Simulated annealing using the classical density distribution

Jianpeng Ma and John E. Straub
Department of Chemistry, Boston University, Boston, Massachusetts 02215

(Received 14 February 1994; accepted 23 March 1994)

Three algorithms for global energy minimization based on the simulated annealing of the classical density distribution are presented. The first is based on annealing the classical density distribution directly in temperature and is the classical analog of imaginary time quantum dynamics. Another two algorithms are based on the approximate solution of the classical Liouville equation for the dynamics of a system coupled to a heat bath using a rigid temperature constraint and Fokker–Planck dynamics. These three methods are compared with standard simulated annealing based on molecular dynamics. The results for a series of Lennard-Jones clusters demonstrate that by annealing the continuous density distribution (representing a volume of phase points) the likelihood of finding the global minimum is dramatically enhanced.

I. INTRODUCTION

Many important questions in chemical physics, ranging from the structure of a molecular cluster[1] to the conformation of a protein,[2] can only be answered through the solution of a complex nonlinear optimization problem involving a multidimensional potential energy hypersurface. The number of degrees of freedom in the system is often large, and the number of minima on the associated potential energy surface is so great that a straightforward, exhaustive search of the potential minima is intractable. In addition, the potential energy landscape of systems such as glasses and proteins is very rough, with minima typically separated by high energy barriers which are not easily crossed by a straightforward molecular dynamics or Monte Carlo exploration at physical temperatures.[3]

A variety of numerical algorithms have been proposed to solve complex optimization problems. The most widely applied method is "simulated annealing."[4] In the standard application, the system is simulated using molecular dynamics (MD) or Monte Carlo starting at a high temperature. The system is slowly cooled to a low temperature at which point it has been annealed into a low lying, hopefully global, energy minimum. While simulated annealing using molecular dynamics is easily implemented and at times highly successful, there are many optimization problems involving complex molecular systems where it is impotent.

Recently, a new class of algorithms has been developed and proved to be more effective than MD simulated annealing for the minimization of atomic and molecular systems. What is common to these algorithms is that at first the complex potential function is "smoothed" to remove high lying minima. The global minimum of the smoothed potential function is then found by a steepest descent energy minimization or Monte Carlo search on the smoothed potential. Finally, the minimum of the smoothed potential surface is mapped back onto the original potential by an inverse transformation. The hope is that the global minimum of the smoothed surface maps onto the global minimum of the original potential energy function. Examples include the antlion method of Stillinger and Weber,[5] the diffusion equation method of Piela, Kostrowicki, and Scheraga,[6,7] and the potential shift method of Pillardy, Olszewski, and Piela.[8]

In this paper we present three algorithms for the simulated annealing of the continuous classical density distribution. These algorithms display all of the positive aspects of the algorithms mentioned earlier (use of a smoothed potential surface with fewer minima and lower barriers). However, there are important differences. (1) Each method is based on the approximate solution of a fundamental equation of motion (either the classical Liouville equation or the Bloch equation) and can be improved upon by systematically improving the representation of the density distribution. (2) While classical simulated annealing is based on molecular dynamics of a single point particle, the classical density distribution provides information on an ensemble of systems. (3) Course graining of the potential energy surface arises naturally from an average of the potential surface over the extent of the ensemble of systems contributing to the density distribution rather than an *ad hoc* transformation.

We describe the optimization methods and provide a detailed comparison both between methods and with classical MD simulated annealing. Our test system is a series of Lennard-Jones clusters which has become a standard benchmark for algorithms developed for global optimization of chemical systems. Each algorithm based on annealing the classical density distribution proves to be far more successful than standard MD simulated annealing with only a small increase in computational overhead.

II. METHODS

We first present the Gaussian density annealing (GDA) algorithm for performing simulated annealing directly in temperature rather than time, and then outline the Gaussian phase packet (GPP) dynamics algorithm based on approximate solution of the Liouville equation, which incorporates a rigid constant temperature constraint. We also present an alternative constant temperature Fokker–Planck dynamics method, where the system temperature is controlled by coupling to a constant temperature heat bath.

A. Gaussian density annealing

In simulated annealing one must define a practical cooling schedule which will lead to a well annealed system. For

 0021-9606/94/101(1)/533/9/$6.00

most problems involving the global optimization of molecular systems there is a range of energy scales which makes the optimal cooling schedule computationally forbidden. One needs to run initially at a high temperature where the largest barrier $E^{\ddagger}_{\max}$ in the system is easily crossed. At the same time, there is an energy scale which marks small differences between the global energy minimum and other low lying minima (ΔE). The larger the ratio $E^{\ddagger}_{\max}/\Delta E$, the more demanding the simulation and this is the case for many systems of chemical interest such as molecular clusters and proteins. As such it seems desirable to avoid the use of real time MD and a cooling schedule and perform the annealing directly in temperature. The GDA algorithm is related to the simulated annealing method but is not based on running real time dynamics (or Monte Carlo). Rather, it is based on simulating the annealing of the equilibrium density distribution directly in temperature.

