PHYSICAL REVIEW A VOLUME 37, NUMBER 4 FEBRUARY 15, 1988

Statistical mechanics of combinatorial optimization

James D. Nulton
Department of Mathematics, San Diego City College, San Diego, California 92101

Peter Salamon
Department of Mathematical Sciences, San Diego State University, San Diego, California 92182

(Received 20 November 1986)

A theoretical criterion is offered for the design of a temperature schedule for simulated annealing. It is based on a measure of distance in probability space. Implementation requires a knowledge of the heat capacity and the relaxation time. A method of calculating these quantities for a combinatorial problem is outlined. The theoretical structure involved seems to point to an information-theoretic measure of computational effort in certain probabilistic algorithms.

I. INTRODUCTION

Simulated annealing as a method of obtaining near-optimal solutions of complex combinatorial optimization problems was first brought to the attention of the general scientific community by Kirkpatrick, Gelatt, and Vecchi.[1] That article discussed the analogy between equilibrium statistical mechanics and the Metropolis[2] algorithm as a probabilistic-search method. In this paper we extend that analogy by presenting a theoretical idea for the design of the annealing schedule, a notion of central importance to the efficient application of simulated annealing.

In a typical combinatorial optimization problem, one is confronted with a very large space of configurations, each with an associated cost. The task is to find a configuration with the lowest possible cost using a certain number of function evaluations. A classical example[3] is the traveling salesman problem: Given a list of N cities and a cost of traveling between any two cities, one must find a route for the salesman which will pass through each city once and return to the starting point while minimizing the total cost. Another classical example, treated further below, is a form of the graph partitioning problem: One partitions the vertices of a given graph into two equal-sized subsets in such a way as to minimize the number of edges which run between vertices in opposite subsets. A concrete instance of this with considerable industrial importance concerns the problem of placing 6000 circuit elements (vertices) onto two chips, each holding 3000, so as to minimize the number of wires (edges) running between the chips.

The central idea in the paper of Kirkpatrick *et al.* was to identify the cost in a combinatorial problem with the energy of an associated physical system. At a low temperature, a physical system in equilibrium samples predominantly its low-energy states. This is the aspect of the analogy which one would like to exploit to sample the low-cost configurations of a combinatorial problem.

In the case of a complex physical system whose low-energy states are hard to access, true equilibrium is often hard to reach. Instead, one sees "frozen-in imperfections" which correspond to the system getting trapped in local minima of the energy. A carefully controlled lowering of the temperature known as an annealing schedule can significantly increase the probability that the system will in fact reach equilibrium. In Sec. II we will show how low-cost states of a combinatorial problem can be sampled by simulating a corresponding physical process. In fact, it will be seen that a phenomenon analogous to freezing in imperfections demands a careful scheduling of temperatures.

In Sec. III we offer a theoretical criterion for the design of the schedule. Section IV addresses the technical problem of testing the criterion. Section V gives two more characterizations of the criterion and connects it with other theory.

II. STOCHASTIC DESCRIPTION OF A COMBINATORIAL PROBLEM

We begin by laying out the mathematical apparatus needed to apply simulated annealing to a combinatorial problem. We then offer the graph-partitioning problem as an illustration.

Let Ω be the space of configurations of a combinatorial problem and let $E(\omega)$ be the cost function on Ω. Following Kirkpatrick *et al.*, we will refer to this function as the energy. Let $\{X_K\}$ be a sequence of Ω-valued random variables, forming a regular Markov chain M with stationary transition probabilities

$$P(\omega,\omega')=\mathrm{Prob}\{X_{K+1}=\omega' \mid X_K=\omega\},\quad \omega,\omega'\in\Omega\ . \qquad (2.1)$$

The requirement of regularity guarantees that each configuration can be reached dynamically from any other, and that there is a unique invariant distribution (equilibrium).

