Application of the Renormalization Group to Deterministic Global Minimization of Molecular Conformation Energy Functions

DAVID SHALLOWAY
Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, U.S.A.

(Received: 11 November 1991; accepted: 21 November 1991)

Abstract. We outline a new global minimization method in which the Gibbs distribution of the objective function is deterministically annealed by tracing the evolution of a multiple-Gaussian-packet approximation. Solutions are reached by iterative approximations with decreasing coarse-graining of both objective-function and spatial scales. Results from application of a partial implementation to the atomic-microcluster conformation problem are presented.

Key words. Renormalization group, global optimization, conformation, microcluster, annealing.

1. Introduction

1.1. GLOBAL MINIMIZATION AND THE PROTEIN-FOLDING PROBLEM

The critical need for methods for theoretical prediction of protein 3-dimensional conformations from amino acid sequences, the "protein folding problem", is well appreciated [18, 27]. While amino acid sequences can be created and modified at will, methods to predict the conformational changes that will be induced are lacking. Most predictive methods are based on: (1) amino acid pattern matching using the database of experimentally determined protein structures [2]; or (2) global energy-minimization based on the hypothesis that native protein conformations correspond to global minima of free-energy functions [28]. These methods are complementary since pattern-matching approaches provide approximate starting points for energy minimizations.

The most immediate obstacle to global energy-minimization approaches is the extraordinarily large number of local minima of typical conformational energy functions. This results from the high dimensionality of the conformation spaces which are parametrized by the spatial coordinates of 10^3–10^4 atoms [4]. While additional obstacles may exist (e.g., current conformational energy functions may not be sufficiently accurate), they can be more effectively addressed once the global minimization problem has been solved. Present methods are practically limited to problems containing relatively few degrees of freedom [e.g. $N \sim 0(10^2)$]; methods specially designed for problems with many degrees of freedom are needed.

Journal of Global Optimization **2**: 281–311, 1992.

1.2. ANNEALING AND SPATIAL SCALE

Many proteins rapidly and spontaneously thermally migrate to their lowest energy state as temperature is lowered. In effect, they are analog computers that can solve the global minimization problem. This important clue motivates the development of computational methods that can efficiently emulate this process.

We begin such a development by considering $p_T(R)$, the probability distribution at temperature T of a protein or other physical system with objective (energy) function $H(R)$ defined over a continuous domain of conformations parametrized by R. This is the Gibbs distribution

$$p_T(R) = e^{-H(R)/k_B T}$$

$$(k_B \equiv \text{Boltzmann's constant}). \tag{1.1}$$

It converges at sufficiently low temperature T_{lo}, to the simple form[1]

$$\tilde{\tilde{p}}_{T_{\text{lo}}}(R) \approx e^{-|(R-R_g^*)/\Lambda_{\text{lo}}|^2} \tag{1.2}$$

which is concentrated near the global minimum R_g^*. The simulated annealing algorithm [19], when applied to continuous systems [3, 5, 31], can be viewed as a method for stochastically tracing this convergence by Monte Carlo simulation. As the number of objective function evaluations becomes large, it explores conformational space with probability density given by (1.1). $\Lambda(T)$, the size of the random jumps that are used in the stochastic search at temperature T, is a decreasing function of T in most implementations. That is, large areas of conformation space are explored at high T and progressively smaller regions at lower T. A central, though often unstated, assumption of simulated annealing is that the global minimum can be identified by a progressively narrowing search where the region explored at temperature $T2$ ($T2 < T1$) is within the dominant, most probable, region identified at $T1$. This assumption, which is required for rapid convergence to the global minimum, is justified by the expectation that the conformation of a physical system will thermally fluctuate through a large region of conformation space at high temperatures and through progressively smaller regions as temperature is lowered.

This behavior is demonstrated in Figure 1 which displays $p_T(R)$ for a hypothetical 2-dimensional objective function at different temperatures. At high temperature, $p_T(R)$ can be roughly modeled by a large Gaussian packet (Fig. 1A). The packet radius decreases with decreasing temperature (Fig. 1B) until a point is reached (Fig. 1C) where it branches into multiple packets. At yet lower temperatures [when (1.2) is valid], $p_T(R)$ is dominated by a single packet centered at R_g^* (Fig. 1E). Simulated annealing in effect traces the approximate locations of the packets as they converge towards R_g^* by tracing the path of a sampling point through multiple stochastic runs. But its application to continuous problems of high dimensionality is limited by its requirement for large numbers of objective

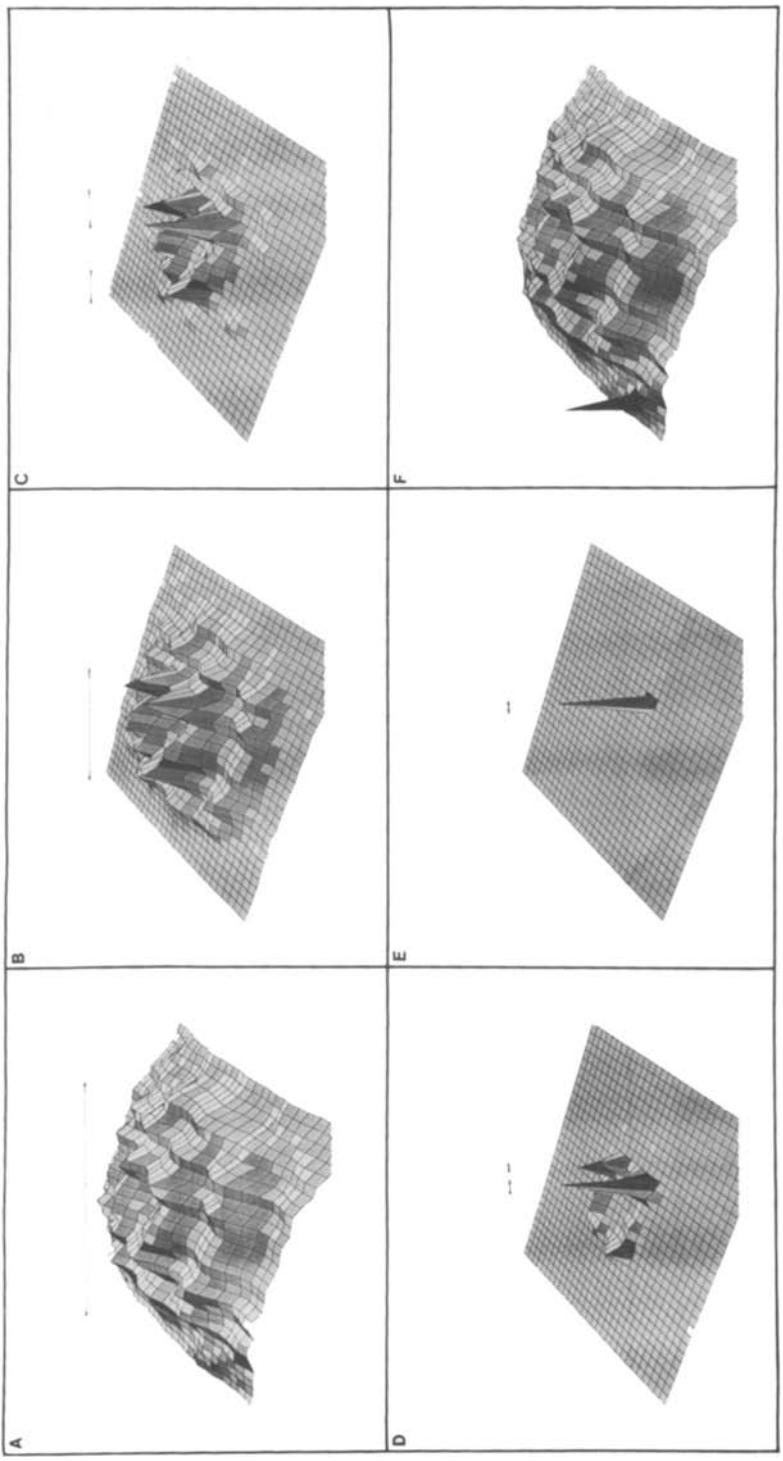

Fig. 1. **Temperature and Size Scales in $p_T(R)$.** The Gibbs distributions $p_T(R)$ of a hypothetical 2-dimensional system at decreasing temperatures are displayed in panels A–E (A = highest temperature, E = lowest temperature). The bars approximate the sizes of the appropriate packets at each temperature. (F) An anomalous objective-function that frustrates annealing methods.

function evaluations. Our goal is to construct a deterministic annealing algorithm that converges to R_g^* with fewer objective function evaluations.

1.3. RENORMALIZATION GROUP APPROACH TO GLOBAL MINIMIZATION

We have begun the application of renormalization group (RG) ideas, which have been exceptionally effective in other areas of statistical physics [33], to this problem. Global minimization is significantly different from the continuum and lattice physics problems where RG methods have previously been applied, so our mathematical formulation is unique. For example, analysis is in coordinate rather than momentum space, cutoffs go from high-to-low rather than from low-to-high momentum, and fixed points are not relevant. However, many of the basic concepts and principles are the same.

RG analysis focuses on the relationship between descriptions of a physical system at different size scales $\Lambda(T)$. In this context, $p_T(R)$ can be viewed as a (positive definite) field of independent variables over the N-dimensional space parametrized by R. It describes the system at infinite spatial resolution (corresponding to $\Lambda = 0$). But, at every T, its predominant variations occur at finite spatial scales. The $\Lambda = 0$ description is inappropriate and contributes unnecessary computational complexity. Instead, we represent the behavior of the predominant components using a smaller number of independent "effective" variables, p_α, R_α^0, and Λ_α, which parametrize partially-averaged field fluctuations or "packets" which extend over finite regions of conformation space. These variables define $\tilde{\tilde{p}}_T(R)$, an approximation to $p_T(R)$ as a finite sum of Gaussian packets:

$$\tilde{\tilde{p}}_T(R) = \sum_{\{\alpha\}_T} p_\alpha(T) C[\Lambda_\alpha(T)] e^{-|[R-R_\alpha^0(T)]/\Lambda_\alpha(T)|^2} \tag{1.3}$$

where C is a normalization constant defined by

$$C(\Lambda) \int e^{-|R/\Lambda|^2} \, dR = 1 .$$

(For simplicity we assume that there are no constraints and that multidimensional integrations are over the entire domain.) Each packet is assigned a unique index α. The sum spans the complete set of indices $\{\alpha\}_T$ for the packets that represent the dominant structure of $p_T(R)$. As illustrated in Figure 1, this set grows and diminishes as T decreases. Each packet α is described by three parameters: R_α^0 (packet center), Λ_α (packet width), and p_α (packet occupation probability). For simplicity, we begin the discussion by considering only isotropic packets even though this overestimates the spread of $\tilde{\tilde{p}}_T(R)$.

This decomposition of $p_T(R)$ dissects its fluctuations into low spatial-frequency components described by the packet variables and high frequency components which kinetically equilibrate rapidly within each packet. We will see that packet motions are governed by an "effective potential" $\tilde{H}_{\Lambda,T}$ which is derived from the original energy function H by integrating out the contributions from spatial fluctuations of size scale Λ. By this means the time-averaged effects of the

high-frequency fluctuations on the low-frequency motions can be calculated. The resultant separation of spatial-frequency components in position space is equivalent to the more customary RG separation in momentum space.

As in simulated annealing, the procedure begins at temperature T_{hi}, where $k_B T_{hi}$ is large relative to the objective function scale. It is initialized by constructing a single-packet approximation to $p_{T_{hi}}(R)$ that spans the dominant region of R (e.g., of size corresponding to the arrowhead bar in Fig. 1A). Packet size and position are initially determined by iterative solution of a set of self-consistent constraints. The same constraints are used to adjust packet size and position as the temperature is lowered in small steps. The Λ_α will generally decrease and packets will often subdivide (branch) as T is reduced (cf Figs. 1B and 1C). The development of $\tilde{\tilde{p}}_T(R)$ will be traced by calculating $p_\alpha(T)$, $R^0_\alpha(T)$, and $\Lambda_\alpha(T)$, for all the packets until $\tilde{\tilde{p}}_T(R)$ reduces to a single packet centered about R^*_g at low temperature T_{lo}. To control computational effort, packets with small $p_\alpha(T)$ will be discarded. Thus, the effective potential $\tilde{H}_{\Lambda,T}$ ultimately converges to H and an exact minimum is obtained. The previous discussion suggests that this will be the global minimum.

