Simulated annealing in the microcanonical ensemble

John R. Ray and Richard W. Harris
Department of Physics and Astronomy, Kinard Laboratory of Physics, Clemson University, Clemson, South Carolina 29634-1911
(Received 13 January 1997)

In earlier work we have extended the microcanonical Monte Carlo method, which had been introduced for systems described by continuous potentials [Phys. Rev. A **44**, 4061 (1991)] to discrete lattice systems [Phys. Rev. E **53**, 3402 (1996)], such as the Ising model. This microcanonical ensemble Monte Carlo method is rigorously based on statistical mechanics and one has available the entire structure of equilibrium statistical mechanics, such as the full set of fluctuation formulas, which are useful in numerical estimates of the equilibrium properties of the system. In the present paper we explore the use of the microcanonical ensemble Monte Carlo probability distribution to study combinatorial optimization problems using simulated annealing. In particular, we present the results of a detailed study of a particular 20-city traveling salesman problem in both the canonical and microcanonical ensembles. [S1063-651X(97)09105-8]

PACS number(s): 05.20.−y, 05.70.−a, 02.50.−r

I. INTRODUCTION

Since the introduction of the canonical-ensemble Monte Carlo method into statistical mechanics by Metropolis *et al.* [1], most Monte Carlo simulations have used the canonical ensemble. Earlier we have introduced a rigorous microcanonical Monte Carlo procedure [2] and applied it to condensed matter systems with a continuous potential. More recently, we have shown that this same microcanonical ensemble (probability distribution in configuration space) can also be used to describe discrete lattice systems [3]. In related work, we have also developed other constant energy ensembles and carried out Monte Carlo simulations in these ensembles; these include the isoenthalpic-isotension ensemble [4] and new ensembles that describe open systems [5]. Since the microcanonical ensemble describes an isolated system in equilibrium, this ensemble is at least as fundamental as the canonical ensemble but has, we believe, been inappropriately neglected in simulation work. The canonical ensemble is more popular in analytic work, where the exponential probability function is often simpler to use; in simulations, there is no difference in difficulty between using the canonical or microcanonical ensemble.

The use of statistical-mechanics methods to study combinatorial optimization problems originated with Kirkpatrick, Gelatt, and Vecchi [6], where they described a method of minimizing a function of many variables, the so-called cost function, using the metropolis Monte Carlo method in the canonical ensemble. They interpreted the cost function as the potential energy of a fictitious system and carried out metropolis Monte Carlo calculations in the canonical ensemble to generate a sequence of configurations of the system. As with any physical system at a given temperature described by the canonical ensemble, the fictitious system will come to equilibrium, where its Helmholtz free energy is minimized. As the temperature of the system is lowered and the system attains equilibrium at the lower temperature, the Helmholtz free energy has a lower value, and in the limit as the temperature goes toward zero, the system ends up, one hopes, in one of a group of states near the ground state of the system, where the potential energy is approximately equal to its Helmholtz free energy and is near its global minimum. This procedure was termed ''simulated annealing'' by Kirkpatrick, Gelatt, and Vecchi and is referred to by this name in the vast combinatorial optimization literature [7]. This method and numerous variations are routinely used in both Monte Carlo and molecular-dynamics simulations to study the low-temperature (-energy) configurations of condensed matter systems.

In the present paper, we present detailed Monte Carlo calculational results in both the canonical and microcanonical ensembles for a 20-city traveling salesman problem (TSP). The basic theory associated with the microcanonical ensemble is presented in detail in Refs. [2] and [3] and will not be repeated. In Sec. II we outline the canonical and microcanonical simulation methods. In Sec. III we present the results of canonical and microcanonical ensemble calculations for the 20-city TSP. In Sec. IV we present our conclusions, along with suggestions for further work.

II. TRAVELING SALESMAN PROBLEM AND ENSEMBLES

In the two-dimensional TSP, we are given a set of N cities at positions (x_i, y_i) and we are asked to determine the minimal path length that starts and ends at any city and passes through each of the other cities exactly one time; such a path is called a tour and is a polygon with N vertices. In the example discussed in this paper, the positions of the cities are chosen randomly inside the unit square. If we label the cities with integers $i=1,2,\ldots,N$, then each path length can be calculated if we give the ordered sequence of cities associated with going around the path, starting from any city; each such tour is associated with a permutation of the integers $1,2,\ldots,N$. For $N \geqslant 3$ cities, there are, in general, $(N-1)!/2$ distinct sequences and, therefore, path lengths. Since the number of distinct paths grows faster than any power of N, the problem has no known direct solution method that works for arbitrary N. Even for our small 20-city problem, there are around 6×10^{16} distinct paths lengths. Simulated annealing is one of the methods used as a practical method to obtain approximate solutions to such problems. Numerous other methods used to study the TSP, which will not be discussed in this paper, may be found in the extensive