The X_K can be thought of as a scheme for exploring the space Ω by a random walk. Typically X_K and the associated transition probabilities P arise in the following way. For a given combinatorial problem, there is often a convenient way to specify when one configuration is a neighbor of another.[4] This confers to the space Ω the structure of a connected, undirected graph in which configurations are vertices and neighboring pairs are edges. Possible one-step transitions in M are then required to be transitions between neighboring configurations, i.e., $P(\omega,\omega')=0$, unless ω and ω' are neighbors. A typical approach is to further require all neighboring transitions be equally likely, i.e.,

$$P(\omega,\omega')=\begin{cases} |N(\omega)|^{-1}, & \omega'\in N(\omega) \\ 0, & \omega'\notin N(\omega) \end{cases} \tag{2.2}$$

where $N(\omega)$ is the set of neighbors of ω, and

$$P(\omega,\omega')=P(\omega',\omega) . \tag{2.3}$$

In the physical language, (2.3) is known as microscopic reversibility. If Eq. (2.2) is adopted, then this follows if the graph with vertices Ω is regular, i.e., $N(\omega)$ is the same for all $\omega\in\Omega$. Equation (2.3) is sufficient to guarantee that the Markov chain M is doubly stochastic and therefore has the uniform distribution on Ω as its equilibrium distribution. In this sense, it is analogous to a master equation model of the approach of a physical system to equilibrium at infinite temperature. Next, we combine this high-temperature chain M with temperature-dependent Metropolis factors to obtain a one-parameter family of Markov chains $M(T)$ with stationary transitions, each of which corresponds to a stochastic model of approach to equilibrium at some temperature, T. The transition probabilities for $M(T)$ are defined by

$$P(T;\omega,\omega')=\begin{cases} P(\omega,\omega')\exp(-\Delta E/T) & \text{when } \Delta E>0 , \\ P(\omega,\omega') & \text{when } \omega\neq\omega' \text{ and } \Delta E\leq 0 , \end{cases} \tag{2.4a}$$

and

$$P(T;\omega,\omega)=1-\sum_{\omega'\,(\neq\omega)} P(T;\omega,\omega') , \tag{2.4b}$$

where

$$\Delta E=E(\omega')-E(\omega) .$$

It can be seen from (2.4) that the Metropolis factor, $\exp(-\Delta E/T)$, severely attenuates the probability of uphill transitions at low temperatures. Since the sets $N(\omega)$ are usually not large, it may happen that all possible transitions from ω are uphill. In this case ω is called a local minimum. Such local minima threaten to trap the process $M(T)$ for small T; whence, the idea of gradually lowering the temperature, or annealing.

Mathematically the annealing process is described by a Markov chain $M\{T_K\}$ with *non*stationary transitions

$$\text{Prob}\{X_{K+1}=\omega' \mid X_K=\omega\}=P(T_K;\omega,\omega') , \tag{2.5}$$

where P is as in (2.4) and $\{T_K\}$ is a decreasing sequence of temperatures, called the annealing schedule.

The distribution of X_K we denote by

$$p_K(\omega)=\text{Prob}\{X_K=\omega\} . \tag{2.6}$$

In practice, the ensemble is usually "prepared" at high temperature, so, to be definite, we may suppose the initial distribution $p_0(\omega)$ to be uniform , i.e., equilibrium at infinite temperature.[5]

Finally, regarding $M(T)$, it can be shown[2] that if (2.3) holds then the equilibrium distribution of (2.4) is

$$p^0(T;\omega)=\exp[-E(\omega)/T]/Z(T) , \tag{2.7}$$

where

$$Z(T)=\sum_{\omega}\exp[-E(\omega)/T] .$$

This fact makes available the formalism of equilibrium statistical mechanics and, as we will see, that formalism is far from vacuous in this context.

To illustrate the machinery of this section, we use the example of the graph partitioning problem mentioned in Sec. I. For a given undirected graph G, let $V(G)$ denote the set of vertices and $X(G)\subset V(G)\times V(G)$, the set of edges.