Branching and merging of packets is naturally accommodated and will not interfere with the tracing procedure. However, for some objective functions, the global minimum may be located in a region where the Gibbs distribution is, on-average, small (e.g., Fig. 1F). The packet containing this minimum will not emerge from subdivision or shrinking of a larger packet and may not be detected. This difficulty is not particular to the packet approximation method; simulated annealing is also likely to fail (or to require inordinate time to detect the minimum) in such cases. Nonetheless, there are many problems, particularly those where the objective function is the partially-separable sum of a large number of partially-independent terms, where it is unlikely that such anomalous behavior will occur or where anomalous minima are not of primary interest.

1.4. OVERVIEW

In this paper we outline the "packet annealing" method of global minimization and show how it can be applied to a simple molecular conformation problem.

In Section 2 we describe $\tilde{H}_{\Lambda,T}(R)$, the effective potential that governs $\tilde{\tilde{p}}_T(R)$. A physical interpretation is presented in Section 3 and the packet annealing algorithm is outlined in Sections 4 and 5. The molecular conformation problem and the microcluster model is described in Section 6. We show how packet-annealing can be applied to this problem in Section 7 and present preliminary numerical results in Section 8.

2. Packet Expansion of $p_T(R)$

$\tilde{\tilde{p}}_T(R)$ is an effective distribution that does not approximate $p_T(R)$ in a uniformly convergent sense: $|\tilde{\tilde{p}}_T(R) - p_T(R)| \not< \varepsilon$. Instead, it provides a method for dissect-

ing the spatial variations of $p_T(R)$ by size-scale. The expansion is constructed so that packets do not significantly overlap; that is,

$$|R_\alpha^0 - R_\beta^0| > \Lambda_\alpha + \Lambda_\beta \,. \tag{2.1}$$

[The procedures that ensure that (2.1) is satisfied for all α and β are discussed in Section 5.] Within each region $|R - R_\alpha^0| \leqslant \Lambda_\alpha$ variations on spatial scales $<\Lambda_\alpha$ are absorbed into a single Gaussian packet approximation; variations on scales $>\Lambda_\alpha$ are represented by the positions and amplitudes of the packets. Only the regions where $p_T(R)$ is significant are modeled since packets having small p_α are deleted. $\tilde{\tilde{p}}_T(R)$ is self-consistently defined by the requirement that, when sampled in region α by a function $S_\Lambda(R)$ having length scale $\Lambda \geqslant \Lambda_\alpha$, it yields results close to those obtained by sampling $p_T(R)$ directly:

$$\begin{aligned} &\int p_T(R')S_\Lambda(R - R')\,\mathrm{d}R \approx \int \tilde{\tilde{p}}_T(R')S_\Lambda(R - R')\,\mathrm{d}R' \\ &|R - R'| < \Lambda_\alpha \,, \Lambda \geqslant \Lambda_\alpha \,, \end{aligned} \tag{2.2}$$

where the sampling function $S_\Lambda(R)$ has the properties

$$S_\Lambda(R) \to 0 \text{ for } |R|/\Lambda \to \infty$$

$$\int S_\Lambda(R)\,\mathrm{d}R = 1$$

We use

$$S_\Lambda(R) = C(\Lambda)\, e^{-|R/\Lambda|^2} \tag{2.3}$$

The integral of $p_T(R)$ over all conformation space gives the partition function $Z(T)$. We define the spatially-localized integral in the left-hand-side of (2.2) as the "local partition function" $\tilde{Z}_{\Lambda,T}(R)$ and, following Wilson [33], define the "effective energy function" $\tilde{H}_{\Lambda,T}(R)$ by

$$\begin{aligned} e^{-\tilde{H}_{\Lambda,T}(R)/k_BT} \equiv \tilde{Z}_{\Lambda,T}(R) &\equiv \int p_T(R')S_\Lambda(R - R')\,\mathrm{d}R' \\ &= C(\Lambda) \int e^{-H(R')/k_BT}\, e^{-|(R-R')/\Lambda|^2}\,\mathrm{d}R' \,. \end{aligned} \tag{2.4}$$

[The right-hand-side of (2.4) is obtained from (1.1) and (2.3).] Renormalization group transformations are usually defined in momentum space, but transformations in configuration space are more appropriate here. (2.4) corresponds to multiplying the Fourier transform of $p_T(R)$ by $\exp(-4k^2\Lambda^2)$, where k is the transform momentum variable. That is, it corresponds to a suppression of spatial-frequencies above cutoff $\sim O(2/\Lambda)$.

$\tilde{Z}_{\Lambda,T}(R)$ satisfies the boundary condition

$$\lim_{\Lambda \to \infty} \tilde{Z}_{\Lambda,T}(R)/C(\Lambda) = Z(T) \,. \tag{2.5}$$

$\tilde{H}_{\Lambda,T}(R)$ is a non-rescaled spatial-domain renormalization group transform of $H(R)$ and satisfies the boundary condition

$$\lim_{\Lambda \to 0} \tilde{H}_{\Lambda,T}(R) = H(R) . \tag{2.6}$$

A complete renormalization group transformation would also rescale R and $\tilde{H}$. Rescalings are incorporated in the computational implementation but, for simplicity, are omitted in this discussion.

(2.2), with (1.3), can be expanded in a Taylor's series in $(R - R_\alpha^0)$ for each α. When restricted to contributions from a single packet α, the first three terms yield constraints on p_α, R_α^0, and Λ_α in terms of $\tilde{H}_{\Lambda,T}(R)$:

$$\tilde{Z}_{\Lambda,T}(R_\alpha^0) \equiv e^{-\tilde{H}_{\Lambda,T}(R_\alpha^0)/k_B T} \approx p_\alpha C(\Lambda) C(\Lambda_\alpha) / C[(\Lambda^{-2} + \Lambda_\alpha^{-2})^{-1/2}] \tag{2.7}$$

$$-k_B T \frac{\partial \tilde{Z}_{\Lambda,T}(R)/\partial R}{\tilde{Z}_{\Lambda,T}(R)} \bigg|_{R=R_\alpha^0} = \partial \tilde{H}_{\Lambda,T}(R)/\partial R \big|_{R=R_\alpha^0} \approx 0 \tag{2.8}$$

$$\text{MINIMUM}\left\{ -k_B T \frac{\partial^2 \tilde{Z}_{\Lambda,T}(R)/\partial R^2}{\tilde{Z}_{\Lambda,T}(R)} \bigg|_{R=R_\alpha^0} \right\}$$

$$= \text{MINIMUM}\{\partial^2 \tilde{H}_{\Lambda,T}(R)/\partial R^2 \big|_{R=R_\alpha^0}\} \approx \frac{2k_B T}{\Lambda^2 + \Lambda_\alpha^2} \tag{2.9}$$

($\Lambda \geq \Lambda_\alpha$; single-packet approximation. The dependence of p_α, R_α^0, and Λ_α on T is implicit. MINIMUM and MAXIMUM operators refer to the eigenvalues of the matrix).

If non-isotropic Gaussian packets were used, Λ_α^{-2} would be the symmetric Gaussian coefficient matrix and (2.9) (without the MINIMUM operator) would represent a matrix of constraints. However, if we restrict to isotropic packets, only one condition can be fixed. We fix Λ_α according to the minimum eigenvalue of the Hessian of $\tilde{H}_{\Lambda,T}(R)$ so that the size of the Gaussian packet matches $p_T(R)$ in the least-localized direction. This conservative choice assures that an anisotropic concentration of $p_T(R)$ in region α will be contained within isotropic Gaussian packet α.

If $p_T(R)$ were well-approximated by a single packet, (2.7)–(2.9) would be satisfied for all $\Lambda \geq \Lambda_\alpha$. However, in general, when Λ is large, S_Λ will overlap multiple packets and (2.7)–(2.9) will not be valid. But, because of the non-overlap condition (2.1), (2.7)–(2.9) will be approximately valid even in the presence of multiple packets when $\Lambda = \Lambda_\alpha$. Restricting (2.7)–(2.9) by this condition generates a set of self-consistent conditions that fix p_α, R_α^0 and Λ_α:

$$\tilde{Z}_{\Lambda_\alpha,T}(R_\alpha^0) \equiv e^{-\tilde{H}_{\Lambda_\alpha,T}(R_\alpha^0)/k_B T} = p_\alpha 2^{-N/2} C(\Lambda_\alpha) \tag{2.10}$$

(N = dimensionality of the domain)

$$-k_B T \frac{\partial \tilde{Z}_{\Lambda_\alpha,T}(R)/\partial R}{\tilde{Z}_{\Lambda_\alpha,T}(R)} \Big|_{R=R_\alpha^0} = \partial \tilde{H}_{\Lambda_\alpha,T}(R)/\partial R \,|_{R=R_\alpha^0} = 0 \tag{2.11}$$

$$\text{MINIMUM}\left\{-k_B T \frac{\partial^2 \tilde{Z}_{\Lambda_\alpha,T}(R)/\partial R^2}{\tilde{Z}_{\Lambda_\alpha,T}(R)} \Big|_{R=R_\alpha^0}\right\}$$

$$= \text{MINIMUM}\{\partial^2 \tilde{H}_{\Lambda_\alpha,T}(R)/\partial R^2 \,|_{R=R_\alpha^0}\} = \frac{k_B T}{\Lambda_\alpha^2} . \tag{2.12}$$

In practice, (2.11) and (2.12) will be solved by alternating iteration. Once an initial solution has been found, it can be iteratively propagated through small downward steps in T (see Section 4). Except for isolated discontinuities caused by branching and merging of packets (see Section 5), the changes in $R_\alpha^0(T)$ and $\Lambda_\alpha(T)$ at each step in T will be small and only one or a few iterations will be required.

3. Physical Interpretation

Consider the problem of finding the lowest point on the surface of the Earth. The simulated annealing procedure corresponds to shaking the earth (earthquake!) with progressively lower intensities (temperature) and tracking the position of a very small test-object (e.g., a marble) as it stochastically migrates to the lowest point. The process is inefficient because the test-object samples the energy function (height) over regions that are unnecessarily small compared to the sizes of its stochastic jumps. The more efficient RG procedure would be to start with a "soft" test-object or packet with diameter comparable to the sizes of the continents and oceans (e.g., a 10,000 kilometer beach-ball) and to iteratively minimize its positions as both temperature and packet size are progressively reduced. If the packet diameter is properly matched to the temperature, most thermal fluctuations will be absorbed internally and the (essentially deterministic) motion of its center-of-mass will be governed by steepest descent minimization of the effective energy function. In this analogy, the packet will first move to the Pacific Ocean; later, as temperature and diameter are reduced, to the Western Pacific, and finally to the Marianna's Trench. At points where equivalent downhill choices are available, it may be necessary to split the search and to track packets part way down multiple paths. (In the example of Figure 1, this occurs at the temperature transition between panels B and C.) This algorithm is not expected to find the actual lowest point on the surface of the Earth (a narrow test-hole in Siberia) because it only finds minima that are apparent at every size scale. The test-hole is apparent only at small size scales: it is an anomalous minimum which would not be present in the absence of the local-entropy-reducing activities of *Homo sapiens*.

