Simulated annealing using coarse grained classical dynamics: Smoluchowski dynamics in the Gaussian density approximation

John E. Straub,[a] Jianpeng Ma, and Patricia Amara
Department of Chemistry, Boston University, Boston, Massachusetts 02215

(Received 12 December 1994; accepted 5 April 1995)

A dynamical annealing algorithm for global optimization based on approximate solution of the Smoluchowski equation is presented. The equations of motion in the Gaussian density approximation are interpreted as a steepest descent quench on a time dependent effective potential energy surface. A relation between the convexity condition for the effective potential surface and the size of thermal fluctuations provides a definition of the critical temperature above which the distribution is delocalized and the effective potential is smooth and convex during an annealing run. This critical temperature may be significantly less than the temperature characteristic of escape from a local energy minimum. © *1995 American Institute of Physics.*

I. INTRODUCTION

Simulated annealing (SA) remains a paradigm for global energy minimization of molecular systems.[1,2] Part of the appeal of SA is its ease of application. A computer program written for Newtonian molecular dynamics or Monte Carlo is easily converted into a program for global optimization by incorporating a temperature control and an annealing schedule for the decrease of the temperature with time. A second reason for the popularity of SA is its success, although not complete, in solving many optimization problems. A third reason is that the analogy with the familiar statistical mechanical process of annealing provides a means of picturing the optimization process and gaining insight into its mechanism.

Nevertheless, there are reasons to believe that simulated annealing is not the optimal optimization method for applications to molecular systems. One characteristic of biomolecular systems is the broad distribution of energy scales in the problem.[3] Configuration space is partitioned by very high barriers (on the order of tens of kcal/mol) corresponding to dihedral angle transitions while the relative stability of conformers can be determined by weak hydrogen bonding forces and van der Waals contacts (which are often tenths of a kcal/mol). In the simulated annealing protocol the trajectory begins at a temperature which initially is large compared with the highest barrier in the system, $E_{\max}$, and ends at a temperature which is small compared with the separation ΔE between the global energy minimum and the next lowest lying minimum. If the trajectory is to sample the equilibrium distribution at each intermediate temperature we expect that the cooling should be logarithmically slow. It follows that the length of the optimal annealing run t_{sim} scales exponentially with the ratio $E_{\max}/\Delta E$. Therefore, problems with a wide range of energy scales represent very hard problems for the standard simulated annealing method. As the ratio of the largest and smallest energy scales grows, the problem becomes exponentially harder. For this reason alone, we expect protein folding (finding the lowest energy state of a protein), and many other molecular global energy minimizations, to be hard optimization problems.

A second paradigm for global energy minimization of atomic and molecular clusters and macromolecules is potential smoothing. It has been known for some time that short range potentials produce potential energy hypersurfaces with a large number of local energy minima.[4,5] Global energy minimization on such landscapes can be frustrating. It is also known that by increasing the range of interaction in the potential energy function it is possible to greatly reduce the number of local minima while preserving the global character of the potential hypersurface.[4,5] Of course, physical potentials are determined for us and the potential surface one must work with may have the character of a rugged energy landscape with a large number of local minima making the optimization problem ''hard.'' While it has been shown that smoothing can merely transform one hard optimization problem into another,[6] in practice smoothing algorithms have had considerable success. Smoothing algorithms involve a transformation of the interaction potential from a rugged surface to a smooth one. The simpler problem of global minimization on the transformed surface is then solved. The transformation is then inverted and the solution is mapped on to the untransformed (physically interesting) potential surface.

Consider an example provided by Stillinger and Weber involving an *ad hoc* local Gaussian coarse graining of the potential energy.[7] Starting from the physical potential $V(\mathbf{r})$ in d dimensions, the Gaussian coarse grained potential energy is defined as

$$\langle V\rangle = (2\pi L^2)^{-d/2}\int d\mathbf{r}'\ V(\mathbf{r}')e^{-||\mathbf{r}-\mathbf{r}'||^2/2L^2}. \tag{1}$$

When the length scale of the coarse graining $L\to 0$, the coarse grained potential reduces to the physical potential $\langle V\rangle \to V(\mathbf{r})$. When the coarse graining length scale L is increased, the potential is smoothed. Minima separated by distances small compared with L will be joined into a single minimum on the smoothed potential surface. An effective coarse graining will result in a simplified potential function with relatively few minima, and these surviving minima will

[a]Author to whom correspondence should be addressed.

 0021-9606/95/103(4)/1574/8/$6.00

often represent the most important minima (deepest with largest volume) on the physical potential surface.

In this paper, we discuss algorithms based on approximate solutions to the Fokker–Planck and Smoluchowski equations. Explicit equations of motion are derived for the Gaussian density approximation. We find that the coarse grained classical dynamics combines the best properties of potential smoothing methods and classical density annealing algorithms.

II. DERIVATION OF COARSE GRAINED SMOLUCHOWSKI DYNAMICS

The Langevin equation represents the dynamics of a system (a single point in phase space) coupled to a heat bath of well defined temperature T. The Fokker–Planck equation (for the phase space density distribution) which is equivalent to the Langevin equation is the Kramers equation[8]

$$\frac{\partial}{\partial t}\,\rho(\mathbf{r},\mathbf{p},t)=-\left[\frac{\mathbf{p}}{m}\cdot\nabla_{\mathbf{r}}+\mathbf{F}(\mathbf{r})\cdot\nabla_{\mathbf{p}}-\gamma[\nabla_{\mathbf{p}}\cdot[\mathbf{p}+mk_BT\nabla_{\mathbf{p}}]]\right]\rho(\mathbf{r},\mathbf{p},t). \tag{2}$$

