Open Systems & Information Dynamics
Vol. 2 No 1 (1993) 1–12

Analytic Constant Thermodynamic Speed-Cooling Strategies in Simulated Annealing

B. ANDRESEN
Physics Laboratory, University of Copenhagen
Universitetsparken 5, DK–2100 Copenhagen Ø, Denmark

J. M. GORDON
Center for Energy and Environmental Physics
Jacob Blaustein Institute for Desert Research
Ben Gurion University of the Negev, Sede Boqer Campus, Israel
and
The Pearlstone Center for Aeronautical Engineering Studies
Department of Mechanical Engineering
Ben Gurion University of the Negev, Beersheva, Israel

(Received March 22, 1993)

Several key problems previously analyzed by simulated annealing are well described by a simple two-state statistical mechanical model, equivalent to the one-dimensional Ising model, in conjunction with an Arrhenius type rate equation. With this model we develop an analytic solution for the optimal cooling strategy which computationally is much faster than the previously developed adaptive algorithm. In the long time limit the analytic solution reduces to the one cooling schedule which is sure to reach the ground state. Our model predictions are in good agreement with earlier numerical simulation results for optimal cooling schedules as well as for stationary and dynamic properties of system energy, heat capacity, and relaxation time.

1 Introduction

Simulated annealing is a stochastic simulation procedure for dealing with complex combinatorial optimization problems [1]. It is particularly well suited for systems characterized by an energetic structure with many local minima and for which relax-

ation times grow exceedingly long as lower energies are sampled, i.e. as the system is 'cooled'.

The optimization method is based on a formal analogy between purely mathematical combinatorial problems and statistical mechanical systems. This analogy permits analysis in terms of: the system energy (or cost function) E, heat capacity C, relaxation time ε, and annealing time t (usually measured in computer iterations), and with the environment temperature $T(t)$ as the only control variable. The basic iteration is usually carried out with the time-honoured Metropolis algorithm [2] which prescribes that the probability of accepting a proposed next step which changes the system energy by ΔE is given by

$$P_{\text{accept}} = \begin{cases} 1 & \text{if } \Delta E \leq 0, \\ e^{-\Delta E/T} & \text{if } \Delta E > 0, \end{cases} \tag{1.1}$$

where the Boltzmann constant k has been set equal to unity throughout.

The control variable in the problem is the cooling strategy or time dependence of $T(t)$. The objective is to find a strategy for $T(t)$ that minimizes entropy production in the entire process for given initial and final states, since entropy production is equivalent to the number of questions asked or the number of computer iterations performed. The problem is often alternatively expressed in terms of arriving at the lowest energy state in a given time.

The aim has usually been to develop an algorithm which treats the system being cooled as an effective black box with unknown internal structure, i.e. treating simulated annealing as a general optimization algorithm as opposed to the highly specialized procedures available, e.g. for the traveling salesman problem [3]. Many *ad hoc* annealing schedules have been proposed, e.g.,

$$T(t) = ae^{-t/b}, \tag{1.2}$$

$$T(t) = \frac{a}{b+t}, \tag{1.3}$$

$$T(t) = \frac{a}{\ln(b+t)}, \tag{1.4}$$

where a and b are constants. However, the most efficient general schedule [4] is based on information-theoretic arguments and proceeds at constant thermodynamic speed [5,6]

$$\frac{dT(t)}{dt} = -\frac{v\,T(t)}{\varepsilon(T)\sqrt{C(T)}} \tag{1.5}$$

or, equivalently,

$$\frac{\langle E\rangle - E_{\text{eq}}(T)}{\sigma} = v, \tag{1.6}$$

where v is the (constant) thermodynamic speed, $\langle E\rangle$ and σ the mean energy and standard deviation of the natural fluctuations of the system, and finally $E_{\text{eq}}(T)$ is the

internal energy the system would have if it were in equilibrium with its surroundings at temperature T. In eq. (1.5) $C(T)$ and $\varepsilon(T)$ are estimated based on the entire past history of the annealing [7].