We start from the normalized equilibrium density distribution

$$\rho_{\rm eq}(r,p)=\frac{1}{Q(\beta)}\,e^{-\beta H(r,p)}, \tag{1}$$

where $H(r,p)$ is the classical Hamiltonian and $Q(\beta)=\int dr\,dp\,\exp(-\beta H)$. The variation of the normalized equilibrium density distribution is described by the first order equation[9]

$$\frac{\partial \rho_{\rm eq}}{\partial \beta}=-(H-\langle H\rangle)\rho_{\rm eq}. \tag{2}$$

Similarly, for any property $A(r,p)$ the statistical thermodynamic average will be

$$\langle A\rangle=\frac{\int dr\,dp\,e^{-\beta H}A(r,p)}{\int dr\,dp\,e^{-\beta H}}=\int dr\,dp\,\rho_{\rm eq}(r,p)A(r,p). \tag{3}$$

The general equation for the temperature dependence of $\langle A\rangle$ is found by differentiating Eq. (3) with respect to β

$$\frac{d\langle A\rangle}{d\beta}=-\langle HA\rangle+\langle A\rangle\langle H\rangle. \tag{4}$$

To obtain an estimate of the global energy minimum we seek an approximate solution for the equilibrium density distribution at zero temperature. We employ a mobile basis of Gaussian density functions whose parameters are determined by a solution of Eq. (4). Throughout, we assume that we have integrated over the momenta and work with the reduced density distribution function $\rho(r,\beta)$. For each particle, we approximate the density distribution function as a single Gaussian packet in d dimensions

$$\rho(r,\beta)=(2\pi\sigma^2)^{-d/2}\exp\left[-\frac{1}{2}\left(\frac{r-r_0}{\sigma}\right)^2\right]. \tag{5}$$

More generally, $\rho(r,\beta)$ could be represented in a basis set expansion. We neglect the explicit temperature dependencies of r_0 and σ for notational simplicity. The equations of motion for the Gaussian approximation to the density distribution of each particle are determined by specifying the equations of motion for the packet center, $r_0=\langle r\rangle$, and second moment, $M_2=\langle (r-r_0)^2\rangle=d\sigma^2$. Direct substitution of r and $(r-r_0)^2$ into Eq. (4), and integration by parts, leads to a set of first order differential equations for the variation of the Gaussian density distribution in β (reciprocal temperature)[10]

$$\frac{\partial r_0}{\partial \beta}=-\frac{1}{d}M_2\nabla_{r_0}\langle V\rangle,\quad \frac{\partial M_2}{\partial \beta}=-\frac{1}{d^2}M_2^2\nabla^2_{r_0}\langle V\rangle, \tag{6}$$

where $\langle V\rangle$ is the pair potential averaged over the density distribution

$$\langle V\rangle=\int dr\rho(r)V(r). \tag{7}$$

In many-body systems the total N-body density distribution $\rho(r^N,\beta)$ may be approximated as a Hartree product of the distributions $\rho_k(r_k,\beta)$ of the N individual particles

$$\rho(r^N,\beta)=\prod_{k=1}^{N}\rho_k(r_k,\beta). \tag{8}$$

Similarly, a pairwise additive potential energy can be written as a sum over the individual pair interaction energies

$$\langle V\rangle=\sum_{i>j=1}^{N}\int dr_i\int dr_j\rho_i(r_i)\rho_j(r_j)V(|r_i-r_j|). \tag{9}$$

Equations (6), (8) and (9) define what we will refer to as the Gaussian density annealing algorithm (GDA).

For $\beta=0$ (infinite temperature) the density distribution is flat—all phase points are equally likely and $\rho(r,\beta=0)=1/Q(\beta=0)$. Therefore, one may choose as an initial condition the known infinite temperature ($\beta=0$) distribution where M_2 is effectively infinite for any r_0. The equations of motion for the center and width of the distribution are then integrated from $\beta=0$ to a finite temperature where the resulting $\rho(r,\beta)$ will be an approximation to the equilibrium density distribution at that temperature.[11]

When the description of the density distribution is complete, a solution will provide the exact global energy minimum geometry (or set of degenerate minima) in the limit $\beta=\infty$. When the density distribution is approximated as a product of single particle Gaussian distribution functions, for some potentials the result will be approximate. Note that the integration may be stopped at intermediate β where the density distribution will be an approximation to the equilibrium thermal density distribution at that temperature. In the following two sections we will present modifications of the simulated annealing protocol where the continuous classical density distribution is dynamically annealed.

B. Gaussian phase packet dynamics

Ma, Hsu, and Straub[12] have developed a dynamic simulation method using a Hartree product of Gaussian phase packets to represent the many body phase space density distribution. This method approximately solves the classical Liouville equation

$$\frac{\partial}{\partial t}\rho(r,p,t)=-L\rho(r,p,t), \tag{10}$$

where L is the Liouville operator. The equations of motion for the GPP's were derived for both constant energy and constant temperature dynamics. Each degree of freedom in d dimensions was described by a single spherically symmetric GPP

$$\rho(r,p,t)=\left(\frac{(1-\alpha^2)^{1/2}}{2\pi\sigma_a\sigma_b}\right)^d \exp\left[-\frac{1}{2}\left(\frac{r-r_0}{\sigma_a}\right)^2 -\frac{1}{2}\left(\frac{p-p_0}{\sigma_b}\right)^2-\alpha\left(\frac{r-r_0}{\sigma_a}\right)\left(\frac{p-p_0}{\sigma_b}\right)\right]. \quad (11)$$

In our notation we neglect the explicit time dependence of r_0, p_0, σ_a, σ_b, and α. The time dependence of the GPP parameters is completely specified by the time evolution of the average position $\langle r\rangle$ and momentum $\langle p\rangle$ and the second-order moments in d dimensions.[13]

The temperature was rigidly constrained using Gauss's principle[14] following the work of Evans and Hoover. We write the generalized Liouvillian as

$$L=L_0+L_c=L_0-\gamma\frac{\partial}{\partial p}\cdot p, \quad (12)$$

where L_0 is the "streaming" operator and L_c is the "collision" operator which couples the system to a heat bath of temperature T. The system-bath coupling is modulated by the parameter γ which is determined by the external constraint equation

$$\frac{d\langle p^2\rangle}{dt}=0 \quad (13)$$

which guarantees that the temperature remains constant at every instant in time.