A configuration is constituted by assigning each vertex to either set A or set B so that both sets receive an equal number of vertices. In the language of graph theory a configuration ω is a disjoint partition, $V(G)=A\cup B$ with $|A|=|B|$, where $|\ |$ denotes cardinality of the set. The energy $E(\omega)$ is the number of edges crossing the partition, i.e., the number of pairs $(v_1,v_2)\in X(G)$ such that $v_1\in A$ and $v_2\in B$. For $|V(G)|=100$, $|\Omega|=10^{29}$, so that Ω is typically an enormous set.

ω' is a neighboring configuration of ω if it is obtained from ω by exchanging two vertices across the partition, i.e., $A'=A\cup\{v_2\}\setminus\{v_1\}$ and $B'=B\cup\{v_1\}\setminus\{v_2\}$ for some $v_1\in A$ and $v_2\in B$. P is now defined by (2.2) and clearly satisfies (2.3). For $|V(G)|=100$, $|N(\omega)|=2500$ for every $\omega\in\Omega$.

III. THEORETICAL CRITERION FOR CHOOSING ANNEALING SCHEDULE

The fact that annealing schedules will vary in the efficiency with which they bring the process to low energy follows heuristically from the possibility that local minima may become traps or frozen states. However, we will examine a general point of view which leads to a natural criterion for choosing an annealing schedule.

We will call the time-varying distribution p in (2.6) the state of the *system*. It can be thought of as a nonequilibrium statistical-mechanical ensemble, and represents the actual time evolution of the process $M\{T_K\}$. There is another time-varying distribution, which, in a sense, represents the ideal evolution of the process: p^0 in (2.7), where the time dependence enters only through the annealing schedule $\{T_K\}$. This p^0 will be called the state of the *target*, as it is the equilibrium distribution toward which the system would relax, if the value of T were held fixed at the current value T_K.

For simplicity in the following discussion we will denote the time by t, a continuous variable, i.e., $X(t)$, $T(t)$, $p(\omega,t)$, and $p^0(\omega,t)$, instead of X_K, T_K, $p_K(\omega)$, and $p^0(T_K;\omega)$. Also, for distributions, one or both of the variables ω and t may be suppressed for emphasis.

A useful view of the annealing process is to focus on the time-varying separation between these two probability distributions: the system at p following the target at p^0.

It is p^0, as an equilibrium distribution, that preserves the functional link between mean energy and temperature on which the principle of annealing is based. In the limit of an infinitely slow process, p and p^0 are equal, so that link is never broken. When annealing becomes limited in real time, a delicate compromise must be effected—moving the target as rapidly as possible while maintaining the link to within some tolerance. That is, the effectiveness of annealing implicitly presupposes an approximate identification of the distributions p and p^0; we propose below to quantify that approximation by means of a measure of statistical uncertainty along a line of argument inspired by Wootters.[6]

For the moment let π be any distribution on Ω. Imagine a multinomial experiment performed by extracting a sample of size N from the ensemble in the state π. Let $\gamma(\omega)$ be the random variable representing the fraction of occurrences of the state ω in the sample. If we suppose that N is large enough for the Gaussian approximation to the multinomial distribution to be valid, then it can be shown[7] that for any $a>0$, a number $r>0$ can be found, which depends on a and $|\Omega|$ but which is independent of the distribution π and the sample size N and which has the following property: For any sampling experiment with the ensemble, the probability that the $\gamma(\omega)$ satisfies the following inequality is $1-a$, i.e., virtual certainty when a is small:

$$\sum_{\omega}[\gamma(\omega)-\pi(\omega)]^2/\pi(\omega)\le r/N \ . \tag{3.1}$$

This provides a setting in which to discuss quantitatively the approximate identity of the system and the target. In the following we assume a value has been specified for a and fixed once and for all.