review paper on the TSP [8], along with the extensive set of references in this paper. The cost function $U(x)$, which is identified with the potential energy of the fictitious system, is just the path length for the given sequence of cities visited. We shall use units so that the energies and temperatures are all measured in the same dimensionless units. In the canonical ensemble, we have the configurational probability distribution $P_T(x)$,

$$P_T(x) = Ce^{-U(x)/T}, \tag{2.1}$$

where C is a normalization constant, x represents a specific sequence of the cities visited and completely specifies the system configuration, and T is the temperature. The average value of $U(x)$, along with the temperature T, can be used to define the total energy of the system as

$$E = NT + \langle U(x) \rangle. \tag{2.2}$$

We are assuming two-dimensional Euclidean space; the changes for arbitrary dimension d are simple and given in Ref. [3]. By the brackets $\langle\ \rangle$, we mean an equilibrium average over the configurations generated by the metropolis Monte Carlo calculation, using the probability distribution equation (2.1). Other quantities for the fictitious system can also be calculated from the configurations generated in the Monte Carlo calculation. The specific heat of the fictitious system $c_T = (1/N)(\partial \langle U \rangle / \partial T)$ can be obtained from the fluctuations in U, as given in Eq. (6) of Ref. [3].

In the microcanonical ensemble, we have the configurational probability distribution $P_E(x)$,

$$P_E(x) = C'(E - U(x))^{N-1} \Theta(E - U(x)), \tag{2.3}$$

where $\Theta(x)$ is the unit step function, which is 1 for $x > 0$ and zero otherwise, and C' is another normalization constant; the Θ function arises because the kinetic energy of the system $K = E - U$ is positive. It may seem unusual to think of the TSP as having a kinetic energy, but as explained in Ref. [3], for lattice systems such as the TSP or the Ising model, we take the limit as the mass of the points of the discrete system goes to infinity and the velocity goes to zero, such that the kinetic energy is finite. The necessary limit is $m \to \infty$, $\dot{x} \to 0$, such that $\sqrt{m}\dot{x} \to$ finite, and the system has finite kinetic energy, but will not have any spatial motion. Other microcanonical ensemble averages can be calculated for the fictitious system using the configurations generated from the Monte Carlo calculation using Eq. (2.3). The microcanonical ensemble specific heat, $c_E = (1/N)(\partial \langle U \rangle / \partial T)$, can be calculated by using the fluctuation formulas given in Ref. [3]. Calculating the average of the cost function $U(x)$ in the microcanonical ensemble and using the specified value of E, we can use Eq. (2.2) to find the temperature T of the fictitious system corresponding to this energy. Thus, the information from the two ensembles is complementary; in the canonical ensemble we specify a temperature and calculate a corresponding average value of the energy, whereas in the microcanonical ensemble we specify a value of the energy and calculate a corresponding average value of the temperature. By feeding the calculated energy or temperature values back into the complementary ensemble, we obtain a consistency (and code) check on our calculational results.

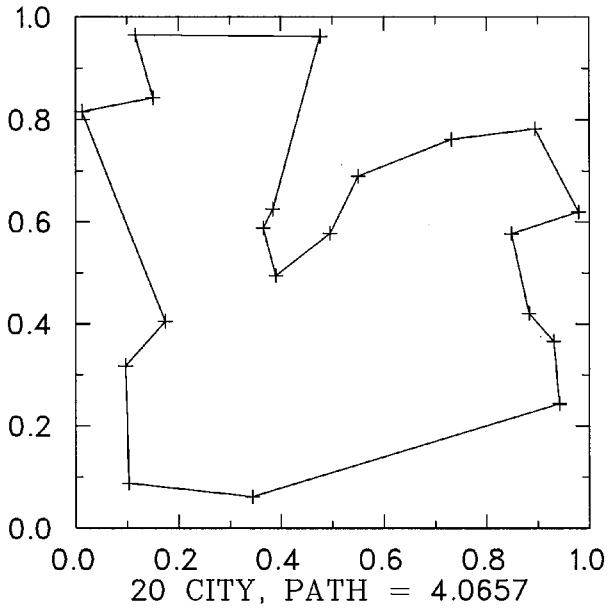

FIG. 1. The randomly selected 20-city problem discussed in this paper. The minimum length path is shown.