Consequently, the time evolution of the first and second-order moments is given by

$$\dot r_0=\frac{p_0}{m},\quad \dot p_0=-\nabla_{r_0}\langle V\rangle-\gamma p_0,\quad \dot M_{2,0}=\frac{2}{m}M_{1,1},$$

$$\dot M_{1,1}=\frac{1}{m}M_{0,2}-\frac{1}{d}M_{2,0}\nabla^2_{r_0}\langle V\rangle-\gamma M_{1,1},$$

$$\dot M_{0,2}=-\frac{2}{d}M_{1,1}\nabla^2_{r_0}\langle V\rangle-2\gamma M_{0,2}. \quad (14)$$

γ is determined by the temperature constraint equation

$$\gamma=\frac{p_0\cdot F_0-M_{1,1}\nabla^2_{r_0}\langle V\rangle/d}{p_0^2+M_{0,2}}, \quad (15)$$

where we have used the fact that $\langle p^2\rangle=p_0^2+M_{0,2}$.

We can also define the annealing algorithm based on the general constraint equation

$$\frac{dT}{dt}=-\eta(t). \quad (16)$$

The value of $\eta(t)$ determines the cooling schedule in the annealing protocol where the generalization of Eq. (15) is

$$\gamma=\frac{p_0\cdot F_0-M_{1,1}\nabla^2_{r_0}\langle V\rangle/d-\eta}{p_0^2+M_{0,2}}. \quad (17)$$

For example, to follow an exponential cooling schedule, $\eta=T/\tau$, where τ is the cooling rate and the temperature $T=d(p_0^2+M_{0,2})/2k_B$.

C. Fokker–Planck dynamics

An alternative means of applying the constant temperature constraint is based on coupling the system to a heat bath using the weak collision Fokker–Planck (FP) operator[15–17]

$$L_c=-\gamma\frac{\partial}{\partial p}\cdot\left(p+mk_BT\frac{\partial}{\partial p}\right). \quad (18)$$

The equations of motion for the FP dynamics are identical to Eq. (14) where the equation for $M_{0,2}$ is replaced by

$$\dot M_{0,2}=-\frac{2}{d}M_{1,1}\nabla^2_{r_0}\langle V\rangle-2\gamma(M_{0,2}-dmk_BT). \quad (19)$$

A second important difference is that the value of γ is not governed by a constraint equation, rather it is a friction constant which determines the coupling strength between the system and the bath as in the Langevin equation. In a simulated annealing procedure the value of γ may be chosen to obtain the desired relaxation time of the bath kinetic energy or a diffusion coefficient.

In the $M_{0,2}$ equation of motion the term proportional to γ represents a restoring force which constrains $M_{0,2}$ about dmk_BT. For example, in the absence of an external potential, deviations of $M_{0,2}$ from dmk_BT are exponentially damped in time with a relaxation rate of 2γ. The equation of motion for p_0 has the form of a Langevin equation with constant dissipation which will damp out p_0 and stall r_0. The temperature becomes proportional to $\langle p^2\rangle=M_{0,2}$ which relaxes to the equilibrium value $M_{0,2}\to dmk_BT$. Therefore, the density distribution relaxes to a static steady state distribution on the time scale $1/\gamma$.

D. Variable rate annealing schedules

In simulated annealing it is typical and reasonable to choose an exponential or even linear cooling schedule. However, a phase change often intervenes. When an increase in the heat capacity is detected, the cooling process can be adjusted to spend more time in the region of larger fluctuations and configurational rearrangements. We have incorporated this idea in a variable rate cooling schedule which in a given temperature range T to $T+dT$ spends a time proportional to $dT/C_v(T)$. The cooling schedule is derived from the constraint that the system spend the same time ζdt in any energy increment dE, or that

$$\frac{dE}{dt}=\zeta, \quad (20)$$

where ζ is a constant which determines the rate of energy decrease. Rewriting this equation as

$$\frac{dE}{dT}\frac{dT}{dt}=\zeta \quad (21)$$

and using $dE/dT=C_v$, we find

$$\eta(T)=-\frac{dT}{dt}=\frac{\zeta}{C_v(T)}, \tag{22}$$

where η is the cooling rate as a function of temperature. The value of the constant ζ, which is the heat evolved from the system per unit time, can be calculated by integrating Eq. (22) from the initial temperature T_i to final temperature T_f over a specified total simulation time Δt

$$\zeta=\frac{1}{\Delta t}\int_{T_i}^{T_f} C_v dT. \tag{23}$$

With this strategy, one can guarantee that the system will spend more time in the temperature region where the rate of change in the energy is greatest, i.e., the region of a transition. This protocol is easily implemented by first calculating $E(T)$ using a linear cooling schedule from which $C_v(T)$ may be calculated leading to $\eta(T)$.

III. APPLICATION TO GLOBAL OPTIMIZATION

The GDA, GPP, and FP methods were applied to the global energy optimization of a series of Lennard-Jones clusters. In each case, we employed a four Gaussian decomposition potential.[18] The parameters (a_k,b_k) for the fit of the form $V(r)=\Sigma_k a_k \exp(-b_k r^2/2)$ given in Lennard-Jones reduced units are (8.467 067×10^5,30.928 81), (2.713 651 ×10^3,14.963 75), (−0.715 442 0,1.279 242) and (−9.699 172,3.700 745). It was found that this fit provides a good representation of the Lennard-Jones 12-6 potential and the corresponding force.

To prevent dissociation and restrict our search to the set of compact clusters, a harmonic pair potential of the form

$$V_{\text{conf}}=\sum_{i>j=1} \kappa r_{ij}^2/2 \tag{24}$$

was employed. The value of κ was optimized by trial and error. The (x,y,z) coordinates for each atom were chosen from a uniform distribution inside a cubic box with side 5 to 10σ. For each algorithm, 100 independent initial configurations were generated and optimized. The equations of motion were integrated using the Bulirsch–Stoer method[19] with an internal relative error tolerance of 10^{-8}. This elegant integrator enables one to take large and variable time steps while integrating the equations of motion to a specified accuracy.

