JOURNAL OF COMPUTATIONAL PHYSICS 39, 128–136 (1981)

On the Selection of the Particle to Be Perturbed in the Monte Carlo Method

MIHALY MEZEI

Department of Chemistry,
Hunter College of the City University of New York,
New York, New York 10021

Received March 4, 1980; revised July 2, 1980

A new method is proposed for the selection of the particle to be perturbed in the single-particle move Metropolis Monte Carlo method which combines many advantages of the previously used random and cyclic methods. Ergodicity and relative efficiency are discussed. A numerical comparison is presented on a system of 45 Lennard-Jones particles.

1. INTRODUCTION

The present paper deals with the introduction of a new method—called the shuffled cyclic method—for selecting the particle to be perturbed in a single particle move Metropolis Monte Carlo method and its analysis, along with the analysis of the two conventional methods: the cyclic and random methods [1].

In general, the discussion of a new algorithm requires two points: correctness and competitiveness. For the case of a stochastic algorithm, besides proving that the algorithm gives the correct answer in some infinite limit, the problem of ergodicity must be dealt with as well. Yet, the competitive comparison of algorithms is a problem that is in general ill-defined since: (a) Different criteria usually lead to different ranking–reflecting the fact that their strong points lie at different aspects of the problem; (b) the comparison depends on the actual choice of the system's parameters; (c) the modification of one algorithm in a system of algorithms may effect the efficiency of other algorithms in the system not under discussion.

The organization of the paper is as follows. Section 2 presents a general description of the three methods under discussion, Section 3 contains the analysis of the ergodicity of the cyclic and shuffled cyclic methods, and Section 4 compares different aspects of the three methods and presents a numerical comparison. The results of the paper are discussed in Section 5.

Throughout the paper we call the attempted moves perturbations and reserve the term move for actual (accepted) moves.

* This research was supported by NIH Grant 1-ROH-NS12149-03.

0021-9991/81/010128-09$02.00/0

2. The Description of the Three Methods

In the Monte Carlo calculations that are in the literature, the particle to be perturbed is selected either randomly (most of the time, but not always, with uniform distribution) or cyclicly, that is, repeatedly in a fixed sequence. It has been shown, by Woods, that the cyclic method leads to the same expectation value as the random method [1]. The question of its ergodicity will be discussed in the next section.

It should be pointed out first, as a drawback of the cyclic method, that it cannot be used for grand-canonical ensemble simulations, where the number of particles varies during the simulation.

An attractive feature of the cyclic method is that it ensures that all particles are perturbed the same number of times, and, after k cycles, each particle has been perturbed exactly k times. As a result, near-neighbour tables (an important device to speed up the calculation on large systems) will be required much less frequently, and averages computed after k cycles only (another, although less important time-saving device) will be more accurate. Furthermore, an approximate argument, presented in Section 4.2, and the numerical example presented in Section 4.3, will demonstrate that the correlation between the successive moves of the same particle, called cage-effect [2], will be reduced. On the other hand, successive steps in the Markov chain will necessarily be more correlated as a result of the fixed sequence of the perturbations.

The compromise solution proposed here is the shuffled cyclic method: for each cycle generate a random permutation of N elements (N being the number of particles considered) and use this sequence. Thus, with minimal extra work, much of the correlation between the successive steps of the Markov chain can be removed while the advantages of the cyclic method can still be retained.

3. The Ergodicity of the Cyclic Methods

The proposed shuffled cyclic method has the further advantage over the cyclic method that it is always ergodic on a system where the random procedure is ergodic. We will prove the following:

Theorem 1. *If the random procedure is ergodic on a given system then the shuffled cyclic procedure is ergodic as well.*

Proof. The ergodicity of the random procedure implies the existence of a sequence of configurations

$$S_R = \langle \mathbf{X}_0, \mathbf{X}_1^{i_1}, \ldots, \mathbf{X}_n^{i_n} \rangle, \qquad \mathbf{X}_n^{i_n} = \mathbf{X}_n, \tag{1}$$

for any two configurations $\boldsymbol{X}_0$, $\mathbf{X}_n$, where $\mathbf{X}_k^{i_k}$ can be obtained from $\mathbf{X}_{k-1}^{i_{k-1}}$ by an allowed move of the i_kth particle. In order to prove the theorem, we will construct a

sequence S_{SC} that has S_R as a subsequence and is realizable by a shuffled cyclic procedure. The sequence S_{SC} will be of the form

$$S_{SC} = \langle \mathbf{X}_0, \mathbf{X}_1^{i_1}, S_1, \mathbf{X}_2^{i_2}, S_2, \ldots, S_{n-1}, \mathbf{X}_n^{i_n} \rangle, \tag{2}$$

where S_i is a sequence of configurations on the empty sequence. The proof is by induction.