The density distribution $\rho(\mathbf{r},\mathbf{p},t)$ is completely specified by the center in phase space $(\langle\mathbf{r}\rangle,\langle\mathbf{p}\rangle)=(\mathbf{r}_0,\mathbf{p}_0)$ and the moments in position and momentum $\mathbf{M}_{n,k}=\langle(\mathbf{r}-\mathbf{r}_0)^n(\mathbf{p}-\mathbf{p}_0)^k\rangle$ are $n+k$ rank symmetric tensors, defined by averages over permutations of the tensor indices of the position and momentum components, where $\langle\cdot\rangle$ represents a phase space average over the distribution $\rho(\mathbf{r},\mathbf{p},t)$. To derive equations of motion for these moments, we use the fact that the time derivative of the average of a quantity $A(\mathbf{r},\mathbf{p},t)$ defined

$$\langle A\rangle=\int d\mathbf{r}\,d\mathbf{p}\,A(\mathbf{r},\mathbf{p},t)\rho(\mathbf{r},\mathbf{p},t) \tag{3}$$

is given by

$$\frac{d}{dt}\,\langle A\rangle=\int d\mathbf{r}\,d\mathbf{p}\left[\rho\,\frac{\partial A}{\partial t}+A\,\frac{\partial\rho}{\partial t}\right]. \tag{4}$$

The partial time derivative acts only on the explicitly time dependent quantities (the moments) while the time dependence of the density distribution is given in this case by Eq. (2).

Evaluation of Eq. (4) leads to

$$\frac{d\mathbf{r}_0}{dt}=\frac{\mathbf{p}_0}{m},\quad \frac{d\mathbf{p}_0}{dt}=\mathbf{F}_0-\gamma\mathbf{p}_0 \tag{5}$$

for the motion of the distribution center. For the equations of motion of the moments, integration by parts of Eq. (4), assuming the surface terms equal to zero because the Gaussian distribution vanishes at infinity, provides

$$\frac{d\mathbf{M}_{n,k}}{dt}=\frac{n}{m}\mathbf{M}_{n-1,k+1}+k\mathbf{W}_{n,k-1}-\gamma[k\mathbf{M}_{n,k}-k(k-1)mk_BT\mathbf{L}_{n,k-2}], \tag{6}$$

where the moments $\mathbf{W}_{n,k}=\langle(\mathbf{r}-\mathbf{r}_0)^n(\mathbf{p}-\mathbf{p}_0)^k(\mathbf{F}-\mathbf{F}_0)\rangle$ and $\mathbf{L}_{n,k}=\langle\mathbf{I}(\mathbf{r}-\mathbf{r}_0)^n(\mathbf{p}-\mathbf{p}_0)^k\rangle$ where $\mathbf{I}$ is the identity matrix. In our shorthand notation, it is implicit that $\mathbf{M}_{n,k}$, $\mathbf{W}_{n,k-1}$, and $\mathbf{L}_{n,k-2}$ are $n+k$ rank symmetric tensors. For the special case of a Gaussian packet representation of the density distribution the equations of motion have been presented previously:[9]

$$\dot{\mathbf{r}}_0=\frac{\mathbf{p}_0}{m},\quad \dot{\mathbf{p}}_0=-\nabla_{\mathbf{r}_0}\langle V\rangle-\gamma\mathbf{p}_0,$$

$$\dot{M}_{2,0}=\frac{2}{m}\,M_{1,1},\quad \dot{M}_{1,1}=\frac{1}{m}\,M_{0,2}-\frac{1}{d}\,M_{2,0}\nabla^2_{r_0}\langle V\rangle-\gamma M_{1,1}, \tag{7}$$

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

where each second order scalar moment is the trace of the corresponding second rank tensor. For the spherically symmetric Gaussian basis, the second rank tensor moments are isotropic. Coupling to the heat bath manifests itself by harmonically constraining the width of the momentum distribution $M_{0,2}$ about the correct temperature constant dmk_BT. As the system evolves in time, the motion of the center of the distribution in momentum $\mathbf{p}_0$ is damped out to zero while the width in momentum space $M_{0,2}$ approaches the correct temperature proportionality. Thus, at a fixed temperature the density distribution of the system will relax to an approximation to the static equilibrium distribution.

For many systems one may assume that the momenta relax quickly to the Maxwell distribution relative to the time scale for the coordinate relaxation. The adiabatic elimination of the momenta is achieved by assuming that instantaneously $d\mathbf{p}/dt=0$. The dynamics of the reduced configurational space distribution function $\hat{\rho}(\mathbf{r},t)$ is given by the Smoluchowski equation[10]

$$\frac{\partial}{\partial t}\,\hat{\rho}(\mathbf{r},t)=\frac{1}{m\gamma}[\nabla_{\mathbf{r}}\cdot[-\mathbf{F}(\mathbf{r})+k_BT\nabla_{\mathbf{r}}]\hat{\rho}(\mathbf{r},t)]. \tag{8}$$

Following the procedure described in the previous section, we derived moment equations for the time evolution of the configurational density distribution which are

$$\frac{d\mathbf{r}_0}{dt}=\frac{1}{m\gamma}\mathbf{F}_0,\quad \frac{d\mathbf{M}_n}{dt}=\frac{1}{m\gamma}[k_BTn(n-1)\mathbf{L}_{n-2}+n\mathbf{W}_{n-1}], \tag{9}$$