The packet annealing method embodies this approach. We imagine constructing

a packet by constraining, with massless linear springs of spring constant $2k_BT/\Lambda_\alpha^2$, a large number N_p of point masses (masses $= m$; positions $= \{R^i\}$, $i = 1..N_p$) to a central point R. (Similar results would be obtained if the point masses were connected together by springs between each pair.) The energy function for this packet in the gravitational field is

$$MgH_p(\{R^i\}, R) = \sum_{i=1}^{N_p} \left[mg\, H(R^i) + \frac{k_BT}{\Lambda_\alpha^2} |R^i - R|^2 \right] \tag{3.1}$$

where g is the gravitational constant, $M = N_p m$ is the total mass of the packet, and H is the height function. The width of the packet is determined by the spring constant and temperature; we have selected units so that this is Λ_α. R, the point where the springs are connected together, will be located at the center-of-mass where the total force exerted by the springs vanishes:

$$-\frac{\partial H_p}{\partial R} = 0 \Rightarrow R = \frac{1}{N_p} \sum_{i=1}^{N_p} R^i .$$

In the limit $N_p \to \infty$, the R^i fluctuate independently, so, by the Central Limit Theorem, the mean-square fluctuation of R will be much less $[O(1/N_p)]$ than the mean-square fluctuation of the individual R^i. Thus, as in the Born-Oppenheimer method for nuclear motions [23], we integrate the Gibbs distribution for mgH_p over the R^i to get the effective free energy as a function of R:

$$\begin{aligned} e^{-Mg\tilde{H}_{\Lambda_\alpha,T}(R)/k_BT} &\propto \lim_{N_p\to\infty} \int e^{-mgH_p(\{R^i\},R)/k_BT}\, \delta\left(R - \frac{1}{N_p}\sum_{i=1}^{N_p} R^i\right) \prod_{i=1}^{N_p} dR^i \\ &\propto \lim_{N_p\to\infty} \left[\int e^{-[mgH(R')/k_BT + |R'-R|^2/\Lambda_\alpha^2]}\, dR' \right]^{N_p} \\ \Rightarrow e^{-mg\tilde{H}_{\Lambda_\alpha,T}(R)/k_BT} &\propto \int e^{-[mgH(R')/k_BT + |R'-R|^2/\Lambda_\alpha^2]}\, dR' . \end{aligned} \tag{3.2}$$

If the factors of mg are absorbed into the definitions of H and $\tilde{H}$ to convert them from height to energy functions, then the definition of $\tilde{H}$ in (3.2) is identical to that in (2.4) except for an unimportant additive constant.

$\tilde{H}_{\Lambda_\alpha,T}(R)$ governs the stochastic motion of the packet in the same way that $H(R)$ governs the stochastic motion of a point mass. The central difference is that we can choose Λ_α so that the stochastic fluctuations of R will be relatively small. Then the packet will move according to

$$\frac{dR}{dt} \propto \frac{-\partial \tilde{H}_{\Lambda_\alpha,T}(R)}{\partial R} \tag{3.3}$$

(i.e., overdamped motion along a steepest-descent trajectory) until equilibrium is reached at a local minimum R_α^0

$$\frac{\partial \tilde{H}_{\Lambda_\alpha,T}(R)}{\partial R}\Big|_{R=R_\alpha^0} = 0\,. \tag{3.4}$$

This is identical to (2.11).

The condition that the stochastic fluctuations are relatively small is

$$\text{MAXIMUM}\langle (R - R_\alpha^0)^2 \rangle_{\Lambda_\alpha,T} \leqslant \Lambda_\alpha^2 \tag{3.5}$$

where

$$\langle (R - R_\alpha^0)^2 \rangle_{\Lambda_\alpha,T} \equiv \frac{\int (R - R_\alpha^0)^2 e^{-\tilde{H}_{\Lambda_\alpha,T}(R)/k_BT}\,dR}{\int e^{-\tilde{H}_{\Lambda_\alpha,T}(R)/k_BT}\,dR}\,.$$

Approximating

$$\tilde{H}_{\Lambda_\alpha,T}(R) \approx \tilde{H}_{\Lambda_\alpha,T}(R_\alpha^0) + \frac{1}{2}\frac{\partial^2 \tilde{H}_{\Lambda_\alpha,T}(R)}{\partial R^2}\Big|_{R=R_\alpha^0}(R - R_\alpha^0)^2 \tag{3.6}$$

we have

$$\langle (R - R_\alpha^0)^2 \rangle_{\Lambda_\alpha,T} \approx \frac{k_BT}{\dfrac{\partial^2 \tilde{H}_{\Lambda_\alpha,T}(R)}{\partial R^2}\Big|_{R=R_\alpha^0}}\,. \tag{3.7}$$

(3.5) and (3.7) yield a relationship between Λ_α and the Hessian of $\tilde{H}_{\Lambda_\alpha,T}$ at the stability position R_α^0:

$$\text{MAXIMUM}\left\{\frac{1}{\dfrac{\partial^2 \tilde{H}_{\Lambda_\alpha,T}(R)}{\partial R^2}\Big|_{R=R_\alpha^0}}\right\} \leqslant \Lambda_\alpha^2/k_BT\,. \tag{3.8}$$

(3.8) provides a lower bound[2] on Λ_α. Since maximal information is provided when the packet is as small as possible without significant fluctuation in R, the equality in (3.8) fixes Λ_α. This is equivalent to (2.12).

Because (3.6) is only approximate, the fluctuations in R about R_α^0 will only be small for a characteristic time period Δt that is small compared to the period Δt_{global} required for equilibration with nearby local minima. That is, (3.3) and (3.4) are valid for

$$\Delta t_{\text{point}} \ll \Delta t \ll \Delta t_{\text{global}} \tag{3.9}$$

where Δt_{point} is the characteristic period required for local equilibration of the small test masses that compose the packet. After longer times ($\Delta t > \Delta t_{\text{global}}$), "tunneling" of the packet to other local minima will occur. Thus, the R_α^0 defined by (3.4) define metastable "states" of the system.

4. Packet Annealing Algorithm

4.1. ENERGY-LEVEL TRAJECTORIES

The overall development of the annealing process can be followed by studying the trajectories described by the set of $F_\alpha(T) \equiv -k_BT \log[p_\alpha(T)]$. These are the

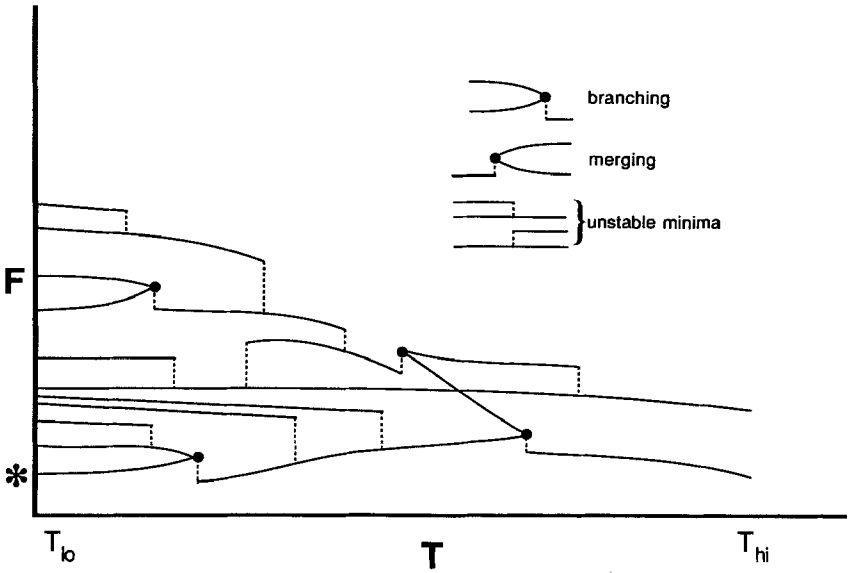

Fig. 2. **Hypothetical Packet-Annealing Energy Level Diagram.** (A) The "free-energies" of packets, given by $F_\alpha(T) = -k_B T \log [p_\alpha(T)]$, for a hypothetical objective function are plotted as a function of T. In this example the "ground state" (global minimum) at $T = T_{lo}$ (marked with an asterisk) is continuously connected to the ground state at all T. Vertical dotted lines show the discontinuous trajectory connections that occur when states become unstable. In practice, trajectories with $F_\alpha(T)$-$F^*(T) > -k_B T \log(p_{\min})$ [where $F^*(T)$ is the lowest energy state at T and $p_{\min} \ll 1$] will not be calculated. (B) The primary types of energy level trajectory singularities involving bifurcations. (Branches and mergers between three or more packets are also possible but are omitted here for simplicity.)

"free-energy levels" of the metastable "states" (self-consistent packet solutions) at different T. Figure 2 displays a hypothetical energy level diagram where the trajectories of all states are displayed. The salient points are: (1) there tend to be fewer states at higher T because the smoother effective potential surface supports fewer local minima, (2) states may merge or branch, and (3) states may become unstable and disappear both as T increases and decreases. In this example, a low-lying trajectory can be traced from the global minimum at T_{hi}, through two branch points, to the global minimum at $T_{\rm lo}$. Such "traceability" (which is not guaranteed) is required for success. It is not affected by mergers and branches (see Section 5) but may be disrupted by discontinuities associated with instabilities. States can become unstable with increasing T when a narrow local minimum in $\tilde{H}_{\Lambda_\alpha(T),T}(R)$ is averaged out as $\Lambda_\alpha(T)$ increases. Figure 3, provides an example of such an instability. These points of instability as T increases are the points where new states spontaneously appear as T decreases. The new states may not be detected, so success requires that a continuous path to $R^*(T_{\rm lo})$ can be found without them. This requirement is not unreasonable since the new states will have relatively high energies when they first appear (see Figure 3).

States can also become unstable as T [and $\Lambda_\alpha(T)$] decreases. This destabiliza-

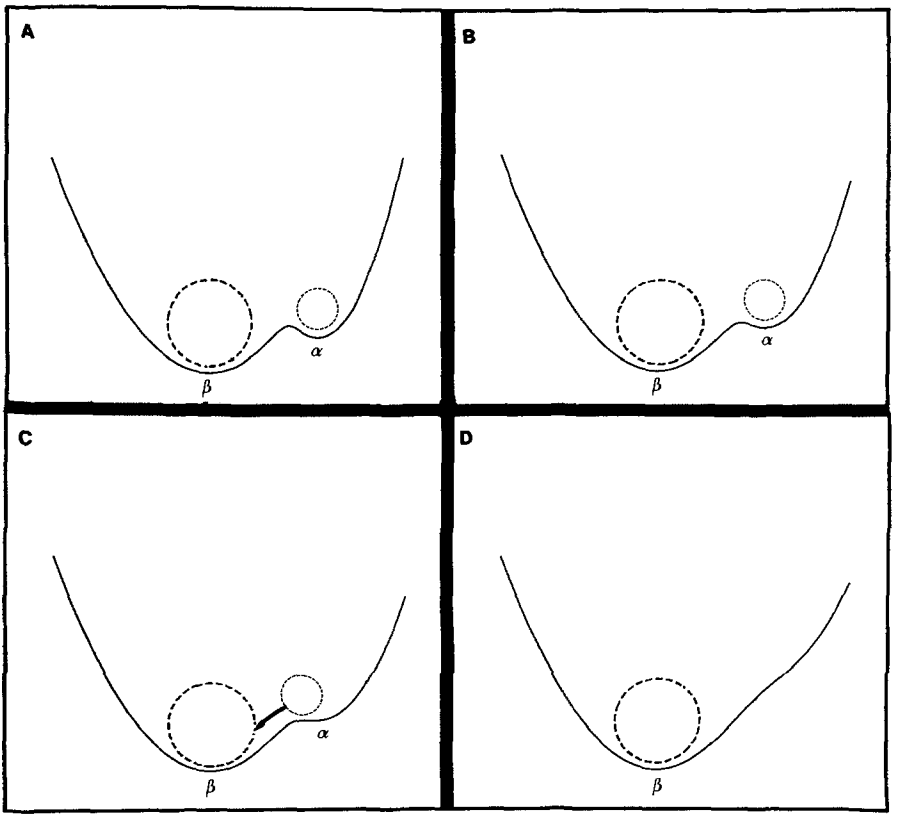

Fig. 3. **Instability in 1-Dimension with Increasing $\Lambda_\alpha(T)$.** Panels A–D show $\tilde{H}_{\Lambda,T}(R)$ as T and Λ increase. The sizes of the packets (finite-test objects) are indicated by dotted circles. The instability in state α at the temperature shown in (C) induces a discontinuity in the trajectory as packet α merges with packet β.

tion does not interfere with the trajectory tracing process since the trajectory can be (discontinuously) traced through the singularity (e.g., following the dotted lines in Figure 2).

