Monte Carlo dynamics in global optimization

C. N. Chen,[1] C. I. Chou,[2] C. R. Hwang,[3] J. Kang,[1] T. K. Lee,[1] and S. P. Li[1]
[1]*Institute of Physics, Academia Sinica, Taipei, Taiwan*
[2]*Computing Center, Academia Sinica, Taipei, Taiwan*
[3]*Institute of Mathematics, Academia Sinica, Taipei, Taiwan*
(Received 26 January 1999)

Several very different optimization problems are studied by using the fixed-temperature Monte Carlo dynamics and found to share many common features. The most surprising result is that the cost function of these optimization problems itself is a very good stochastic variable to describe the complicated Monte Carlo processes. A multidimensional problem can therefore be mapped into a one-dimensional diffusion problem. This problem is either solved by direct numerical simulation or by using the Fokker-Planck equations. Above certain temperatures, the first passage time distribution functions of the original Monte Carlo processes are reproduced. At low temperatures, the first passage time has a path dependence and the single-stochastic-variable description is no longer valid. This analysis also provides a simple method to characterize the energy landscapes. [S1063-651X(99)06808-7]

PACS number(s): 02.60.Pn, 05.10.−a, 87.15.−v

Over the past several decades there have been dramatic increases in the number of global optimization problems. The problems arise in many diverse areas, such as sciences, engineering, and business [1]. The major difficulty with this kind of problem is the exponential increase of search space with system size. Hence most of the research effort has been devoted to inventing algorithms to speed up the search process. A number of generic methods, such as the simulated annealing (SA) [2], genetic algorithm (GA) [3], and neural network [4] have been used widely to study all sorts of global optimization problems. Based on these generic methods, specific algorithms are developed for particular problems. These methods are usually evaluated only in terms of their efficiency against other methods and there are no more quantitative and objective analyses employed to understand such methods.

The difficulty in quantitatively analyzing the optimization processes is due to the large number of variables involved and the lack of physical guidance to replace these variables by a few quantities, in analogy to the order parameters used in describing the phase transitions of real materials with an infinite number of degrees of freedom. In the protein folding problem (PFP), Socci *et al.* [5] have proposed two order parameters. Knowledge of the optimal state is essential in calculating the order parameters. Hence this is unlikely to be useful for general optimization problems where the optimal state is unknown in most cases. On the other hand, we must continue to look for a general way to characterize the optimization process with a few variables if we intend to have any semiquantitative understanding. We will show below that the cost function or the energy is a very natural stochastic variable to use. For most cases there is no need to introduce other variables. Using the energy as a basic physical quantity has an additional advantage where it provides a means to quantitatively characterize the ''energy landscapes'' of the problem that one is studying.

In this paper, we use fixed-temperature Monte Carlo (MC) dynamics to study four very different optimization problems: the traveling salesman problem (TSP) [6], the lattice version of the protein folding problem (PFP) [7], x-ray data analysis (XDA), and a multivariable function. XDA will be defined more clearly below. Both TSP and PFP involve discrete operations, while XDA and the multivariable function have continuous variables. Despite the very different nature of each problem, they share many common features. There exists a temperature $T_{\rm op}$ in each of the above cases at which the average time required to reach the global minimum is the shortest. At temperatures greater than $T_{\rm op}$, first passage time distribution functions (FPTDF) are all reasonably well fitted by inverse Gaussian distribution functions. This result suggests that an optimization process may be viewed as a biased diffusion in the one-dimensional space of the energy. This is explicitly demonstrated by a simulation of a one-dimensional diffusion. We also show that for cases with continuous variables we could use the Fokker-Planck (FP) equation [8] to reproduce FPTDF. The drift velocity and diffusion constant used in the FP equation or the one-dimensional simulation provide a nice way to characterize the energy landscape in an optimization process.

In the following, the function to be optimized in the MC process, namely, the cost function, will be denoted as the energy E. Just like SA [2], we will introduce an effective temperature T (T is in the same unit as E) and use the Boltzmann factor $\exp(-E/T)$ and the Metropolis algorithm [9] to determine the transition probability. Notice that T is not a real temperature, its use is to parametrize the transition probability. In algorithms without using the Boltzmann factor it is possible to find a similar parameter. Our MC process is similar to the Glauber dynamics [10] used to study Ising model.

