JOURNAL OF COMPUTATIONAL PHYSICS 108, 353–356 (1993)

Features of the Histogram Monte Carlo Method: Application to N_2 Monolayer Melting on Graphite

B. KUCHTA AND R. D. ETTERS

Department of Physics, Colorado State University, Fort Collins, Colorado 80523

Received March 2, 1992; revised December 14, 1992

Previously developed methods that determine properties of a system over a range of thermodynamic points, using information accumulated at a single point, are summarized, as is an extension called the multiple histogram method. These strategies are applied to Monte Carlo calculations of melting for N_2 adlayers on graphite. The results are used to show the utility of the multiple histogram method and the inadequacy of the single point method for this application to a physical system.

I. INTRODUCTION

It has been shown [1–4] that information from a single thermodynamic point of a Monte Carlo simulation can provide valuable knowledge about neighboring points. These works examined the thermodynamic properties of liquid argon [1], the Ising model near the critical point [2, 3], and a generic Lennard–Jones crystal below 75% of the melting temperature [4]. The results were promising. It is noted that the breadth and accuracy of the probability distribution at a point influences the magnitude of the neighborhood around it, where reliable data can be obtained. If this is insufficient, the authors [3] have formulated a method, where calculations are performed at multiple points, that extends the range over which accurate information can be attained. Here the overlapping distribution between neighboring points enhance the quality of the calculated probability over the entire range of interest. Indeed, Bowen *et al.* [5] have demonstrated this feature in studying the results from an Ising model.

In this work we report on an application of the above-mentioned strategies to a physical system, namely the melting of fractional and monolayer deposits of N_2 on graphite. It will be shown that information from single Monte Carlo points is insufficient to characterize the entire transition region, but multiple points do provide a prudent strategy for calculating properties of the phase transition.

II. THE PHYSICAL MODEL

Monolayers of molecular nitrogen deposited on graphite are known to form an in-plane herringbone orientational structure at low temperatures, with the molecular centers forming a triangular $\sqrt{3}\times\sqrt{3}$ arrangement [6], commensurate with the substrate lattice. At a temperature $T \simeq 27$ K, the N_2 molecules undergo an orientational order–disorder transition [7], but the mass center lattice positions are undisturbed until melting [6, 8, 9] at $\simeq 85$ K. The work presented here is part of a study of the melting transition for N_2 surface densities $0.2 \leqslant \rho \leqslant 1.0$, where the upper limit corresponds to monolayer coverage.

The N_2–N_2 interactions are given by our site–site representation of ab initio results [10], and calculations on bulk fluid and solid N_2 indicate that it is quite accurate. The N_2–substrate interactions include the overlap–van der Waals terms, expressed using the Fourier decomposition of Steele [11], the image charge, and the substrate mediated dispersion terms. This expression has also proved to be quite satisfactory. Details are given in Ref. (12).

III. THE HISTOGRAM METHOD

To see how the probability distribution at one point can be used to determine it at another point, recall that for a canonical ensemble

$$P_\beta(E) = N_\beta(E)/n_\beta = W(E)\exp[-\beta E + f], \qquad (1)$$

where the probability distribution is $P_\beta(E)$, the temperature is $\beta = (kT)^{-1}$, $W(E)$ is the density of states at energy E, and $F = f/\beta$ is the Helmholtz free energy. In a Monte Carlo calculation $N_\beta(E)$ is the number of configurations in the interval E to $E + \Delta E$ out of n_β total configurations. Note that in equilibrium $P_\beta(E)$ peaks at some energy $E_0(\beta)$ and tails off for energies away from this value. It is these tails

0021-9991/93 $5.00

that give information at energies relevant to neighboring points β'. Here temperature is used to identify the thermodynamic points, but other variables may be useful alternatives. As pointed out by Ferrenberg and Swendsen [2], $P_{\beta'}(E)$ is mathematically related to $P_\beta(E)$ by the exact expression

$$P_{\beta'}(E) = \frac{P_\beta(E)\exp[-(\beta'-\beta)E]}{\sum_E P_\beta(E)\exp[(-(\beta'-\beta)E]}. \tag{2}$$

An examination of the numerator shows that $P_{\beta'}(E)$ is most accurate when $|\beta'-\beta|$ is small and $P_\beta(E)$ is broad and well described. Clearly, these are reasonable expectations.