A. Gaussian density annealing

We have applied the GDA algorithm to the optimization of a series of Lennard-Jones n-mer clusters. Aside from the choice of initial conditions, the GDA algorithm represents a fully deterministic set of equations for finding the global energy minimum. First, a few points concerning the implementation of the algorithm.

(1) We assumed that the energy decays exponentially as $E(\beta)=A+B\exp(-c\beta)$ and adjusted the time step such that $\ddot{E}/\dot{E}=$const. The constant was chosen to be 0.1. The simulation was stopped when $M_2 \leqslant 10^{-10}$. The conjugate gradient algorithm with the exact Lennard-Jones potential was then used to refine the energy of the minimum.

TABLE I. Results for global energy optimization using the GDA algorithm. Energies are given in Lennard-Jones reduced units. The table shows the number of times the global minimum was located from 100 random initial configurations.

Number of atoms	Global	First excited	GDA	DEM
8	−19.822	−19.765	26	local
9	−24.113	−23.270	41	local
10	−28.420	−27.555	39	local
11	−32.765	−31.915	46	global
12	−37.967	−36.306	60	local
13	−44.327	−41.472	70	global
14	−47.847	−46.163	63	global
15	−52.322	−51.418	39	global
16	−56.815	−55.943	29	
17	−61.317	−61.307	18	
18	−66.531	−65.842	23	
19	−72.660	−71.082	13	local
55	−279.248	−276.603	11	global

(2) The initial values of M_2 should be chosen such that $\sqrt{M_2}/d$ is large compared with the interparticle separations, or that the infinite temperature density distribution $\rho(r,\beta=0)$ is approximately constant over the extent of the cluster. In this work we found our results to be independent of the exact value of the initial width when $3.0-3.8 < \sqrt{M_2}/d < 6.0-8.5$. The calculations reported here were performed with initial $M_2=150$ for all clusters.

(3) During the simulation, the force constant for the harmonic confining potential was taken to be $\kappa=0.02$ for all clusters.

The GDA algorithm was applied to Lennard-Jones n-mer clusters with $n=8-19$ and 55. For each cluster size we examined 100 independent initial configurations. The probability of finding the global energy minimum for each cluster is listed in Table I. A comparison is also made with the results of the "diffusion equation method" (DEM) of Scheraga and co-workers.

In Fig. 1(a) the energy is plotted as a function of the logarithm of the temperature for a typical run with $N=13$ which successfully located the global energy minimum. The run clearly shows that the optimization occurs in two stages. There exists a very rapid decrease in energy at smaller β followed by a second steep decay before the system converges to the final global minimum. A fast decay in energy is associated with a rapid decrease in the squared width M_2 of the particles. In Fig. 1(b) the M_2 and gyration radius R_g of the cluster, which is defined as the RMS distance between the particles and the center of mass, are plotted as functions of the logarithm of the temperature. There exists an abrupt decrease in the gyration radius which coincides with the second sudden drop in energy. At this point the cluster is localized in the global minimum structure. The peak in the R_g curve at smaller β indicates that the cluster expands at the beginning due to the large initial M_2 which "vaporizes" the cluster. After the two rapid drops in energy there is a more gradual decrease occurring in the final convergence stage.

Abrupt changes in the energy are associated with important structural changes such as global particle rearrangement.

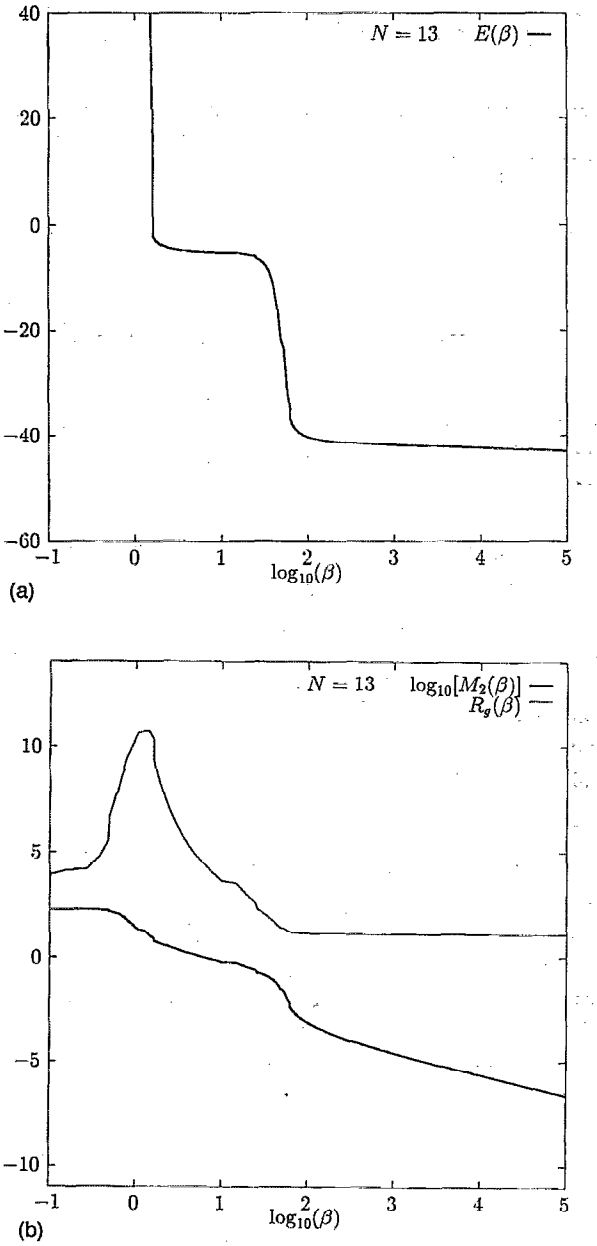

FIG. 1. (a) The energy as a function of β for the cluster $N=13$. (b) A typical M_2 and R_g as functions of β for the cluster $N=13$.

