Classification
Physics Abstracts
02.70 — 05.40 — 02.50

On lumped models for thermodynamic properties of simulated annealing problems

Bjarne Andresen (1), Karl Heinz Hoffmann (2), Klaus Mosegaard (3), Jim Nulton (4), Jacob Mørch Pedersen (1) and Peter Salamon (1)

(1) Physics Laboratory, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark

(2) Institute of Theoretical Physics, University of Heidelberg, 6900 Heidelberg, F.R.G.

(3) Geophysical Institute, University of Copenhagen, Haraldsgade 6, 2200 Copenhagen N, Denmark

(4) Department of Mathematics, San Diego City College, San Diego, California 92101, U.S.A.

(*Reçu le 24 novembre 1987, accepté sous forme définitive le 29 avril 1988*)

Résumé. — Nous décrivons une nouvelle méthode pour estimer les propriétés thermodynamiques dans des problèmes de recuit simulé, méthode qui utilise des données acquises durant le recuit simulé. La méthode est basée sur l'estimation des probabilités de transition énergie-énergie et est bien adaptée à des simulations de type recuit simulé, dans lesquelles le système n'est jamais à l'équilibre.

Abstract. — The paper describes a new method for the estimation of thermodynamic properties for simulated annealing problems using data obtained during a simulated annealing run. The method works by estimating energy-to-energy transition probabilities and is well adapted to simulations such as simulated annealing, in which the system is never in equilibrium.

1. Introduction.

Simulated annealing, the recent technique introduced by Kirkpatrick *et al.* [1], exploits an analogy between physical systems and combinatorial optimization problems. The analogy gives rise to an algorithm for finding near optimal solutions to NP problems by simulating the cooling of the corresponding physical system. The algorithm has proved most effective in a number of contexts of scientific and industrial import [2-6], being superior to the best known heuristics on many problems.

The basic analogy arises by identifying states of the combinatorial problem with states of a statistical mechanical system and identifying the cost function with the energy of such a system. The system is allowed to equilibrate by random walk according to the Metropolis algorithm [7] for progressively lower temperaures T_k with the aim of reaching the equilibrium distribution at a sufficiently low T such that the ground state, i.e the solution of the combinatorial optimization problem, has appreciable probability.

This paper introduces an effective technique for the estimation of the thermodynamic properties of the system over a whole range of temperatures from data gathered during an annealing run. As such it is a competitor to the method of McDonald and Singer [8, 9]. Our technique works by estimating a lumped transition probability matrix and is more effective for use with data collected while varying the temperature. Our method allows one to estimate the density of states, the mean energy, the heat capacity, and the relaxation time for a range of temperatures over which there has been an adequate sampling of *attempted* moves. We will refer to these properties collectively as a thermodynamic portrait of the system. In section 6 we present two examples comparing exact thermodynamic portraits with those measured during simulations. However, it is beyond the scope of this paper to attempt physical interpretations of the optimal solutions found.

Such a portrait is useful for implementing schedules $\{T_k\}$ which depend on thermodynamic properties [10-14, 3], including the schedule with constant thermodynamic speed which keeps the mean energy at a fixed number of standard devia-

tions from the equilibrium energy [3, 11-13]. Such schedules have been shown to yield a marked improvement of the performance of the simulated annealing algorithm [3, 11, 13, 14]. In particular, they have yielded a factor of 7 improvement on a seismic deconvolution problem [13], a factor of 3 improvement on logic minimization [3], and a factor of 100 improvement (with modified move class as well) on the traveling salesman problem [3]. While this does not show that knowing thermodynamic properties can help for simulated annealing problems generally, it strongly suggests a closer examination of schedules which exploit such properties. The present paper does not argue for such methods nor does it attempt to discuss the significance of the thermodynamic properties found. Its scope encompasses only the description of a noise resistant method for observing such properties.

2. Dynamics for simulated annealing.

Simulated annealing is based on the Monte Carlo simulation of physical systems. As such it uses the Metropolis algorithm [7] which involves a random walk through the state space of the system. Let $\Omega = \{\omega\}$ represent this state space, let $E : \Omega \to R$ be the cost function (energy) defined on the state space, and let T be the adjustable parameter in the algorithm representing the temperature of the heat bath in which the corresponding physical system is immersed. At each step of the algorithm a neighbour ω' of the current state ω_t is selected at random to become the candidate for the next state. It actually becomes the next state only with probability

$$P_{\text{acceptance}} = \begin{cases} 1 & \text{if } \Delta E \leqslant 0 \\ \exp(-\Delta E/T) & \text{if } \Delta E > 0 , \end{cases} \quad (1)$$

where $\Delta E = E(\omega') - E(\omega_t)$. If this candidate is acceptated, then $\omega_{t+1} = \omega'$, else the next state is the same as the old state, $\omega_{t+1} = \omega_t$. Thus to complete the definition of the dynamics for the algorithm, we must specify (i) the schedule of temperatures as a function of time $T(t)$, and (ii) a definition of which states are to be considered neighbours.

