RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 40, NUMBER 10 1 OCTOBER 1989

Improved Monte Carlo distribution

P. B. Bowen,* J. L. Burke,* P. G. Corsten,* K. J. Crowell,* K. L. Farrell,* J. C. MacDonald,* R. P. MacDonald,* A. B. MacIsaac,* S. C. MacIsaac,* P. H. Poole,† and Naeem Jan
Physics Department, St. Francis Xavier University, Antigonish, Nova Scotia, Canada B2G 1CO
(Received 11 April 1989)

The Monte Carlo technique of Ferrenberg and Swendsen [Phys. Rev. Lett. **61**, 2635 (1988)] is improved by efficiently determining the tails of the Boltzmann distribution at the appropriate temperature. This is achieved by combining several distributions generated at different temperatures to form a composite distribution. The composite distribution leads to values of the specific heat and energy which are accurate over the entire temperature range of interest. Results illustrating these improvements are reported for the square two-dimensional Ising model.

INTRODUCTION

A recent Letter by Ferrenberg and Swendsen[1] (FS) reported a sophisticated method of extracting information over a wide range of temperature from the Monte Carlo data obtained at a single temperature (the critical temperature of the infinite Ising model). This method promises to significantly reduce the computational effort required to determine thermodynamic properties over a wide range of temperature. We were intrigued by the results reported. Figure 1 of Ref. 1 shows excellent agreement of the extrapolated specific heat with the exact results[2] in the vicinity of the critical temperature [$T_c = 2/\ln(1+\sqrt{2})$] but deviates somewhat at lower (~ 1.2) and higher temperatures (~ 2.8).

We review the essential elements of the method below and point out that inadequate sampling of the tails of the Boltzmann distribution is responsible for systematic deviations as the temperature is varied. Simulations of the order of 10^9 configurations would be necessary to obtain any data in these energy ranges. This error, or noise due to insufficient sampling in the extremities, can be reduced with modest computing efforts.

THE FS METHOD

A canonical system is described by

$$P_K(S) = \frac{e^{KS} N(S)}{Z}, \tag{1}$$

where

$$-\beta H = K \sum_{\langle ij \rangle} \sigma_i \sigma_j = KS. \tag{2}$$

K is the reduced coupling constant (J/kT), $\sigma_i = \pm 1$ is the spin value of the ith spin, Z is the partition function, $N(S)$ is the number of configurations with energy S, and $P_K(S)$ is the Boltzmann distribution giving the probability of finding the system with energy S.

FS focused on the histogram $N(S)$ produced during a Monte Carlo run at a specific temperature. They noted that since the configurations are generated with probability e^{KS}, $N(S)$ is therefore proportional to $P_K(S)$

$P_K(S)$ can be used to determine the distribution at any other temperature K' by

$$P_{K'}(S) = \frac{P_K(S) e^{(K'-K)S}}{\sum_S P_K(S) e^{(K'-K)S}}. \tag{3}$$

These equations are exact and errors, if any, are due to the determination of the histogram $N(S)$ by the Monte Carlo method. Figure 1 shows the energy and specific heat versus temperature for a two-dimensional system of linear length $L = 16$, determined from 100 K configurations generated by the standard Monte Carlo method.[3] Note that there are substantial deviations from the exact results as we move away from the temperature at which the histogram is generated. Figure 2 shows the histogram $N(S)$ generated at T_c (○) and also the histogram at a nearby higher temperature $T = 2.5$ (●) obtained from the T_c distribution and Eq. (3). The peak of this new distribution, which occurs at $S/2 = 135$, relies heavily on values of S which occur in the tail of the original distribution.[4] These values in the tail of the T_c distribution are more prone to noise, as $N(S)$ in these tails are small. While a small fluctuation caused by the Monte Carlo simulations would appear as an insignificant quantity in the regions near the peak of the histogram, these fluctuations represent large percentage errors in the extremities of the distribution. These errors would be magnified and transferred to the new distribution formed at $T = 2.5$. Thus the specific heat and average energy calculated at $T = 2.5$ from this error-ridden distribution would be less accurate than the corresponding values obtained at the temperature of the original distribution. Of course, these error effects would increase as one moved away from the temperature of the original Monte Carlo simulation.

A simple means of increasing the accuracy of the tails of the distribution is by generating more configurations. In fact, FS generated 4000 K configurations by the highly efficient Swendsen-Wang[5] algorithm and their results for the specific heat, although better than Fig. 1, still suffer somewhat from the same defect.