The multiple histogram method [3] is a straightforward generalization of the above-mentioned arguments. If Monte Carlo calculations are performed at R different temperatures, the normalized probability distribution is

$$P_\beta(E) = D_\beta(E) \Big/ \sum_E D_\beta(E) \tag{3}$$

$$D_\beta(E) = \frac{\sum_{i=1}^R N_i(E)}{\sum_{i=1}^R n_i \exp[-(\beta_i-\beta)E+f_i]} \tag{4}$$

where

$$\sum_E D_{\beta_i}(E) = \exp(-f_i). \tag{5}$$

Clearly Eq. (5) is simply the partition function (normalization) for $P_{\beta_i}(E)$. The free energies $\{f_i\}$, $i = 1, 2, ..., R$, can be determined self-consistently from Eqs. (3) and (4), or from the intersection of neighboring histograms $\{N_i(E)\}$. That is,

$$P_{\beta_i}(E)/P_{\beta_{i+1}}(E) = \exp[-(\beta_i-\beta_{i+1})E+(f_i-f_{i+1})].$$

If these distributions overlap, there is one energy $\bar{E}$, where $P_{\beta_i}(\bar{E}) = P_{\beta_{i+1}}(\bar{E})$. Then

$$f_i - f_{i+1} = (\beta_i-\beta_{i+1})\bar{E}. \tag{6}$$

It is sufficient to determine the set $\{f_i\}$ to within an additive constant so one value is set equal to zero and all others are measured with respect to it.

IV. APPLICATION TO N_2 MELTING ON GRAPHITE

The melting of N_2 monolayers on graphite was examined using the constant volume, histogram Monte Carlo technique, with individual points calculated at $T = 75$, 80, 81, 82, 84, 90, and 95 K. A cell with $N = 256$ molecules and periodic boundary conditions was used. Between 5×10^5 and 10^6 steps were performed at each temperature, requiring some 3500 cpu hours on an IBM 6000 RISC work station. Each step consists of randomly sampling all 5N degrees of freedom. Step averages were used to calculate various thermodynamic expectation values and the histograms $N_i(E)$ in Eqs. (1) and (4).

The solid circles, shown on Fig. 1, represent the specific heat per molecule, in units of the Boltzmann constant, calculated from the multiple histogram method [3]. The peak temperature signifies melting at $T_M = 80.5$ K. This value, the peak height, and the width at half maximum, are in good agreement with experiment [8, 9]. The specific heat varied negligibly over the last 10^5 steps, giving confidence that convergence had been reached. The curve represented by triangles with the highest peak value is a result of the single histogram method initiated at 80 K. The specific heat at this temperature is, of course, identical to the results obtained without the histogram method. Away from this point the results are a consequence of Eq. (2). The second triangular curve, initiated at 81 K, is similarly described. While they show a small signature of the transition, they are quantitatively inadequate as indicated by their behavior over the last 10^5 steps. Values in the vicinity of the transition show fluctuations almost an order of magnitude larger than comparable points using the multiple histogram method and are worse away from the transition and away from the initiation temperature. The solid circles on the top graph of Fig. 2 show the internal energy, calculated from the multiple histogram method [3]. This curve is believed to be quite accurate because the results are remarkably stable over the last 10^5 steps and because the energies calculated at the seven separate temperatures agree almost exactly with the multiple histogram [3] results at those points. These points have a statistical uncertainty in energy of less than 0.5%. However, the application of these points in the single histogram method [2], represented by the seven curves formed from the triangles, do not individually provide enough information to accurately span the transition region, a conclusion in concert with that drawn from

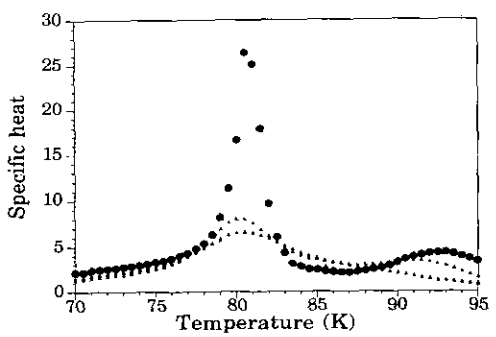

FIG. 1. The solid circles give the calculated specific heat per N_2 molecule near melting from the multiple histogram [3] method. The statistical uncertainty in these points is within ± 1.0 K. The curves with triangles are results from the single point histogram method [2], based on calculated points at 80 and 81 K.