The latter is known as the move class and typically takes the form of an undirected graph structure on the state space. We will denote by $N(\omega)$ the set of neighbours of a state ω in this graph. We remark that for certain problems one may have several alternative move classes available. An important family of such problems involves minimization of a function on real m-space R^m. Here it is common to make moves which change one variable at a time and cycle through the m variables [6, 13]. We discuss the modifications necessary when we define a move class corresponding to each variable in turn in section 6 below in connection with a seismic deconvolution problem.

The matrix $\mathbf{\Pi} = [\Pi_{\beta\alpha}]$ of infinite-temperature transition probabilities from state α to β is defined by

$$\Pi_{\beta\alpha} = \begin{cases} 0 & \text{if } \beta \notin N(\alpha) \\ 1/|N(\alpha)| & \text{if } \beta \in N(\alpha) , \end{cases} \quad (2)$$

where $|N(\alpha)|$ is the number of neighbours of α. These are the transition probabilities if the algorithm automatically accepts each attempted move, i.e., if $T = \infty$. At a finite temperature the acceptance decision is superimposed on $\mathbf{\Pi}$ in equation (2) to give $\mathbf{\Gamma}(T)$ defined by

$$\Gamma_{\beta\alpha} = \begin{cases} \Pi_{\beta\alpha} \exp(-\Delta E/T) & \text{if } \Delta E > 0 , \ \alpha \neq \beta \\ \Pi_{\beta\alpha} & \text{if } \Delta E \leqslant 0 , \ \alpha \neq \beta \\ 1 - \sum_{\xi \neq \alpha} \Gamma_{\xi\alpha} & \text{if } \alpha = \beta , \end{cases} \quad (3)$$

where now $\Delta E = E(\beta) - E(\alpha)$. Equation (3) defines what we will call the *Boltzmannization* operator $\mathbf{B}_T$ which sends a transition probability matrix on a state space equipped with an energy function to a finite temperature version of the transition matrix :

$$\mathbf{B}_T(\mathbf{\Pi}) = \mathbf{\Gamma}(T) . \quad (4)$$

Before any function evaluations, the algorithm begins with a uniform level of ignorance concerning the location of the optimum. This corresponds to a uniform distribution. According to the analogy, this is represented by infinite temperature. An easy way to force the stationary distribution to be uniform [15, 16] at $T = \infty$ is to make $|N(\omega)|$ constant over ω, i.e. to make each vertex have the same number of neighbours. This property is expressed in the language of graph theory [17] as making the graph regular. $\mathbf{\Pi}$ given by equation (2) with $N(\omega)$ corresponding to a regular undirected graph guarantees that $\mathbf{\Pi}$ is symmetric. This is the present counterpart of microscopic reversibility.

3. Lumping.

Since the number of states $|\Omega|$ is astronomical for the problems of interest (of the order of $10^{1\,000}$ for graph partitioning problems of industrial interest), it is not feasible to deal with the $|\Omega| \times |\Omega|$ matrix $\mathbf{\Pi}$ directly. By the physical analogy we will refer to the description at the level of Ω as *microscopic*. Rather than deal with a microscopic description of the process, we introduce a new *coarse-grained* state space whose points consist of subsets of Ω [18, 19]. The simplest and most important case is when each subset contains all states with the same energy, i.e. $A_i = \{\omega \,|\, E(\omega) = E_i\}$. We will refer to this descrip-

tion as *macroscopic* and say that it lumps all states at each energy. (Here and in the following we use latin letters to index macroscopic quantities and greek letters to index microscopic quantities.) Intermediate meso descriptions are also possible and can be physically interesting [20].

In lumping the description at some temperature T to any meso- or macroscopic level, we introduce transition probabilities $p(\mathrm{B}|\mathrm{A})$ between any two subsets A and B of microstates in Ω. To assign reasonable values to such transition probabilities working from the $\Gamma_{\beta\alpha}$, consider the likelihood of a transition from subset A to subset B of Ω. If $\alpha \in \mathrm{A}$, a transition to B occurs from α if the transition occurs to some β in B, i.e. with probability

$$p(\mathrm{B}|\alpha) = \sum_{\beta \in \mathrm{B}} \Gamma_{\beta\alpha} . \tag{5}$$

This just expresses the fact that a move occurs from α to B if and only if a move occurs to some β in B. Given only that we start from some α in a subset A, the likelihood of reaching the disjoint subset B depends on the distribution for starting from the different α's in A. Denoting this distribution by $p_{\mathrm{A}}(\alpha)$, we have

$$p(\mathrm{B}|\mathrm{A}) = \sum_{\alpha \in \mathrm{A}} p(\mathrm{B}|\alpha)\, p_{\mathrm{A}}(\alpha) . \tag{6}$$