4.2. ALGORITHM

An important computational characteristic of the method is that the scales of significant variation of both spatial and objective-function values are regulated at all steps by Λ_α and T. These parameters (multiplied by small factors ε_R, ε_Λ, ε_H) provide natural measures for the accuracies needed in (iterative) solutions of (2.11) and (2.12), so computational effort is controlled. The procedure terminates when a putative global minimum has been identified to accuracy $\leqslant \Lambda_{lo}$.

For convenience we redefine p_α, R^0_α and Λ_α as functions of a discrete iteration parameter $\tau(\tau = 1, 2, \ldots)$ rather than direct functions of T and use superscripts

for an independent iteration index which cycles at each τ. The central features of the algorithm are:

(*INITIALIZE*)
$T = T_{hi}$ (*T_{hi} is sufficiently high to "melt" all structure in $p_{T_{hi}}(R)$*)
${}^{(0)}R = 0$ (*an arbitrary point in the domain is selected*)
${}^{(0)}\Lambda = \Lambda_{hi}$ (*Λ_{hi} > the largest size scale of $p_{T_{hi}}(R)$; see Section 7*)

(*solve Eqns. (2.11) and (2.12) for the initial solution by iteration*)
$i = 0$
repeat

$$i = i + 1$$

$${}^{(i)}\Lambda^2 = \mathrm{MAX}\left\{\left[\frac{\partial^2 \tilde{H}_{({}^{(i)}\Lambda, T}(R)}{\partial R^2}\bigg|_{R={}^{(i-1)}R}\right]^{-1}\right\} \tag{4.1}$$

local minimization of ${}^{(i)}R$ starting from ${}^{(i-1)}R$ until

$$\frac{\partial \tilde{H}_{({}^{(i)}\Lambda, T}(R)}{\partial R} \leqslant \frac{\varepsilon_H T}{\Lambda^{(i)}} \tag{4.2}$$

until $|{}^{(i)}R - {}^{(i-1)}R| \leqslant \varepsilon_R {}^{(i)}\Lambda$ AND $|\log {}^{(i)}\Lambda / {}^{(i-1)}\Lambda| < \varepsilon_\Lambda {}^{(i)}\Lambda$
$R_1^0(0) = {}^{(i)}R$
$\Lambda_1(0) = {}^{(i)}\Lambda$

(* ITERATE: decrement T, adjust Λ, locally minimize R^0, branch/merge *)
repeat

$$\tau = \tau + 1$$

$$T(\tau) = [1 - \varepsilon_T(\tau)]T(\tau - 1)$$

for $\alpha \in \{\alpha\}_T$ *do*
begin
(* update $\Lambda_\alpha(\tau)$ using (2.12) *)

$$\Lambda_\alpha^2(\tau) = \mathrm{MAX}\left\{\left[\frac{\partial^2 \tilde{H}_{\Lambda_\alpha(\tau-1), T(\tau)}(R)}{\partial R^2}\bigg|_{R=R_\alpha^0(\tau-1)}\right]^{-1}\right\} \tag{4.3}$$

local minimization starting from $R_\alpha^0(\tau - 1)$ until

$$\frac{\partial \tilde{H}_{\Lambda_\alpha(\tau), T(\tau)}}{\partial R}\bigg|_{R=R_\alpha^0(\tau)} \leqslant \frac{\varepsilon_H T}{\Lambda(\tau)}\,. \tag{4.4}$$

Repeat iterations of (4.1) and (4.2) if necessary (see below)
Test and branch/merge packets if necessary (*see Section 5*)
(* update $\mathbf{p}_\alpha(\tau)$ using (2.10) *)

$$p_\alpha(\tau) = \frac{2^{-N/2}}{C[\Lambda_\alpha(\tau)]}\, e^{-\tilde{H}_{\Lambda_\alpha(\tau),T(\tau)[R^0_\alpha(\tau)]}}$$

end

(* discard low probability packets *)

$P_{\text{tot}} = \sum_{\{\alpha\}_T} p_\alpha(\tau)$ (*calculate probability normalization factor*)

for $\alpha \in \{\alpha\}_T$ *do*

if $p_\alpha(\tau) < p_{\min} p_{\text{tot}}$ *then* $\{\alpha\}_T = \{\alpha\}_T - \alpha$

(* test for termination *)

until $\{\alpha\}_T =$ unique AND $|R^0_\alpha(\tau) - R^0_\alpha(\tau - 1)| \leqslant \Lambda_{lo}$.

The maximum decrease in T at each step [determined by $\varepsilon_T(\tau)$] is limited by the requirement that $R^0_\alpha(\tau - 1)$ lies within the catchment region of $R^0_\alpha(\tau)$. This ensures that $R^0_\alpha(\tau)$ can be found by local minimization. The maximum number of packets in the expansion at any time will be $\leqslant 1/p_{\min}$.

Multiple iterations of (4.1) and (4.2) during initialization are required to get the initial one-packet approximation. In some cases, this packet will be unstable and will divide into subpackets when first tested for branching (see Section 5). During the main iterative loop, (4.3) and (4.4) update Λ_α and R^0_α towards solutions of (2.12) and (2.11). If the steps in T are sufficiently small, one update per cycle will be sufficiently accurate. This can be checked empirically and additional alternating iterations of (4.3) and (4.4) can be applied if necessary. Multiple iterations will be required when a state becomes unstable and the packet moves discontinuously. This can be detected by testing for $|R^0_\alpha(\tau) - R^0_\alpha(\tau - 1)| > \Lambda_\alpha(\tau - 1)$. In this case, the packet will move to a lower energy state where it may merge with a pre-existing packet. This process is naturally handled by the algorithm and poses no difficulty.

5. Branching and Merging of Packets

The possibility that packets α and β may merge can be tested by checking for

$$|R^0_\alpha(T) - R^0_\beta(T)| \sim O[\Lambda_\alpha(T) + \Lambda_\beta(T)]\,. \tag{5.1}$$

Packet merging is naturally accommodated by (2.11) and (2.12). When the packets come sufficiently close, their separate solutions $[R^0_\alpha(T), \Lambda_\alpha(T)]$ and $[R^0_\beta(T), \Lambda_\beta(T)]$ will become unstable; only a single solution centered between them [at $R^0_\gamma \approx (R^0_\alpha + R^0_\beta)/2$ with width $\Lambda_\gamma \approx (\Lambda_\alpha + \Lambda_\beta)$] will exist. Conversely, packet branching can be detected by testing for the appearance of new solutions of (2.11) and (2.12).

The merged (γ) packet remains a solution of (2.11) and (2.12) throughout branches and mergers. However, when separate (α and β) solutions exist, they

are included in expansion (1.3) and the parental γ solution is deleted. When this occurs, the size scale for approximation changes discontinuously. While (2.11) and (2.12) can continuously propagate solutions between branch/merger points, special procedures are needed to handle these discontinuities.

The packet merging procedure is straightforward: alternating iteration of (2.11) and (2.12) will propagate the separated solutions to the merged solution when the former become unstable. The branching procedure is more complex. In this case, we first test if a branched solution can exist at an appropriately reduced $\Lambda = \Lambda_b$. If so, iterative alternation of (2.11) and (2.12), using Λ_b and $R^0_\pm = R^0_\gamma \pm \Delta R^0_b$, is used to seek the branched packets. A prescription for specifying Λ_b and ΔR^0_b is needed.

These parameters can be explicitly calculated in the 1-dimensional case shown in Figure 4 where a probability distribution that is well-represented by a single packet at large T changes to a distribution that must be represented by two packets at lower T. For analytic simplicity, we model this function as

$$p_T(x) = [e^{-[(x-a)/w(T)]^2} + e^{-[(x+a)/w(T)]^2}]/[2\sqrt{\pi}w(T)] \qquad (5.2)$$

$$(\mathrm{d}w/\mathrm{d}T > 0)\,.$$

(For the potential shown in Figure 4, $w(T) \sim 2.25a\sqrt{T/T_0}$, where T_o is the temperature in panel A. The exact form of $w(T)$ is irrelevant.) The effective energy can be analytically calculated from (2.4) and (5.2):

$$e^{-\tilde{H}_{\Lambda,T}(x)} \equiv \tilde{Z}_{\Lambda,T}(x) = \frac{e^{-(x-a)^2/[\Lambda^2+w^2(T)]} + e^{-(x+a)^2/[\Lambda^2+w^2(T)]}}{2\sqrt{\pi}\sqrt{\Lambda^2 + w^2(T)}}\,. \qquad (5.3)$$

The derivatives of $\tilde{H}_{\Lambda,T}$ are

$$\frac{\partial \tilde{H}_{\Lambda,T}(x)}{\partial x} = \frac{2k_BT}{\Lambda'^2}\left[x - a\tanh\left(\frac{2ax}{\Lambda'^2}\right)\right] \qquad (5.4)$$

$$\frac{\partial^2 \tilde{H}_{\Lambda,T}(x)}{\partial x^2} = \frac{4k_BT}{\Lambda'^4}\left\{\frac{\Lambda'^2}{2} - x^2 - a^2 + 2ax\tanh\left(\frac{2ax}{\Lambda'^2}\right)\right\} \qquad (5.5)$$

where $\Lambda'^2 \equiv \Lambda^2 + w^2$.

Thus, replacing $R \rightarrow x$ for this one-dimensional case, (2.11) and (2.12) become

$$\frac{2k_BT}{\Lambda'^2}\left[x^0_\alpha - a\tanh\left(\frac{2ax^0_\alpha}{\Lambda'^2}\right)\right] = 0 \qquad (5.6)$$

$$\frac{4k_BT}{\Lambda'^4_\alpha}\left[\frac{\Lambda'^2_\alpha}{2} - x^{02}_\alpha - a^2 + 2x^0_\alpha\tanh\left(\frac{2ax^0_\alpha}{\Lambda'^2_\alpha}\right)\right] = \frac{k_BT}{\Lambda^2_\alpha} \qquad (5.7)$$

where $\Lambda'^2_\alpha \equiv \Lambda^2_\alpha + w^2$.