For the protein-folding lattice model it is known [7] that in the MC dynamics the temperature dependence of the average first passage time to reach the global minimum $E_{\rm min}$ is a U-shape curve. This is also true for other optimization problems as shown in Fig. 1. In this figure the average MC steps required to reach $E_{\rm min}$ is plotted as a function of the temperature, Fig. 1(a) is for TSP of 24 cities [11], Fig. 1(b) is for XDA of 32 atoms and Fig. 1(c) is for an analytical function [Eq. (2)] with 20 variables. In obtaining the curves, we have taken the following strategy. In the case of the PFP,

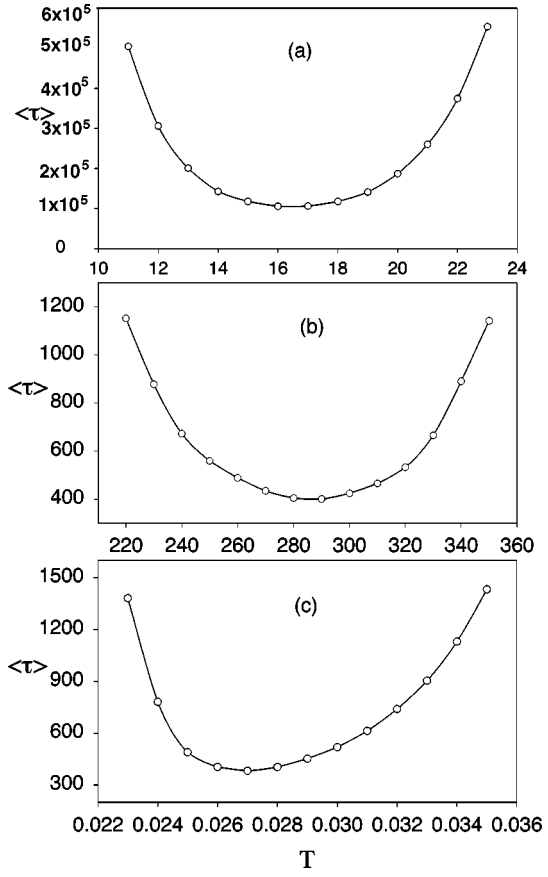

FIG. 1. Temperature dependence of the average MC steps required to reach a global minimum for (a) TSP, (b) XDA, and (c) Eq. (2).

since it is a discrete problem, we continue our simulation until it finds the global minimum. In the case of the XDA and Eq. (2), since they are problems in continuous spaces, the probability that the system is exactly at the global optimum is zero unless the temperature is zero. Therefore, we require instead the time to reach the global optimum to be the time that the system reaches a certain energy $E_{\min}$ which is low enough so that the system is already in the well that contains the global minimum. Since the system is already deep inside the well that contains the global minimum, the Gaussian fluctuation will eventually bring it down to the global minimum. If one uses a different $E_{\min}$, one would in this case alter the average MC steps to reach that energy but the U-shape curve is still maintained.

Many examples of the TSP are tabulated in the TSP library [11]. We have worked on several of them and obtained similar results. Here we shall only discuss our result for gr24 [11]. The cost function is the round trip distance traveled through these 24 cities. In each MC step we randomly select two out of the 24 cities and evaluate its possibility of interchanging paths irrespective of whether the move is accepted or not. At each temperature about several million samples are used to calculate the average MC steps required to reach $E_{\min}$.

The purpose of the XDA is to find the best possible crystal structure that fits the x-ray diffraction data [12]. The cost function is

$$E_{\mathrm{XDA}}(\{\mathbf{r}_j\},\lambda)=\sum_{\mathbf{k}}\ [\lambda|F(\mathbf{k})|-|F(\mathbf{k})_{\mathrm{obs}}|]^2, \qquad (1)$$

where $\mathbf{r}_j$ is the coordinate of the atom j, $\mathbf{k}$ is the Miller index of a reflection, $F(\mathbf{k})$ is the structure factor calculated from the atomic scattering factor $f_j(\mathbf{k})$ by $F(\mathbf{k})=\Sigma_j f_j(\mathbf{k})\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_j)$, $F(\mathbf{k})_{\mathrm{obs}}$ is the square root of the intensity observed or measured, and λ is the scale factor between $F(\mathbf{k})$ and $F(\mathbf{k})_{\mathrm{obs}}$. The result in Fig. 1 is for the case of a hypothetical crystal with 32 carbon atoms in a cubic cell. The optimization process is carried out by moving the atoms around until $E_{\mathrm{XDA}}=0$. Here the MC step represents the attempt to move all 32 atoms. Usually only a few atoms will move in each sweep.