The matrix $G(T)$ of lumped transition probabilities at temperature T between subsets A_i, $i = 1, \ldots n$, of a partition of Ω, i.e.

$$G_{ji} = p(\mathrm{A}_j|\mathrm{A}_i) , \tag{7}$$

can be computed using two matrices V and U of size $n \times |\Omega|$ and $|\Omega| \times n$ respectively such that [18, 19]

$$G = V\Gamma U = \mathbf{L}(\Gamma), \tag{8}$$

where $\mathbf{L}$ is a lumping operator, and V and U are defined by

$$V_{i\omega} = \begin{cases} 1 & \text{if } \omega \in \mathrm{A}_i \\ 0 & \text{else} \end{cases} \tag{9}$$

$$U_{\omega i} = \begin{cases} p_{\mathrm{A}_i}(\omega) & \text{if } \omega \in \mathrm{A}_i \\ 0 & \text{else} . \end{cases} \tag{10}$$

Thus $G(T)$ is the lumped analog of $\Gamma(T)$. Note that this definition requires a distribution p_{A_i} on each of the subsets A_i in the partition.

The stochastic matrix G now defines a new Markov process whose states are the subsets A_i. This new Markov process depends on the p_{A_i} and will be a coarse grained equivalent to the dynamics of the microscopic process provided a double transition gives the same result for both processes [18] :

$$V\Gamma^2 U = (V\Gamma U)^2 . \tag{11}$$

Except for very special Γ (e.g. where $\Gamma_{\beta\alpha}$ has the same value for all $\alpha \in \mathrm{A}$ and all $\beta \in \mathrm{B}$, see Ref. [18]), this will not hold for any choice of the p_{A_i}. On the other hand, one can show [18] that if equation (11) works for any p_{A_i}, then it works for p_{A_i} equaling the equilibrium distribution relativized to A_i,

$$p_{\mathrm{A}_i}(\omega) = \gamma(\omega) / \sum_{\xi \in \mathrm{A}_i} \gamma(\xi) , \tag{12}$$

where γ is the equilibrium distribution for Γ on the full microscopic space Ω. In general the microscopic Markov process viewed at the macroscopic level will not be Markoffian ; instead it will show apparent memory effects. If however the process has reached equilibrium, and the p_{A_i} are taken to equal the equilibrium distribution relativized to A_i (henceforth assumed for any *lumping* effort), then on the macroscopic level of description the processes described by Γ and G are equivalent [18].

For the present paper, the all important special application of the above lumping procedure consists of lumping the infinite-temperature transition matrix Π using the energy subsets $\mathrm{A}_i = \{\omega | E(\omega) = E_i\}$ in which case the stationary distribution is changed from uniform to being the density of states $\mathbf{p}$ with $p_i = |\mathrm{A}_i| / |\Omega|$. In this case we will denote the lumped matrix by P. Our estimating procedures below work by collecting runtime estimates of P.

4. A commutative diagram.

Starting from the microscopic infinite-temperature description Π of the dynamics, which was defined from the graph structure on the set of states, we defined a finite-temperature version Γ and a lumped version P. It is natural at this stage to consider the finite-temperature version of the lumped process and its relation to the lumped version G of the finite-temperature process. Formally, the question concerns the commutativity between the Boltzmannization operator $\mathbf{B}_T$ and the lumping operator $\mathbf{L}$. Provided the energy is constant on all sets in the partition $\{\mathrm{A}_i, i = 1, \ldots n\}$ defining the lumping operator $\mathbf{L}$,

$$\mathbf{L}(\mathbf{B}_T(\Pi)) = \mathbf{B}_T(\mathbf{L}(\Pi)) \tag{13}$$

if and only if Π satisfies microscopic reversibility.

Microscopic reversibility is the microscopic counterpart of detailed balance [21], and in physical terms it states that, at equilibrium, the flows in the two directions between any pair of states are equal. Mathematically for a stochastic matrix M and a vector $\mathbf{m}$, we say that M and $\mathbf{m}$ satisfy microscopic reversibility provided

$$M_{\beta\alpha}\, m_\alpha = M_{\alpha\beta}\, m_\beta . \tag{14}$$

Note that if a vector **m** satisfies microscopic reversibility with respect to a stochastic matrix M then **m** must be the stationary distribution of M, i.e.,