Define a quantity $\Xi(p^0,p)$, called the p^0 discrepancy of p, by

$$\Xi(p^0,p)\equiv\sum_{\omega}[p(\omega)-p^0(\omega)]^2/p^0(\omega) \ . \tag{3.2}$$

This quantity and the associated sample size

$$N(p^0,p)\equiv r/\Xi(p^0,p) \tag{3.3}$$

can be given an operational meaning as follows. Suppose the exact probabilities $p(\omega)$, associated with the system, were instead reinterpreted as relative frequency data from a sampling of p^0. Would they be acceptable? Evidently, if the sample were of size $N\le N(p^0,p)$, they would be; otherwise, they would not.

Thus the annealing process may be viewed as follows. At each instant of time the system p may be identified with the target p^0, so long as that identification is subject to the statistical uncertainty associated with samples of size $N(p^0,p)$.

When viewed in this way, a natural condition to impose on the annealing process is that the level of this statistical uncertainty remain bounded by some constant throughout the process.[8] But, for any schedule which seeks to lower the temperature as rapidly as possible subject to this constraint, it is clear that the condition "bounded by some constant" can be replaced by the apparently stronger condition "equal to some constant." The supposition here is that any time gained by pushing the target so far ahead of the system as to dissolve the identification would be more than offset by the time-consuming relaxation required to reestablish it.

We do not propose to prove the optimality of this constancy criterion here; it is offered as a promising theoretical conjecture, which, as we show below, is testable when translated into thermodynamic terms describing the relaxation dynamics.

For purposes of deriving the consequences of the constancy criterion, the equivalent condition of constant discrepancy $\Xi(p^0,p)$ is the most convenient mathematical starting point. Indeed, as we show in Sec. V, Ξ is related to a number of interesting theoretical structures.

With the view of putting this criterion in a testable form, let us make the approximation that the system is itself a canonical distribution and a small perturbation of the target, i.e., $p=p^0+\delta p$. It is a direct calculation from (2.7) to show that

$$\Xi(p^0,p)=\sum_{\omega}[\delta p(\omega)]^2/p^0(\omega)=(\delta E)^2/\sigma(E^0)^2 \ , \tag{3.4}$$

where

$$E^0=\sum_{\omega}E(\omega)p^0(\omega),\quad E=\sum_{\omega}E(\omega)p(\omega)$$

and

$$[\sigma(E^0)]^2=\sum_{\omega}[E(\omega)-E^0]^2p^0(\omega) \ .$$

In view of (3.4) the criterion of constant discrepancy can be written

$$\delta E=E-E^0=v\sigma(E^0) \ , \tag{3.5}$$

where v is a constant equal to the thermodynamic distance, $[\Xi(p^0,p)]^{1/2}$, between the system and the target (see Sec. V). We note here that it is given by the energy discrepancy δE measured in the natural units of $\sigma(E)$. In Sec. V we will see that v is also equal to the *thermodynamic speed* of the process.[8]

Now let dE/dt denote the time rate of change of mean energy of the system during the annealing process. We take the following equation as the definition of ϵ:

$$dE/dt=(-1/\epsilon)(E-E^0) \ , \tag{3.6}$$

and proceed to its interpretation.

For each T, $M(T)$ is a stationary Markov chain whose relaxation to equilibrium is known to be asymptotically exponential. That is, mean energy E satisfies an equation like (3.6) asymptotically, where $\epsilon=\epsilon(T)$ is the relaxation time characteristic of the temperature T.

Now in this section we have recast the annealing process so that each instant can be viewed as an instant in a virtual process, namely, relaxation to the (target) temperature $T(t)$. For that reason we identify ϵ, defined in equation (3.6), with the asymptotic relaxation time of this virtual process.

How justified we are in this identification depends on how close these virtual relaxations are to asymptotic mode at the instants in question. For an annealing process in which $T(t)$ varies slowly this seems quite plausible. In any case, this assumption is the device by which we relate the (stationary) relaxation dynamics of $M(T)$ at each T to the (nonstationary) dynamics of annealing, where T is a function of time.