III. RESULTS

A. Computational strategy

In order to omit repeated statements in the discussion of the data generated in our calculations, we shall refer to the temperature, energy, and potential energy of the fictitious system without pointing out every time that these quantities do not have a meaning in terms of a real system; we are applying statistical mechanics ideas to a fictitious physical system.

We constructed a 20-city TSP problem by randomly choosing 20 x,y pairs in the unit square. In Fig. 1 we show this city arrangement, as well as the minimum path length for this arrangement. In order to carry out the Monte Carlo procedure, we need a method of making configurational changes (moves) in the TSP, that is, trial configuration changes in the metropolis Monte Carlo procedure. We carry out two types of trial moves with equal weighting: (1) given the present sequence of cities to be visited x, we select a subsequence of random length and reverse the order of this subsequence, thus producing a new configuration of the system x'; (2) given the sequence of cities to be visited x, we select a subsequence of random length and transport this subsequence and insert it between two cities not on the subsequence. These changes are the same as those suggested by Press *et al.* [9]. In the former, ''reversal change,'' two city-city distances are replaced by two other city-city distances in the calculation of the path length; these are the end points of the reversed subsequence. In the latter, ''transport change,'' three city-city distances are replaced by three other city-city distances; these three city-city distances are associated with the ends of the segment before its transport and the position to where it is transported.

We then carried out canonical ensemble Monte Carlo simulations at each of eleven chosen T values. Since we are using a fictitious physical system, an arbitrary but satisfactory method of defining a high (dimensionless) temperature

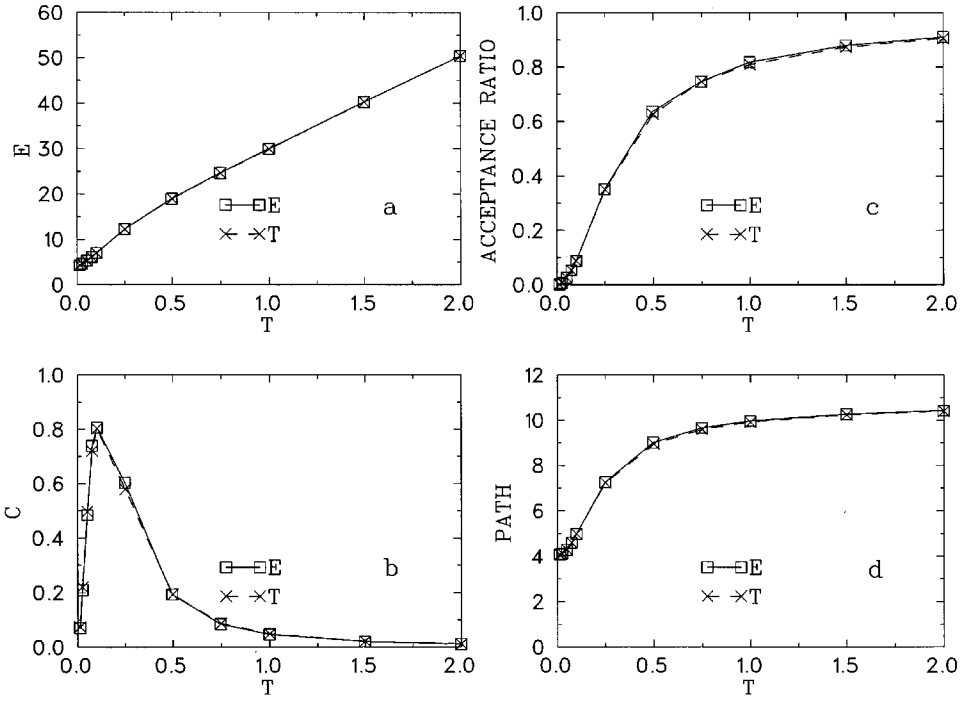

FIG. 2. (a) Total energy versus temperature for the two ensembles for the 11 values selected. In the canonical ensemble the energy is a derived quantity, while in the microcanonical ensemble the temperature is a derived quantity. The solid line with boxes are the microcanonical ensemble values, while the dashed line and crosses are the canonical ensemble values. (b) The specific heat versus temperature for the two ensembles. The solid line with boxes are the microcanonical ensemble values, while the dashed line and crosses are the canonical ensemble values. (c) The acceptance ratio versus temperature for the two ensembles. The solid line with boxes is the microcanonical ensemble values, while the dashed line and crosses are the canonical ensemble values. (d) The average path length versus temperature for the two ensembles. The solid line with boxes are the microcanonical ensemble values, while the dashed line and crosses are the canonical ensemble values. The lines are just to guide the eye. The values used to make these graphs are shown in Tables I and II.