$$m_\alpha = \sum_\beta M_{\alpha\beta}\, m_\beta \,. \tag{15}$$

Then the symmetry of $\mathbf{\Pi}$ implies that the uniform distribution satisfies microscopic reversibility relative to $\mathbf{\Pi}$. It then follows that the lumped P satisfies microscopic reversibility with the density of states **p** defined above, $\mathbf{\Gamma}$ satisfies microscopic reversibility with the distribution $\gamma_\xi = \mathrm{e}^{-E_\xi/T}/\sum_\zeta \mathrm{e}^{-E_\zeta/T}$, and G satisfies microscopic reversibility with the distribution $g_i = p_i\, \mathrm{e}^{-E_i/T}/\sum_j p_j\, \mathrm{e}^{-E_j/T}$. Thus microscopic reversibility is satisfied and we have

$$\mathbf{B}_T(P) = \mathbf{B}_T(\mathbf{L}(\mathbf{\Pi})) = \mathbf{L}(\mathbf{B}_T(\mathbf{\Pi})) = \\ = \mathbf{L}(\mathbf{\Gamma}(T)) = G(T)\,. \tag{16}$$

We extend $\mathbf{B}_T$ and **L** to vectors in the natural way (cf. Eqs. (8) and (3)) using,

$$\mathbf{L}(\mathbf{m}) = V\mathbf{m} \tag{17}$$

and

$$\mathbf{B}_T(\mathbf{m}) = (m_1\, \mathrm{e}^{-E_1/T}/Z(T), \ldots, \\ m_n\, \mathrm{e}^{-E_n/T}/Z(T))^{\mathrm{tr}}\,, \tag{18}$$

where tr signifies transpose, and $Z(T)$ is the associated partition function

$$Z(T) = \sum_j m_j\, \mathrm{e}^{-E_j/T}\,. \tag{19}$$

Letting **F** represent the operator which sends a stochastic matrix to its stationary distribution, we can see that the three operators **L**, $\mathbf{B}_T$, and **F** commute on the subset of stochastic matrices satisfying detailed balance. In particular, the equilibrium energy distribution $\mathbf{g}(T)$ may be computed by any one of several ways, e.g.,

$$\mathbf{B}_T(\mathbf{p}) = \mathbf{B}_T(\mathbf{F}(\mathbf{L}(\mathbf{\Pi}))) = \mathbf{F}(\mathbf{B}_T(\mathbf{L}(\mathbf{\Pi}))) = \\ = \mathbf{F}(G(T)) = \mathbf{g}(T)\,. \tag{20}$$

These relations are illustrated in figure 1.

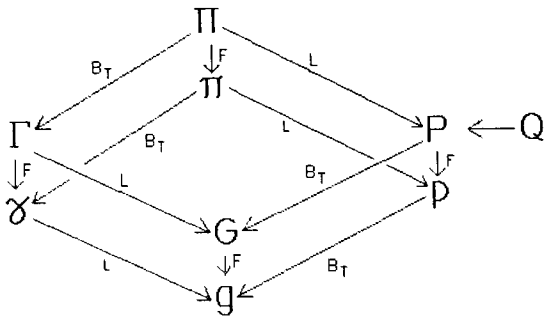

Fig. 1. — Diagram illustrating the commutativity of the lumping **L** and Boltzmannization $\mathbf{B}_T$ operators with each other and with the operator **F** which sends a stochastic matrix to its stationary distribution. The commutativity holds only for matrices satisfying microscopic reversibility. Experimental data is used by the method described in the paper in the form of the frequency of attempted energy changes. These are stored in the matrix **Q** which is used to estimate **P**.

The experimental technique described below will estimate values of $\mathbf{\Pi}$ using data gathered during simulations. These estimates typically do *not* satisfy microscopic reversibility and so one gets slightly different predictions computing **g** different ways.

5. The method.

We are now ready to describe our method. We emphasize that the needs of simulated annealing for accurate estimators of thermodynamic properties are unique in Monte Carlo methods in that the estimates should be collected without spending time on Metropolis moves at equilibrium, i.e. once equilibrium has been approached closely, simulated annealing requires the temperature to be lowered further. The noise levels are very large and, while several implementations reported in the literature use estimated thermodynamic properties to good advantage [3, 11, 13, 14], runtime estimates of such properties without specifically parametrized models have proved impossible to date.

Our method works by estimating the entries in P, the infinite-temperature transition probability matrix on the set of energies. These can be estimated during the Metropolis algorithm by keeping track of attempted moves from the current energy to other energies while the annealing proceeds. The entries are relative probabilities of moving from energy E_i to energy E_j. The actual estimate is taken to be the fraction of the time a move from energy E_i is *attempted* to energy E_j. Specifically, we keep running tallies

$$Q_{ji} = \text{number of attempted moves from } i \text{ to } j \tag{21}$$

which accumulate information and from which one can construct the estimators

$$P_{ji} = Q_{ji}/\sum_k Q_{ki}\,. \tag{22}$$

Once P is known, the equilibrium properties of the system such as mean energy $E(T)$ and the heat capacity $C(T)$ may be computed by finding its stationary distribution **p** which is the density of states and which gives the partition function $Z(T)$ by means of formula (19). From Z all equilibrium properties follow in the usual way, e.g.