We now obtain the annealing schedule. Let $C=dE^0/dT$ denote heat capacity of the target. Combine the well-known relation[9]

$$\sigma(E^0)=T\sqrt{C}\ , \tag{3.7}$$

with (3.5) and (3.6), eliminating E and E^0, to obtain

$$\frac{dT}{dt}=-(vT/\epsilon)(v\theta+\sqrt{C}\,)^{-1}\ , \tag{3.8}$$

where

$$\theta(T)=1+(T/2C)\frac{dC}{dT}\ .$$

Note that if v is small or $C(T)$ is slowly varying, (3.8) simplifies to

$$\frac{dT}{dt}=\frac{-vT}{(\epsilon\sqrt{C}\,)}\ . \tag{3.9}$$

Schedules incorporating various features of (3.9), such as direct variation with T or inverse variation with ϵ or C, have been proposed by several previous authors,[1,10,11] with, however, only rough plausibility.

In any case, (3.9) is testable as soon as $\epsilon(T)$ and $C(T)$ are known. Preliminary experiments in fact support the conjecture that (3.9) gives optimal annealing schedules.[8]

IV. HEAT CAPACITY AND RELAXATION TIME

For most combinatorial problems, the calculations of C and ϵ from first principles along the lines sketched below range from difficult to intractable. The primary purpose of such calculations is to test the effectiveness of schedules such as (3.8) and (3.9), and thereby to clarify the theoretical role of C and ϵ in optimal annealing.

If this role is to have practical significance, C and ϵ must become available by other means. This could be achieved, for example, by determining reliable statistical estimators for C and ϵ, so that their values can be monitored during an annealing process and used to guide its progress by feedback. Alternatively, C's and ϵ's may be calculated and used for a family of problems. In either case, this is a separate theoretical enterprise, which we do not address here.

Finding $C(T)$ and $\epsilon(T)$ from first principles is a two-phase endeavor. The first phase calculates the density of states g and a matrix Π, which represents the density of neighbors. The second temperature-dependent phase, which we sketch below, derives $C(T)$ and $\epsilon(T)$ from this combinatorial information.

The mean energy and heat capacity are calculated in the usual way from the partition function

$$Z(T)=\sum_{\alpha} g(\alpha)\exp(-\alpha/T)\ , \tag{4.1}$$

where the sum is over values of the energy and $g(\alpha)$ is the number of states with energy α. We have

$$E(T)=T^2\frac{\partial \ln Z}{\partial T}\ , \tag{4.2a}$$

$$C(T)=\frac{dE}{dT}\ . \tag{4.2b}$$

The relaxation time, not unexpectedly, depends on the transition matrix P from Sec. II, which represents the "move class" or dynamical framework of the problem. From one point of view ϵ can be obtained from the modulus of the second largest eigenvalue of the matrix P. Since P is so large, a more practical, albeit approximate, approach is to consider the relaxation dynamics of the "macroscopic" observable, mean energy.

Suppose a canonical ensemble with mean energy E undergoes one step of the process $M(T)$ in (2.4). Let $\Delta(E,T)$ denote the average energy change over the ensemble. Since one step is our unit of time, we may identify Δ with dE/dt. We show below how Δ arises from $Z(T)$ and the matrix P, but first, we establish its connection with relaxation time ϵ.

Let $E^0=E(T)$ denote the mean energy of the canonical ensemble at the relaxation temperature, T. Then, by definition, $\Delta(E^0,T)=0$, so a first-order expansion of $\Delta(E,T)$ about E^0 becomes

$$\Delta(E,T)=\left.\frac{\partial\Delta}{\partial E}\right|_{E=E^0}(E-E^0)\ . \tag{4.3}$$

Comparison with (3.6) gives us the formula for $\epsilon(T)$:

$$-\frac{1}{\epsilon(T)}=\left.\frac{\partial\Delta}{\partial E}\right|_{E=E^0}=\frac{\partial\Delta}{\partial E}(E(T),T)\ . \tag{4.4}$$

We now derive $\Delta(E,T)$ in terms of P and g as follows. We obtain in (4.6) the expected value of the energy change during one step of the process $M(T)$ for a configuration with a specified energy.[12] Then, in (4.7a) weights are applied to these expectations based on the energy distribution in a canonical ensemble of temperature, say T'. The T' dependence is then eliminated by means of the thermodynamic equation of state, $E=E(T')$, in (4.7b).