is to call the temperature high when approximately 90% of the moves are accepted; for our 20-city problem, this temperature is around $T=2.00$. After the high temperature was determined, we selected ten lower temperatures and carried out five Monte Carlo calculations of 10^8 configuration change trial moves at each of the eleven temperatures to determine the thermodynamic properties of the fictitious system. Note that in the present study we do not investigate the annealing schedule (temperature- or energy-lowering schedule). Our approach is to study the fictitious system as a thermodynamic system in the two ensembles. Our annealing schedule is to lower the temperature to the next value and carry out five runs of 10^8 moves, and then lower the temperature or energy to the next value and so on until all the temperatures or energies are considered. The values of the average energies obtained in these eleven canonical ensemble temperatures are then used in the microcanonical probability distribution equation (2.3) to carry out eleven microcanonical ensemble runs at each of these energies. In the microcanonical ensemble calculations we again carry out five 10^8-move Monte Carlo runs at each energy. There is, of course, no significance to the fact that we carry out canonical ensemble runs to first determine the energies to be used in microcanonical ensemble calculations; we could just as well reverse this procedure. As mentioned, in our Monte Carlo calculations we carried out five 10^8 Monte Carlo moves, and then lowered the temperature or energy to the next value and carry out another five 10^8 moves and so on, through all of the temperatures and energies studied. We consider the first 10^8 configuration run as the equilibration portion of the calculation and do not use the values calculated in this run for the final error estimate. We use the calculated values in the other four runs to calculate a standard deviation for each quantity of interest to use as an error estimate. For each of the 11 canonical ensemble calculations, we arrive at a value of the energy $\langle E\rangle$ using Eq. (2.2). For example, for the high temperature $T=2.00$ mentioned above, we find $\langle E\rangle = 50.398$; this value of $\langle E\rangle$ is then used in the microcanonical ensemble to calculate the temperature (which turns out to be $T=1.9988$) and the other quantities of interest.

B. Computational results

In Fig. 2 we present various quantities versus temperature as calculated in the two ensembles. Because of the system's small size we would expect small differences in the value of the various quantities. Ensemble theory shows that ensemble differences may be of the order of $(1/N)$ for intensive variables. All of the values comparing the two ensembles are smaller than this value, often much smaller. The microcanonical ensemble results are shown as solid lines with squares while the canonical ensemble results are shown as dashed lines with crosses. Note the very close agreement in Fig. 2(a) between the energy-versus-temperature relation as calculated in the two ensembles. The specific heats are not in as close agreement in Fig. 2(b). For example, at the temperature 0.075, the microcanonical ensemble specific heat is 0.7385 (0.002), while the canonical ensemble specific heat is

TABLE I. Canonical ensemble Monte Carlo values for the temperature, average total energy, average path length, and the specific heat. All the quantities are dimensionless. The average values were determined by carrying out five calculations of 10^8 moves and averaging the last four numbers obtained for each quantity. The temperature is an input value and has no uncertainty in the calculation.

T	$\langle E \rangle$	$\langle U \rangle$	c
0.0125	4.3258	4.0758	0.08318
0.025	4.6114	4.1139	0.21622
0.05	5.2901	4.2901	0.49409
0.075	6.1012	4.6012	0.72135
0.10	6.9860	4.9860	0.80244
0.25	12.224	7.2249	0.57876
0.50	18.9462	8.9462	0.19305
0.75	24.5972	9.5972	0.08709
1.00	29.9214	9.9214	0.04842
1.50	40.2410	10.2410	0.02102
2.00	50.3979	10.3979	0.01163

TABLE II. Microcanonical ensemble Monte Carlo values for the average temperature, total energy, average path length, and specific heat. All the quantities are dimensionless. The average values were determined by carrying out five calculations of 10^8 moves and averaging the last four numbers obtained for each quantity. The energy is an input value and has no uncertainty in the calculation.