Since it is possible to select i_1 as the first particle of the first cycle, the sequence $\langle \mathbf{X}_0, \mathbf{X}_1^{i_1} \rangle$ is realizable with the shuffled cyclic method. Assume that we have constructed the sequence

$$S_{SC}^k = \langle \mathbf{X}_0, \mathbf{X}_1^{i_1}, S_1, \ldots, S_{k-1}, \mathbf{X}_k^{i_k} \rangle. \tag{3}$$

The following possibilities can occur: (a) It is possible to choose i_{k+1} as the next particle to be perturbed in the cycle (the previous cycle was just completed or i_{k+1} has not been perturbed yet in the current cycle). In that case S_k is the empty sequence and S_{SC}^{k+1} is realizable by the shuffled cyclic method. (b) i_{k+1} has already been perturbed in the present cycle, but it is possible to find perturbations for the remaining particles that can be rejected with finite probability. In this case S_k is again the empty sequence and S_{SC}^{k+1} is realizable by the shuffled cyclic method. (c) i_{k+1} has already been perturbed in the present cycle and for some of the remaining particles there is no perturbation that can be rejected. In this case, however, these particles can also be selected as the first particles to be perturbed in the next cycle and their move can be chosen to cancel the previous "forced" move. After the restoration of the configuration $\mathbf{X}_k^{i_k}$, one can select i_{k+1} as the next particle to be perturbed. Thus S_{SC}^{k+1} was found realizable by the shuffled cyclic method in this case too, with S_k the sequence of the "forced" moves and their reversals.

The case of the cyclic method is more involved since the analogous theorem is not true. A simple counterexample is a spin-lattice at infinite temperature [3]. Another counterexample, at finite temperatures, is presented by Friedberg and Cameron [4]. The analogous theorem can be proved, however, for systems in the (continuous) configuration space.

THEOREM 2. *If the random procedure is ergodic for a system of N particles in the configuration space, then the cyclic method will be ergodic too, provided that arbitrary small displacements are allowed in the procedure.*

Proof. If the random procedure is ergodic, then for every pair of orientations $\mathbf{X}_0$, $\mathbf{X}_n$ there is a sequence of orientations

$$S_R = \langle \mathbf{X}_0, \mathbf{X}_1^{i_1}, \ldots, \mathbf{X}_n^{i_n} \rangle, \qquad \mathbf{X}_n^{i_n} = \mathbf{X}_n, \tag{4}$$

such that $\mathbf{X}_k^{i_k}$ can be obtained from $\mathbf{X}_{k-1}^{i_{k-1}}$ by moving the i_kth particle. Furthermore, there is a neighborhood $\varepsilon_k(\mathbf{X}_k^{i_k})$ of $\mathbf{X}_k^{i_k}$ within which any point can substitute $\mathbf{X}_k^{i_k}$—otherwise the path between $\mathbf{X}_0$ and $\mathbf{X}_n$ can be realized only with zero

probability. Again, we will show the existence of an alternative sequence connecting $\mathbf{X}_0$ and $\mathbf{X}_n$ that can be realized by the cyclic method.

A configuration $\mathbf{X}$ will be called non-maximal with respect to particle i if the probability of rejecting a perturbation of i is greater than zero. This requires that there should be another configuration $\mathbf{X}'$ (in a countable space) or a set of measures greater than zero (in the continuous configuration space) that can be reached from $\mathbf{X}$ by a single move of particle i and $U(\mathbf{X})/\mathrm{k}T < U(\mathbf{X}')/\mathrm{k}T$. Here $U(\mathbf{X})$ is the internal energy of the configuration $\mathbf{X}$, T is the absolute temperature, and k is the Boltzmann constant.

Any sequence $\langle \mathbf{X}_{k-1}^{i_{k-1}}, \mathbf{X}_k^{i_k} \rangle$ can be realized by a full cycle if the configuration $\mathbf{X}_{k-1}^{i}$ is non-maximal with respect to the particles $1,..., i_k - 1, i_k + 1,..., N$. If, however, $\mathbf{X}_{k-1}^{i_{k-1}}$ is not non-maximal with respect to particles $m_1,..., m_q$, then these particles have to be moved in the cycle. Since a move can be arbitrarily small, we can choose these moves small enough that for each $i \leqslant q$, the subsequent points of the sequence $\langle \mathbf{X}_k^{i_k},..., \mathbf{X}_{l-1}^{i_{l-1}} \rangle$, $i_l = m_i$, will be changed only with an amount small enough that the modified sequence $\langle \mathbf{X}_k'^{i_k},..., \mathbf{X}_{l-1}'^{i_{l-1}} \rangle$ will still be realizable by the random procedure. Then, when the particle m_i is to be moved again, the move that was made by the random procedure is to be modified so that the coordinates of the m_ith particle coincide with the original configuration $\mathbf{X}_l^{i_l}$ in the sequence.