For the PFP, following Socci *et al.* [7] we only consider the HP model on a $3\times3\times3$ cubic lattice. The amino acid is either hydrophobic (H) or polar (P). A protein with 27 monomers of either H or P type is folded into a compact form on the cubic lattice by choosing one of three operations successively. The energy is determined by the number of HH and PP nearest-neighbor pairs. There is no energy for HP pairs. We obtain the same results as Ref. [7], hence it is not shown in Fig. 1. For our discussion below, we will use the 002 sequence from their paper.

TSP and the lattice version of PFP only involve discrete operations in each MC step and they belong to the class of combinatorial optimization problems. Usually they are considered to be different from problems with continuous variables such as the XDA. To make a more careful comparison we also studied another function with continuous variables. The function is of the form

$$E_{fn}=\left(\sum_{i=1}^{N}x_i^2\right)^{1/2}+1-\prod_{i=1}^{N}\cos(2\pi x_i). \qquad (2)$$

We choose the number of variables N to be 20. The energy has the form of a multidimensional cone as represented by the first term added with an oscillatory function given by the last term. It roughly mimics a funnel energy surface with local minima and barriers.

Besides the four examples discussed above, the results shown in Fig. 1 are also known in many other optimization problems. It is straightforward to come up with a qualitative reason for the U-shape and its generality. At high temperatures there are too many paths available and it would take quite a long time to locate the paths that would reach the global minimum. The situation improves as temperature decreases. However, as temperature decreases further, the potential barriers and energy traps make the system difficult to move out of the local minima. Hence there is a temperature, T_{op}, at which the two factors balance each other and the system approaches the global minimum with the shortest time. If the above reasoning is correct, then above T_{op} the system behaves more ergodic and less path-dependent. Our analysis below using the FP equation reconfirms this understanding.

The common behavior of the average first passage time to reach the global minimum $E_{\min}$ as a function of temperature suggests taking a more careful examination of the *first passage time distribution function* (FPTDF) itself rather than

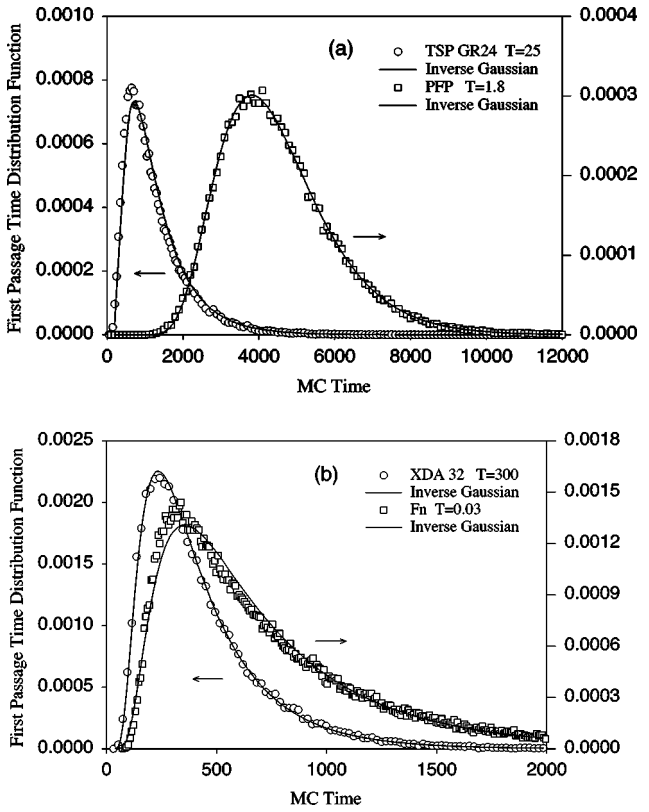

FIG. 2. The first passage time distribution function to reach equilibrium for four optimization problems. In (a) the circles are for TSP-24 at $T=25$ and squares for PFP at $T=1.8$; (b) the circles are for XDA at $T=300$, and squares for E_{fn} of Eq. (2) at $T=0.03$.