(5.6) and (5.7) imply

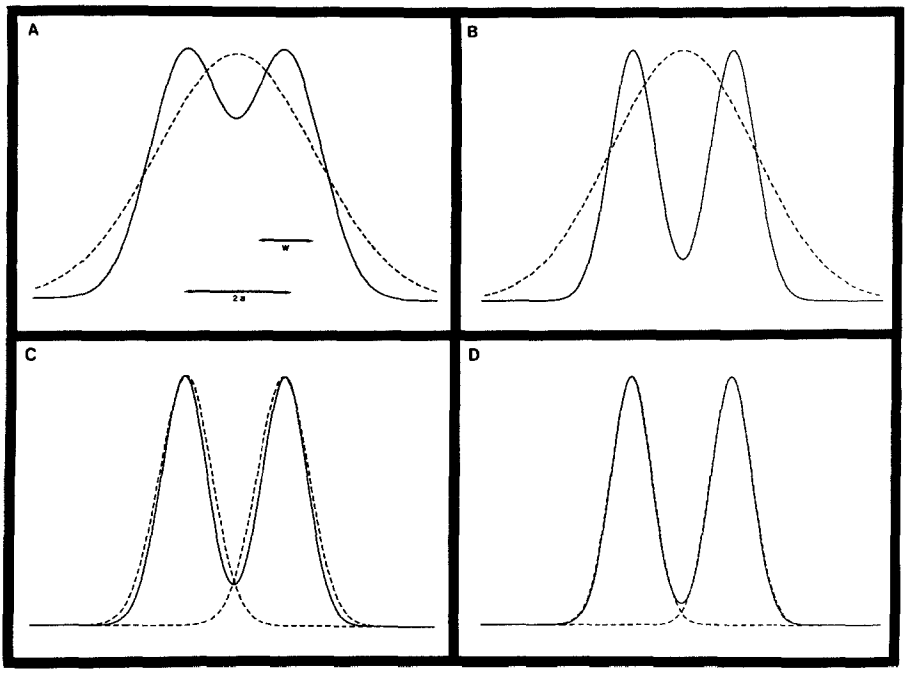

Fig. 4. **Packet Branching**. The bifurcated Gibbs distribution $p_T(x)$ of (5.2) (solid lines) and its packet approximations $\tilde{p}_T(x)$ determined by (5.8) and (5.9) (dashed lines) are displayed at different temperatures. For $w(T) = a$, the single packet ($x_0^0 = 0$, $\Lambda_0 = 2.06\,a$) is the only solution to (5.8) and (5.9) (panel A). When $w(T)$ decreases to $0.63\,a$, the single packet approximation ($x_0^0 = 0$, $\Lambda_0 = 2.01\,a$) (panel B) bifurcates into two subpackets ($x_b^0 = \pm 0.97\,a$, $\Lambda_b = 0.74$) (panel C). These packets closely approximate $p_T(x)$ as T is further reduced [e.g., as in panel D where $w(T) = 0.57\,a$].

$$x_\alpha^0 = \tanh\left(\frac{2ax_\alpha^0}{\Lambda'^2}\right) \tag{5.8}$$

$$\Lambda_\alpha^2 = 2a'^2 + \sqrt{4a'^4 + w^4}$$
$$a'^2 \equiv a^2 - (x_\alpha^0)^2\,. \tag{5.9}$$

For $2a^2 \leqslant w(T)^2$, $p_T(x)$ has a single maximum at $x = 0$ and (5.8) and (5.9) have the unique solution

$$x_0^0 = 0$$
$$\Lambda_0^2 = 2a^2 + \sqrt{4a^4 + w^4} \qquad (2a^2 \leqslant w^2)\,. \tag{5.10}$$

For $2a^2 > w(T)^2$, the maximum of $p_T(x)$ at 0 bifurcates. However, $\tilde{\tilde{p}}_T(x)$ will not bifurcate until $w(T)$ becomes small enough so that solutions of (5.8) and (5.9) with $x_\alpha^0 = \pm \Delta x_b \neq 0$; $\Lambda_\alpha = \Lambda_b$ exist. The condition specifying $w(T_b)$, where T_b is the highest temperature where bifurcated solutions can exist, is obtained by substituting (5.9) into (5.8) to yield

$$\xi = \tanh\{\xi/[v^2 + (1-\xi^2) + \sqrt{(1-\xi^2)^2 + v^4}]\}$$

$$\xi \equiv \frac{\Delta x_b^0}{a}\,, \tag{5.11}$$

$$v \equiv \frac{w^2}{2a^2}\,.$$

Numerical analysis shows that (5.11) has real non-zero solutions for ξ only when

$$v^2 \leqslant 0.20\,. \tag{5.12}$$

That is, as T and $w(T)$ decrease, bifurcation is first possible when

$$w(T_b) = 0.63a\,. \tag{5.13}$$

(5.11) implies that the bifurcated packets will be first located at

$$\pm\,\Delta x_b^0 = \pm\,0.96a \tag{5.14}$$

with [using (5.9)]

$$\Lambda_b = 0.77a\,. \tag{5.15}$$

Using (5.10), (5.13) and (5.15), we find that

$$\Lambda_0(T_b) = 2.0a \tag{5.16}$$

implying that

$$\Lambda_b = 0.38\Lambda_0(T_b) \equiv \gamma\Lambda_0(T_b) \tag{5.17}$$

$$\Delta x_b^0 = 0.48\Lambda_0(T_b) \equiv \chi\Lambda_0(T_B)\,. \tag{5.18}$$

This agrees with the intuitive idea that branching will occur when the width of each subpacket is about half the width of the parental packet. Using (5.3) and (5.16) we get

$$\left.\frac{\partial^2 \tilde{H}_{\gamma\Lambda_0,T_b}(x)}{\partial x^2}\right|_{x=x_0^0=0} = \frac{-8.3k_BT}{\Lambda_0^2(T_b)} \equiv \frac{-\rho k_BT}{\Lambda_0(T_b)}\,. \tag{5.19}$$

In practice, a and $w(T)$ will be unknown, but (5.16)–(5.18) depend only on the known values $\Lambda_0(T_b)$ and $x_0^0(T_b)$ and provide the basis for a test-and-branch in which we check to see if branched solutions to (2.11) and (2.12) can exist with $\Lambda = \gamma\Lambda_0$. In more general form (replacing $x \to R$ and $0 \to \alpha$) the algorithm is:

(* Given $R_\alpha^0(\tau)$ and $\Lambda_\alpha(\tau)$ *)

$$\textit{if MIN}\left\{\left.\frac{\partial^2 \tilde{H}_{\gamma\Lambda_\alpha(\tau),T(\tau)}(R)}{\partial R^2}\right|_{R=R_\alpha^0(\tau)}\right\} \leqslant \frac{-\rho k_BT}{\Lambda_\alpha^2(\tau)} \tag{5.20}$$

then begin

initialize $R^0 \approx R_\alpha^0(\tau) \pm \chi\Lambda_\alpha(\tau)$ (5.21)

$$\Lambda \approx \gamma \Lambda_\alpha(\tau) \tag{5.22}$$

iteratively find branched solutions of (2.10) and (2.11)
end

Although the real $p_T(R)$ will not have the simple form of (5.2), its packet branching will be governed by similar self-consistency considerations. In the multidimensional case, the positions of the branched solutions will be offset from $R^0_\alpha(\tau)$ by a vector of magnitude $\sim \chi \Lambda_\alpha(\tau_b)$ in the direction of the eigenvector of the Hessian of $\tilde{H}_{\gamma\Lambda_\alpha(\tau_b),T(\tau_b)}(R)$ which satisfies (5.20). Approximately the same values of χ, γ and ρ given by (5.17)–(5.19) can be used. High accuracy is not required since χ and γ only provide initial values for iterative solution. Similarly, the precise value of ρ is not critical: if it is too large, branching will not be detected at the earliest possible point, if it is too small, some unnecessary attempts will be made to find iterative solutions that do not exist.

More than one eigenvector of the Hessian will satisfy (5.19) if a packet branches into more than two subpackets. The branching procedure can be extended to accommodate this case. Other branch-testing algorithms can also be devised. For example, instead of examining $\partial^4 \tilde{H}_{\Lambda,T}/\partial R^4$ at $\Lambda = \gamma\Lambda_0$ by (5.20), we can examine $\partial^4 \tilde{H}_{\Lambda,T}/\partial R^4$ at $\Lambda = \Lambda_0$. The computational efficiency of the alternative methods must be compared.

The packet occupation probability for each branched solution, p_b, is smaller than the probability for the merged solution, p_0, since the total probability is split between the two branched solutions. Using (2.10), (5.3), and (5.14)–(5.16) we get

$$p_0 = 1.08 \tag{5.23}$$

$$p_b = 0.56\,. \tag{5.24}$$

However, $2p_b > p_0$, reflecting the fact that the branched solutions provide a better fit to $p_T(x)$ than the merged solution. (Formally, this is the reason that the merged solution is replaced.) These discontinuities in the p_α at branch points are schematically represented in Figure 2.

6. The Molecular Conformation Problem

The packet annealing method depends on the feasibility of evaluating the effective potential $\tilde{H}_{\Lambda,T}(R)$ from (2.4). For some objective functions, the integrations will require as much computation as a local grid-search and will be impractical. However, for some $H(R)$ of simple structure, (2.4) can be analytically approximated. Prediction of molecular conformation involves this type of $H(R)$.

The probability distribution for a system of n atoms in thermal equilibrium with a heat bath at temperature T is proportional to the Gibbs distribution, (1.1),

where R is the set of coordinates needed to specify the atom positions [$R = \{\bar{r}_i \in \mathbf{R}^3; i = 1, n\}$] and H is the energy function. For many problems, $H(R)$ has the form

$$H(R) = \sum_{i=1, j>i}^{n} h_{ij}(|\bar{r}_i - \bar{r}_j|) + \sum_{i=1, j>i, k>j}^{n} h_{ijk}(|\bar{r}_i - \bar{r}_j|, |\bar{r}_j - \bar{r}_k|, |\bar{r}_i - \bar{r}_k|) + \cdots \tag{6.1}$$

That is, H is the sum of interatomic 2-body, 3-body, . . . potentials and is partially separable. For many important problems, $n \sim O(10^2 - 10^4)$. The h_{ij}, h_{ijk}, . . . are relatively simple functions that represent the interactions between different types of atoms and have lower, but not upper bounds (e.g., $-h_{ij} \max < h_{ij} < \infty$); that is, there are limits on attractive but not on repulsive potentials. Since only a small number (<10) of different atom types are usually involved, there are only a relatively small number of independent h_{ij}, h_{ijk} . . . The complexity of the minimization problem arises from the summation and the fact that the $n(n-1)/2$ $|\bar{r}_i - \bar{r}_j|$ terms are not completely independent. $H(R)$ is invariant under rigid-body translations and rotations so it only depends on $3n(n-1)/2$-6 independent coordinates, and the global minimum (for 4 or more atoms) is a 6-dimensional hypersurface. This degeneracy can be removed; for pedagogical simplicity, we ignore it for this discussion.

To focus attention on the novel aspects of the packet annealing method, we consider the problem of finding the lowest energy state of a "microcluster" of identical atoms that interact by a two-body potential:

$$H(R) = \sum_{i=1, j>1}^{n} h(|\bar{r}_i - \bar{r}_j|) . \tag{6.2}$$

Some examples of low energy states are shown in Figure 5. This problem, itself

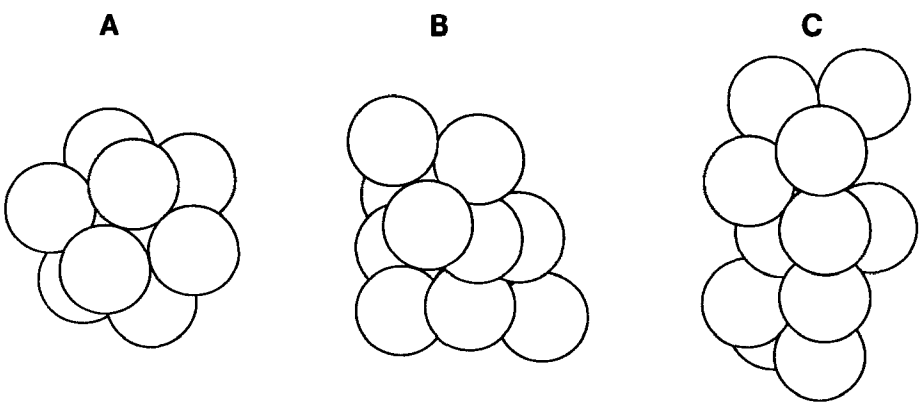

Fig. 5. **Global and Local Microcluster Energy Minima**. Three conformations of 13 Lennard–Jones atoms corresponding to minima of (6.2) and (6.3) are shown. (A) Global minimum with icosahedral symmetry; (B) and (C) higher energy metastable local minima.