One might reason that the second decay in energy during the simulation is the most important stage in determining the final cluster geometry. To examine this idea, we measured the radial distribution functions at different stages (see Fig. 2). From high temperature (small β) to intermediate temperature, the $g(r)$ evolved from a broad gaslike peak to one which displays more liquidlike structure which clearly demonstrates that the system undergoes a gas-liquidlike change from a diffuse to a compact disordered state. Note that the radial distribution functions (RDF) in Fig. 2 represent instantaneous snapshots of a single configuration. The smoothness of the curves results from the fact that the $g(r)$ corresponds to the density distribution—not a single point particle. We also measured the radial distribution function after the second abrupt change in energy (see Fig. 2, $T=4.24$ case). The

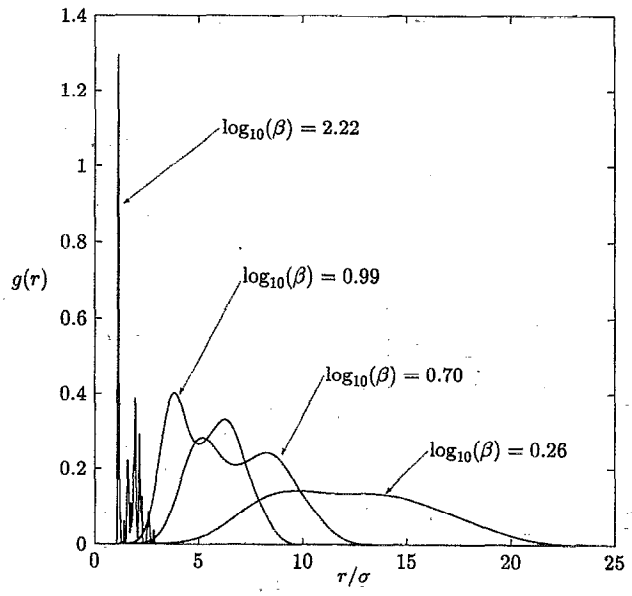

FIG. 2. The radial distribution functions at three intermediate temperatures are presented. The calculations were performed for the $N=13$ cluster. Each distribution function is an instantaneous snapshot for the cluster of packets.

results show a clear solidlike structure evidently indicating that the second sudden drop in energy results from a liquid-solid-like change. We believe that rearrangements which occur during the second drop in energy are the most important in deciding the final optimized structure. We also examined the evolution of trajectories using repeated conjugate gradient quenches along the temperature axis during the simulation (see Fig. 3). The system clearly swaps between different local (and on occasion the global) energy minima in the intermediate temperature range but was confined to the global minimum after the second sudden decay in energy.

B. Parallel comparison of the dynamical annealing algorithms

We applied the GPP, FP, and classical simulated annealing algorithms to find the global energy minimum of the series of Lennard-Jones clusters. For comparison sake, we employed an exponential cooling schedule

$$T(t)=T_0 \exp(-t/\tau) \tag{25}$$

with $\tau=10$ for the GPP, FP, and MD dynamical annealing algorithms. We used the confining potential defined by Eq. (24) with $\kappa=0.2$ for all dynamical annealing algorithms. The results are shown in Table II.

The data demonstrate that those methods based on annealing the continuous density distribution have a significantly higher probability of finding the global minimum than conventional molecular dynamics based annealing. Of course, all of the annealing methods are superior to a local energy minimizer such as the conjugate gradient (CG) method.

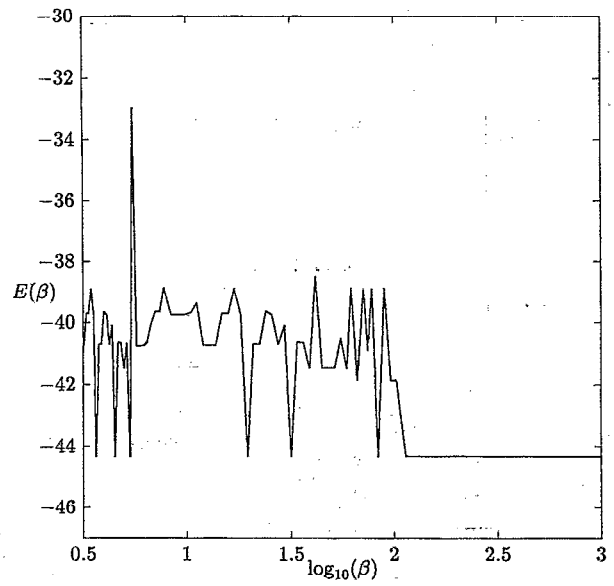

FIG. 3. Illustration of the results of frequent conjugate gradient energy quenches along a trajectory. The results were from a single calculation for the $N=13$ cluster.

Note that the exponential cooling schedule defined by Eq. (25) is both practical and popular but not optimal. The absolute value of the probability for finding the global minimum is a strong function of the cooling schedule. However, the relative success rates are clear and not expected to change dramatically with the cooling schedule. Our data indicate that the GPP and FP provide a similarly enhanced probability of locating the global minimum and that the GPP method is consistently more effective than the FP algorithm. This may result from the use of a variable γ in the GPP dynamics whose fluctuations are fairly large for the smaller clusters. This may add "noise" which helps the trajectory sample conformational space more effectively. In the next section, we will discuss the results of a more refined cooling schedule.

TABLE II. Summary of results for the global energy minimization of Lennard-Jones atomic clusters. Energies are given in Lennard-Jones reduced units. The table shows the number of times the global energy minimum was located out of 100 random initial configurations. For comparison, we also list the results for the conjugate gradient (CG) local energy minimization method.