Following this procedure, the modified sequence either ends a $\mathbf{X}_n$ or at a point in $\varepsilon_n(\mathbf{X}_n)$. In the latter case, two cycles can move this sequence to $\mathbf{X}_n$. In the first cycle thus all particles are to be moved. These moves are now chosen to produce the final configuration $\mathbf{X}_n$.

For the case of systems with finite or countable infinite states only some restricted statements can be made. It can be immediately seen that in this case: (a) The random method is ergodic if (but not only if) $U/\mathrm{k}T$ is finite at all states; (b) the ergodicity of the random method implies the ergodicity of the cyclic method if all states are nonmaximal with respect to all particles.

It can also be shown that

THEOREM 3. *If for any two states S_a^k, S_b^k of the particle k that can be connected by a single move there is a state S_c^k such that the transitions $S_a^k \to S_c^k$ and $S_c^k \to S_b^k$ are both allowed, then the ergodicity of the random method implies the ergodicity of the cyclic method.*

Proof. The theorem is proved when it is shown that any transition $S_a^k \to S_b^k$ that can be realized with the move of a single particle, say k, can also be realized by complete cycle(s). Actually, at most two cycles are needed.

If both the initial and the final states are non-maximal with respect to all of the particles, then it is possible to reject the perturbation of all but the kth particle, thus achieving the transition $S_a^k \to S_b^k$ in one full cycle. If not, all particles that can be moved at all are to be moved in the first cycle. Particle k is to be moved to state S_c^k, whose existence is ensured by the condition of Theorem 3 and the others are moved arbitrarily. In the second cycle particle k is to be moved to S_b^k and the other moved

particles are to be returned to their original state. This achieves the required move in two full cycles.

4. Comparison of the Three Procedures

This section presents three different analyses. First, the Markov chain will be considered as a sequence of independent random configurations. Second, the extent of the cage-effect will be examined for the three methods. Third, a numerical example will be presented.

4.1. *Considerations under the Independence Assumption*

In Section 2 it was pointed out that both the cyclic and the shuffled cycles methods perturb each particle the same number of times. Besides the advantages mentioned earlier it can be shown that under the independence assumption the efficiency of the algorithm is improved by this property.

The extent to which a given calculation covers the configuration space can be charaterized by the displacement of the system in the configuration space. If we assume that the displacement of a single particle i is proportional to $n_i^{1/2}$, n_i being the number of times particle i was perturbed, then it can be immediately seen that

(a) The total system displacement

$$\sum_{i=1}^{N} |\mathbf{X}_i^n - \mathbf{X}_1^0|^2 \tag{5}$$

is the same for all three methods, but

(b) the average displacement of a particle,

$$\left(\sum_{i=1}^{N} |\mathbf{X}_i^n - \mathbf{X}_i^0|\right)\Big/N \tag{6}$$

is maximal if $n_1 = n_2 = \cdots = n_N$, since the Lagrange multiplicator method yields the optimal n_i's as the solution of the equations

$$\frac{\partial}{\partial n_i}\left(\sum_{i=1}^{N} cn_i^{1/2} + \lambda \sum_{i=1}^{N} n_i\right) = 0 \tag{7}$$

that give as a solution for all i

$$n_i = \sum_{i=1}^{N} n_i/N.$$

In fact, the function $cn_i^{1/2}$ can be replaced by any (continuous) function whose second derivative is negative.

The beneficial effect of keeping all n_i equal can be further demonstrated by

showing that under the independence assumption it minimizes the variance of the expectation values computed.

The argument of the theory of group sampling. If a quantity A is the sum of N different and independent quantities A_i with standard deviation D_i then the variance of A is minimal if the number of times A_i is sampled is proportional to $D_i^{1/2}$ [5]. Since in our case the average of a quantity over the Markov chain can be regarded as a sum of N different quantities, each being the sum of contributions when particle i was perturbed and each having the same standard deviations since the particles are identical, under the independence assumption the above argument applies.

In closing, it should be stressed that the above arguments do not prove the superiority of the cyclic methods since the effect of correlation between successive steps in the Markov chain, which tends to increase the variance, was neglected.