In the process M of Sec. II the one-step conditional transition probability from $\omega\in\Omega$ to any state ω' with $E(\omega')=\beta$, given that $E(\omega)=\alpha$, is just

$$\Pi(\alpha,\beta)=[g(\alpha)]^{-1}\sum P(\omega,\omega')\ , \tag{4.5}$$

where the sum is over all pairs (ω,ω') with $E(\omega)=\alpha$ and $E(\omega')=\beta$.

If we start from a state ω with $E(\omega)=\alpha$, the expected one-step energy change for the process $M(T)$ is then

$$\delta(\alpha,T)=\sum_{\beta\,(<\alpha)}(\beta-\alpha)\Pi(\alpha,\beta)+\sum_{\beta\,(>\alpha)}(\beta-\alpha)\Pi(\alpha,\beta)\exp[-(\beta-\alpha)/T]\ , \tag{4.6}$$

whereupon, for a canonical ensemble with temperature T', we must apply Boltzmann probabilities as weights to obtain

$$\Delta(E,T)=[Z(T')]^{-1}\sum_{\alpha} g(\alpha)\exp(-\alpha/T')\delta(\alpha,T)\ , \tag{4.7a}$$

$$E=E(T')=[Z(T')]^{-1}\sum_{\alpha}\alpha g(\alpha)\exp(-\alpha/T')\ . \tag{4.7b}$$

Since (4.7) defines Δ implicitly as a function of E, a convenient way to calculate $\partial\Delta/\partial E$ is to differentiate the identity $\Delta(E(T),T)=0$ with respect to T.

Preliminary results[13] of experimental tests of these formulas for a special class of graphs appear very encouraging. Predicted and observed ϵ's differ by about 10%.

V. CHARACTERIZATIONS AND CONNECTIONS

The form of (3.2) shows that the Riemannian metric on probability space defined by

$$(ds)^2=\Xi(p,p+dp) \tag{5.1}$$

is precisely the one studied originally by Wootters[6,14] and independently by Ĉencov.[15,16] The restriction of this metric to the submanifold of canonical distributions is just the thermodynamic metric studied by Ruppeiner[17] and Salamon and co-workers.[18,14]

The constant v in (3.5) can be written as the ratio of two natural scales. The first is the differential displacement in energy, $ds=dE/\sigma(E)$, measured in units of the local energy standard deviation. The second, $d\xi=dt/\epsilon$, is the differential displacement in time, measured in units of the local relaxation time. Note that ds represents a natural scale for displacements in the set of states, while $d\xi$ represents a natural time scale for the process. Using Eqs. (3.5) and (3.6), we see that

$$v=-\frac{ds}{d\xi}\,, \tag{5.2}$$

i.e., v is the distance traversed per unit time where both distance and time are expressed in natural dimensionless units; hence, the apellation *constant thermodynamic speed* for the criterion (3.5) or (5.2).

For a final characterization of our criterion, two more facts need to be connected. First, it is known that for a (slow) quasistatic process that brings a physical system along a one-parameter family of equilibrium states in a given time and with given initial and final states, minimum entropy production and constant thermodynamic speed are equivalent.[19] To state the second fact, let

$$K(p,p^0)=\sum_{\omega}p(\omega)\ln[p(\omega)/p^0(\omega)] \tag{5.3}$$

be the Kullback[20] separation of the system and the target. A corollary of Schlögl's result[21,22] characterizing entropy production is that the entropy produced in a target-system process is the quantity

$$\int\sum_{\omega}\frac{\partial K(p(\omega,t),p^0(\omega,t))}{\partial p(\omega,t)}dp(\omega,t)\,, \tag{5.4}$$

i.e., the sum over the process of the partial changes in Kullback separation due to the system's instantaneous response to the target.[23]

Thus, it seems that minimizing (5.4) is equivalent to constant thermodynamic speed. This information-theoretic characterization of our criterion is very tantalizing. Is it possible that, when a theory of computational complexity for probabilistic algorithms is fully developed, the quantity (5.4) will turn out to be a natural measure of computational effort?[24] If so, the sense in which constant thermodynamic speed is optimal will have been made precise.