$$E(T) = T^2\, \partial(\ln Z)/\partial T \\ C(T) = \mathrm{d}E/\mathrm{d}T\,. \tag{23}$$

The final part of the thermodynamic portrait, the

relaxation time $\varepsilon(T)$, is an estimate of the local time scale. This can be obtained from the second largest eigenvalue λ_2 of $G = \mathbf{B}_T(P)$; the largest eigenvalue $\lambda_1 = 1$ corresponding to equilibrium. From λ_2 we have

$$\varepsilon = -1/\ln \lambda_2 . \tag{24}$$

An alternative route is the first auto correlation time of the energy which is conveniently computed by [11-13, 3]

$$\varepsilon(T) = T^2 C(T) / \sum_i g_i \sum_{j>i} (E_j - E_i)^2 G_{ji} \tag{25}$$

and which can also be characterized as the relaxation time of one Boltzmann distribution to a neighbouring Boltzmann distribution [12]. Under rather mild assumptions the two estimators in equations (24) and (25) become identical for simulated annealing problems in the limit of low temperatures.

The method outlined above is in certain respects a continuous version of that of McDonald and Singer [8, 9]. They want to infer the equilibrium distribution $\mathbf{g}(T_1)$ at some temperature T_1 from a known equilibrium distribution $\mathbf{g}(T_0)$ by estimating the density of states $\mathbf{p}$. If the two distributions have insufficient overlap to allow the extrapolation with acceptable accuracy, one or more intermediate equilibrium distributions are used to bridge the gap. Thus their method involves generating large samples in order to determine accurate $\mathbf{g}(T)$ at a few temperatures. Our method, on the other hand, operates by collecting a few samples at very many temperatures, none of which need to be at equilibrium. It is evident then that our procedure is tailored to generate the thermodynamic portrait as the calculation progresses, i.e., specific to the needs of simulated annealing.

6. Examples.

Our two examples involve a graph partitioning problem and a problem of seismic deconvolution. Previous direct attempts to collect equilibrium state data during simulated annealing runs have failed for both examples due to large levels of noise and lack of appreciable time spent at each temperature after equilibration [3, 11]. As illustrated in figures 3 and 4 below, our method gives quite good portraits from simulations. In particular annealing schedules $T(t)$ using data from the portraits produced significant improvement in the performance of simulated annealing runs [11, 13].

6.1 GRAPH PARTITIONING. — Graph partitioning involves dividing the vertices of a given graph into two subsets while minimizing the number of edges running from one subset to the other. Our implementation requires the two subsets of the vertex partition to be the same size. Industrially interesting applications include partitioning the elements of a circuit to fit onto two chips while minimizing the number of wires between the chips. Physically interesting realizations include finding the ground state of a spin glass by associating a spin to each vertex. Each spin points up or down according to which subset of the partition the vertex belongs to, and an edge between vertices corresponds to coupling between the spins. Our constraint of an even partitioning of the graph corresponds to requiring a constant magnetization of zero for the spin glass.

The graph used here is a [20, 2] necklace [20, 25] shown in its simplest form in figure 2. In general an $[m, n]$ necklace is a graph consisting of m identical units connected in a circular fashion, where each unit is a completely connected graph of n nodes. This graph was chosen for its high degree of symmetry which allowed a direct combinatorial calculation [25] of its energy-lumped density of states $\mathbf{p}$ and transition probability matrix P for the theoretical portrait as well as its ground state energy (equal to 2). Clearly a necklace is not as complicated as a random graph, and its partitioning is not NP complete, but it has proven a convenient compromise between complexity and analytical feasibility. Subsequent annealing experiments on random graphs have been of comparable difficulty, supporting our belief that the algorithm depends only on thermodynamic properties. The experimental data was obtained during simulated annealing runs, i.e. using the Metropolis [7] algorithm at progressively lower temperatures. All together 1.37×10^8 steps were attempted in a state space with 3.87×10^{11} states. The accumulated counts in Q, equation (21), were finally used to determine the density of states $p(E)$, average energy $E(T)$, heat capacity $C(T)$, and relaxation time $\varepsilon(T)$ using equations (22, 23 and 25).

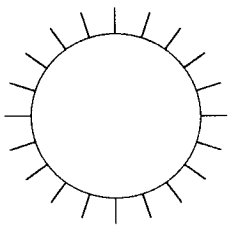

Fig. 2. — [20, 2] necklace.