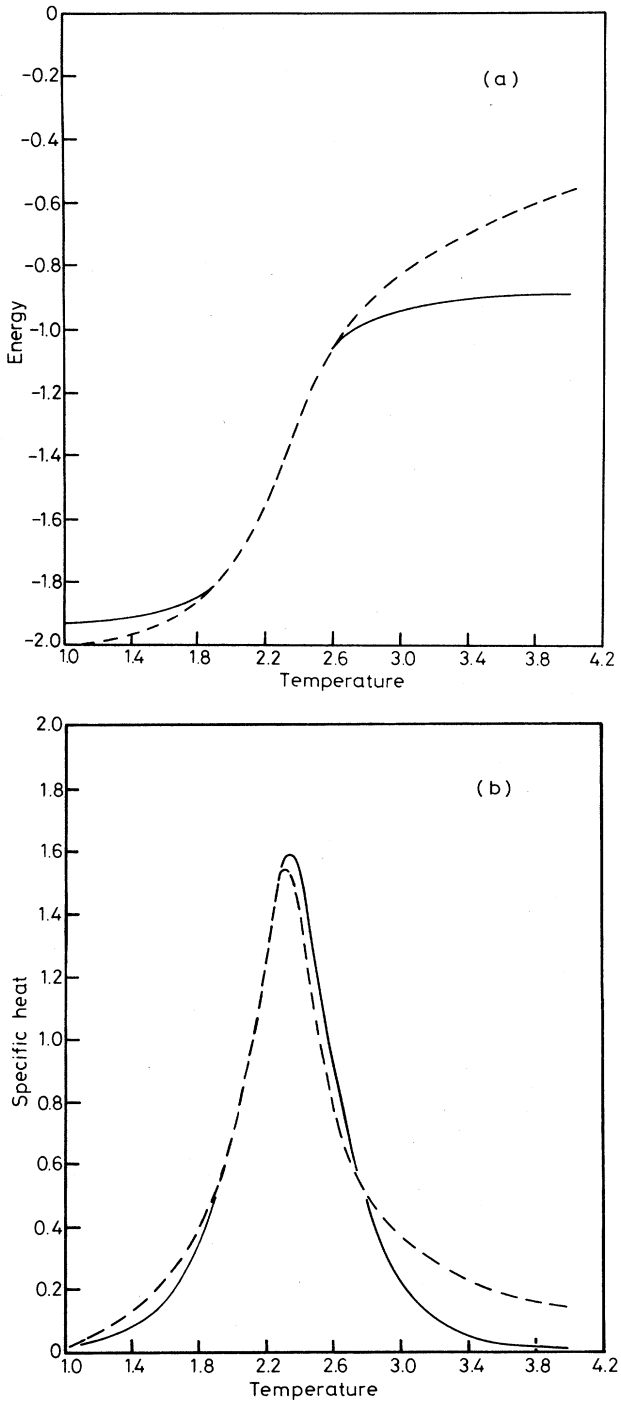

FIG. 1. (a) Energy and (b) specific heat vs temperature with the exact results of Ferdinand and Fisher (Ref. 2) superimposed (dashed line).

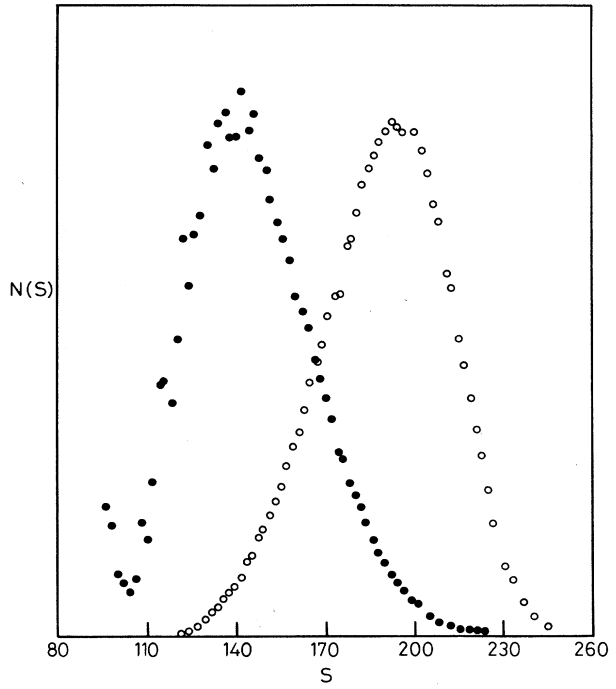

FIG. 2. The histogram of frequency $N(S)$ vs energy (S) generated at T_c (open circle) by Monte Carlo sampling and the extrapolated distribution at $T=2.50$ (solid circle) constructed from the histogram at T_c and Eq. (3).