where $\mathbf{F}_0=\langle\mathbf{F}(\mathbf{r})\rangle$, $\mathbf{M}_n=\langle(\mathbf{r}-\mathbf{r}_0)^n\rangle$, $\mathbf{W}_n=\langle(\mathbf{r}-\mathbf{r}_0)^n(\mathbf{F}(\mathbf{r})-\mathbf{F}_0)\rangle$, and $\mathbf{L}_n=\langle\mathbf{I}(\mathbf{r}-\mathbf{r}_0)^n\rangle$, where $\mathbf{I}$ is the identity matrix. Again, it is implicit that $\mathbf{M}_{n,k}$, $\mathbf{W}_{n,k-1}$, and $\mathbf{L}_{n,k-2}$ are $n+k$ rank symmetric tensors. This hierarchy of equations describes the time evolution of the configurational distribution exactly. However, to make use of these equations we must truncate the moment expansion. For example, suppose we approximate the single particle configurational density distribution as a delta function $\hat{\rho}(\mathbf{r},t)=\delta(\mathbf{r}(t)-\mathbf{r}_0)$. The moments $\mathbf{M}_n=0$ and the dynamics consists of a steepest descent along the potential at a rate defined by the gradient divided by the friction. Using a spherical Gaussian approximation to the configurational distribution

$$\hat{\rho}(\mathbf{r},t)=(2\pi M_2/d)^{-d/2}\exp\left[-\frac{d}{2M_2}(\mathbf{r}-\mathbf{r}_0)^2\right], \tag{10}$$

the dynamics of the distribution are completely defined in terms of the time dependence of the distribution center $\mathbf{r}_0=\langle \mathbf{r} \rangle$ and variance, M_2 is the trace of the second rank tensor $\mathbf{M}_2$ which is isotropic for the spherical Gaussian basis. The resulting equations of motion are

$$\frac{d\mathbf{r}_0}{dt} = -\frac{1}{m\gamma}\nabla_{r_0}\langle V \rangle,$$
$$\frac{dM_2}{dt} = \frac{1}{m\gamma}\left[2dk_BT - \frac{2}{d}M_2\nabla^2_{r_0}\langle V \rangle\right]. \tag{11}$$

Alternatively, these equations of motion may be derived from the Fokker–Planck equations for the special case that (1) $\dot{\mathbf{p}}_0=0$, (2) $M_{0,2}=dmk_BT$, and that (3) $\dot{M}_{1,1}=0$. Note that if these equations are written in terms of the scaled time $\tau=t/m\gamma$ the structure of the dynamical equations are universal and independent of γ.

The equation of motion for the center of the density distribution is a steepest descent equation of motion on the coarse grained effective potential $\langle V \rangle$.[11] The widths of the distribution adjust to the curvature of the effective potential and can reach a static value when a balance is reached between the delocalizing influence of the temperature and the localizing influence of the potential or when

$$k_BT = \frac{1}{d^2}M_2\nabla^2_{r_0}\langle V \rangle. \tag{12}$$

This condition provides the exact distribution width for a harmonic system and an approximate effective harmonic distribution width for anharmonic potentials.[12] We now turn to an analysis of the properties of the effective potential and the dynamical optimization mechanism.

III. GENERAL PROPERTIES OF COARSE GRAINED DYNAMICS

It is useful to explore the general properties of simulated annealing using Fokker–Planck or Smoluchowski packet dynamics in the context of a one-dimensional rough potential

$$V(x) = \tfrac{1}{2}\kappa x^2 + \epsilon\cos[qx+\pi], \tag{13}$$

where we set $\epsilon=2$, $q=10$, and $\kappa=1$. In Fig. 1 we show the time evolution of a phase space density packet during an annealing run. For a conservative dynamics the center would trace an ellipse in phase space. For the dissipative dynamics, as the temperature of the packet is lowered, the packet narrows in both momentum and position and eventually becomes localized at a single point in phase space. Midway in the annealing run, the configuration space variance of the packet $M_{2,0}$ collapses and the packet is trapped in a local minimum of the potential surface. The rise in temperature that coincides with the packet collapse indicates that the localization transition has an associated ''latent heat.'' This leads to a jump in temperature and an increase in momentum fluctuations which are evident on the phase space plot.

An example of the Smoluchowski dynamics in a one-dimensional potential is provided in Fig. 2. The dynamics is characterized by a steady reduction in temperature over the annealing run. The localization transition in which the packet width M_2 collapses is associated with a sharp drop in the average energy of the packet. In contrast to Fokker–Planck packet dynamics, where there is an associated heat of transition which is converted into packet kinetic energy, in Smoluchowski packet dynamics the temperature is strictly controlled and the decrease in temperature during annealing is monotonic. What is the nature of this transition in terms of the packet dynamics and the properties of the effective potential $\langle V \rangle$? What is the temperature associated with the localization transition and how can it be calculated from the potential parameters? A similar localization transition was seen in the Fokker–Planck packet dynamics. How is the localization transition *into* a local minimum related to an escape (delocalization) transition *from* a local minimum for packet dynamics?

A. Critical fluctuations and convexity conditions

Convexity conditions are a useful means of analyzing potentials which are transformed by smoothing. By demanding that the smoothed potential function be everywhere convex, a unique solution to the minimization problem on the transformed potential is guaranteed. To develop an understanding of the dynamics on the Gaussian coarse grained effective potential $\langle V \rangle$ we return to the simple example of a one-dimensional harmonic potential with added sinusoidal roughness:[13]

$$V(x) = \tfrac{1}{2}\kappa x^2 + \epsilon\cos[qx+\pi]. \tag{14}$$

To compute the effective potential $\langle V \rangle$ we must average the potential over the configurational distribution. The effective potential for the center of a one-dimensional Gaussian distribution as a function of its center x_0 and variance M_2 is

$$\langle V \rangle(x_0) = \tfrac{1}{2}\kappa(x_0^2+M_2) + \epsilon\, e^{-M_2q^2/2}\cos[qx_0+\pi]. \tag{15}$$

How does the character of the effective potential change as the width of the Gaussian density is increased? When the distribution variance $M_2=0$, the bare potential is recovered. As the variance is increased, the sinusoidal roughness diminishes in importance. In the regime where the variance of the distribution is large compared to the length scale of the roughness $M_2 \gg 1/q^2$, the roughness becomes exponentially small and the bare harmonic potential is recovered. In that regime, a coarse grained averaging on the length scale of the potential roughness erases the roughness and exposes the broader features of the potential which correspond to longer length scale interactions or correlations.