$\langle E-U \rangle / N = T$	E	$\langle U \rangle$	c
0.01243	4.3234	4.0749	0.0702
0.02513	4.6141	4.1115	0.2098
0.05037	5.2910	4.2837	0.4818
0.07525	6.1005	4.5955	0.7385
0.10	6.9856	4.9833	0.8058
0.2483	12.225	7.2581	0.6021
0.4967	18.946	9.0115	0.1931
0.7472	24.597	9.6540	0.0845
0.9970	29.922	9.9676	0.04640
1.4984	40.241	10.2721	0.02000
2.000	50.398	10.4215	0.01102

0.7214 (0.001), where we show the standard deviation in parentheses after the specific heat values. Note that the maximum in the specific heat is associated with an inflection point in the energy-versus-temperature relation. This is suggestive of a higher-order phase transformation in the system. We shall return later to a discussion of this suggested phase transformation. In Fig. 2(c) we show the acceptance ratio, defined as the ratio of the number of accepted moves to the total number of moves, for calculations in the two ensembles. Finally, in Fig. 2(d) we show the average path length as calculated in the two ensembles. The fact that this relation is very nearly the same for the two ensembles suggests that either ensemble can be used in simulated annealing algorithms to find the lowest value of the cost function for the system.

C. Numerical values and errors

In Table I we give the canonical ensemble values of T, $\langle E \rangle$, $\langle U \rangle$, and c at the temperatures studied, while in Table II we give the microcanonical ensemble values at the energies studied. These are the values used to construct Figs. 2(a)–2(d). The error estimates, except at the lowest temperature, are small. For example, in the canonical ensemble runs at $T=0.075$, the errors in the average energy, average path, and specific heat are 0.0013%, 0.0017%, and 0.16%, respectively. We expect a larger error in the specific heat, since it is related to the derivative of the energy. It is well known that in Monte Carlo and molecular-dynamics simulations such derivative quantities converge more slowly and have larger uncertainties. The errors in the microcanonical ensemble are similar to the canonical ensemble errors. For example, at an energy of $E=6.1005$, which corresponds to a temperature of 0.07525, the errors in the average energy, average path length, and specific heat are 0.003%, 0.0009%, and 0.27%, respectively.

D. Probability distribution of path lengths in the two ensembles

During each simulation we constructed a histogram of path lengths. This histogram is the probability distribution for $U(x)$ at the given temperature or energy for this city arrangement. In Figs. 3(a)–3(d) we show these probability distributions at temperatures 1.50, 0.25, 0.100, and 0.0125, respectively, for the two ensembles. Note the differences in scales on the four figures. These histograms are calculated during the last of the five 10^8-move calculations. Thus, the area under each of these curves is 10^8; to obtain the probability distribution, we would divide by this normalization factor. At the highest temperature/energy, 1.50, the two probability distributions are similar, and it is difficult to tell the difference between them in Fig. 3(a), while in Fig. 3(b) for the temperature 0.25, the two distributions have become quite different, with the microcanonical ensemble distribution being sharper, while the canonical ensemble distribution is broader and shows noticeably more asymmetry about its maximum value. This tendency is even more pronounced in Fig. 3(c) for the temperature 0.100. Note that this value is near the maximum in the specific heat and, hence, near the suggested phase transformation. Note the irregular features in the probability distribution at small distances in the canonical ensemble. Here, the system is able to sample the shorter paths and is starting to spend a significant amount of time in these shorter path configurations; this can also be seen by the density of the curve (histogram) on the shorter distance side. This may be the physical explanation for the suggested phase transformation, namely, it is associated with the system starting to sample the shortest paths. In Fig. 3(d) this behavior is even more exaggerated, with the system spending most of the time in the shortest path, 72% of the time in the microcanonical ensemble and 59% of the time in the canonical ensemble. Note that at this low temperature only a few of the shorter paths are visited a significant fraction of the time and the microcanonical ensemble still has a narrower and more peaked probability distribution.

IV. CONCLUSIONS

We have presented Monte Carlo calculations for a combinatorial optimization problem in the microcanonical ensemble and compared it to the canonical ensemble results.