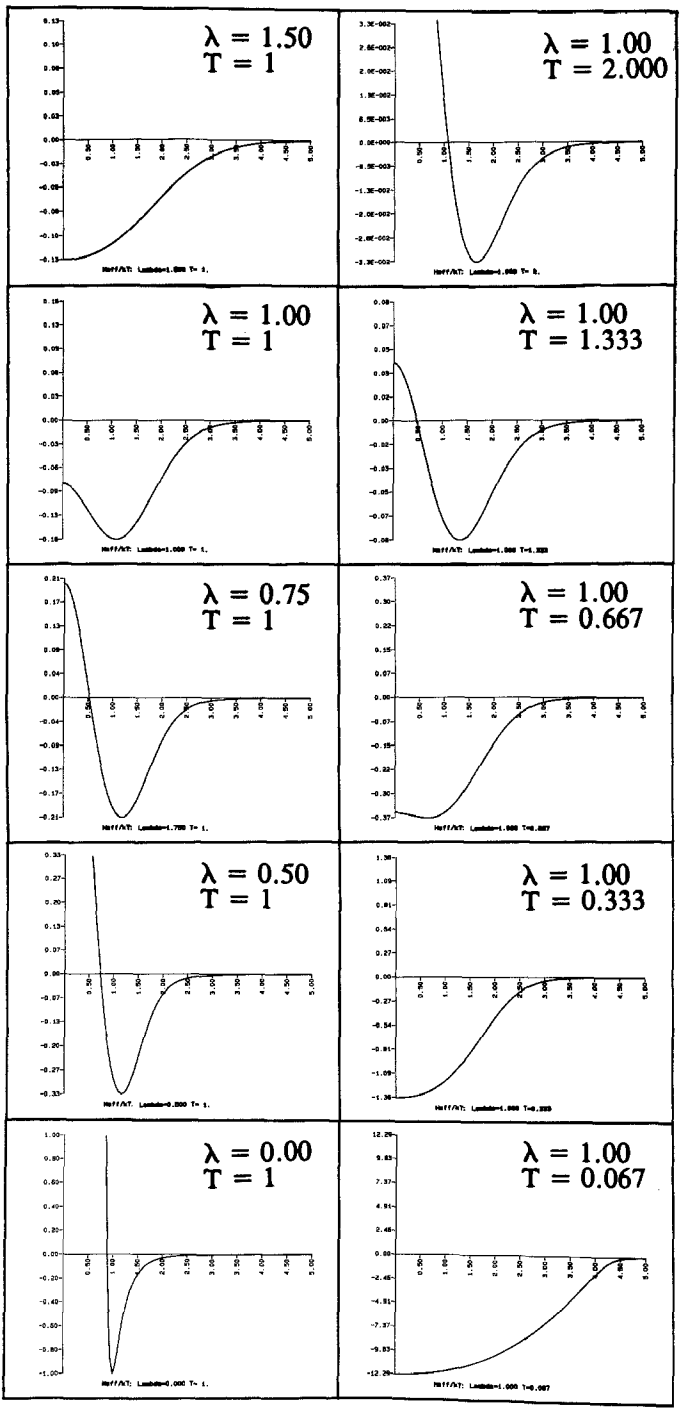

Fig. 6. $\tilde{h}_{\lambda,T}$. This is calculated from the 6–12 Lennard–Jones potential with $E_0 = 1$, $\sigma = 1$ (6.3) using (7.8). Effective potentials for varying λ, constant T (left panels) and varying T, constant λ (right panels are shown (ordinate = $\tilde{h}$, abscissa = r). The lower left panel ($\lambda = 0$, $\tau = 1$) is identical to $h(r)$.

important in chemistry [6, 24, 26], is a model for the more complex protein-folding problem. For inert atoms, h represents the Van der Waal's interaction and is often taken as a Lennard–Jones 6–12 potential:

$$h(r) = E_0\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right]. \tag{6.3}$$

E_0 and σ set the objective (energy) and spatial scales, respectively. This potential is positive and unbounded as $r \geqslant 0$, corresponding to the "hard core" repulsive interaction between adjacent atoms, and has an attractive well of depth E_0 with minimum at $r = \sigma$, the atom pair equilibrium distance (see Figure 6; $\lambda = 0$, $T = 1$).

Most if not all local minima of (6.2) and (6.3) for small microclusters ($n \leqslant 13$) have been identified by exhaustive search of the potential energy surface [13]. It has been suggested that the number of local minima grows roughly like $\exp(-2.52 + 0.36n + 0.029n^2)$ [11], indicating that identification of the global minimum for even moderately large n is a problem of significant complexity. Putative global minima in the range $n \leqslant 46$ have been identified by studying polyhedral growth sequences [7, 14], by Monte Carlo and simulated annealing methods [9, 32] and by molecular dynamics [15]. These results provide standards for evaluating the performance of new algorithms.

Hoare and McGinnes [12, 13] have compared the number of local minima for microclusters interacting with either a Lennard–Jones 6–12 potential or with a slightly smoother $\alpha = 3$ Morse potential $[h(r) = \{1 - \exp[-\alpha(1 - r)]\}^2 - 1]$. They find that even the minor relative smoothing of the Morse potential causes a large reduction in the numbers of local minima. For example, for $n = 13$, the Lennard–Jones potential supports 988 local minima while the Morse potential supports only 36 minima. This supports the intuitive idea that the complexity of the minimization problem will be reduced during the early stages of the annealing algorithm when highly averaged effective energy functions are being used.

7. Packet Annealing of Microcluster Conformation

7.1. APPROXIMATING THE EFFECTIVE POTENTIAL

(2.4) and (6.2) yield

$$\begin{aligned}\tilde{Z}_{\Lambda,T}(R) &\equiv e^{-\tilde{H}_{\Lambda,T}(R)} \\ &= C(\Lambda)\int \exp\Bigg\{-\sum_{i=1,j>i}^{n} h(|\bar{r}'_i - \bar{r}'_j|)/k_B T \\ &\quad - \frac{1}{\Lambda^2}\sum_{i=1}^{n}|\bar{r}'_i - \bar{r}_i|^2\Bigg\}\prod_{i=1}^{n} \mathrm{d}\bar{r}'_i\,.\end{aligned} \tag{7.1}$$

In general, this complex integral can not be exactly evaluated, but it can be evaluated in two special cases:

If h is quadratic,

$$h(r) = \frac{g}{2} r^2 \quad \text{[quadratic case]} \tag{7.2}$$

then (7.1) can be analytically evaluated (see Appendix A) to give

$$e^{-\tilde{H}_{\Lambda,T}(R)} \propto \prod_{i=1,j>1}^{n} e^{-\tilde{h}_{\sqrt{n}\Lambda,T}(|\bar{r}_i - \bar{r}_j|)/k_B T} \quad \text{[quadratic case]} \tag{7.3}$$

where

$$\tilde{z}_{\lambda,T}(|\bar{r}|) \equiv e^{-\tilde{h}_{\lambda,T}(|\bar{r}|)/k_B T} \equiv (\sqrt{\pi}\lambda)^{-3} \int e^{-h(|\bar{r}'|)/k_B T} e^{-|\bar{r}'-\bar{r}|^2/\lambda^2} \, d\bar{r}' \,. \tag{7.4}$$

That is

$$\tilde{H}_{\Lambda,T}(R) = \sum_{i=1,j>i}^{n} \tilde{h}_{\sqrt{n}\Lambda,T}(|\bar{r}_i - \bar{r}_j|) + c_q(n, \Lambda, T) \quad \text{[quadratic } h\text{]} \,. \tag{7.5}$$

Alternatively, if $h(r)$ is a bounded function, then at high temperature T, where $|h(r)|/k_B T \ll 1$, we can approximate $\exp[-h(|\bar{r}_i' - \bar{r}_j'|)/k_B T] \approx 1 - h(|\bar{r}_i' - \bar{r}_j'|)/k_B T$ and evaluate (7.1) to get

$$\tilde{H}_{\Lambda,T}(R) \approx \sum_{i=1,j>i}^{n} \tilde{h}_{\sqrt{2}\Lambda,T}(|\bar{r}_i - \bar{r}_j|) + c_{ht}(n) \quad [h(r)/k_B T \ll 1] \tag{7.6}$$

where we have again used definition (7.4).

Comparing (7.5) and (7.6) we see that in both cases

$$\tilde{H}_{\Lambda,T}(R) \approx \sum_{i=1,j>1}^{n} \tilde{h}_{\lambda,T}(|\bar{r}_i - \bar{r}_j|) + \text{constant} \tag{7.7}$$

where $\lambda = f(n, \Lambda, T)$; $\lim_{\Lambda \to 0} f = 0$.

Thus, it seems reasonable to use an approximation of this sort in the general case. The fact that $f(n, \Lambda, T)$ is unknown is unimportant since Λ will be determined self-consistently by homologs of (2.11) and (2.12). This factorization corresponds to averaging over atom positions in pairs and assuming that the effects of cross-correlations between pairs can be absorbed into Λ without changing the functional form of the effective interaction. It is analogous to using the two-body term in the Mayer cluster expansion of the partition function in statistical mechanics [8]. Even in cases where (7.5) is a poor approximation, it satisfies (2.6), the boundary condition as $\Lambda \to 0$, up to an unimportant constant and provides a natural method for spatial averaging of $H(R)$.

The factorization approximation to $\tilde{H}_{\Lambda,T}(R)$ is actually a special case of the more general "effective harmonic expansion" which is applicable when anisotropic packets are employed (i.e., when Λ is a symmetric matrix rather than a single number). This self-consistent method finds a quadratic approximation to $\tilde{H}$ that correctly reproduces the root-mean-square inter atom pair distance fluctuations about the stability point. It uses the partially-separable structure of $H(R)$ to reduce the multidimensional integrals in (7.1) to one-dimensional integrals which

can be evaluated by numerical integration. This reduction leads to a self-consistent set of equations that replace the packet annealing equations (2.10)–(2.12) and allows for distinct Λ_{ij}. A full exposition of the methods for generating the effective harmonic expansion will be presented later.

7.2. EVALUATION OF $\tilde{h}_{\lambda,T}(r)$

In spherical coordinates (7.4) becomes

$$\tilde{z}_{\lambda,T}(|\bar{r}|) = \frac{e^{-|\bar{r}|^2/\lambda^2}}{(\sqrt{\pi}\lambda)^3} \int_0^\infty e^{-h(r')/k_BT - r'^2/\lambda^2} r'^2 \int_{-1}^1 e^{2r'|r|\cos\theta/\lambda^2} \times \int_0^{2\pi} d\phi \, d\cos\theta \, dr' .$$

The ϕ and $\cos\theta$ integrals can be evaluated to yield:

$$\tilde{z}_{\lambda,T}(r) = \frac{2e^{-r^2/\lambda^2}}{r\lambda\sqrt{\pi}} \int_0^\infty e^{-[h(r')/k_BT + r'^2/\lambda^2]} \sinh(2rr'/\lambda^2) r' \, dr'$$

$$= \frac{\lambda e^{-r^2/\lambda^2}}{r\sqrt{\pi}} \int_0^1 e^{-h[\lambda\sqrt{\log(1/v)}]/k_BT} \sinh\left[\frac{2r\sqrt{\log(1/v)}}{\lambda}\right] dv , \qquad (7.8)$$

The behavior of $\tilde{h}_{\lambda,T}(r)$ for the Lennard–Jones potential (6.3), calculated by numerical integration of (7.8) for varying λ (left panels) and varying T (right panels) is shown in Figure 6. As expected, $\tilde{h}_{\lambda,T}(r)$ becomes progressively smoother as λ increases. The suppression of the repulsive singularity at $r = 0$ and the convexification of $\tilde{h}_{\lambda,T}(r)$ for large λ (e.g., see $\lambda = 1.5$, $T = 1$ panel) deserves special note. This phenomenon occurs because the spatial average in (7.4) is over the probability density, not over the singular objective function itself. The convexification at large λ reflects the fact that, to low spatial resolution, the most probable conformation of two atoms occurs when they are together. $\tilde{h}_{\lambda,T}(r)$ also broadens as T is lowered while keeping λ fixed. This is because lower energy atoms (with lower thermal velocities) do not penetrate and sample much of the repulsive core. Thus, their averaged interaction tends to be more attractive. Conversely, at higher temperatures, the bounded attractive part of the potential becomes less significant than the unbounded repulsive part so the minimum of $\tilde{h}_{\lambda,T}(r)$ is found at increasingly large distances. The panels in Figure 6 do *not* represent a typical packet annealing progression of $\tilde{h}_{\lambda(\tau),T(\tau)}(r)$ since both λ and T will change simultaneously during the process.