just the average time, which is the first moment of the distribution. A more careful consideration also questions the suitability of using $E_{\min}$ as the final state in the first passage time measurement. In the MC dynamics at a fixed temperature, the system tends not to go to the global minimum but most likely to stay in states with the equilibrium energy $E_{\rm eq}$ determined by the particular temperature. As the system relaxes toward its equilibrium energy $E_{\rm eq}$, it would then fluctuate around this equilibrium value. The Gaussian-like energy fluctuation would bring the system from $E_{\rm eq}$ to $E_{\min}$ as long as there are no insurmountable barriers in between. In this paper our FPTDF is always measured by reaching the $E_{\rm eq}$ at the specific temperature. How this is changed when $E_{\min}$ is used as the final state will be discussed later.

In Fig. 2, FPTDF are plotted for the four problems discussed above: TSP-24, PFP, XDA, and Eq. (2). The distribution functions are normalized. All four curves can be fitted quite well by the inverse Gaussian distribution function [13] (IGDF), although it is less spectacular for the two combinatorial optimization problems, TSP and PFP [14,15].

Similar to the Gaussian distribution function, the IGDF is also a solution of the one-dimensional diffusion equation with a constant drift velocity. However, the Gaussian distribution function gives the probability density at time t for a diffusive particle with a net drift velocity to be at a distance d from the starting point. On the other hand, IGDF gives the probability density of the first passage time of the particle to arrive at distance d and it takes the form

$$f(t)=\left(\frac{\lambda}{2\pi t^3}\right)^{1/2}\exp\left(-\frac{\lambda(t-\mu)^2}{2t\mu^2}\right), \tag{3}$$

where $\mu=d/v$ and $\lambda=d^2/\sigma^2$. v is the drift velocity and $D=\sigma^2/2$ is the diffusion constant.

The success of fitting the FPTDF by the IGDF is quite surprising since the original MC process involves many variables while IGDF describes a diffusion motion in one dimension. The result strongly suggests the possibility of using one variable to understand these MC processes. The cost function or energy therefore appears to be the natural variable of choice.

Once one decides to have the energy as the only stochastic variable, the original multivariable MC processes could be mapped into a one-dimensional diffusion equation in the energy space. A well known approximate approach to describe the diffusion process is the FP equation. Hence we shall consider the FP equation first. However, it seems to be unsuitable for combinatorial optimization problems, and a direct MC simulation of one-dimensional diffusion works better.

In the MC process, the system starts from a large E and diffuses toward lower energy states. At a certain energy E, the transition to the state $E+\triangle E$ in the next step is determined by the transition probability $W_E(\triangle E)$. This transition probability is measured during our MC simulation directly. According to the standard derivation of the FP equation [8], the first and second moments of the transition probability $W_E(\triangle E)$ determine the drift velocity and diffusion coefficient $A(E)$ and $B(E)$, respectively. The FP equation for the probability density $P(E,t)$ is then given by

$$\frac{\partial P(E,t)}{\partial t}=-\frac{\partial}{\partial E}\left(A(E)-\frac{1}{2}\frac{\partial}{\partial E}B(E)\right)P(E,t). \tag{4}$$

$A(E)$ and $B(E)$ are plotted as a function of E for TSP-24 [17] at $T=25$ in Figs. 3(a) and 3(b), respectively, and for E_{fn} of Eq. (2) in Figs. 3(c) and 3(d). Clearly $A(E)$ and $B(E)$ are very different for TSP and E_{fn}. The result of XDA is similar to E_{fn} while that of PFP is similar to TSP. We should also mention here that the equilibrium value of the TSP here is about 1390, while that of Eq. (2) is about 0.42 for $T=0.031$ and 0.336 for $T=0.025$, respectively.