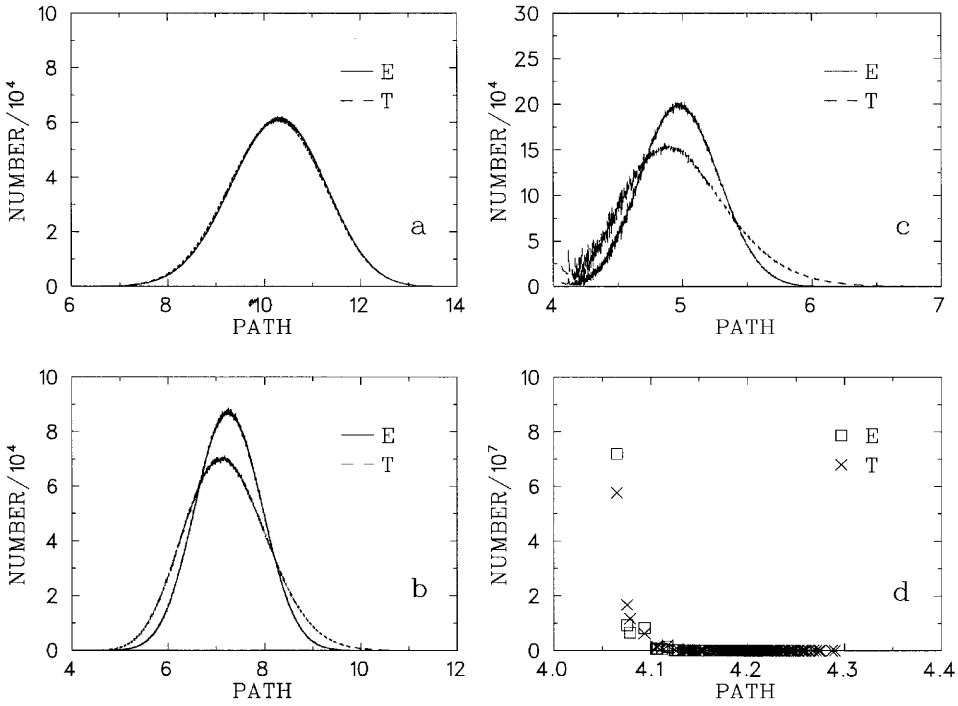

FIG. 3. (a) The probability distributions for a temperature of 1.50. (b) The probability distributions for a temperature of 0.25. (c) The probability distributions for a temperature of 0.100. (d) The probability distributions for a temperature of 0.0125.

For a 20-city TSP we present the results of accurate values of the basic variables given in Tables I and II and shown in Figs. 2(a)–2(d). These figures show that there are only small differences in the calculated thermodynamic quantities in the two ensembles. Thus, one conclusion we reach is that either ensemble may be used with confidence in simulated annealing calculations. In our studies of this TSP, we have not noticed an advantage of one ensemble over the other in reaching the ground state, but a systematic study would have to be done using various system sizes and annealing schedules, that is, rules to lower T in the canonical ensemble or E in the microcanonical ensemble. We have carried out similar calculations on another 20-city TSP with results similar to those in this paper, although, of course, the specific values will be different for a different city arrangement. We have also carried out similar calculations on larger TSPs and again we find a correspondence between the values in the two ensembles similar to those shown in Figs. 2(a)–2(d). For larger system sizes, the maximum in the specific heat shifts toward lower temperatures and is more peaked.

Although the thermodynamic quantities in Figs. 2(a)–2(d) are in close agreement, it is interesting that the probability distributions, shown in Figs. 3(a)–3(d), which measure the frequency of visiting a given configuration in the course of the calculation, are quite different in the two ensembles. At very high and very low temperatures, the distributions are nearly the same [compare Figs. 3(a) and 3(d)], whereas at intermediate temperatures, Figs. 3(b) and 3(c) show that the microcanonical distribution is more sharply peaked, narrower, and more symmetric about the maximum value. These distributions are reproducible from one run to the next and show that there is a real difference in the frequency of visiting different configurations during calculations in the two ensembles. Whether the microcanonical ensemble approach to simulated annealing has advantages over the canonical ensemble approach is an important question for future study. The results of this paper suggest that the microcanonical ensemble is at least as good as the canonical ensemble for simulated annealing applications.

ACKNOWLEDGMENTS

The authors thank Murray Daw and Malcolm Skove for helpful discussions.

[1] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[2] J. R. Ray, Phys. Rev. A **44**, 4061 (1991).
[3] J. R. Ray and C. Freléchoz, Phys. Rev. E **53**, 3402 (1996).
[4] P. J. Fay and J. R. Ray, Phys. Rev. A **46**, 4645 (1992).
[5] J. R. Ray and R. J. Wolf, J. Chem. Phys. **98**, 2263 (1993).
[6] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1982).
[7] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications* (Kluwer, Dordrecht, 1987), and references contained therein.
[8] G. Laporte, Eur. J. Oper. Res. **59**, 231 (1992), and references contained therein.
[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1992).