7.3. CRITICAL λ AND T FOR CONVEXIFICATION

Convexification of $\tilde{h}_{\lambda,T}(r)$ implies convexification of $\tilde{H}_{\Lambda,T}(R)$ and provides a natural condition for choosing λ_{hi} for the initialization step (see Section 4). We

select $\lambda_{hi} > \lambda_c(T_{hi})$, where $\lambda_c(T_{hi})$ is the value at which convexification first occurs. λ_c is determined implicitly by (see Appendix B)

$$\int_0^\infty [e^{-h(u\lambda_c)/kT} - 1] \frac{\mathrm{d}}{\mathrm{d}u} [u^3 e^{-u^2}] \, \mathrm{d}u = 0 . \tag{7.9}$$

λ_c decreases with decreasing T (Figure 6). Applying Laplace's approximation to (7.9) as $T \to 0$, we find that

$$\lim_{T \to 0} \lambda_c/\sigma = \sqrt{\frac{2}{3}} , \tag{7.10}$$

where σ is defined in (6.3). Convexification does not occur for $T > T_c$. T_c is determined (see Appendix B) by

$$\int_0^\infty [e^{-h(u)/kT_c} - 1] u^2 \, \mathrm{d}u = 0 . \tag{7.11}$$

Numerically solving (7.11) for the Lennard–Jones 6–12 potential, we get

$$T_c = \frac{3.27 E_0}{k_B} . \tag{7.12}$$

The physical interpretation of T_c is not understood. It may be related to the physical critical point temperature, but the analysis contains no homolog to the physical critical point pressure. The physical critical temperature for argon is 151°K [16]. Using the argon value $E_0 \sim 120°K \, k_B$ [1], $T_c \sim 390°$K.

8. Numerical Testing

We have not yet tested an implementation of the entire packet annealing algorithm. However, preliminary tests to demonstrate the feasibility of computing and minimizing $\tilde{H}_{\Lambda,T}(R)$ have been completed. Tables of $\tilde{h}_{\lambda,T}(r)$ for discrete values of λ, T and r were pre-calculated and values for arbitrary r were determined by cubic-spline interpolation. This allowed $\tilde{h}_{\lambda,T}(R)$ and its first and second derivatives to be rapidly evaluated. $\tilde{H}_{\Lambda,T}(R)$ was evaluated using the factorization approximation and local minimizations were performed using the conjugate gradient method.

A partial implementation was tested with $5 \leq n \leq 24$ by tracing a single trajectory as λ was decreased in small steps from $\lambda > \lambda_c(T)$ to $\lambda = 0$ at fixed $T < T_c$. Execution required relatively little computation and was carried out on a microcomputer. The overhead for computing the $\tilde{h}_{\lambda,T}$ tables was minimal; most time was used by the conjugate gradient minimizer. Computations took from a few minutes ($N = 5$) to about 20 hrs ($N = 24$) on an IBM PS/2 Model 80. This probably represents an overestimate of computational effort since λ was decremented in conservatively small steps (50 steps/run) and no effort was made to optimize the conjugate gradient minimizer.

The minima identified by this procedure are compared with the best minima

found by other means in Table I. The fact that even this crude method finds many of the putative global minima correctly is encouraging. One of the computed minima ($n = 23$) had lower energy than the lowest minimum identified by studying polyhedral growth sequences [14], although a lower value has recently been identified by both geometric [7] and simulated annealing [32] methods.

The 3-dimensional conformations were graphically examined as λ was reduced. As expected, for $\lambda > \lambda_c$, all atoms were condensed and located at the same point. As λ was decreased, the overlap between atoms decreased as they were "pushed" apart by the increasingly hard core of the potential (re Figure 6, left panels). Spatial symmetry breaking (conformations going from higher to lower symmetry)

Table I. Lowest minima found by partial implementation

Number of Atoms	Previous Minimum Energy	Partial Implementation Energy
5	9.104	=
6	12.303	=
7	16.505	=
8	19.822	19.766
9	24.113	=
10	28.420	=
11	32.765	=
12	37.967	=
13	44.327	=
14	47.845	=
15	52.322	=
16	56.815	55.345
17	61.318	61.095
18	66.531	66.285
19	72.659	=
20	77.177	=
21	81.685	=
22	86.148	=
23	92.844	91.348
24	97.349	93.654

The magnitudes of the energies for the lowest energy states identified for microclusters of 5 to 24 atoms interacting with the Lennard–Jones potential [(6.3) with $E_0 = 1$] are listed. Minimum energies were obtained from studies of polyhedral growth sequences [14] except for $N = 17$ [9], $N = 23$ [7] and $N = 24$ [32]. The partial implementation values were obtained by tracking a single packet as Λ was reduced [ignoring constraint (2.11)] at fixed $k_B T = 0.033\, E_0$. Runs were initiated with $\lambda(0) > \lambda_c$ [re (7.9)] so that $\tilde{h}_{\lambda(0),T}$ was convex. The energies obtained as $\lambda(\tau) \to 0$ are listed. The partial implementation value at $N = 23$ is superior to that obtained by the polyhedral build-up algorithm (90.647) indicating that nonpolyhedral configurations are explored as well. = denotes that the same energy was obtained.

was observed at some values of λ. In the complete algorithm, these events would be associated with trajectory branching. However, only a single, randomly selected, broken-symmetry trajectory was followed in this partial implementation. This may be the reason that the global minimum was not identified in some cases.

9. Discussion

The packet annealing method is a synthesis of both temperature-annealing and spatial-averaging methods which uses variable coarse-graining of both spatial and objective function values. Simulated annealing effectively provides coarse-graining in objective-function but not in spatial values. Conversely, methods that provide spatial coarse-graining alone have also been proposed [34]. Levitt [22] has shown that smoothing the repulsive hard-core part of the Lennard–Jones potential accelerates the rate at which the global minimum can be found in molecular dynamics simulations of protein condensation. This type of modification, which allows the atoms to pass through each other, emerges naturally in the packet annealing method (e.g., see Figure 6). Piela *et al.* [25] have recently proposed a differential "diffusion equation method" for linear smoothing of objective functions. Their method, if recast as an integral method, corresponds to averaging the objective function itself with a Gaussian test-function and is similar to the high-temperature approximation (7.6). Like the partial implementation described in Section 8, this method has had partial success when used for global minimization of microcluster conformations [20].

These approaches can be related to penalty function methods for constrained optimization [10] if we view the hard-core repulsive part of the potential as a constraint which keeps atoms from overlapping. Increasing the coefficient of the penalty function is analogous to decreasing λ in the packet annealing method. Sha [29, 30] has shown that penalty function methods can give good results in global optimization of 2-dimensional component layouts. Components have large overlaps at the start of the procedure and gradually separate as the penalty function coefficient increases. This is analogous to the progressive separation of atoms in the partial implementation.

Novel features of packet annealing are that spatial averages are performed over the Gibbs density rather than over the objective function and that T and Λ_α are coordinately varied. The use of the Gibbs density is heuristically motivated and provides a natural regularization of positive singularities that appear in many physical-model objective (energy) functions. It has the attractive feature that spatial averages are not unduly affected by large positive excursions of the objective function. As in statistical mechanics, there is little discrimination between regions where $H(R) \gg k_B T$; thus, computational effort is focused on discriminations in the important low energy regions. The heart of the approach is that it identifies and focuses computational effort on the spatial scales $\Lambda_\alpha(T)$ that

dominate behavior at each temperature T. This permits simplifications that would not otherwise be apparent.

The primary obstacle to implementation is the evaluation of the effective energy function (2.4). In one sense, (2.4) is simply a method for recasting the global minimization problem: if it could be evaluated exactly, the objective function could be convexified and global minimization would be trivial. (This can be recognized by applying Laplace's approximation to (2.4) at very low temperature. In this limit, $\tilde{Z}_{\Lambda,T}(R)$ is approximately Gaussian and $\tilde{H}_{\Lambda,T}(R)$ is convex.) The significance of (2.4) is that it (and the associated algorithms) provides a structure for novel approximation methods such as the effective harmonic expansion which take advantage of the partially-separable structure of molecular energy functions that can be decomposed as sums of 2- and 3-body interaction terms. This structure greatly restricts the shape of the objective function surface and, when properly utilized, can reduce greatly computational complexity.

As pointed out, the packet annealing method will not converge to the global minimum in all cases. In particular, global minima which occur in regions where the Gibbs distribution is, on average, small will not be covered by packets and will be overlooked (e.g. see Fig. 1F). However, it seems likely that the global minima for many physically-motivated problems will be located in regions where the Gibbs distributions are, on average, large. This is particularly plausible for objective functions that represent extensive physical properties (e.g., energy) that are sums of large numbers of small, partially-independent terms. In the spirit of the Central Limit Theorem, we expect that isolated anomalies like that shown in Fig. 1F will become increasingly rare as the number of terms increases. It will be interesting to determine if there are any classes of problems for which the packet annealing algorithm can be guaranteed to find the global minimum.

The solutions obtained by packet annealing may still be of physical interest even when they are not global minima. Like simulated annealing, packet annealing models the behavior of a cooled physical system as it "seeks" low-energy states. Isolated anomalous global minima which are likely to be missed by both forms of computational annealing, are also likely to be kinetically inaccessible to a physically cooled system if the system condenses in a time period that is short compared to the period required for stochastic sampling of the entire energy surface. It has been suggested that this may be the case for some proteins and macromolecules that condense down kinetically preferred pathways to metastable states (local minima) [21]. Since it is the physically-selected minima (whether global or not) that are generally of primary interest, the potential neglect of anomalous global minima may not be a significant disadvantage.

We have only begun the application of renormalization group ideas to the protein folding problem and much remains to be done. In particular, the convergence properties must be examined, the packet branching algorithm must be implemented and tested, and the behavior of more complex energy functions that accurately model protein interactions must be investigated.

Acknowledgments

This study was initiated during a Sabbatical with the group of Dr. P. Kollman (UCSF) during 1988–1989. I thank Dr. Kollman, his research group, Dr. I. Kuntz (UCSF), Dr. W. Murray (Stanford) and Dr. P. Pardalos (Penn State) for stimulating discussions and Drs. H. Scheraga and J. Kostrowicki (Cornell) for bringing their results to my attention and for sharing their manuscript prior to publication. In particular, I thank Lisa Miller for her excellent preparation of the manuscript. This research was supported in part by a Public Health Service Research Career Development Award (CA01139).