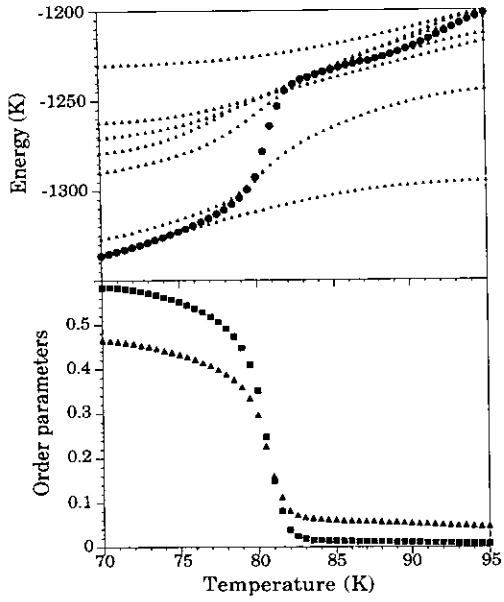

FIG. 2. The circles on the top graph give the internal energies calculated from the multiple histogram method, and the seven curves composed of triangles are results calculated using the single point histogram method from Monte Carlo runs at each of seven different temperatures. The bottom graph shows two different order parameters, sensitive to melting, calculated using the multiple histogram method.

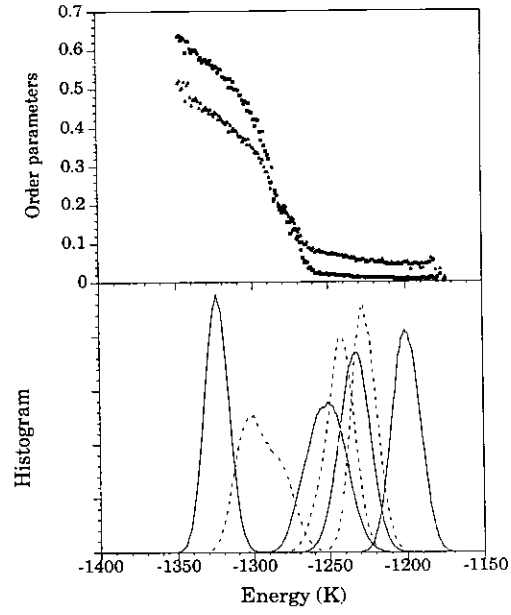

FIG. 3. The top graph shows the same order parameters as in Fig. 2, plotted versus energy. The bottom graph shows the probability distributions calculated at $T = 75$, 81, 84, and 95 K (solid lines) and at $T = 80$, 82, and 90 K (dashed lines).

specific heat results. The two curves on the lower graph of Fig. 2 show order parameters sensitive to melting, calculated from the multiple histogram scheme. The squares are just the structure factor, evaluated at the reciprocal lattice vectors of the graphite substrate. It is designed to equal unity if the molecular centers are all statically located over the center of a graphite hexagon and zero if they uniformly sample all locations above the substrate. The non-zero high temperature tail indicates some preference for the above-mentioned registry sites above melting. The triangles represent

$$O_2 = [3N(N-1)]^{-1} \sum_{i<j}^{N} \left\langle \sum_{s=1}^{6} \exp(i\mathbf{k}_s \cdot \mathbf{r}_{ij}) \right\rangle, \qquad (7)$$

where $\mathbf{k}_s$ are the reciprocal lattice vectors of the $\sqrt{3} \times \sqrt{3}$ center of mass structure [12], stable at low temperature, and $\mathbf{r}_{ij}$ is the vector connecting molecular centers (i, j). The brackets indicate a thermal average. It is designed to be unity for a static $\sqrt{3} \times \sqrt{3}$ structure and zero if there is no vestige of it. The top graph on Fig. 3 shows the same two order parameters shown on Fig. 2, but plotted against energy. It is shown to dramatize the energy interval over which the transition occurs. The bottom graph shows the probability distributions calculated from the seven individual points. From left to right the solid curves are for $T = 75$, 81, 84, and 95 K, and the dashed lines are for $T = 80$, 82, and 90 K. Note that their energy widths at half maximum are about 20 K outside the transition region and about 35 K inside.

