wilson method, we run the chain from successively longer times in the past with random numbers as a function of simulation time, starting from an initial state that is entirely ?. When no variables remain in ? states, the algorithm has converged, and we may take a sample at $t=0$.

For the case of attractive distributions, this procedure is exactly equivalent to the Propp and Wilson scheme. The value ? denotes variables that differ between the maximal and minimal states, and removal of all ? states corresponds to coalescence of the bounding chains. However, using a single summary state, there is no requirement that the states be ordered in any way. Thus the summary state method can also be applied to nonattractive distributions—for example, the antiferromagnetic Ising model.

Although the samples returned by this method are exact, the algorithm does not necessarily converge after a reasonable amount of time. Huber has shown that, for antiferromagnetic spin systems at sufficiently high temperature, the expected running time of the algorithm is polynomial in the number of spins [6]. However, for systems with a phase transition, the convergence time diverges as a power law at the critical temperature, a phenomenon known as critical slowing down [8].

For the attractive case, Propp and Wilson showed that the convergence time of coupling from the past is linear in the mixing time of the Markov chain, so that there is a sense in which there is no additional cost for producing exact samples [4]. Thus, for example, coupling from the past based on the heat bath algorithm should diverge no faster at the critical temperature than the heat bath algorithm alone. However, this is not necessarily the case for the more general summary state algorithm—indeed, we will show that it is possible for the algorithm to diverge at a strictly higher temperature.

III. THE ISING AND POTTS MODELS

Consider the Hamiltonian

$$E(\sigma) = -\frac{1}{2}\sum_{m,n} J_{mn}\sigma_m\sigma_n - \sum_m H_m\sigma_m, \tag{3}$$

where J_{mn} is the coupling between spins m and n and H_m is the value of an external magnetic field at spin m. The appropriate Markov chain update rule is Eq. (2) with

$$P(\sigma_i=\pm 1|\overline{\sigma_i}) = \frac{e^{-\beta E(\sigma_i=\pm 1)}}{e^{-\beta E(\sigma_i=+1)}+e^{-\beta E(\sigma_i=-1)}}. \tag{4}$$

In the Ising model [9], J_{mn} is taken to be zero unless spins m and n are adjacent, in which case it is some constant J. Cases of particular interest are the square lattice, in which every spin has four neighbors, and the triangular lattice, with six neighbors per spin. In general, the behavior of Ising systems can vary with their spin connectivity. For both kinds of lattice, we use periodic boundary conditions.

In this paper, we use the normalization $J=\pm 1$. $J=+1$ corresponds to the ferromagnetic case, in which spins prefer to point in the same direction; $J=-1$ corresponds to the antiferromagnet. As mentioned previously, the ferromagnetic case is attractive under single-spin-flip dynamics. The antiferromagnet on a square lattice is a special case, because its properties are isomorphic to those of a square ferromagnet. However, for a triangular lattice, there is no such isomorphism. With six neighbors per spin, there is no way to minimize the energy locally at all sites: we say the system is *frustrated*.

It is well known that a two-dimensional ferromagnetic Ising model exhibits a phase transition [10,11]. Below a critical temperature β_c^{-1}, there is spontaneous symmetry breaking, and the system develops a preferred spin orientation in the absence of any magnetic field. For a square lattice, $\beta_c = \frac{1}{2}\ln(1+\sqrt{2}) \approx 0.440\,687$. At this temperature, the relaxation time of the dynamic system diverges, a phenomenon known as critical slowing down [8]. Correspondingly, there is a divergence in the convergence time for some Markov chain Monte Carlo algorithms, such as coupling from the

past, and exact samples cannot be generated for lower temperatures. Note that the triangular antiferromagnet has a critical point at zero temperature [12], so we expect critical slowing down only as $\beta \to \infty$.

To circumvent the problem of nonconvergence below the critical temperature, Propp and Wilson actually used a related system, the Fortuin-Kasteleyn random cluster model, to generate ferromagnetic Ising samples [4,13]. Unfortunately, this model has no obvious analog in the antiferromagnetic case.