VI. CONCLUSION

We have presented a system-target view of the annealing process which relates the problem of choosing a schedule to that of determining an appropriate measure of distance in probability space. We have chosen a natural one, which quantifies the *separation between* the system and the target as the measure of statistical uncertainty associated with their *distinguishability.*

Expressing our criterion for the schedule in testable form, as constant thermodynamic speed, brings into play two quantities, heat capacity and relaxation time, whose importance had been recognized on heuristic grounds. It is gratifying to see these quantities emerge in a natural way from a single theoretical idea. As part of a program to test the criterion experimentally, a method was outlined for calculating these quantities from first principles for a combinatorial problem.

Finally, we have pointed out that if our criterion is in some sense optimal, an intriguing connection exists between computational effort of simulated annealing and what would be entropy production or dissipation for a physical process.

ACKNOWLEDGMENTS

We are grateful to our fellow workers J. Robinson, G. Ruppeiner, J. Pedersen, B. Andresen, and L. Liao for thoughts and criticisms, and to the stimulating environment of the Telluride Summer Research Center.

---

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

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

[3] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, N.J., 1982).

[4] A convenient criterion here is to require easy evaluation of the cost function at the neighboring state.

[5] Our discussion in Secs. III and IV reveals annealing as a slow process dominated at each instant by a target distribution, and for which the initial state of the ensemble is effectively forgotten.

[6] W. Wootters, Phys. Rev. D **23**, 357 (1981).

[7] B. V. Gnedenko, *The Theory of Probability* (Chelsea, New York, 1962), p. 85.

[8] P. Salamon, J. Nulton, J. Robinson, J. Pedersen, G. Ruppeiner, and L. Liao (unpublished).

[9] See, e.g., Ref. 1, p. 673.

[10] R. Otten and L. van Ginneken, *Stepwise Layout Refinement*, in IEEE, Proceedings of the International Conference on Computer Design, Port Chester, New York, 1984.

[11] F. Romeo, A. Sangiovanni Vincentelli, and C. Sechen, *Simulated Annealing Research at Berkeley,* in IEEE, Proceedings of the International Conference on Computer Design, Port Chester, New York, 1984.

[12] This is just the average one-step energy change for a microcanonical ensemble.

[13] G. Ruppeiner (unpublished).

[14] P. Salamon, J. Nulton, and R. S. Berry, J. Chem. Phys. **82**, 2433 (1985).

[15] N. N. Ĉencov, *Statistical Decision Rules and Optimal Inference,* Translations of Mathematical Monographs 53 (American Mathematical Society, Providence, RI, 1982).

[16] L. L. Campbell, Info. Sci. **35**, 199 (1985).

[17] G. Ruppeiner, Phys. Rev. A **27**, 1116 (1983).

[18] P. Salamon and S. Berry, Phys. Rev. Lett. **51**, 1127 (1983).

[19] J. Nulton, P. Salamon, B. Andresen, and Q. Anmin, J. Chem. Phys. **83**, 334 (1985).

[20] S. Kullback, *Information Theory and Statistics* (Dover, New York, 1968).

[21] F. Schlögl, Z. Phys. **191**, 81 (1966); **198**, 559 (1967).

[22] I. Procaccia and R. D. Levine, J. Chem. Phys. **65**, 3357 (1976).

[23] F. Schlögl, Z. Phys. B **59**, 449 (1985).

[24] This conjecture was first proposed by J. Robinson (private communication).