At small values of M_2 there are many local minima on the potential surface. The number of these minima can be found by evaluating the solutions of $\partial_{x_0}\langle V \rangle = 0$ where $\partial_{x_0x_0}\langle V \rangle > 0$. The number of extrema is given by the number of solutions of

$$x_0 = \frac{\epsilon q}{\kappa}\, e^{-M_2q^2/2}\sin[qx_0+\pi]. \tag{16}$$

As M_2 increases, the right-hand side of Eq. (16) becomes increasingly small. When M_2 exceeds a critical value there will be a single solution to this equation which corresponds to the global energy minimum. The zeros are the points of intersection of the linear (left-hand) and the sinusoidal (right-

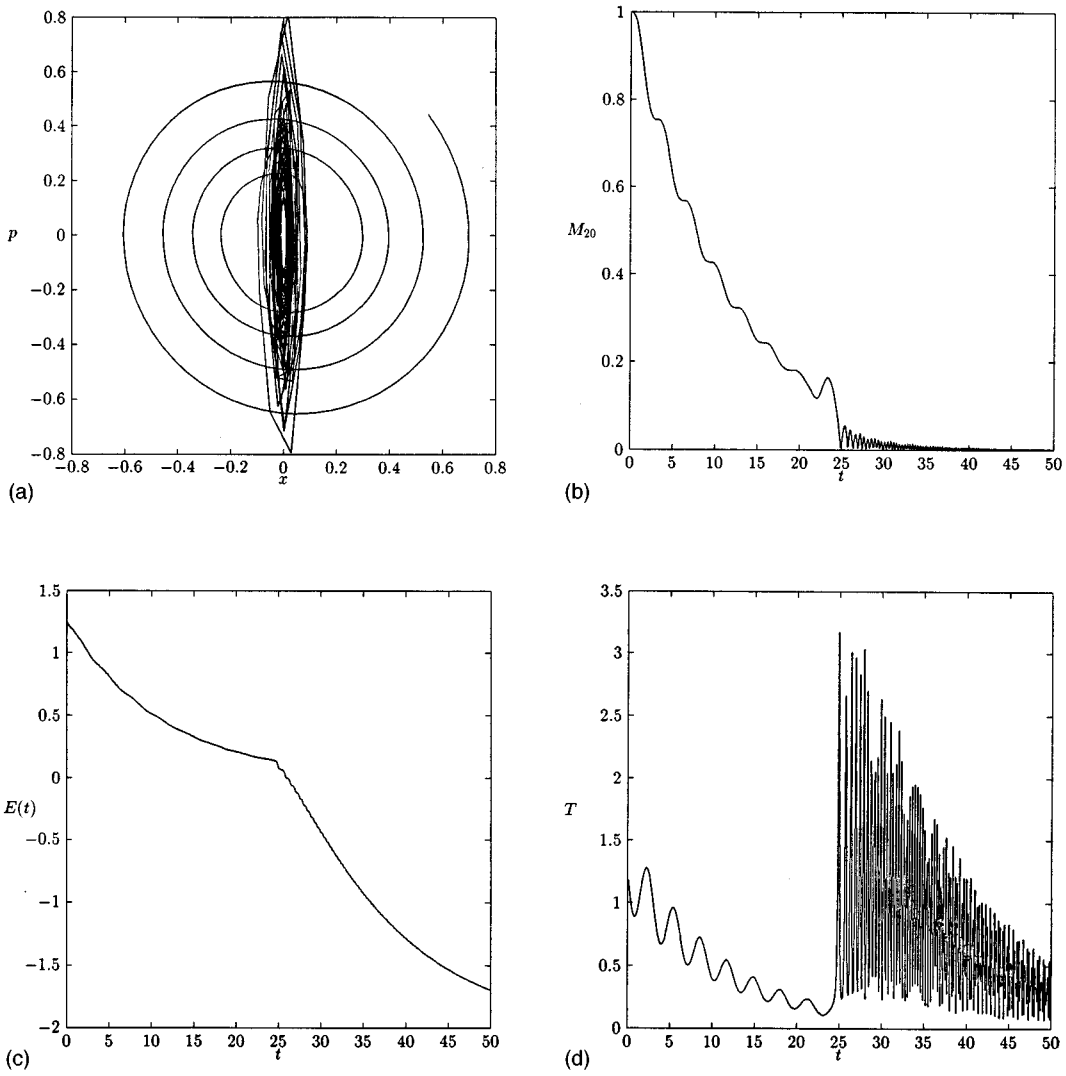

FIG. 1. Time evolution of the phase space density distribution following the approximate Fokker–Planck equations for a Gaussian density distribution in the one-dimensional potential energy $V(x)=x^2/2+2\cos[10x+\pi]$. For the run, we set the bath temperature $T=0$ and $\gamma=0.09$ and show (a) the trajectory of the packet center including the dynamics before (smooth spiral) and after (dense elliptical spiral) the packet is localized, (b) the configuration space variance $M_{2,0}$, and the time evolution of the packet's (c) total energy and (d) temperature.

hand) terms whose limits of oscillation are determined by the prefactor which becomes exponentially small. In fact, if the slope of the right-hand side is less than unity, there will be only one zero corresponding to a single extremum. In other words, if

$$\frac{\epsilon q^2}{\kappa}\, e^{-M_2 q^2/2} < 1 \tag{17}$$

there will be a single minimum on the effective potential surface. Therefore, we can define a critical value of the Gaussian packet variance

$$M_2^c = -\frac{2}{q^2}\ln\left(\frac{\kappa}{\epsilon q^2}\right). \tag{18}$$

If the packet is broader than this critical value, the effective potential is convex and there is a single global energy minimum on the surface. This is a ''convexity condition'' commonly employed in the study of optimization problems. Once a smoothing transformation is defined, the convexity condition can be used to define an optimal smoothing such that all local minima are annihilated and a single energy minimum remains.