The Potts model is a generalization of the Ising model wherein spins may take on q different values $\{0,1,\ldots,q-1\}$ [14]. Spins interact only with others of the same type. The Hamiltonian is

$$E(\sigma) = -\frac{1}{2}\sum_{m,n} J_{mn}\delta_{\sigma_m,\sigma_n} - \sum_{m,k} H_m^k \delta_{\sigma_m,k}. \tag{5}$$

Specifically, we consider the antiferromagnetic Potts model with $q=3$ on a square lattice with zero magnetic field. Like the triangular Ising antiferromagnet, this model also has a critical point at zero temperature [15].

IV. RESULTS

A. Ising model

By implementing the summary state method, we have produced exact samples from the Ising and Potts models. For example, Fig. 1 shows a sample from a triangular Ising antiferromagnet consisting of $120^2 = 14\,400$ spins at $\beta = 4.9^{-1} \approx 0.2041$ with zero applied magnetic field.

We find that the number of iterations required for the algorithm to converge, where one iteration consists of updating each spin in the lattice, diverges at a threshold temperature. We have studied this divergence using a lattice of $N = 63^2 = 3969$ spins. Simulations using larger N (e.g., $N = 99^2$) suggest that the outcome is not significantly affected by choosing a larger grid size. Figure 2 shows the divergence, to which we have fitted a power law of the form

$$t = \frac{a}{(\beta^{-1} - \beta_t^{-1})^b} + c. \tag{6}$$

We find that the time diverges with an exponent $b = 1.03 \pm 0.01$ at the threshold temperature $\beta_t^{-1} = 4.839 \pm 0.005$ (or $\beta_t = 0.2067 \pm 0.0002$, corresponding to a correlation length of $\xi_t \approx 1.6$ lattice spacings, using the relationship $\xi = -\ln\tanh\beta$ which can be inferred from the analysis of Stephenson [16]).

This divergence is an important feature of the summary state method. It is qualitatively similar to critical slowing down, but note that the threshold temperature is above the critical temperature, so the divergence does not correspond to a physical phase transition. In divergent situations the augmented Markov chain has a metastable set of distributions with many ?'s, such that it is very unlikely for it to enter a state with no ?'s.

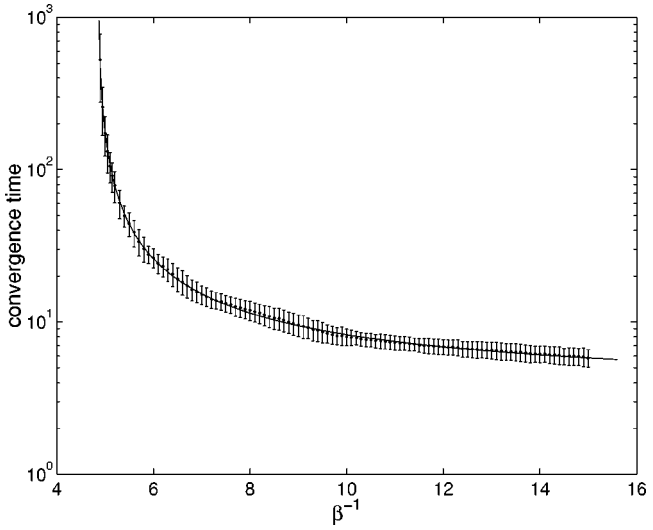

FIG. 2. Variation with temperature of the number of iterations required for convergence of the summary state algorithm for the triangular, antiferromagnetic Ising model. Each point corresponds to either 500, 1000, or 1500 exact samples, with more samples taken at lower temperature. The solid line shows the fit to Eq. (6).

We can arrive at a very rough estimate of the threshold temperature based on a simplified description of the dynamics. To draw an exact sample using the summary state method, the system must go from a completely uncertain state to a completely certain state. Thus, it must pass through a state with only a few scattered ?'s. For temperatures sufficiently near the threshold, where we know that such a sparse configuration *can* be reached, we might expect that the limiting factor is the probability that an isolated ? can cause divergence.

Therefore, we might suppose that the divergence occurs when the probability of a single ? turning one of its six neighbors into a ? rises above $\frac{1}{6}$. We expect that the neighbors of any given spin σ_i should be (on average) half up and half down. Replacing one of these neighbors by a ?, we may assume the configuration $(\uparrow\uparrow\uparrow\downarrow\downarrow?)$ without loss of generality. Then the threshold temperature is determined by

$$1 - \frac{1}{1+e^{4\beta}} - \frac{1}{2} = \frac{1}{6}, \tag{7}$$

which has the solution $\beta = \frac{1}{4}\ln 2 \approx 0.17$.