In Fig. 3(a), $A(E)$ is approximately a straight line and crosses zero at the equilibrium energy. If we approximate $A(E)=-g(E-E_{\rm eq})$ and assume $B(E)=\sigma^2$ to be independent of E, then the FP equation describes the famous Ornstein-Uhlenbeck process [8] and it can be solved analytically. Starting from E_0, the FPTDF to reach $E_{\rm eq}$ is of the form

$$f(t)=2\frac{yd}{\sigma\pi^{1/2}}\left(\frac{g}{1-y^2}\right)^{3/2}\exp\left(-\frac{gy^2d^2}{\sigma^2(1-y^2)}\right), \tag{5}$$

where $y=\exp(-gt)$ and $d=E_0-E_{\rm eq}$. For very small and very large t this equation and the inverse Gaussian distribution function of Eq. (3) are very similar. However, we can also solve the FP equation numerically without making the approximations to $A(E)$ and $B(E)$. The numerical result for the FPTDF is plotted in Fig. 4(a). The circles are the result of Monte Carlo simulation for the original TSP-24 problems.

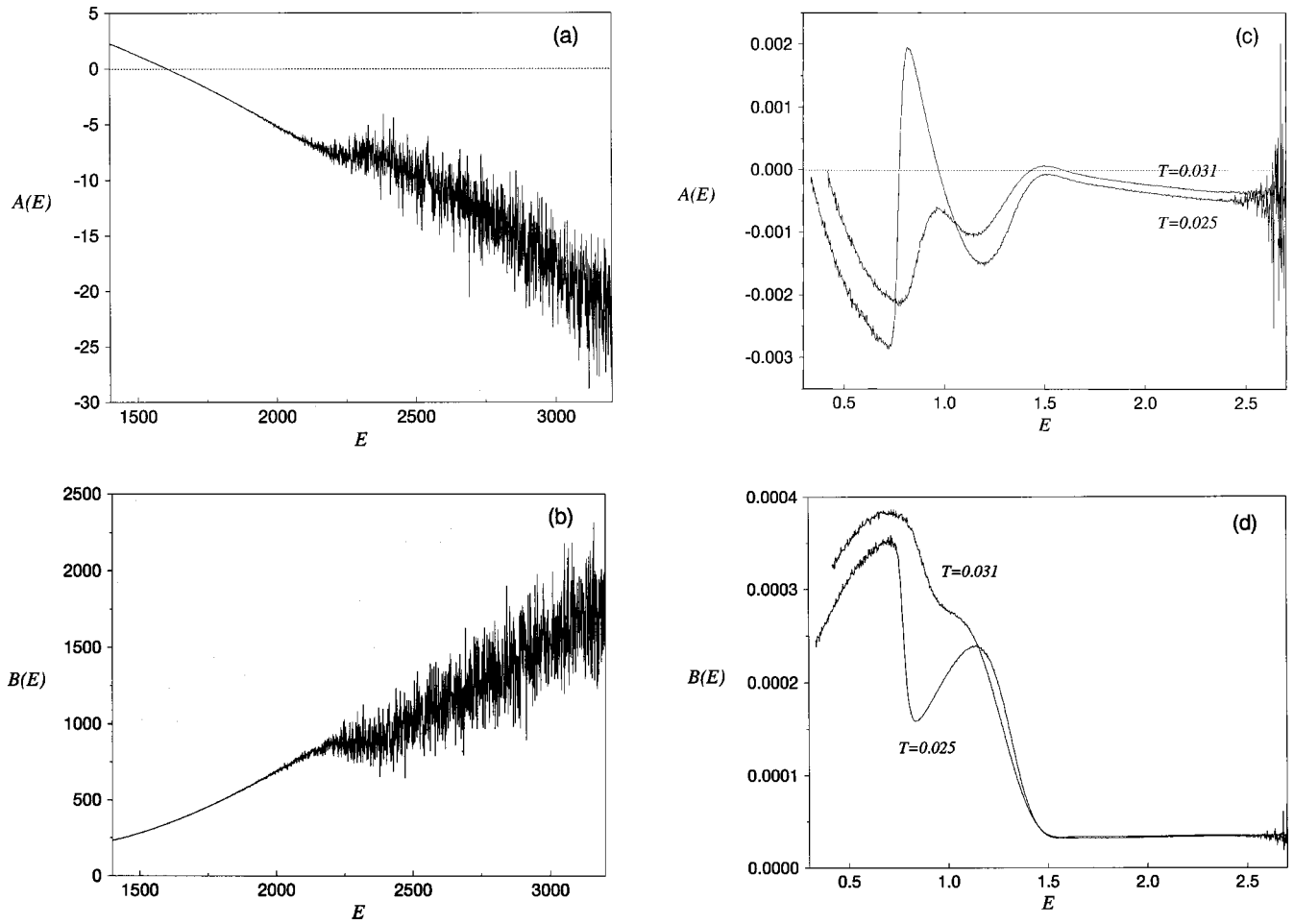

FIG. 3. Functions $A(E)$ and $B(E)$ of Eq. (4) as a function of E (a) and (b) for TSP-24 at $T=25$; and (c) and (d) for E_{fn} at $T=0.025$ and $T=0.031$.

In Figs. 3(c), and 3(d), the drift velocity $A(E)$ is fairly small and with little variation until it reaches about $E=1.5$. The temperature dependence indicates that there are two barriers around $E=0.8$ and $E=1.5$ at $T=0.025$. Similar scenario also appears at $T=0.031$. FPTDF obtained by integrating FP equation numerically is plotted as the solid line in Fig. 4(b). It agrees very well with the original simulation result shown as the circles. The fact that there are barriers and local wells in these optimization problems is expected. Take, for example, the multivariable function [Eq. (2)] that we are studying. Let us take the number of variable N to be 1 instead of 20. It is a quadratic function plus a cosine function. With the cosine function alone, it is easy to see that there are local wells when x is equal to $\pi, 3\pi$, etc. One can thus expect that the function E_{fn} has local wells and barriers as the energy decreases in this case. As one increases the number of variables, there