Number of Atoms	Energy	CG	MD	GPP	FP
8	−19.822	2	24	35	52
9	−24.113	0	9	72	52
10	−28.420	1	2	80	63
11	−32.765	0	8	92	86
12	−37.967	0	2	100	97
13	−44.327	0	10	100	86
14	−47.847	0	12	100	88
15	−52.322	0	10	86	43
16	−56.815	0	5	60	31
17	−61.317	0	1	21	26
18	−66.531	0	1	46	42
19	−72.660	0	2	49	25
55	−279.248	0	0	36	20

TABLE III. The results of simulated annealing using the cooling schedule defined by Eq. (22) for the clusters $N=17$ and 19. The numbers inside the parentheses are the results for a linear cooling schedule using the same total simulation time.

Algorithm	$N=17$	$N=19$
MD	20	38
	(5)	(11)
GPP	42	63
	(15)	(20)
FP	39	33
	(10)	(12)

C. Comparison of linear and variable rate annealing schedules

The results of annealing according to the cooling schedule defined by Eq. (22) are displayed in Table III. We calculated $C_v(T)$ by standard molecular dynamics simulation. Employing a linear cooling schedule $T(t)=T_0-\alpha t$ with $\alpha=0.1$ and an initial temperature $T_0=2.0$ the average energy was calculated as a function of temperature. In Fig. 4 we show a typical energy curve for the $N=19$ system. This curve was calculated by averaging over 30 independent initial configurations in order to generate better statistics for this small system. The heat capacity in Fig. 5 was calculated by numerically smoothing the energy and taking the first derivative $C_v(T)=dE/dT$. We expect the width of the region over which the phase change occurs to scale as $\Delta T/T\sim N^{-1}$.[20,21] Using the heat capacity data to define the cooling schedule in Eq. (22), we performed simulated an-

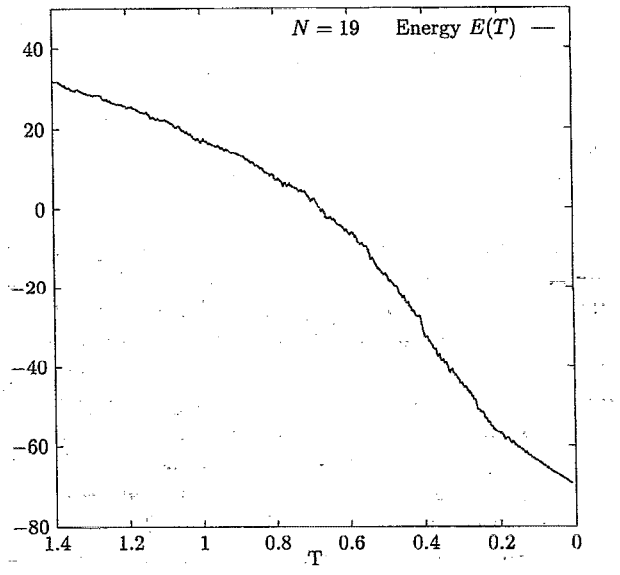

FIG. 4. The energy as a function of temperature for the $N=19$ cluster. The data were obtained from an average over 30 independent initial configurations.

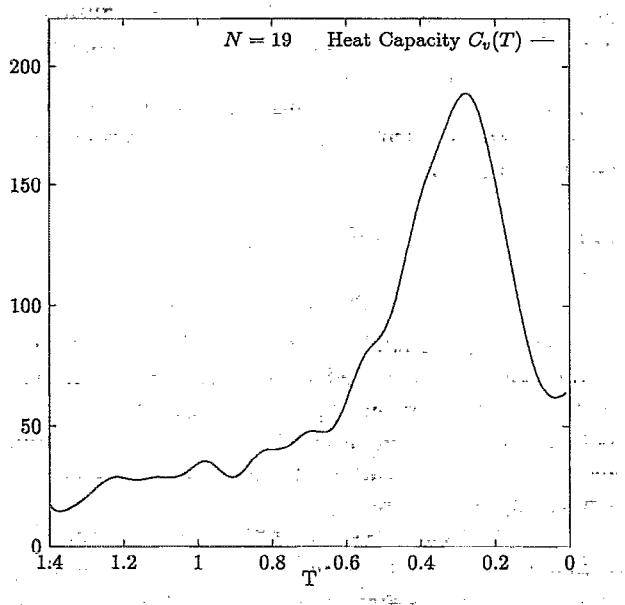

FIG. 5. The heat capacity as a function of temperature for the $N=19$ cluster. The data were obtained by numerically smoothing the energy curve in Fig. 4 and differentiating that curve to estimate $C_v(T)$.

nealing for cluster sizes $N=17$ and 19. The annealing simulations were significantly enhanced by spending an increased time in the regions of largest energy fluctuation. For comparison, we also list the results for a linear cooling schedule with the same total simulation time. The variable cooling rate leads to a significantly enhanced probability of finding the global minimum in the same computational time. The variable rate annealing schedule is easily automated and provides a simple means of significantly improving the cooling schedule from a monotonic linear or exponential temperature decrease. Variations of Eq. (22) are easily imagined using, for example, a higher power of $1/C_v(T)$.

IV. DISCUSSION AND CONCLUSION

We have presented new optimization algorithms and compared them with each other and classical simulated annealing based on molecular dynamics. As global optimization methods, each algorithm has its own merits and shortcomings. Perhaps the most important question one must answer concerns the computational costs of the algorithm. Using the Gaussian packets approach, there is a cost associated with (1) the slight increase in the number of first order equations of motion from conventional molecular dynamics and (2) the computation of second derivatives of the potential. In Table IV, we list the number of equations one must integrate for each algorithm. We also present the typical timing results for the global optimization of selected clusters. Apparently the FP algorithm is faster than GPP, although they each have the same number of equations to integrate. This may indicate that the equations of motion in the GPP dynamics are stiffer, perhaps due to the variation in γ and the rigid constraint on the temperature. This rigid constraint becomes more difficult to satisfy as the number of particles decreases. The GDA algorithm leads to the best timing results of any Gaussian packet algorithm although it is slightly slower than MD. However, the small additional cost in CPU time required by the GDA algorithm is more than compensated for by the significantly enhanced probability of success in locating the global energy minimum.