EFFICIENT SAMPLING OF THE EXTREMITIES OF THE DISTRIBUTION

To improve the accuracy of the tails of the distribution, we turn the FS technique around. We generate histograms not only at T_c (now referred to as the primary distribution) but also at a lower temperature ($T=1.7$) and at a higher temperature ($T=3.0$). These histograms are referred to as the secondary distributions. These secondary distributions have an energy range of high accuracy located in areas of low accuracy in the primary distribution. Therefore this information may be used to accurately determine the tails of the primary distribution. This is

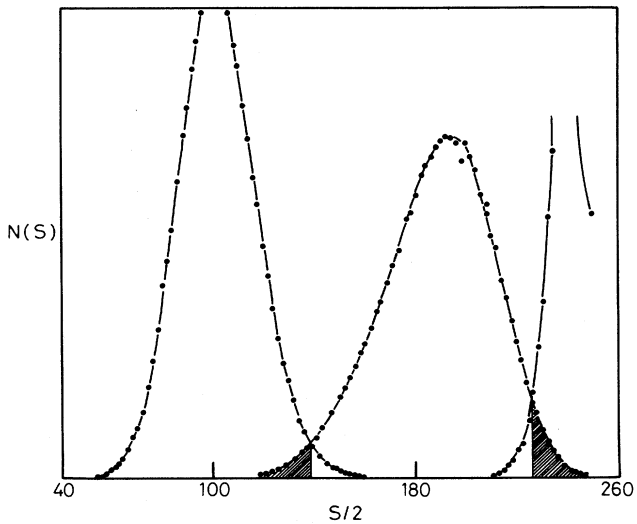

FIG. 3. The tails of the primary distribution shaded regions are replaced by constructed sections from $T=3.0$ (the distribution to the left) and $T=1.7$ (the distribution to the right).

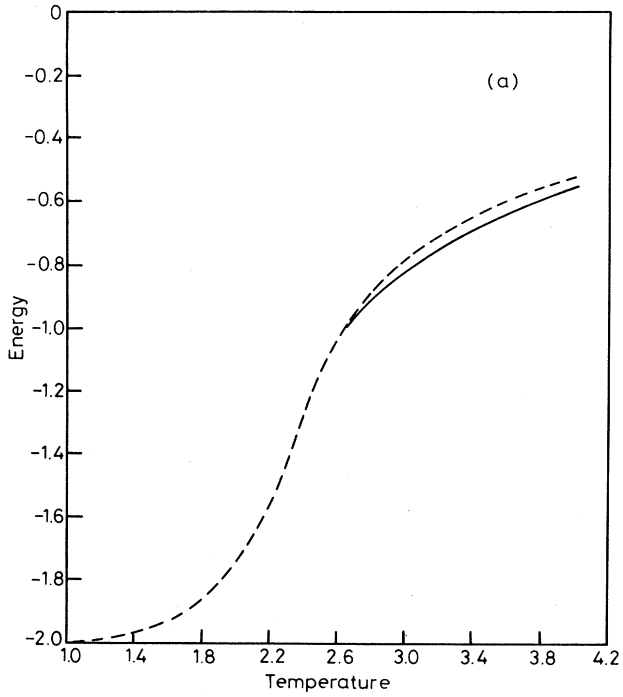

FIG. 4. Energy and specific heat vs temperature. The solid line is the computed results while the dashed line is the exact values from Ref. 2.

done in the following manner: Equation (3) is used to predict values in the energy range of the tail of the primary distribution (Fig. 3). The shaded regions in the tails of the primary distribution are replaced by equivalent areas constructed from the secondary distributions. The