Figure 3 compares the essentially exact theoretical and the energy-lumped experimental results. The density of states shows excellent agreement, and the average energy shows only a slight discrepancy at low energies, leading to a 15 % overestimation of the maximum of the heat capacity but at the correct temperature. This modest build up of error is to be expected since C is a derivative of E which in turn is

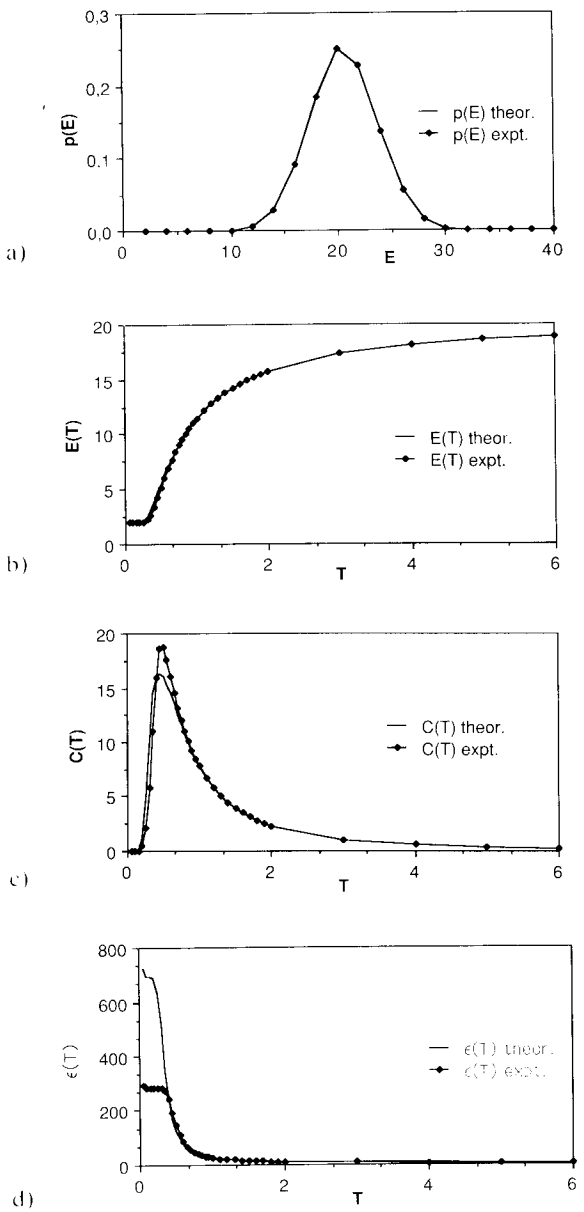

Fig. 3. — Thermodynamic portrait of the partitioning of a [20, 2] necklace with states lumped to the 20 possible energy levels. Full curves are theoretical results from an almost exact combinatorial analysis of the problem, points are experimentally evaluated based on a compilation of attempted steps during simulated annealing runs. $p(E)$ is the density of states as a function of state energy ; $E(T)$ the average energy as a function of heat bath temperature ; $C(T)$ the heat capacity ; and $\varepsilon(T)$ the relaxation time.

a derivative of Z. Also the relaxation times agree nicely down to a temperature of about 0.4, just under the maximum of C, below which any ε calculated from P becomes essentially constant. The true relaxation time ε calculated from $\mathbf{\Pi}$ of course keeps growing. This effect is due to the lumping together of local minima with other more transient states. (This will be discussed further in Ref. [20].)

6.2 SEISMIC DECONVOLUTION. — The seismic reflection method attempts to find the thicknesses and sound velocities of subsurface strata using data on the recordings of ground motion set up by a seismic energy source. The associated deconvolution problem is of great industrial significance in assessing the prospectivity of an area in the exploration for oil and gas. In essence it works as follows : from the surface a signal $w(t)$ is sent vertically down into the earth's crust which is modeled as a stack of layers. The signal is reflected at each interface resulting in a composite signal $s(t)$ at the surface. The unknowns are the travel times of the signal between the surface and the layer interfaces, and the reflection coefficients of these interfaces. Assuming a linear relation between signal and response, the response is

$$s(t) = \int \mathrm{d}t'\, r(t')\, w(t - t') + \xi\ , \qquad (26)$$

where ξ is a noise term and

$$r(t) = \sum_k a_k\, \delta(t - t_k) \qquad (27)$$

is the reflectivity of the earth's crust. Here t_k is twice the travel time from the surface down through layer k, and a_k is the strength of the reflection at the bottom of layer k. The deconvolution estimates the parameters t_k and a_k which minimize the difference between the monitored signal and the predicted one. Traditional methods [26] seek local minima, but such problems typically have many local minima [13, 27], and it is imperative to locate the global minimum wich is exactly the intended purpose of simulated annealing.

The small example considered here involves two reflecting interfaces and thus contains four unknown parameters : two a's and two t's. It was chosen so that a direct evaluation of **p** by exhaustive examination of all states in a discretized grid was feasible. (This is a highly unusual situation for simulated annealing problems.) The experimental data was collected during simulated annealing runs with a total of 5×10^3 attempted steps in a state space with 2×10^5 states.