TABLE IV. Typical timing results for various energy minimization algorithms for clusters $N=13$ and $N=19$. The simulations were implemented using identical temperature intervals and time step sizes. The timings are in units of minutes on a single processor of a Titan 3040 computer.

Algorithm	Equations to integrate	$N=13$	$N=19$
GDA	$(d+1)N$	2.1	3.2
MD	$2dN$	1.5	2.5
GPP	$(2d+3)N$	11.5	27.0
FP	$(2d+3)N$	6.5	11.3

For the Lennard-Jones clusters we found a correlation between the probability of successfully locating the global energy minimum and the energy gap, $\Delta E=E_1-E_0$, between the global minimum, E_0, and the next higher energy state, E_1, found in our study. This correlation strongly suggests that the success of the algorithm is enhanced by a significant energy gap between the ground and first excited state (see Fig. 6). We believe this is an important feature for the global optimization of Lennard-Jones clusters and all molecular systems. The melting and freezing phase dynamics of Lennard-Jones clusters have been carefully studied by Honeycutt and Anderson.[22] They reported that those clusters with larger energy gaps have the best defined melting and freezing phase changes. If there is a smaller energy separation, it is more likely that the system may be trapped in a glassy state making it less likely that the global minimum will be located.

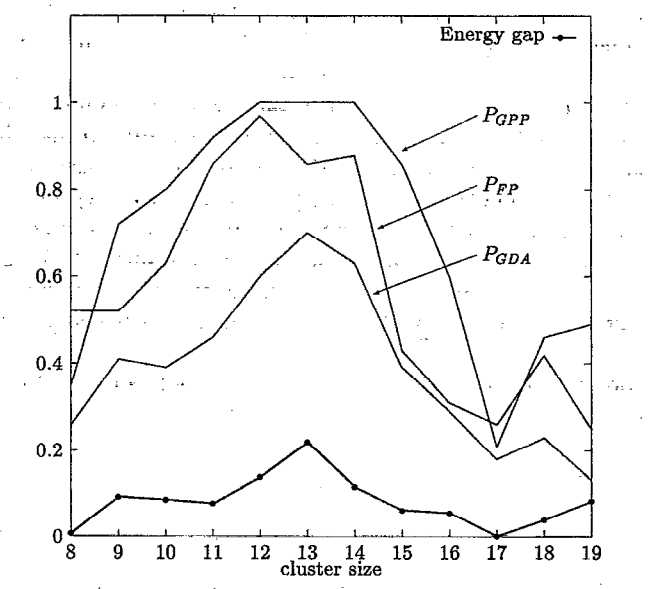

FIG. 6. Illustration of the correlations between the probabilites of locating the global energy minimum and the energy gap ΔE (heavy line) between the ground state and first excited state. Data are shown for the P_{GPP}, P_{FP}, and P_{GDA}.

TABLE V. The relative magnitude of the energy gap, ΔE, and the well volume ratio V_0/V_1 for all clusters using $T=0.2$.

Number of atoms	ΔE	$-k_BT\ln(V_1/V_0)$
8	0.0071	0.0042
9	0.0927	−0.0055
10	0.0856	−0.0153
11	0.0773	−0.0065
12	0.1384	−0.0115
13	0.2196	−0.0338
14	0.1156	−0.0217
15	0.0603	−0.0137
16	0.0545	−0.0102
17	0.0006	−0.0030
18	0.0382	0.0040
19	0.0831	−0.0094
55	0.0481	−0.0071

This behavior is clearly demonstrated in our annealing simulations.

From our annealing results, we found that the systems were trapped either in the global minimum or the nearest local minimum most of the time. Suppose we approximate the cluster as a two level system. The probability of finding the global minimum should be well correlated with the free energy difference between the two states, which we take in the classical approximation to be $\Delta A=\Delta E-k_BT\ln(V_1/V_0)$, where V_0 and V_1 are the well "volumes" given by the product of normal mode well frequencies ν^i as

$$V_0=\prod_{i=1}^{3N-6}\frac{1}{\nu_0^i} \tag{26}$$

for the global minimum and similarly for the first excited state.

The normal mode frequencies were calculated by diagonalizing the force constant matrix in the two states. We calculated the quantity $-k_BT\ln(V_1/V_0)$ at a temperature near the transition point which in this case we took to be $T=0.2$. From the results in Table V, one can see that the contribution from the difference in well volumes is much smaller than the energy gap. As we have seen (Fig. 6), the effectiveness of the annealing simulation is strongly correlated with the energy gap which is a good estimate of the free energy difference.

In closing, we comment on the relation of the GDA algorithm presented here and the "diffusion equation method" (DEM) developed by Piela, Kostrowicki, and Scheraga.[7] In the DEM the interaction potential surface $V(r)$ is transformed, as if it were a concentration gradient subjected to a diffusion process for some time, t, into

$$\langle V\rangle_{\mathrm{DEM}}(r,t)=(4\pi t)^{-d/2}\int dr'V(r')\times\exp(-\|r_0-r'\|^2/4t). \tag{27}$$

This has the effect of filling in wells (raising minima) and lowering barriers, reducing the number of potential minima. When the diffusion time is large enough, there are few or, as is the case for their treatment of Lennard-Jones clusters, one surviving minimum.[6] The surviving minimum is found and the diffusion process is reversed so that the diffusion time is gradually reduced to zero and the minimum is mapped back to the original potential surface.