should be more local wells and barriers but the phase space that one can wander around also increases. It is therefore conceivable that most of the barriers and wells at higher energies would have fewer effects because there are more paths for one to go down the hill. As the energy gets lower, effects of large local barriers and wells will begin to contribute. This is indicated in Figs. 3(c) and 3(d). In the case of the TSP, the smoothness of the functions $A(E)$ and $B(E)$ suggests that the local barriers and wells are still relatively small as compared to the energy scale that one can hop in the energy range plotted in Figs. 3(a) and 3(b).

Notice that the TSP result shown in Fig. 4(a) does not agree so well with the FP result as for E_{fn}. Indeed, the numerical data for this figure are the case where we have truncated the steps where they involve long jumps, i.e., large ΔE. The data for this figure involve ΔE always less than 100. We have tried several other truncations and the fit gets worse as the allowed ΔE gets larger. We believe that this is mainly due to the discrete nature of the MC operation in TSP. In the TSP MC process, there is a possibility for very large change in energy due to a single exchange of two cities. Hence the transition probability $W_E(\triangle E)$ has very long tails with small weight, which is similar to the famous Levy flight [8]. The second moment $B(E)$ used in the FP approach cannot fully account for this behavior. It is worth mentioning here that as one truncates the large ΔE, one makes these discrete models closer to the continuous models, as supported by our study of the sets of data with various truncations. Since combinatorial optimization problems usually involve large change of energy, the FP approach is therefore not appropriate.

Since we know the transition probability $W_E(\triangle E)$ at all the energy E, it is then not necessary to use the FP equation which only uses the first and second moments of the transition probability. We could calculate FPTDF by just doing a one-dimensional MC simulation. In Fig. 5, FPTDF obtained