The physical interpretation of eq. (1.6) is that the environment should at all times be kept v standard deviations ahead of the system. Similarly, eq. (1.5) indicates that the annealing should slow down where internal relaxation is slow and where large amount of 'heat' has to be transferred out of the system. When C and ε do not vary with temperature, eq. (1.5) integrates to eq. (1.2). The constant thermodynamic speed algorithm continuously updates or educates itself during the annealing process based on all prior events in the simulation. That is, at any given moment the simulation finds the temperature-dependent functions $C(T)$ and $\varepsilon(T)$ in a manner prescribed by statistical mechanics and adopts the new values in the continuation of the cooling schedule [7].

The goal in the present paper is not to propose an alternative black-box algorithm, but rather to examine the contents of the black box to see if a relatively simple model for its contents can lead to an optimal cooling strategy for that particular system — and possibly systems of a similar structure — which is less computer intensive than the procedure outlined in ref. [7]. We will demonstrate that this appears to be the case for several problems which have been central in initial simulated annealing studies. For these cases we will derive an *analytic* cooling strategy, one that does not require continual updating during the annealing process, but rather is a specified function of temperature only.

2 Two-state model

2.1 Basic concepts

We seek the optimal path described by eq. (1.5), but instead of compiling statistical information about the state space and hence C and ε, we assume that the simple two-state model, equivalent to the one-dimensional Ising model [8], is a good approximation to the microscopic picture of a number of complex optimization problems. We should emphasize that this approximation is used only for determining the optimal temperature schedule $T(t)$; the annealing itself is carried out with the full energy function.

The combinatorial necklace or integrated chip problem [4,7,9], partitioning of random graphs [10], and certain spin glass systems [11] are obvious candidates for the model, since they involve individual vertices which can be placed in either of two sets. In general the two-state model should be useful for any system consisting of a set of weakly interacting two-level systems.

Spin glasses [11,12] as well as real glasses [13,14] of higher than one dimension have also been successfully treated by a two-state model. However, our two-state model is not the same as the one used in these last references, where the energy difference

between the two levels is distributed according to Gaussian statistics. Those authors were motivated by trying to explain a linear dependence of the heat capacity on temperature which was observed in experiments on real glasses. Because our simple two-state model does not incorporate a distribution of energy separations, it cannot predict this characteristic, nor certain other properties, of real glasses. However, neither is it intended nor expected to do so. Rather, recognizing its simplicity and limitations, we concentrate on appropriate classes of optimization problems, amenable to treatment with simulated annealing, for which we can compare theoretical predictions with computer-experimental results.

2.2 Basic equations

The model used here states that each particle can be in one of two states only: a lower state of energy $-J$ and a higher state of energy $+J$. For ease of notation we define the dimensionless temperature variable

$$x = \frac{T}{2J}. \tag{2.1}$$

It is then straightforward to derive the partition function for a macroscopic system comprised of N such particles ($N \gg 1$) at temperature x [8] from which the system energy and heat capacity follow directly,

$$E(x) = -NJ \tanh \frac{1}{2x}, \tag{2.2}$$

$$C(x) = \frac{N\left(\frac{1}{2x}\right)^2}{\cosh^2 \frac{1}{2x}}. \tag{2.3}$$

$E(x)$ and $C(x)$ are plotted in Fig. 1.

In addition, we assume that the system relaxation time is well represented by the classical Arrhenius expression

$$\varepsilon(x) = Ae^{B/x}, \tag{2.4}$$

where A is a (constant) collision frequency, and B is the apparent barrier height of the transition state. Even though the energy landscape of the system will generally contain several different barriers, only the highest will be effective in the long-time limit when all the faster relaxations have died out, i.e. close to equilibrium. This assumption is consistent with the derivation of eq. (1.5). Simulations [15] on the travelling salesman problem indicate that such an Arrhenius-type law is a good description of the relaxation behaviour.