Now consider the GDA method. The effective potential for the center r_0 of the Gaussian density distribution in the GDA algorithm is

$$\langle V\rangle_{\mathrm{GDA}}(r_0,M_2)=(2\pi M_2/d)^{-d/2}\int dr'V(r')\times\exp(-d\|r_0-r'\|^2/2M_2). \tag{28}$$

This effective potential is identical to the transformed interaction potential employed in the DEM where the squared width of the Gaussian $M_2=2td$. Therefore, the appealing potential deformation used in the DEM results from our more fundamental treatment for the special case of a Gaussian density distribution.

In the GDA algorithm we initially choose large M_2's and then allow the centers to follow a steepest descent trajectory to the minimum potential energy conformation. However, in the DEM there is a single diffusion time meaning that all particles are constrained to have the same M_2, and as the diffusion time is reduced, corresponding to $\beta\to\infty$, the effective widths decrease monotonically. In the case of the GDA algorithm the widths are variationally optimized depending on the environment of each particle. Thus the DEM can be thought of as a special case of the GDA algorithm where (1) the second-moment M_2 for each packet is identical, and (2) the magnitude of M_2 is initially set to something large and then monotonically decreased to zero. Moreover, the GDA algorithm can be generalized to a more complete representation for the density distribution by replacing spherical Gaussians with anisotropic Gaussians or by adding orthogonal polynomials. It also holds the promise of generating estimates of the equilibrium thermal distribution at nonzero temperatures.

ACKNOWLEDGMENTS

We thank Patricia Amara for many helpful discussions. This work was supported in part by a grant from the American Chemical Society's Petroleum Research Fund and in part by a grant from the National Science Foundation (CHE-9306375).

[1] M. R. Hoare and P. Pal, Adv. Phys. **20**, 161 (1971).
[2] K. D. Gibson and H. A. Scheraga, in *Structure and Expression: From Proteins to Ribosomes*, edited by M. H. Sarma and R. H. Sarma (Adenine, Schenectady, NY, 1988), Vol. 1.
[3] J. E. Straub and D. Thirumalai, Proc. Natl. Acad. Sci. USA **90**, 809 (1993); J. E. Straub and D. Thirumalai, Proteins **15**, 360 (1993).
[4] S. Kirkpatrick, Jr., C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[5] F. H. Stillinger and T. Weber, J. Stat. Phys. **52**, 1429 (1988); F. H. Stillinger and D. K. Stillinger, J. Chem. Phys. **93**, 4266 (1990).
[6] J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga, J. Phys. Chem. **95**, 4113 (1991); J. Kostrowicki and H. A. Scheraga, J. Phys. Chem. **96**, 7442 (1992).
[7] L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem. **93**, 3339 (1989).
[8] J. Pillardy, K. A. Olszewski, and L. Piela, J. Phys. Chem. **96**, 4337 (1992).
[9] The unnormalized equilibrium density distribution $\hat{\rho}(r,\beta)=\exp(-\beta H)$ satisfies the Bloch equation $\partial\hat{\rho}/\partial\beta=-H\hat{\rho}$. [See R. W. Zwanzig, in *Lectures in Theoretical Physics, Vol. III*, edited by W. E. Britton, B. W.

Downs, and J. Downs (Wiley Interscience, New York, 1961), p. 106.] Integration of a normalized distribution of a thermodynamic average quantity avoids the numerical difficulties associated with integration of unnormalized quantities, such as $\hat{\rho}(r,\beta)$, which can become diminishingly small for large values of β.

[10] Equation (2) is the classical analog ($\hbar=0$ limit) of the imaginary time Schrödinger equation. Elsewhere we have presented an optimization algorithm based on approximate solution of the imaginary time Schrödinger equation using Gaussian wavepackets. The algorithm presented here can be thought of as special case of that more general work where $\hbar=0$. See P. Amara, D. Hsu, and J. E. Straub, J. Phys. Chem. **97**, 6715 (1993).

[11] Related to this work is an intriguing method of global optimization based on the classical density distribution expanded as a Gaussian by D. Shalloway, in *Recent Advances in Global Optimization*, edited by A. Floudas and P. M. Pardalos (Princeton University, Princeton, N.J., 1992), p. 433.

[12] J. Ma, D. Hsu, and John E. Straub, J. Chem. Phys. **99**, 4024 (1993).

[13] J. Grad, Y. J. Yan, and S. Mukamel, Chem. Phys. Lett. **134**, 219 (1987); J. Grad, Y. J. Yan, A. Haque, and S. Mukamel, J. Chem. Phys. **86**, 3441 (1987).

[14] W. G. Hoover, A. J. C. Ladd, and B. Moran, Phys. Rev. Lett. **48**, 1818 (1982); D. J. Evans, J. Chem. Phys. **78**, 3297 (1983).

[15] J. L. Skinner and P. G. Wolynes, J. Chem. Phys. **69**, 2143 (1978).

[16] J. L. Skinner and P. G. Wolynes, J. Chem. Phys. **72**, 4913 (1980).

[17] H. Risken, *The Fokker-Planck Equation,* (Springer-Verlag, Berlin, 1989).

[18] P. Amara, D. Hsu, and J. E. Straub, J. Phys. Chem. **97**, 6715 (1993).

[19] S. A. Teukolsky W. H. Press, B. P. Flannery, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University, Cambridge, 1986).

[20] T. Hill, *The Thermodynamics of Small Systems* (Benjamin, New York, 1963/64).

[21] R. S. Berry, T. L. Beck, and H. L. Davis, Adv. Chem. Phys. **70B**, 75 (1988).

[22] J. D. Honeycutt and H. C. Andersen, J. Phys. Chem. **91**, 4950 (1987).