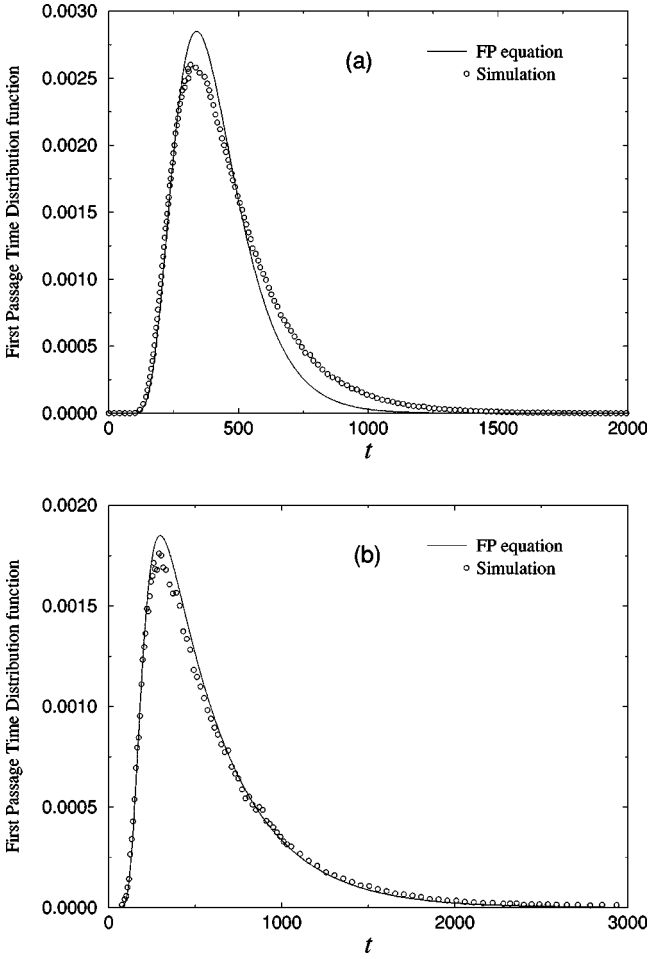

FIG. 4. FPTDF obtained by solving FP [Eq. (4)] with the input of Fig. 3. In (a) the original MC result for TSP-24 is shown by circles and solid lines are FP results; (b) is for E_{fn} at $T=0.031$.

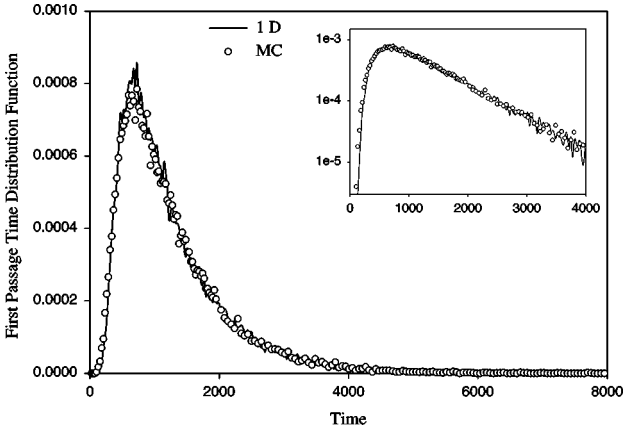

FIG. 5. FPTDF obtained from the one-dimensional diffusion in the energy space (the solid line) and the original multivariable MC result (circles) for TSP-24 at $T=25$.

by using the transition probability $W_E(\triangle E)$ (solid line) in a one-dimensional MC simulation is compared with the original FPTDF obtained for TSP-24. Excellent agreement supports the idea of using energy as the only stochastic variable.

Another consequence of the success of the simple one-dimensional model is the realization that at $T>T_{\mathrm{op}}$, the information about the particular path is unimportant and the statistical information about all the paths represented by $W_E(\triangle E)$ is enough. This is usually referred to as the ergodic property in statistical physics. It is surprising to see that the ergodic assumption actually works.

Since E_{eq} is the most likely state that the system will stay at the corresponding temperature, the system is more likely to go up to the equilibrium state than to go down to lower energy states when it is at energy $E<E_{\mathrm{eq}}$. Hence at $E<E_{\mathrm{eq}}$ the drift velocity or $A(E)$ is positive. This is very different from the case with $E>E_{\mathrm{eq}}$. Thus at finite temperatures the FPTDF to reach the global minimum E_{min} does not have the form of IGDF and it has more paths with longer first passage time.