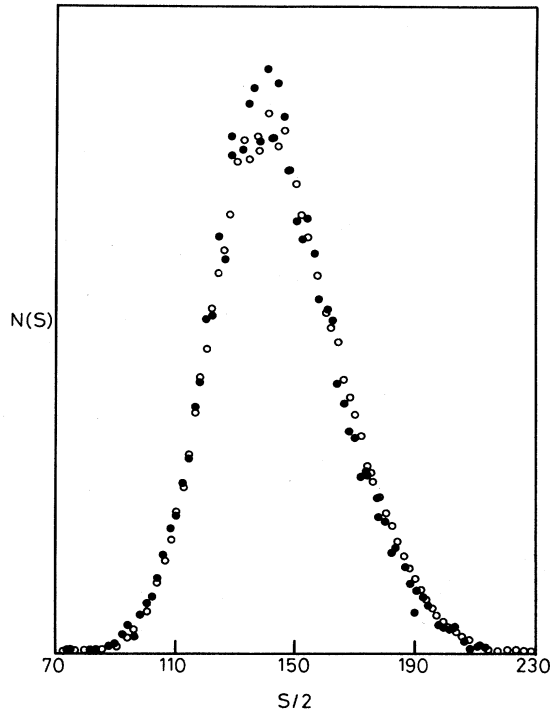

FIG. 5. The distribution generated at $T=2.50$ by Monte Carlo sampling (solid circle) and from the composite distribution and Eq. (3) (open circle).

concept of equal areas is necessary because of the inaccuracy also present in the tails of the secondary distribution. The inaccurate areas of the secondary distribution contribute significantly to Z when one makes use of Eq. (3), making the rescaling method necessary. In this manner we were able to correctly determine events which occur with a probability of 10^{-9}: insignificant for the determination of any thermodynamic property at T_c but whose contribution becomes more important as the temperature is varied. This method is similar in spirit to the "multistage sampling" technique of Valleau and Card,[6] who considered overlapping energy density distributions in the calculation of the free energy of a system of hard spheres with Coulombic forces.

The energy and specific heat are shown in Fig. 4 together with the exact values for finite systems.[2] There is no systematic deviations of these qualities with temperature—the difference observable at high temperatures is due to the periodic boundary conditions and this effect is observable elsewhere in the comparison of Monte Carlo results with the exact results of Onsager.[7] The histogram $P_K(S)$ obtained from the composite distribution and Eq. (3) is shown at $T=2.5$ (Fig. 5). The actual distribution obtained from a Monte Carlo simulation at $T=2.5$ is superimposed. The agreement is excellent. The direct Monte Carlo simulation leads to an energy of -1.122 and a specific heat of 1.134, while the composite distribution predicts -1.132 and 1.181. The corresponding exact values are -1.131 and 1.065.

CONCLUSION

The powerful technique introduced by Ferrenberg and Swendsen, whereby thermodynamic properties for a wide range of temperatures are determined from Monte Carlo data at a specific temperature, is analyzed in detail. Errors are due to insufficient occurrence of events in the tails of the original distribution, which were generated in the first place by "important sampling." We have used the FS algorithm to reconstruct the tails of the primary distribution from two secondary distributions at different temperatures—thus reducing the noise and improving the accuracy of the extremities. The composite distribution leads to improved results for the energy and specific heat over the complete temperature range by providing information on events as rare as 10^{-9}. We have obtained accurate results with a total of 300 K configurations, a factor of 10 less than FS. We refer interested readers to FS for the advantages and flexibility of this approach for the calculation of thermodynamic properties of systems. Note that the method in the present form will be extremely efficient for determining crossover and other subtle effects from nearby critical points and multicritical points.

ACKNOWLEDGMENTS

This research is funded by Natural Sciences and Engineering Research Council of Canada. After the completion of this work we learned of unpublished work by Swendsen and Ferrenberg outlining an alternative method for combining several distributions.

---

*Physics 320, St. Francis Xavier University.

†Present address: Center for Polymer Studies and Physics Department, Boston University, 590 Commonwealth Ave., Boston, MA 02215.

[1]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988). For earlier work using similar techniques for self-avoiding walks and interacting self-avoiding walks see, e.g., K. Kremmer, A. Baumgartner, and K. Binder, J. Phys. A **15**, 2879 (1982); V. Privman, J. Phys. A **19**, 3287 (1986); A. Coniglio, N. Jan, I. Majid, and H. E. Stanley, Phys. Rev. B **35**, 3617 (1987); and also Ref. 4.

[2]A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).

[3]*Monte Carlo Methods in Statistical Physics,* edited by K. Binder (Springer-Verlag, Berlin, 1979).

[4]P. H. Poole, A. Coniglio, N. Jan, and H. E. Stanley, Phys. Rev. B **39**, 495 (1989). See the Appendix for an in depth discussion on the effects of noise from the tails of distributions.

[5]R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[6]J. P. Valleau and D. N. Card, J. Chem. Phys. **57**, 5457 (1972).

[7]L. Onsager, Phys. Rev. **67**, 117 (1944); N. Jan and D. A. Pink, J. Phys. A **16**, 4325 (1983).