Using the Smoluchowski dynamics algorithm, the critical variance M_2^c can be given a physical, thermodynamic meaning in terms of a critical fluctuation size or temperature. Our equation for the time evolution of M_2 in the approximate Smoluchowski dynamics [Eqs. (11)] provides a means of defining the critical temperature. Suppose that the center x_0 is fixed at a particular position. The variance of the distribution will evolve in time until a balance is reached between the kinetic energy contribution and the curvature of the effective potential. For a free particle distribution, the variance will increase linearly in time as $M_2 = 2d(k_BT/m\gamma)t = 2dDt$ where the diffusion constant is $D=k_BT/m\gamma$. For a particle distribution in a convex potential, a steady state value of the variance is reached by bal-

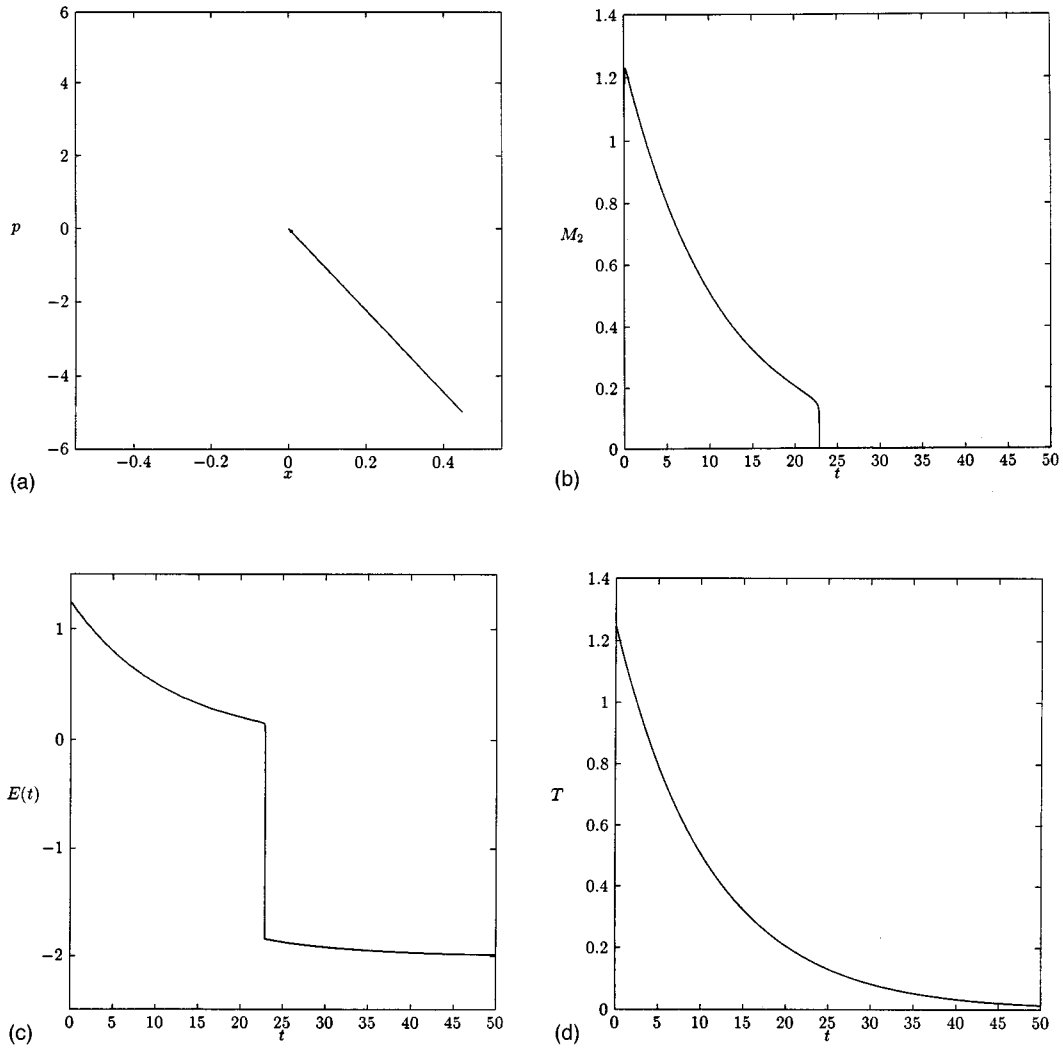

FIG. 2. Time evolution of the configuration space density distribution following the approximate Smoluchowski equations for a Gaussian density distribution and the one-dimensional potential energy $V(x)=x^2/2+2\cos[10x+\pi]$. For the run we set the bath temperature $T=0$ and $\gamma=0.09$ and show (a) the trajectory of the packet center (a linear steepest descent), (b) the configuration space variance M_2, and the time evolution of the packet's (c) total energy and (d) temperature.

ancing the localizing influence of the potential and the delocalizing influence of the kinetic energy (and temperature).