To examine the validity of a threshold temperature analysis based on the persistence of single ?'s, we compiled statistics on the stability of an equilibrium system with a single ? added. Because we cannot create exact samples for much of the temperature range of interest, we generated approximate samples by simulating for fixed time (100 iterations) a random (infinite temperature) initial state. We then set one spin to ? and simulated the system forward. If any uncertainty remained after 500 iterations, we said the system diverged. Figure 3 shows the fraction of divergent trials for various temperatures. As one would expect, this fraction goes to zero very near the threshold temperature.

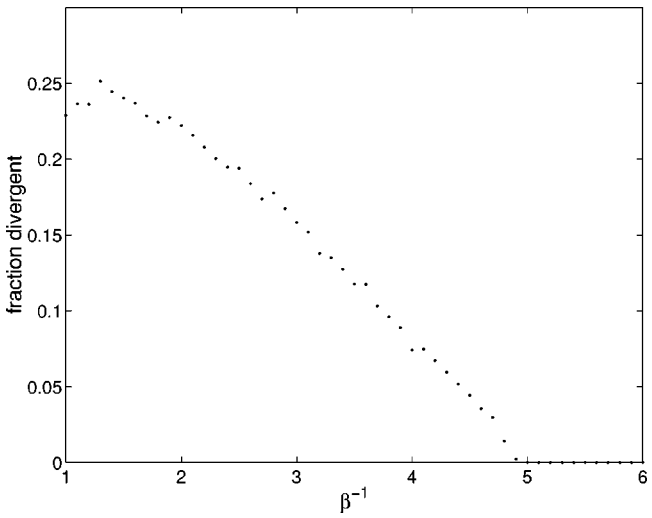

FIG. 3. Fraction of equilibrium systems that diverged after a single ? was added. The data are from 9000 trials at each temperature.

It is also interesting to consider how the algorithm behaves when a uniform nonzero magnetic field H is applied. Biasing the spins makes it easier for them to choose a particular orientation, so we would expect convergence to be easier. Figure 4 shows the region of convergence in the (β^{-1},H) plane.

B. Potts model

In addition, we have implemented exact sampling of the Potts model for arbitrary q. Figure 5 shows an exact sample with $q=3$ for a square antiferromagnetic lattice of $100^2=10\,000$ spins.

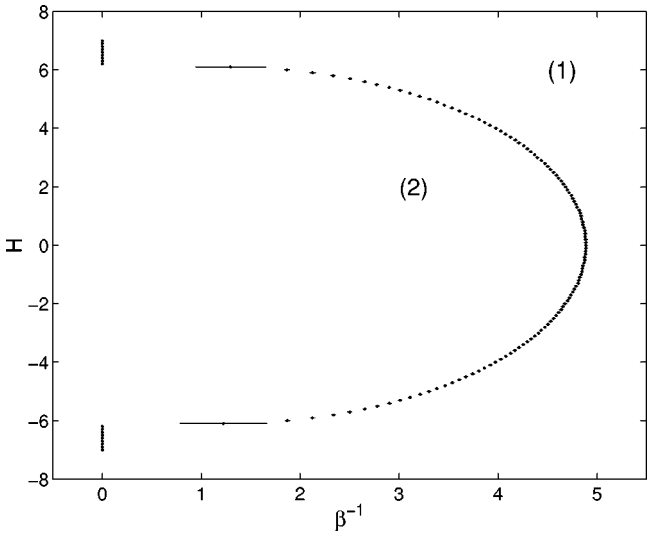

FIG. 4. Region of convergence for the summary state algorithm on the triangular antiferromagnetic Ising model. Points in region (1) allow convergence, whereas points in region (2) are inaccessible to the algorithm. Each data point corresponds to 45 searches for the threshold, each using a different set of random numbers.