Appendix A: Effective Potential for Quadratic Energy Function

Using (7.2) and

$$\sum_{i=1,j>i}^{n} |\bar{r}'_i - \bar{r}'_j|^2 = n \sum_{i=1}^{n} |\bar{r}'_i - \bar{r}'_0|^2 \tag{A.1}$$

where

$$\bar{r}'_0 \equiv \frac{1}{n} \sum_{i=1}^{n} \bar{r}'_i \,. \tag{A.2}$$

(7.3) can be rewritten as

$$\tilde{Z}_{\Lambda,T}(R) = C(\Lambda) \int_{-\infty}^{\infty} e^{-\phi} \prod_{i=1}^{n} \mathrm{d}\bar{r}'_i \tag{A.3}$$

where

$$\phi \equiv \frac{gn}{2k_B T} \sum_{i=1}^{n} |\bar{r}'_i - \bar{r}'_o|^2 + \sum_{i=1}^{n} \frac{|\bar{r}'_i - \bar{r}_i|^2}{\Lambda^2}$$

$$C(\Lambda) \equiv (\sqrt{\pi}\Lambda)^{-3n} \,.$$

We release constraint (A.2) and treat $\bar{r}'_0$ as an independent variable by introducing a Dirac δ function and integrating over $\bar{r}'_0$:

$$\tilde{Z}_{\Lambda,T}(R) = C(\Lambda) \int_{-\infty}^{\infty} e^{-\phi}\, \delta\left(\bar{r}'_0 - \frac{1}{n} \sum_{i=1}^{n} \bar{r}'_i\right) \mathrm{d}\bar{r}'_0 \prod_{i=1}^{n} \mathrm{d}\bar{r}'_i \,. \tag{A.4}$$

Using

$$\delta(\bar{x}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} e^{i\bar{\alpha}\cdot\bar{x}}\, \mathrm{d}\bar{\alpha}$$

we have

$$\tilde{Z}_{\Lambda,T}(R)$$

$$= \frac{C(\Lambda)}{(2\pi)^3} \int_{-\infty}^{\infty} \exp\left\{-\phi + i\bar{\alpha}\cdot\left(\bar{r}'_0 - \frac{1}{n}\sum_{i=1}^{n} \bar{r}'_i\right)\right\} \mathrm{d}\bar{\alpha}\, \mathrm{d}\bar{r}'_0 \prod_{i=1}^{n} \mathrm{d}\bar{r}'_i \tag{A.5}$$

The integrand in (A.5) can now be factored into a product of terms of the form $\exp[-a(\bar{r}_i' - b)^2]$ and the $\bar{r}_i'$ integrals can be analytically evaluated. The α and $\bar{r}_0'$ integrals then take a similar form and can be evaluated to yield

$$\tilde{Z}_{\Lambda,T}(R) = \left(\frac{gn\Lambda^2}{2k_BT} + 1\right)^{-\frac{3(n-1)}{2}} \times \exp\left\{-\left[\frac{g}{(gn\Lambda^2 + 2k_BT)} \sum_{i=1,j>i}^{n} |\bar{r}_i - \bar{r}_j|^2\right]\right\}. \tag{A.6}$$

By similar means, we calculate $\tilde{z}_{\lambda,T}(r)$ from (7.2) and (7.4):

$$\tilde{z}_{\lambda,T}(|\bar{r}|) = \left(\frac{g\lambda^2}{2k_BT} + 1\right)^{-3/2} \exp\left\{-\left[\frac{g|\bar{r}|^2}{(g\lambda^2 + 2k_BT)}\right]\right\}. \tag{A.7}$$

Comparing (A.6) and (A.7) gives

$$\tilde{Z}_{\Lambda,T}(R) = \left(\frac{gn\Lambda^2}{2k_BT} + 1\right)^{3(n-1)(n-2)/4} \sum_{i=1,j>1}^{n} \tilde{z}_{\sqrt{n}\Lambda,T}(|\bar{r}_i - \bar{r}_j|). \tag{A.8}$$

Taking the logarithm of both sides yields (7.5).

Appendix B: Convexification of $\tilde{h}_{\lambda,T}(r)$

Examination of Figure 6 shows that $\tilde{h}_{\lambda,T}(r)$ will be convex when

$$\left.\frac{\partial^2 \tilde{h}_{\lambda,T}(r)}{\partial r^2}\right|_{r=0} \geqslant 0 \quad \text{[convexity condition]}. \tag{B.1}$$

Since

$$\frac{\partial^2 \tilde{h}_{\lambda,T}(r)}{\partial r^2} = -k_BT\left\{\frac{\partial^2 \tilde{z}_{\lambda,T}(r)}{\partial r^2} \Big/ \tilde{z}_{\lambda,T}(r) - \left[\frac{\partial \tilde{z}_{\lambda,T}(r)}{\partial r} \Big/ \tilde{z}_{\lambda,T}(r)\right]^2\right\} \tag{B.2}$$

and [from (7.4)]

$$\left.\frac{\partial \tilde{z}_{\lambda,T}(r)}{\partial r}\right|_{r=0} = 0. \tag{B.3}$$

(B.1) is equivalent to

$$\left.\frac{\partial^2 \tilde{z}_{\lambda,T}(r)}{\partial r^2}\right|_{r=0} \leqslant 0 \quad \text{[convexity condition]}. \tag{B.4}$$

Expanding (7.8) about $r = 0$ yields

$$\tilde{z}_{\lambda,T}(r) = \frac{4}{\sqrt{\pi}} \int_0^\infty e^{-h(u\lambda)/k_BT} e^{-u^2} \times \left[1 + \frac{r^2}{\lambda^2}\left(\frac{2}{3}u^2 - 1\right) + 0(r^4)\right] u^2 \, du. \tag{B.5}$$

Substitution of (B.5) into (B.4) yields

$$\int_0^\infty e^{-h(u\lambda)/k_BT}e^{-u^2}u^2\left(\frac{2}{3}u^2-1\right)\mathrm{d}u \leqslant 0 \quad \text{[convexity condition]} \tag{B.6}$$

The u terms in the integrand of (B.6) can be rewritten as

$$e^{-u^2}u^2\left(\frac{2}{3}u^2-1\right) \propto -\frac{\mathrm{d}}{\mathrm{d}u}(u^3e^{-u^2}) \tag{B.7}$$

and a constant can be subtracted from the $\exp[-h]$ term [so that it approaches 0 as $u\lambda/\sigma \to \infty$; σ defined in (6.3)] without affecting the value of the integral. This yields (7.9).

From Figure 6 we see that $\tilde{h}_{\lambda,T}(r)$ tends towards convexity as λ increases. The highest temperature, T_c, for convexification is determined by evaluating (7.9) for $\lambda/\sigma \to \infty$. In this limit, the only contribution to the integral occurs for $u \ll 1$ and the $\exp[-u^2]$ term can be ignored. After rescaling $u\lambda \to u$, (7.9) reduces to (7.10).

Notes

[1] Matrix notation is implicitly assumed throughout. That is

$$|R/\Lambda|^2 \equiv \sum_{i,j,k=1}^{N} r_i\Lambda_{ik}^{-1}r_j\Lambda_{jk}^{-1}$$

where N is the dimensionality of the space. In subsequent equations, R^2 denotes the outer product r_ir_j. The squared vector magnitude is denoted $|R^2|$; $|R^2| = \mathrm{Tr}(R^2)$.

[2] In addition, using (3.2) we can show that

$$\Lambda_\alpha^2/2k_BT \leqslant \text{MINIMUM } \{[\partial^2\tilde{H}_{\Lambda_\alpha,T}(R)/\partial R^2|_{R=R_\alpha^0}]^{-1}\}$$

is always satisfied.

References

1. Allen, M. P. and D. J. Tildesley (1987), *Computer Simulation of Liquids*, Clarendon Press: Oxford, p. 9.
2. Blundell, T. L., B. L. Sibanda, M. J. E. Sternberg, and J. M. Thornton (1987), Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules, *Nature* **326**, 347–352.
3. Bohachevsky, I. O., M. E. Johnson, and M. L. Stein (1986), Generalized Simulated Annealing for Function Optimization, *Technometrics* **28**, 209–217.
4. Brooks, C. L., III, M. Karplus, and B. M. Pettitt (1988), *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, J. Wiley and Sons: New York.
5. Corana, A., M. Marchesi, C. Martini, and S. Ridella (1987), Minimizing Multimodal Functions of Continuous Variables with the "Simulated Annealing" Algorithm, *ACM Trans. on Math. Software* **13**, 262–280.
6. Duncan, M. A. and D. H. Rouvray (1989), Microclusters, *Sci. Am.* **261**, 110–115.
7. Farges, J., M. F. DeFeraudy, B. Raoult, and G. Torchet (1985), Cluster Models Made of Double Icosahedron Units, *Surface Sci.* **156**, 370–378.
8. Feynman, R. P. (1972), *Statistical Mechanics*, W. A. Benjamin: Reading, Mass., pp. 105–110.
9. Freeman, D. L. and J. D. Doll (1985), Quantum Monte Carlo Study of the Thermodynamic

Properties of Argon Clusters: The Homogeneous Nucleation of Argon in Argon Vapor and "Magic Number" Distributions in Argon Vapor, *J. Chem. Phys.* **82**, 462–471.
10. Gill, P. E., W. Murray, and M. H. Wright (1981), *Practical Optimization*, Academic Press: London, pp. 207–219.
11. Hoare, M. R. (1979), Structure and Dynamics of Simple Microclusters, *Adv. Chem. Phys.* **40**, 49–135.
12. Hoare, M. R. and J. McInnes (1976), Statistical Mechanics and Morphology of Very Small Atomic Clusters, *Faraday Discussions Chem. Soc.* **61**, 12–24.
13. Hoare, M. R. and J. McInnes (1983), Morphology and Statistical Statics of Simple Microclusters, *Advan. Phys.* **32**, 791–821.
14. Hoare, M. R. and P. Pal (1971), Physical Cluster Mechanics: Statics and Energy Surfaces for Monatomic Systems, *Adv. Phys.* **20**, 161–196.
15. Honeycutt, J. D. and H. C. Andersen (1987), Molecular Dynamics Study of Melting and Freezing of Small Lennard–Jones Clusters, *J. Phys. Chem.* **91**, 4950–4963.
16. Kauzmann, W. (1966), *Kinetic Theory of Gases*, W. A. Benjamin: New York, p. 31.
17. Kawai, H., T. Kikuchi, and Y. Okamoto (1989), A Prediction of Tertiary Structures of Peptide by the Monte Carlo Simulated Annealing Method, *Protein Eng.* **3**, 85–94.
18. King, J. (1989), Deciphering the Rules of Protein Folding, *Chem. Eng. News* **67**, 32–54.
19. Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983), Optimization by Simulated Annealing, *Science* **220**, 671–680.
20. Kostrowicki, J., L. Piela, B. J. Cherayil and H. A. Scheraga (1991), Performance of the Diffusion Equation Method in Searches for Optimum Structures of Clusters of Lennard–Jones Atoms, *J. Phys. Chem.* **95**, 4113–4119.
21. Levinthal, C. (1968), Are There Pathways for Protein Folding? *J. Chim Phys.* **65**, 44–45.
22. Levitt, M. (1983), Protein Folding by Restrained Energy Minimization and Molecular Dynamics, *J. Mol. Biol.* **170**, 723–764.
23. Messiah, A. (1961), *Quantum Mechanics*, Vol. 2, John Wiley: New York, pp. 781–793.
24. Phillips, J. C. (1986), Chemical Bonding, Kinetics, and the Approach to Equilibrium Structures of Simple Metallic, Molecular, and Network Microclusters, *Chem. Rev.* **86**, 619–634.
25. Piela, L., J. Kostrowicki, and H. A. Scheraga (1989), The Multiple-Minima Problem in the Conformational Analysis of Molecules. Deformation of the Potential Energy Hypersurface by the Diffusion Equation Method, *J. Phys. Chem.* **93**, 3339–3346.
26. Pool, R. (1990), Clusters: Strange Morsels of Matter, *Science* **248**, 1186–1188.
27. Richards, F. M., (1991), The Protein Folding Problem, *Sci. Am.* **264**(1), 54–63.
28. Scheraga, H. A. (1987), Conformational Analysis of Polypeptides and Proteins for the Study of Protein Folding, Molecular Recognition, and Molecular Design, *J. Prot. Chem.* **6**, 61–80.
29. Sha, L. (1989), A Macrocell Placement Algorithm Using Mathematical Programming Techniques. Ph.D. Thesis, Department of Electrical Engineering, Stanford University.
30. Sha, L. and T. Blank (1987), ATLAS – A Technique for Layout Using Analytic Shapes. Digest of Technical Papers. IEEE Conf. on Computer-Aided Design, pp. 84–86.
31. Wille, L. T. (1986), Searching Potential Energy Surfaces by Simulated Annealing, *Nature* **324**, 46–48.
32. Wille, L. T. (1987), Minimum-Energy Configurations of Atomic Clusters: New Results Obtained by Simulated Annealing, *Chem. Phys. Lett.* **133**, 405–410.
33. Wilson, K. G. (1975), The Renormalization Group: Critical Phenomena and the Kondo Problem, *Rev. Mod. Phys.* **47**, 773–840.
34. Zakharov, V. V. (1970), The Method of Integral Smoothing in Many Extremal and Stochastic Problems, *Eng. Cybernetics* **4**, 637–642.