V. SUMMARY AND DISCUSSION

It is understood that both the single and multiple point histogram methods will accurately represent a physical system over any finite range of thermodynamic points, given sufficiently long Monte Carlo sequences. The issue is, therefore, relative efficiencies. The advantage of the multiple point analysis is that neighboring points can be chosen so that the overlap in the probability distribution between them is large, indicating that the entire thermodynamic interval is well sampled. The advantage of the single point histogram method is that only one distribution need be calculated and that the total computational expenditure can be concentrated on it. The problem with this method is that the probability distribution decreases rapidly away from the point at which it was taken. The expenditure of computational effort to accurately determine these tails is large, particularly if the thermodynamic range of interest is also large. As for the traditional method of calculating properties at a number of separate, individual points, this method cannot be as efficient as the multiple histogram method unless the points are so separated that their probabilities have negligible overlap.

It is evident that the single point histogram method is not satisfactory when applied to the melting of N_2 on graphite, as evidenced by the specific heat and energy results on Figs. 1 and 2. This failure can be understood by examining the internal energy and probabilities on Figs. 2 and 3. The energy change across the transition is approximately 80 K,

but the width of the probability distribution at half maximum is only about 35 K in the transition region and only 20 K on the edge of it. This is simply not enough, even though as many as 10^6 steps (10^9 configurations) were run at each Monte Carlo point. Note that, at each temperature at which the distribution was calculated and used independently to determine the specific heat, energy, and the order parameters, the results were consistent with those from the single histogram method, initiated at each of these points.

The multiple histogram method has proved to be most satisfactory for this problem because it accurately characterizes the probability distribution over the entire transition region, as shown on Fig. 3. Comparing the energies deduced from this method with those derived from Eq. (2), using the single point histograms, Fig. 2 shows that the results are quite different. However, at the temperature about which a distribution is generated, the average energy determined by both methods differs by less than 1%. Away from this temperature the results become increasingly dissimilar. Qualitatively similar results occur for single point histograms generated at all different temperatures, as is evident on Fig. 2. The order parameters behave in the same way. However, the two methods do not give comparable specific heats anywhere, as evidenced on Fig. 1. One could have guessed this by examining the energies on Fig. 2. Such a result is not surprising since the specific heat is a fluctuation quantity, the solution of which requires high statistical accuracy.

Finally, there are difficulties in applying this method to an actual physical system that do not occur for model systems, where exact solutions are already known. The primary problem is that a priori knowledge about the optimal location of Monte Carlo points to be used is generally not available, nor is the range of energies that must be sampled to span the transition. The multiple point analysis helps to ameliorate this problem and we conclude that it is a valuable and economical tool.

REFERENCES

1. I. R. McDonald and K. Singer, *Discuss. Faraday Soc.* **43**, 40 (1967).
2. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **23**, 2635 (1988).
3. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
4. S. R. Phillpot, *Phys. Rev. Lett.* **66**, 349 (1991).
5. P. Bowen, J. Burke, P. Corsten, K. Crowell, K. Farrel, J. MacDonald, R. MacDonald, A. MacIsaac, P. Poole, and N. Jan, *Phys. Rev. B* **40**, 7439 (1989).
6. J. K. Kjems, L. Passell, H. Taub, G. Dash, and A. Novaco, *Phys. Rev. B* **13**, 1446 (1976).
7. R. D. Diehl, M. F. Toney, and S. Fain, *Phys. Rev. Lett.* **48**, 177 (1982).
8. T. Chung and G. Dash, *Surf. Sci.* **66**, 559 (1977).
9. M. Chan, A. Migone, K. Miner, and Z. Li, *Phys. Rev. B* **30**, 2681 (1984).
10. R. Berns and A. van der Avoird, *J. Chem. Phys.* **72**, 6107 (1980).
11. W. A. Steele, *Surf. Sci.* **36**, 317 (1973).
12. B. Kuchta and R. D. Etters, *Phys. Rev. B* **36**, 3400 (1987).