4.2. *Analysis of the Cage-effect*

The cage-effect is due to te fact that in a dense fluid a move of the particle will bring it to the proximity of another particle; thus the next move of the same particle is most likely to be in the reverse direction.

Define P_r as the probability that when particle i made a move to the proximity of particle j the particle remains at its position until the next perturbation of particle i. The conditional probability of particle j remaining unmoved after exactly k perturbations, P_r^k, is given by the simple expression

$$P_r^k = r^k, \qquad r < 1. \tag{8}$$

Here r is the probability that a perturbation is rejected. We then obtain

$$P_r = \langle P_r^k \rangle = \sum_{k=0}^{\infty} P_r^k P_p^{ij}(k), \tag{9}$$

where $P_p^{ij}(k)$ is the probability that particle j was perturbed k times between two successive perturbations of i. The expressions for $P_p^{ij}(k)$ for the three methods can be obtained from simple combinatorial arguments. For the random method.

$$P_k^{ij}(k) = 2^{-(k+1)}. \tag{10}$$

For the cyclic method,

$$\begin{aligned} P_p^{ij}(k) &= 1 \qquad \text{for} \quad k = 1, \\ &= 0 \qquad \text{for} \quad k \neq 1. \end{aligned} \tag{11}$$

For the shuffled cyclic method,

$$\begin{aligned} P_p^{ij}(k) &= \tfrac{1}{4} \qquad \text{for} \quad k = 0, 2, \\ &= \tfrac{1}{2} \qquad \text{for} \quad k = 1, \\ &= 0 \qquad \text{for} \quad k > 2. \end{aligned} \tag{12}$$

Substituting Eqs. (10)–(12) into Eq. (9) we obtain

$$\begin{aligned} P_r &= 1/(2-r) && \text{for random,} \\ &= r && \text{for cyclic,} \\ &= \tfrac{1}{4} + r/2 + r^2/4 && \text{for shuffled cyclic.} \end{aligned} \tag{13}$$

Simple considerations show that for all $r < 1$,

$$P_r(\text{random}) > P_r(\text{shuffled cyclic}) > P_r(\text{cyclic}). \tag{14}$$

The quantity P_r, however, is closely related to the magnitude of the cage-effect since the ability of the particle j to induce a reverse move of particle i is decreased if it is moved. This follows from the fact that in a dense fluid both energetic and geometric factors favor a move of particle j pointing away from particle i over a move pointing toward particle i. Thus it is plausible to assume that the magnitude of the cage-effect, which can be characterised by $\langle \cos \gamma \rangle$, where γ is the angle between two successive moves of the same particle, follows the inequalities analogous to (14):

$$\langle \cos \gamma \rangle \text{ (random)} < \langle \cos \gamma \rangle \text{ (shuffled cyclic)} < \langle \cos \gamma \rangle \text{ (cyclic).} \tag{15}$$

4.3. *Numerical Example*

The three methods were compared on a system of 45 Lennard–Jones particles at reduced density $\zeta^* = 0.9$ and at reduced temperature $T^* = 2.0$, that is, in the supercritical region. Face-centered cubic periodic boundary conditions were used with a spherical cutoff of 2.067σ. The maximum displacement in the x, y, z directions was 0.3σ, resulting in a 33–34% acceptance rate. Starting from a random configuration, about 100K steps were discarded. From that point, 900K were run using the random method. The 600K-th configuration was the starting point of a 700K run using the shuffled cyclic method. Furthermore, both runs were divided into 100K segments and the first configuration of each segment was used as the starting configuration of a 100K run using the respective other two methods. Further examination of the data led us to exclude the first two segments of the random run from the averaging.

The following quantities are used to characterize the runs:

(a) the difference between the average of all the runs, -187.18, and the average over the segment considered, $\langle \Delta E \rangle$;

(b) the error estimate 2σ, computed from 5K block averages by the prescription of Wood [1]; the block averages were found to satisfy statistical randomness test;

(c) the average of the cosine of the angle between two accepted moves, $\langle \cos \gamma \rangle$;

(d) the averages of the above quantities over all the runs, over the segments 3–9 and 10–16.

The results are collected in Table I. Based on these numbers, the following three conclusions appear to us warranted:

(1) None of the three methods have been found to be consistently more accurate than the others.

(2) Inequality (15) appears to hold on this system rather consistently, although the difference is rather small. Also, the difference between the shuffled cyclic and cyclic cage-effect is less than the difference between the random and shuffled cage-effect.