A move class was adopted for which the four parameters were visited cyclically. For each visit, the considered parameter was perturbed and a Metropolis decision was made. While this move class is commonly used in seismic applications of simulated annealing [6, 13, 27], in the language of section 2 this amounts to a different move class for each parameter in the objective function. In particular it requires a separately estimated P_s for each of the move classes and yields a proper Markov process only on using the combined transition matrix $P = P_1 P_2 P_3 \ldots P_s$ for a full cycle [28]. For the present method it is probably advisable to adopt a scheme where the next

parameter is randomly rather than cyclically selected in which case estimating one P suffices.

Figure 4 compares the exact **p**, E, and C with the corresponding experimental functions based on the accumulated Q and equations (22)-(24). No theoretical ε was available since it would have required knowledge of the transition matrix P (equation (24) or (25)) which would have entailed an exhaustive examination of all 4×10^{10} possible transitions, not just the 2×10^5 states. Although agreement is not quite as good as for the graph partitioning problem in spite of the larger coverage of the state space, dynamic estimation of the thermodynamic portrait during the simulated annealing run is a very attractive improvement over *a priori* estimated quantities. In particular the double-maximum form of C would have been hard to predict.

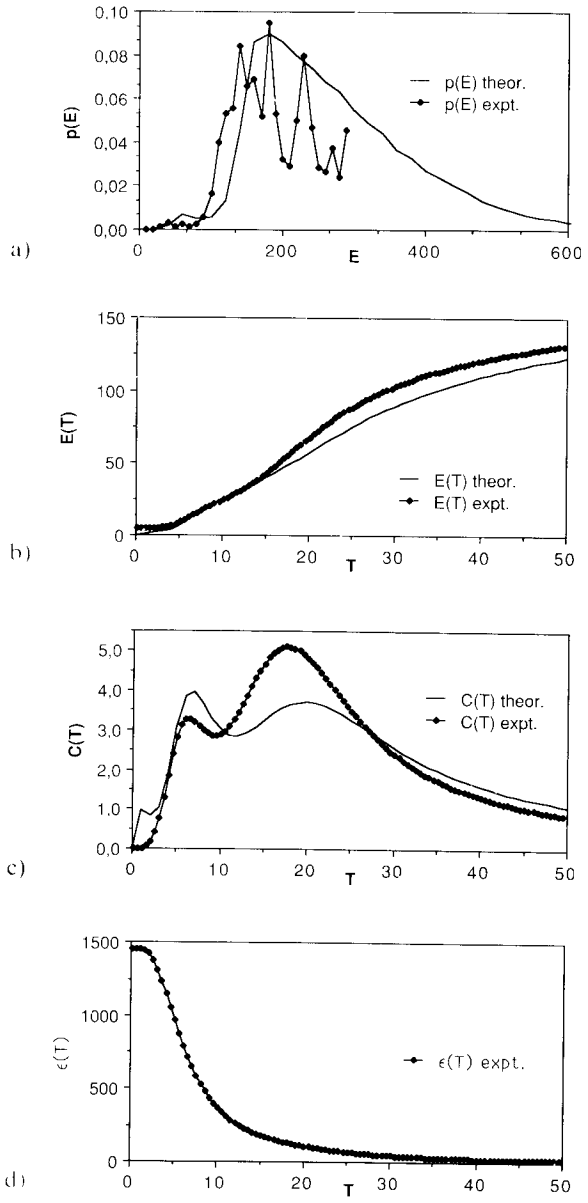

Fig. 4. — Thermodynamic portrait of the deconvolution of seismic reflection data in a two-layer model. Signatures are the same as in figure 3.

7. Conclusions.

The present paper presents a new competitor to the technique of MacDonald and Singer [8, 9] for the estimation of thermodynamic properties over a range of temperatures. Whereas their method was adapted to using a few well sampled temperatures, the present method is based on a sliding temperature, and it appears to be rather insensitive to the temperature schedule used. While the present data was collected using runs assessing the efficacy of competitor schedules, real runtime implementations are made possible by careful application of the techniques outlined in this paper and are presently under development.

The problems of interest tend to be highly non-ergodic at $T = 0$ and, on the time scale of the observation, at higher temperatures as well. Therefore the system stays for extended times in certain parts of the state space. Thus, on shorter time scales, only the thermodynamic properties of that local region of state space are of interest for the behaviour of the system. The method presented here can take care of this fact automatically.

Our method involved accumulating the matrix Q which is the size of the state space squared. Without using lumped states, that would have been entirely out of the question. Although, as discussed in section 3, exact lumpability is rare, judiciously chosen lumping can provide a very good approximation. A good rule of thumb is that each subset should be chosen as homogeneous as possible. One may lump at many levels thereby obtaining much or little detail. At any level, the point of the present article holds true : lumping provides a workable, noise resistant technique for runtime evaluation of thermodynamic properties.