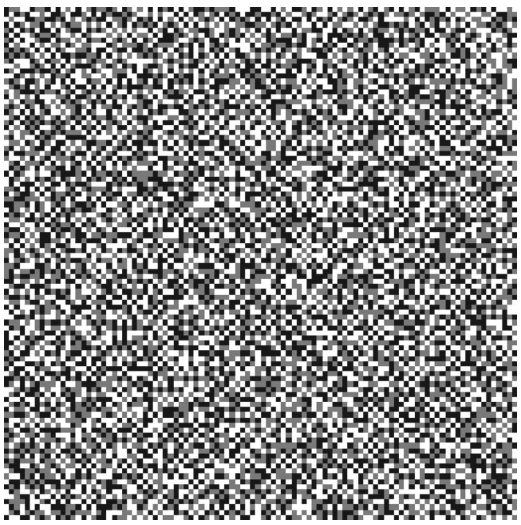

FIG. 5. An exact sample from the $q=3$ antiferromagnetic Potts model for a 10 000 spin square lattice at $\beta=1.2^{-1}\approx0.833$.

A naive implementation of the summary state method would augment the possible spin values with a single ?. We refer to this method as algorithm *A*. However, it is possible to retain more information about uncertain spins: for each spin, we store a binary q-bit vector $(b_1,b_2,\ldots,b_q)$, $b_i\in\{0,1\}$. Bit b_i is set to 1 if it is possible for the spin to take on the value i: thus the initial state of each spin is $b=(1,1,\ldots,1)$. In updating the state of the system, we set $b_i=0$ only when the spin cannot take on the value i for any allowed configuration of its neighbors. We refer to the latter method as algorithm *B*.

To demonstrate the advantage of retaining more information in the summary state, we have studied the convergence properties of both algorithms. This comparison is shown in Fig. 6, based on data for a square $64^2=4096$ spin lattice. As in the Ising study, both algorithms lead to a power law di-

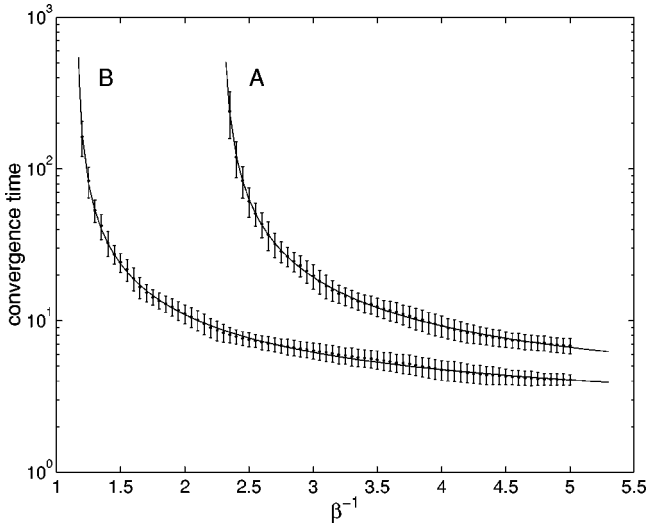

FIG. 6. Temperature dependence of the convergence time for the square, antiferromagnetic $q=3$ Potts model under algorithms *A* and *B*. Each point corresponds to 1600 exact samples on a 4096 spin lattice. The solid lines give fits to Eq. (6).

vergence with an exponent of 1 ($b_A = 1.04 \pm 0.03, b_B = 0.99 \pm 0.02$). However, the threshold temperatures for the two algorithms are quite different: $\beta_{t,A}^{-1} = 2.293 \pm 0.005$ (or $\beta_{t,A} = 0.436 \pm 0.001$, corresponding to a correlation length $\xi_{t,A} \approx 0.24$ lattice spacings, as extrapolated from the $\beta \lesssim 3.4$ behavior of ξ seen in the Monte Carlo data of Ferreira and Sokal [15]), whereas $\beta_{t,B}^{-1} = 1.157 \pm 0.004$ ($\beta_{t,B} = 0.864 \pm 0.003, \xi_{t,B} \approx 0.58$). As in the Ising example above, neither of the divergences corresponds to a physical phase transition.