What is an upper bound on the value of the critical annealing temperature T_a? We define T_a to be the temperature above which the variance of the distribution in the steady state limit will necessarily be above the critical value M_2^c for all x_0. The steady state variance is defined by the solution to $\dot{M}_2=0$ or

$$T=\frac{M_2}{k_B}[\kappa-\epsilon q^2\ e^{-M_2q^2/2}\cos[qx_0+\pi]]. \tag{19}$$

Noting that the right-hand side is largest when the cosine term is minus unity provides an upper bound on the critical temperature

$$T_a=\frac{M_2^c}{k_B}[\kappa+\epsilon q^2\ e^{-M_2^cq^2/2}]. \tag{20}$$

Substituting the critical value of M_2^c we find

$$T_a=-\frac{4\kappa}{q^2k_B}\ln\left(\frac{\kappa}{\epsilon q^2}\right). \tag{21}$$

This is a critical temperature for simulated annealing using the Smoluchowski Gaussian packet dynamics. For $T<T_a$, it is likely that the Smoluchowski dynamics will converge to a local energy minimum on the effective potential. For any $T>T_a$, the dynamics of the Smoluchowski equation should converge from the high temperature distribution to a single global minimum on the potential energy surface. Therefore, this critical temperature is the thermodynamic equivalent of a convexity condition. These properties are displayed in Fig. 3 where the effective potential is drawn for four values of the packet variance, in each case using a steady state value of M_2 at each x_0. In simulated annealing, it is possible to reduce the temperature rapidly to a value slightly larger than T_a without loss of precision. As we will see in the next section, this T_a may be significantly higher than the characteristic temperature of escape from a local minimum.

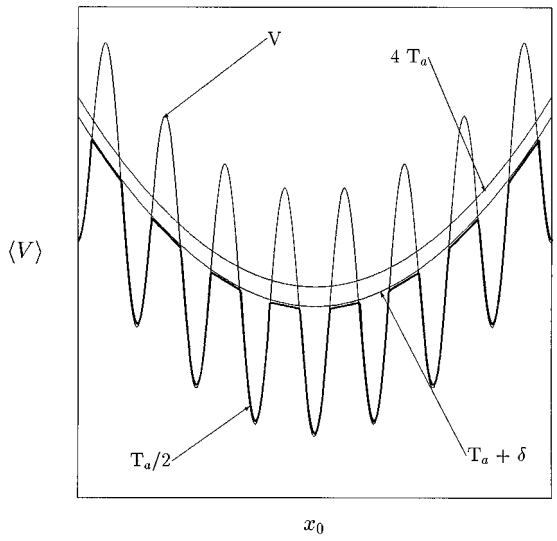

FIG. 3. The effective potential energy $\langle V \rangle$ defined by Eq. (15) using a steady state value of M_2 at each x_0 for four values of temperature $T=0, T_a/2,\ T_a+\delta$ (δ as small as you like) and $4T_a$ demonstrating that for values of $T>T_a$ the effective potential is convex and there is a single energy minimum on $\langle V \rangle$.

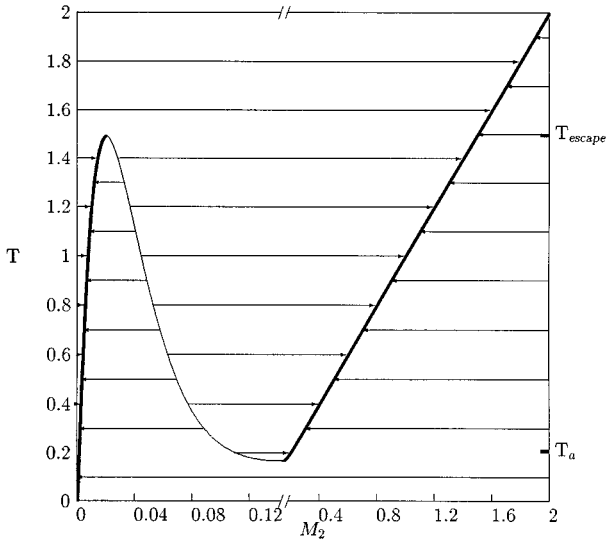

FIG. 4. The space of temperature T and packet variance M_2 for fixed x_0 in the potential $V(x)=x^2/2+2\cos[10x+\pi]$. The line indicates the solutions of the equation $dM_2/dt=0$ which are the steady state packet widths. The heavy lines are stable steady state fixed points while the lighter connecting line represents unstable steady state fixed points. The arrows show how trajectories at a fixed temperature and initial value of M_2 evolve in time, draining to the stable fixed points (steady state widths). The T_a marker signifies the upper bound provided by Eq. (21).

B. Critical temperatures for escape T_{escape} and annealing T_a

In the previous section we found that when the density distribution is annealed from a high temperature (where the variance M_2 is large) to a low temperature (above T_a) there will be a single minimum on the effective potential surface. This implies that the annealing can be performed rapidly to $T \simeq T_a$ without a concern that the trajectory will be kinetically trapped and quenched in a high energy local minimum. Annealing below the critical annealing temperature T_a must proceed more slowly since the effective potential will in general have multiple minima.

In this section, we develop a detailed understanding of the dependence of M_2 on T. At a given temperature, there are a set of fixed points for M_2 which correspond to the zeros of $dM_2/dt=0$. The Gaussian packet will evolve in time to a fixed point of the variance M_2 at a given center position x_0. Consider again the one-dimensional harmonic potential with sinusoidal roughness. Suppose that we set the center of the packet to be $x_0=0$. Due to the symmetry of the potential, the packet center will not move, but the widths will evolve from some initial value to a steady value that is a solution of $dM_2/dt=0$ or

$$T=\frac{M_2}{k_B}[\kappa+\epsilon q^2\ e^{-M_2 q^2/2}]. \tag{22}$$

The solutions of this equation are displayed in Fig. 4. There are two types of solutions to this nonlinear equation. One set of solutions are *stable* fixed points (heavy lines) and the other set are *unstable* fixed points (light line). At a given temperature, any initial value of M_2 which is not itself a steady state value will evolve to a steady state solution corresponding to a stable fixed point. (The evolution of M_2 at a fixed temperature is shown by the arrows.) If the initial width and temperature correspond to an unstable fixed point, M_2 will remain constant in time. However, if the value of M_2 is perturbed, it will evolve to a steady state solution corresponding to a stable fixed point.

When the initial packet is wide, as in an annealing run, the width will decrease in time until reaching its steady state value [a solution of Eq. (22)]. For $T>T_a$, this width will correspond to the width of the broad harmonic potential. For $T<T_a$, the steady state width is narrower and corresponds to the width of the sinusoidal wells of potential roughness. For this potential $T_a \simeq 0.2$.