Acknowledgements.

We are grateful for inspiring conversations with Drs. Lars Kai Hansen and Stig Steenstrup. Peter Salamon's visit at the University of Copenhagen was supported by the Danish Natural Science Research Council.

References

[1] KIRKPATRICK, S., GELATT, C. D., Jr and VECCHI, M. P., *Science* **220** (1983) 671.

[2] JOHNSON, D. S., ARAGON, C. A., MCGEOCH, L. A. and SCHEVON, C., Optimization by Simulated Annealing : an Experimental Evaluation (Part I) to appear in *Oper. Res.* :
Private communication from JOHNSON, D. S. (1988).

[3] LAM, J., and DELOSME, J.-M., Yale University Report 8608 (1987) ; *IEEE, Proc. ICCAD* (1986) p. 348.

[4] GEMAN, S. and GEMAN, D., *IEEE Trans., PAMI* **6** (1984) 721.

[5] GREST, G. S., SOUKOULIS, C. M. and LEVIN, K., *Phys. Rev. Lett.* **56** (1986) 1148.

[6] ROTHMAN, D. H., *Geophysics* **50** (1985) 2784 ; **51** (1986) 332.

[7] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E., *J. Chem. Phys.* **21** (1953) 1087.

[8] MCDONALD, I. R. and SINGER, K., *J. Chem. Phys.* **47** (1967) 4766.

[9] VALLEAU, J. P. and CARD, D. N., *J. Chem. Phys.* **57** (1972) 5457.

[10] OTTEN, R. and VAN GINNEKEN, L., *IEEE, Proc. ICCAD* (1984) p. 96.

[11] SALAMON, P., NULTON, J., ROBINSON, J., PEDERSEN, J. M., RUPPEINER, G. and LIAO, L., Simulated annealing with constant thermodynamic speed, *Comp. Phys. Commun.* (to appear).

[12] NULTON, J. and SALAMON, P., *Phys. Rev. A* **37** (1988) 1351.

[13] JAKOBSEN, M. O., MOSEGAARD, K. and PEDERSEN, J. M., Global model optimization in reflection seismology by simulated annealing. Proc. of The Mathematical Geophysics Fifth International Seminar on Model Optimization in Exploration Geophysics (Berlin) 1987.

[14] REES, S. and BALL, R. C., *J. Phys. A* **20** (1987) 1239.

[15] While this condition is not quite necessary (see Ref. [16], p. 5 and the literature there cited), relaxing this assumption has been of no practical importance so far [16].

[16] BINDER, K., Monte Carlo Methods in Statistical Physics (Springer Verlag, Berlin) 1979.

[17] HARARY, F., Graph Theory (Addison-Wesley, Reading, MA) 1969.

[18] KEMENY, J. G. and SNELL, J. L., Finite Markov Chains (D. Van Nostrand, Inc., Princeton, NJ) 1960.

[19] BURKE, C. J. and ROSENBLATT, M., *Ann. Math. Stat.* **29** (1958) 1112.

[20] RUPPEINER, G., PEDERSEN, J. M. and ANDRESEN, B., Experimental properties of necklaces (in preparation).

[21] The two notions are often confused [16], but strictly speaking [22-24] microscopic reversibility is the statement, equation (14) about transition rates between individual states of the system at equilibrium, whereas detailed balance is the same property at the macroscopic level. Thus detailed balance is a slightly weaker property in that circular reactions are possible [22-24] violating equation (14).

[22] REIF, F., Foundations of Statistical Thermal Physics (McGraw-Hill, New York, NY) 1965.

[23] KREUZER, H. J., Nonequilibrium Thermodynamics and its Statistical Foundations (Clarendon Press, Oxford) 1981.

[24] DE GROOT, S. R. and MAZUR, P., Non-Equilibrium Thermodynamics (North-Holland, Amsterdam) 1962.

[25] NULTON, J., PEDERSEN, J. M. and ANDRESEN, B., Combinatorics on necklaces (in preparation).

[26] DUIJNDAM, A. J. W., VAN DER SCHOOT, A. and BERKHOUT, A. J. Seismic Inversion by Parameter Estimation and Depth Migration, paper presented at the Sixth MINTROP seminar (Kassel, FRG) 1986.

[27] MOSEGAARD, K., Inversion of Seismic Field Data by Monte Carlo Optimization, paper presented at the Forty Eighth EAEG meeting in Oostende (1986) ; Stochastic Model Optimization in Reflection Seismology, Ph. D. thesis presented to the University of Copenhagen (1987).

[28] CREUTZ, M. and FREEDMAN, B., *Ann. Phys.* **132** (1981) 427.