V. CONCLUSIONS

We have demonstrated the use of the summary state method for exact sampling from nonattractive distributions [17]. However, the algorithm will require substantial improvement before it can be used for practical sampling. In both the antiferromagnetic Ising and Potts models, the method works only above a certain threshold temperature, with a power law divergence in the coalescence time at the threshold. Because most numerical studies are interested in the low-temperature behavior of these models—especially near the critical points at zero temperature—divergence of the running time of the algorithm presents a serious difficulty.

Although similar to the phenomenon of critical slowing down, the divergence does not occur at a physical phase transition. Furthermore, as the Potts example shows, the location of the divergence is a feature of the specific implementation of the summary states, not of the underlying distribution. We have shown that retaining more information in the summary state allows convergence at lower temperatures.

Based on this result, we may propose an improved algorithm for the triangular antiferromagnetic Ising model. At lower temperatures, the system should be increasingly ordered, and tracking this order might make it easier to gain incremental knowledge of the state of the system. One idea is to keep track of correlations between spins by grouping them into hexagonal clumps of seven, which can be used to tile the triangular lattice. Each tile has $2^7 = 128$ possible states. In analogy to the Potts method presented earlier (in which a q-bit vector represents the uncertainty about a spin), representing each tile with a 128-bit vector would allow individually tracking the possible arrangements of those seven spins. Within a tile, the summary state can track anticorrelation, which we expect to arise at low temperatures. Each edge of a tile can be easily summarized in a four-bit vector for comparison with its neighboring tiles. Each tile can then update its summary state by considering the possible states of the neighboring edges.

Despite the difficulties we have encountered in generating exact samples from antiferromagnetic spin systems, the technique of summary state sampling in general remains appealing as a way of producing provably exact samples. We hope that further study of the method will ultimately permit efficient exact sampling from thermodynamic systems at arbitrarily low temperatures.

ACKNOWLEDGMENTS

We wish to thank Radford Neal and David Wilson for several helpful discussions. D.J.C.M. is supported by the Gatsby Charitable Foundation and by IBM Zürich.

[1] For reviews of Monte Carlo sampling, see, for example, *Monte Carlo Methods in Statistical Physics*, 2nd ed., edited by K. Binder (Springer-Verlag, Berlin, 1979), 1986; *Applications of the Monte Carlo Method in Statistical Physics*, 2nd ed., edited by K. Binder (Springer-Verlag, Berlin, 1984), 1987; A.D. Sokal, in *Functional Integration: Basics and Applications*, Proceedings of the 1996 Cargèse Summer School, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci (Plenum, New York, 1997).

[2] N. Metropolis, A. W. Rosenbluth, N. M. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[3] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721 (1984).

[4] J. G. Propp and D. B. Wilson, Random Struct. Alg. **9**, 223 (1996).

[5] O. Häggström and K. Nelander, Stat. Neerland. **52**, 360 (1998).

[6] M. Huber, in *Proceedings of the 30th ACM Symposium on the Theory of Computing*, (ACM, New York, 1998), p. 31.

[7] M. Harvey and R. M. Neal, in *Uncertainty in Artificial Intelligence: Proceedings of the 16th Conference*, edited by C. Boutilier and M. Goldszmidt (Morgan Kaufmann, San Francisco, 2000), p. 256.

[8] P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).

[9] E. Ising, Z. Phys. **21**, 613 (1925).

[10] L. Onsager, Phys. Rev. **65**, 117 (1944).

[11] For a general overview of the theory of phase transitions and critical phenomena, see, for example, J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Clarendon Press, Oxford, 1992); N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, MA, 1992); C. Domb, *The Critical Point: A Historical Introduction to the Modern Theory of Critical Phenomena* (Taylor & Francis, London, 1996).

[12] G. H. Wannier, Phys. Rev. **79**, 357 (1950).

[13] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. **26**, 11 (1969); C. M. Fortuin and P.W. Kasteleyn, Physica (Amsterdam) **57**, 536 (1972); C. M. Fortuin, *ibid.* **58**, 393 (1972); **59**, 545 (1972).

[14] R. B. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1952); F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[15] S. J. Ferreira and A. D. Sokal, J. Stat. Phys. **96**, 461 (1999).

[16] J. Stephenson, J. Math. Phys. **11**, 413 (1970).

[17] Further information on exact Ising and Potts samples can be found at http://wol.ra.phy.cam.ac.uk/mackay/exact.