At temperatures below T_{op}, E_{eq} is very low. Before the system reaches E_{eq}, its drift velocity either vanishes or becomes positive as shown in Fig. 3(c). We do not expect the approach to remain valid here. At very low temperatures, different paths will have very different time scales involved. The system behaves nonergodic in the time scale of the simulation. The statistical information stored in $W_E(\triangle E)$ is thus not so useful. The functions $A(E)$ and $B(E)$ for XDA are very similar in nature to those of Eq. (2) as shown in Figs. 3(c) and 3(d), where one can see large barriers as the system goes to lower energy while those of the PFP are similar to the ones corresponding to the TSP in Fig. 3(a) and 3(b). The nonergodic nature can also be reflected in the plots of FPTDF for the cases studied here where the curves corresponding to those in Fig. 2 cannot be fitted by IGDF anymore, though they are not shown here. They all have long time tails, which is related to the fact that the system is trapped in local minima as the system searches for its global minimum.

Figure 3(b) shows that the effective diffusion coefficient $B(E)$ becomes very small at low energy. Vanishing drift velocity and small diffusion coefficient imply that the system is confined to move in a relatively small region. Unless the path to global minimum is within this region, it would be very difficult for the system to locate it and this essentially becomes a trapped state.

We notice that Figs. 3(a) and 3(b) demonstrate a typical funnel energy surface [16]. The drift velocity $A(E)$ decreases as energy is lowered and similarly the diffusion coefficient $B(E)$ or the size of the neighborhood also shrinks.

In summary, we have shown that four different optimization problems have many common features in the way they approach their equilibrium states. The average first passage time to reach the global minimum at a fixed temperature Monte Carlo process has a minimum at a certain temperature T_{op}. At $T>T_{\mathrm{op}}$, FPTDF are approximated well by the IGDF defined by Eq. (3). This suggests that we could use the cost function or the energy as the stochastic variable in describing these complicated MC processes. With the energy as the only stochastic variable, we further show that FPTDF could be derived more systematically by solving the FP equation. This approach works well for the optimization problems involving continuous variables. On the other hand, for combinatorial optimization problems, we need to use all the information contained in the transition probability $W_E(\triangle E)$.

Nevertheless, the original multivariable MC processes in either case are mapped into a one-dimensional diffusion problem.

This result also indicates that at temperatures above T_{op} the MC processes are dominated by entropy and the system is more or less ergodic. Below T_{op}, there are traps and barriers that not only slow down the search for the global minimum but also make the system behave in a nonergodic way in the time scale of the simulation. Since at T_{op} we need the smallest number of MC steps to find the global equilibrium, it is very advantageous for the optimization algorithm to quickly locate this temperature. The failure of treating the multivariable MC processes as a one-dimensional diffusion problem near or below T_{op} might be useful information for this purpose.

This work was partially supported by a grant from Academia Sinica, Republic of China. Parts of the computations were performed at the National Center for High-Performance Computing in Taiwan. We are grateful for their support. T.K.L. would like to thank R.K.P. Zia and R. Swendsen for many stimulating and important discussions.

[1] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization* (Kluwer Academic Publishers, Boston, 1995).

[2] P. J. M. van Laarhovan, *Simulated Annealing, Theory and Applications* (Kluwer Academic Publishers, Boston, 1989).

[3] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison Wesley, Reading, MA, 1989).

[4] B. Muller and J. Reinhardt, *Neural Networks: An Introduction* (Springer-Verlag, New York, 1990).

[5] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **32**, 136 (1998).

[6] G. Reinelt, *The Traveling Salesman: Computational Solution for TSP Applications* (Springer-Verlag, New York, 1994).

[7] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994); **103**, 4732 (1995).

[8] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).

[9] N. Metroplis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[10] R. J. Glauber, J. Math. Phys. **4**, 294 (1963).

[11] TSP library, http://www.crpc.rice.edu/softlib/tsplib.

[12] W.-P. Su, Acta Crystallogr., Sect. A: Found. Crystallogr. **51**, 845 (1995); Physica A **221**, 193 (1995).

[13] V. Seshadri, *The Inverse Gaussian Distribution* (Oxford University Press, New York, 1993).

[14] J. W. Haus and K. W. Kehr, Phys. Rep. **150**, 263 (1987); A. N. Borodin and P. Salminen, *Handbook of Brownian Motion - Facts and Formulae* (Birkhäuser Verlag Basel, Basel, 1996).

[15] Similar results also have been found in other optimization problems such as the quadratic assignment problem, spin glass, etc.

[16] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).

[17] Here we only consider processes with an energy change of less than 100.