When the initial packet is narrow, as for a localized delta function distribution, the packet will spread in time in evolving to its steady state value. For $T<T_{\text{escape}}$, the steady state width is narrow and the packet remains trapped in the local sinusoidal potential minimum. When $T>T_{\text{escape}}$, the packet may "escape" from the well and the width will evolve to a value which reflects filling the broad harmonic potential. For this potential $T_{\text{escape}} \simeq 1.5$.

For temperatures $T_a<T<T_{\text{escape}}$ and intermediate initial values of M_2, there is a set of steady state M_2 values corresponding to unstable fixed points. These are solutions of Eq. (22). However, if the values of M_2 at these unstable fixed points are slightly perturbed to larger M_2 (to the right) they will evolve to the stable fixed points corresponding to a steady state value in the broad harmonic potential. A slight perturbation to smaller M_2 (to the left) will cause the variance to narrow until the packet is localized in a narrow well on the sinusoidal roughness potential.

This analysis provides several conclusions. (1) For an initially localized packet, there is a critical temperature T_{escape} which must be reached before the Gaussian packet can delocalize beyond the local potential well. Reasonably, the value of the escape temperature is comparable to the

depth of the roughness potential well $2\epsilon=4$. (2) For an initially delocalized packet, there is a critical temperature T_a above which there is a single minimum on the effective potential surface. The system may be rapidly annealed by setting the temperature to a value slightly greater than T_a followed by a slower annealing to $T=0$. (3) While the value of T_{escape} is comparable to the depth of the wells on the rough potential, the value of the critical annealing temperature T_a is nearly an order of magnitude smaller. Using the Gaussian density annealing through the Smoluchowski equation, it is possible to anneal rapidly to a temperature significantly smaller than the well depth without loss of precision. The disparity between T_{escape} and T_a represents a hysteresis. With Gaussian density annealing, there will be a single energy minimum for all temperatures above T_a making it possible to anneal *by quench* to this low temperature, significantly below T_{escape}.

For the simulated annealing of a point particle, frustration due to trapping in local minima occur at a temperature comparable to the barrier height of the roughness potential. In general, it seems preferable for the packet to remain delocalized to as low a temperature as possible during the annealing process. However, this does not guarantee that when the packet does collapse that it is more likely to localize in the global minimum. Orešič and Shalloway have considered this point in detail using a method related to that proposed here.[14] We now turn to a more general discussion of the coarse graining inherent to the Gaussian density simulated annealing.

C. Coarse graining and potential smoothness

We can explore in greater detail the nature of the approximate Smoluchowski dynamical equations by focusing on a one-dimensional example. Zwanzig has described a coarse grained diffusion of a particle in a rough one-dimensional potential,[13] extending a result of Lifson and Jackson.[15] While the coarse grained dynamical equations are arrived at by conjecture, the mean first passage time is preserved. Consider the potential $V(x)=V_0(x)+V_1(x)$ where $V_0(x)$ is a smooth "background" potential which varies on a length scale which is long compared to the more rapidly varying "roughness" $V_1(x)$. D is the diffusion constant. The effective Smoluchowski equation for coarse grained dynamics is defined[13] by motion in an effective potential:

$$V_{\text{eff}}(x)=V_0(x)-\frac{1}{\beta}\ln[\langle e^{-\beta V_1}\rangle] \tag{23}$$

with an effective diffusion constant:

$$D_{\text{eff}}=D/(\langle e^{\beta V_1}\rangle\langle e^{-\beta V_1}\rangle). \tag{24}$$

The coarse graining average is written $\langle\cdot\rangle$.

For example, Zwanzig considered diffusion in a rough potential with the form of Eq. (13).[13] In this case $\langle\exp(\pm\beta V_1)\rangle=I_0(\beta\epsilon)$ where I_0 is a modified Bessel function. In the low temperature limit, $\beta\epsilon$ is large and

$$D_{\text{eff}}=D(2\beta\epsilon)e^{-2\beta\epsilon}, \tag{25}$$

which has the suggestive Arrhenius form indicating that diffusion on a length scale long compared with $1/q$ is activated involving the crossing of barriers on the order of 2ϵ. The effective potential V_{eff} and effective diffusion constant D_{eff} correspond to motion coarse grained with a resolution of $\Delta x=\pi/q$. How do the approximate equations of motion for the Gaussian density compare with this more rigorous coarse grained treatment?

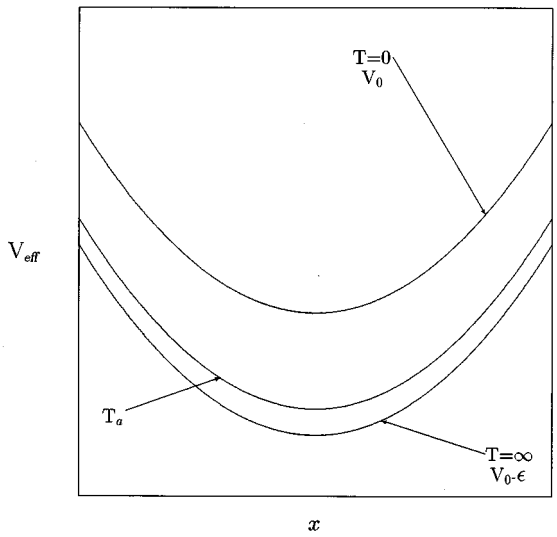

FIG. 5. The effective potential energy V_{eff} defined by Eq. (23) for the high temperature limit ($V_{\text{eff}}=V_0$) and low temperature limit ($V_{\text{eff}}=V_0-\epsilon$) as well as at the critical annealing temperature T_a.