Introducing these assumptions into the optimal rate annealing schedule eq. (1.5) yields

$$\frac{dx}{dt} = -v'x^2\left(\exp\left(\frac{1}{x}\right) + 1\right) \exp\left(-\frac{B + 1/2}{x}\right), \tag{2.5}$$

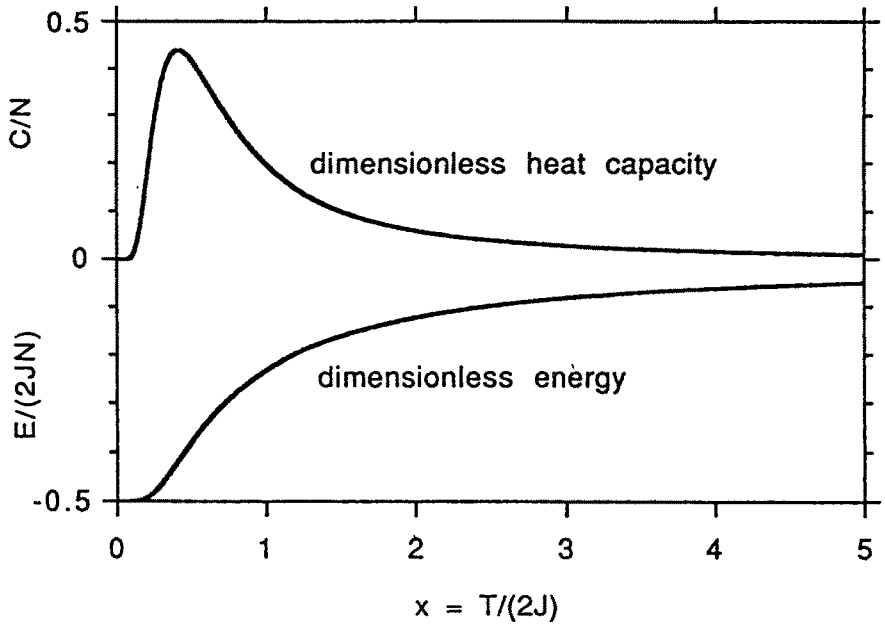

Fig. 1. Dimensionless heat capacity C/N and dimensionless energy $E/(2JN)$ as functions of the similarly dimensionless system temperature $x = T/(2J)$ for the two-state model (see eqs. (2.1)–(2.3)).

where v' is a constant. This is the differential equation defining the optimal temperature path $x(t)$ for a two-state model with an Arrhenius-type relaxation. It can, at least in principle, be solved analytically before any annealing begins and thus replaces the much more time-consuming procedure of collecting statistical information along the way [7]. The 'price' for this faster procedure clearly is less generality and knowledge about the system required in advance.

Equation (2.5) has closed form solutions only for half-integer values of B, i.e. for $B = 0.5, 1.0, 1.5$, etc. For example, for $B = 1.5$ we find

$$v't = \exp\left(\frac{1}{x}\right) - \exp\left(\frac{1}{x_0}\right) - \ln\left(\frac{1+\exp(1/x)}{1+\exp(1/x_0)}\right), \tag{2.6}$$

where we write $x_0 = x(t = 0)$. Equation (2.6) specifies the annealing path $x(t)$ implicitly. For non-half-integer values of B eq. (2.5) must be solved numerically, but only once, before the start of annealing.

2.3 Comparison with previous simulated annealing studies

It is not surprising that the energy and heat capacity for the initial combinatorial optimization problems analyzed [4,7,9] are in good agreement with eqs. (2.2) and (2.3), for an appropriate choice of J, including a maximum in C at $x = 0.42$. However, excellent agreement of the solution of eq. (2.5) is also found with the *dynamics* of the optimal annealing schedule [4] when the barrier height $B = 1.2$ is used. This B value is somewhat less than the rigorously derived [16] barrier height of 2 for $[n, 2]$ necklaces. Since B is an averaged quantity, it also contains paths with no barrier and thus must be less than the combinatorial value.