(3) The comparison of the averages over the two types of runs suggests that there are small but subtle differences between the Markov chain generated by different particle selection methods.

TABLE I

Numerical Comparison of the Three Methods on LJ Fluid

	Random			Cyclic			Shuffled cyclic		
	$\langle \Delta E \rangle$	2σ	$\langle \cos\gamma \rangle$	$\langle \Delta E \rangle$	2σ	$\langle \cos\gamma \rangle$	$\langle \Delta E \rangle$	2σ	$\langle \cos\gamma \rangle$
1	0.03	1.49	−0.158*				−2.69	1.94	−0.152
2	2.51	2.87	−0.157*	1.20	2.35	−0.146	0.89	1.70	−0.148
3	1.78	1.93	−0.155*	−0.35	2.49	−0.149	−0.40	2.01	−0.148
4	−0.46	1.70	−0.153*	0.20	2.94	−0.152	1.44	1.93	−0.156
5	0.75	1.68	−0.151*	−0.48	2.08	−0.148	−0.50	1.94	−0.147
6	−0.03	2.03	−0.154*	0.13	2.48	−0.150	0.61	2.53	−0.150
7	−0.58	2.17	−0.161*	−2.42	1.81	−0.148	−2.16	2.36	−0.146
8	−0.04	2.17	−0.149*	−1.79	2.44	−0.150	0.19	1.94	−0.151
9	1.68	2.61	−0.150	−1.51	1.94	−0.148	1.48	1.84	−0.149
10	0.49	2.59	−0.157	−0.87	1.92	−0.147	0.27	2.41	−0.154*
11	3.55	2.52	−0.154	0.30	2.35	−0.153	−1.73	2.61	−0.149*
12	0.91	2.37	−0.157	−1.47	2.27	−0.146	0.89	1.88	−0.153*
13	−0.06	2.55	−0.159	0.55	2.38	−0.148	0.77	2.02	−0.148*
14	2.09	2.12	−0.156	−2.43	2.16	−0.147	−0.37	2.39	−0.148*
15	−1.63	2.62	−0.156	−0.70	2.83	−0.152	1.38	1.89	−0.148*
16	1.07	2.10	−0.163	0.34	1.67	−0.143	−0.80	2.79	−0.156*
RN	0.76	2.05	−0.154	0.98	2.41	−0.150	0.97	2.08	−0.150
SC	1.40	2.41	−0.157	0.95	2.22	−0.148	0.89	2.28	−0.151
ALL	1.08	2.23	−0.156	0.96	2.32	−0.149	0.93	2.18	−0.150

Notes. (a) Segments 1, 2 are excluded from all averaging; (b) the last three lines refer to averages over segments 3–9, 10–16, and 3–16, respectively; (c) the last configuration of the run, marked with *, was used as the starting configuration for the next run with the three methods.

5. Discussion

The present paper introduced a new method for the selection of the particle to be moved in the Metropolis Monte Carlo method. The reason for the introduction of the new method is that while the cyclic method possesses several advantages over the random method, it introduces certain unwanted correlations into the Markov chain and in some instances it is not ergodic, in spite of the fact that the random method is ergodic on the same system. The suggested shuffled cyclic method removes part of the unwanted correlation and was proven to be always ergodic whenever the random method is ergodic on the same system. Furthermore, the extra computational work involved in the implementation of the shuffled cyclic method is negligible since a random permutation can be generated in cN steps [6]. The accuracy of the averages obtained using the different methods is difficult to compare since, on the one hand, the fact that the particles are moved the same number of times is in itself a variance reducing factor, and on the other, the increased correlations in the cyclic method are working in the opposite direction. Its effect, however, is difficult to assess quantitatively. The numerical example examined here has failed to show significant differences in the accuracy of averages.

Acknowledgments

The author is grateful to one of the referes of an early version of this paper for thorough reading and very helpful comments.

References

1. W. W. Wood, *in* "Physics of Simple Liquids" (H. N. V. Temperley, J. S. Rowlinson and G. S. Rushbrooke Eds.), Chap. 5, pp. 115–230, North-Holland, Amsterdam, 1968.
2. M. Rao, C. Pangali and B. J. Berne, *Mol. Phys.* **37** (1979), 1773.
3. J. Miyazaki, personal communication of W. W. Wood.
4. R. Friedberg and J. E. Cameron, *J.Chem. Phys.* **52** (1970), 6049.
5. Yu. A. Schreider (Ed.), "The Monte Carlo Method," p. 103, Pergamon Press, Oxford, 1966.
6. D. E. Knuth, "The Art of Computer Programming," Addison–Wesley, Reading, Mass., 1968.