The Smoluchowski equations of motion for the Gaussian density packet do not have a renormalized diffusion constant. At all temperatures, the diffusion constant is $D=k_BT/m\gamma$. However, as we have shown above, the effective potential changes drastically as the temperature is reduced. At temperatures $T>T_a$ the potential is convex while at $T<T_a$ the potential is rough. In the coarse grained dynamics the effective potential V_{eff} is always convex. Changing the temperature merely shifts the potential in energy by a position independent constant (see Fig. 5). The high temperature limit of V_{eff} is V_0 while at low temperatures it is $V_0-\epsilon$. However, the diffusion constant D_{eff} is D at high temperatures while D_{eff} decreases exponentially with decreasing temperature [see Eq. (25)].

Therefore, at high temperatures the dynamics of the Gaussian packet is consistent with the coarse grained effective Smoluchowski equation. As the temperature is lowered, each coarse grained dynamical theory treats the slowing of the motion differently. (1) The coarse graining intrinsic to the Gaussian density packet leads to an effective potential $\langle V\rangle$ which changes dramatically between the low and high temperature limits; the form of the diffusion constant remains unchanged (the high temperature limit of D_{eff}). (2) In the effective Smoluchowski equation of Zwanzig, the potential is smoothed and changes with temperature only through a constant shift (the high temperature limit of $\langle V\rangle$); the slowing of diffusion due to the roughness potential is manifest in the effective diffusion constant D_{eff} which decreases dramatically at low temperatures.

IV. CONCLUSIONS

We have presented algorithms for dynamical simulated annealing of a continuous classical density distribution. At the level of the Gaussian density approximation, the equations of motion have the simple interpretation that the distribution center moves on an effective ''coarse grained'' potential energy surface. These methods share the favorable properties of many potential smoothing[16,17] or ''continuation'' methods[18] developed for many body optimization problems. However, the algorithm remains a classical simulated annealing algorithm and preserves the useful insights provided by the analogy with statistical mechanical annealing.

Our explanation of packet annealing in terms of two critical temperatures implies that there is a ''hysteresis'' involved.[19] When the packet is initially localized, there is a high temperature T_{escape} which must be reached before the packet can delocalize beyond the confines of the local minimum. However, in approaching this localization/delocalization transition as an initially delocalized packet, there is a much lower critical temperature T_a in the annealing below which the packet is localized. For the one-dimensional model potential studied here $T_a \ll T_{\text{escape}}$. For a classical point particle, there will be a single characteristic temperature which characterizes this transition, regardless of whether it is approached from low or high temperature. The analysis of the one-dimensional system provides insight into the mechanism of packet annealing. It appears that the hysteresis effect can be a positive aspect of the packet annealing which allows a rapid quench to a low temperature without danger of trapping at high energies.

ACKNOWLEDGMENTS

We thank Ioan Andricioaei and David Hsu for helpful discussions. This work was supported by a grant from the National Science Foundation (CHE-9306375).

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).
[2] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (Wiley, New York, 1990).
[3] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991); H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Chem. **17**, 451 (1988).
[4] M. R. Hoare and J. McInnes, Faraday Discuss. Chem. Soc. **61**, 12 (1976).
[5] R. S. Berry, Chem. Rev. **93**, 2379 (1993).
[6] E. B. Baum, Phys. Rev. Lett. **57**, 2764 (1986).
[7] F. H. Stillinger, Phys. Rev. B **32**, 3134 (1985).
[8] H. Risken, *The Fokker–Planck Equation* (Springer-Verlag, Berlin, 1989).
[9] J. Ma and J. E. Straub, J. Chem. Phys. **101**, 533 (1994).
[10] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1983).
[11] The transformed effective potential for the distribution center has the property that roughness with a length scale L is eliminated. For example, the Fourier transform of the averaged potential $\langle V \rangle$ is
$$\langle \hat{V} \rangle(\mathbf{k}) = \hat{V}(\mathbf{k}) e^{-k^2 M_2/2}, \tag{26}$$
where $\hat{V}(\mathbf{k})$ is the Fourier transform of the physical potential energy function. The wave vector $\mathbf{k}$ defines a convenient length scale for the smoothing $L = 1/|\mathbf{k}|$ (Ref. 18).
[12] These equations, combined with a cooling schedule, define an optimization algorithm. If T is held constant, the dynamics evolve to a static distribution $\hat{\rho}(r|T_0)$. The temperature may then be reduced and the equations of motion integrated again until convergence and a new static distribution $\hat{\rho}(r|T_1 < T_0)$ is reached. Following such a protocol makes this algorithm similar to the diffusion equation method of Scheraga and coworkers (Ref. 16).
[13] R. Zwanzig, Proc. Natl. Acad. Sci. U. S. A. **85**, 2029 (1988).
[14] M. Orešič and D. Shalloway, J. Chem. Phys. **101**, 9844 (1994). They consider a Gaussian packet coarse graining of the Gibbs–Boltzmann distribution rather than the energy. They have examined two one-dimensional potentials to provide examples of when packet annealing will and will not find the global minimum. The model potential considered here corresponds to what they refer to as the ''strong scaling'' case when the global minimum at T_{escape} maps to the global minimum at T_a (temperatures they refer to as T_{hi} and T_{low}).
[15] S. Lifson and J. L. Jackson, J. Chem. Phys. **36**, 2410 (1962).
[16] L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem. **93**, 3339 (1989).
[17] D. Shalloway, in *Recent Advances in Global Optimization*, edited by C. A. Floudas and P. M. Pardalos (Princeton University, Princeton, 1992).
[18] Z. Wu, Cornell Theory Center Technical Report, CTC93TR143 (July 1993).
[19] A hysteresis similar to that noted here can be seen in Fig. 3 (region II) of P. Amara, D. Hsu, and J. E. Straub, J. Phys. Chem. **97**, 6715 (1993) for a quantum mechanical annealing study of an asymmetric double well potential.