We find that the much more complicated problem of seismic inversion [7] turns out to be qualitatively approximated by a superposition of just two autonomous two-state models, which reproduces the relative positions and values of the two maxima in the heat capacity as well as its approach to zero. This agreement is probably achieved because the two-way travel times and the reflection coefficients are independent quantities. Unfortunately, the original raw data from ref. [7] are no longer available and hence we cannot present a more quantitative comparison.

When C and ε are independent of T, eq. (1.5) easily integrates to the simple *ad hoc* annealing schedule eq. (1.2). However, the only schedule for which it has rigorously been proven [17] that the procedure will with certainty find the ground state is the extremely slow eq. (1.4), with the constant a equal to the largest barrier height. It is interesting to note that this schedule is the solution of the differential equation

$$\frac{dx}{dt} = -v''x^2 \exp\left(-\frac{a}{x}\right) \tag{2.7}$$

which bears a close resemblance to the optimal eq. (2.5), except that the Geman-Geman barrier height is one half unit less than Arrhenius's,

$$a = B - \frac{1}{2}. \tag{2.8}$$

In other words, the optimal dynamic annealing schedule approaches the logarithmic schedule for sufficiently low temperatures (relative to the barrier heights). Thus it will also with certainty eventually find the true solution without spending as much time at higher energies as eq. (1.4). This is nicely borne out by Ruppeiner's very carefully and slowly annealed random graphs [18], where the optimal cooling schedule at constant thermodynamic speed is very accurately fit by a Geman-Geman schedule. One should keep in mind that energies as defined in Ruppeiner's studies [18] are twice as large as our energies, eq. (2.1). We find that a barrier height of $B = 0.86$ very accurately reproduces the data and fit in his Fig. 4. In contrast, Salamon *et al.*'s comparison among several schedules [4] has a much shorter time horizon and thus shows constant thermodynamic speed annealing superior to the Geman-Geman schedule *on that time scale.*

At low temperatures (corresponding to long times), when $\exp(1/x) \gg 1$, the optimal path eq. (2.5) may be integrated to yield

$$\exp\left(\frac{1}{x}\right) = gt^{1/(B-1/2)}, \tag{2.9}$$

where g is some constant. Next note that the ground-state energy of the two-state model is $E_g = -NJ$ so that by eq. (2.2)

$$E(x) - E_g = \frac{2NJ}{1+\exp(1/x)} \tag{2.10}$$

and hence

$$(E(x) - E_g)\left[\frac{1}{g} + t^{1/(B-1/2)}\right] = h, \tag{2.11}$$

where h is another constant.

Our simple model does not contain any statistical variance, so any power n of eq. (2.11) is also valid. In the long-time limit, where one can neglect $1/g$ in eq. (2.11), this equation becomes identical to eq. (13) in the study of Sibani *et al.* [19],

$$\langle (E - E_g)^n \rangle (1 + t^\alpha)^n = \text{const.} \tag{2.12}$$

which is a consequence of their scaling ansatz for tree structures, and to which they also apply the long-time limit before actually using it to extrapolate to the ground state. However, in their ansatz α is simply a parameter, $0 < \alpha < 1$, whereas our two-state model relates it to the barrier height of the system:

$$\alpha = \frac{1}{B - \frac{1}{2}} \quad \text{or} \quad B = \frac{1}{\alpha} + \frac{1}{2}. \tag{2.13}$$

Thus any α found by the method of Sibani *et al.* can easily be translated into an apparent barrier height. The best fit to the traveling salesman problem they consider is $\alpha = 0.13$ corresponding to $B = 8.2$ which, considering the complexity of the problem, is a large but reasonable barrier height. Since in general there is no lower bound on B (except zero, of course), our model does not restrict α to be less than one.

When contrasted with the constant thermodynamic speed schedule of the comparative study by Salamon *et al.* [4], our two-state model with a barrier height of $B = 1.20$ provides a very accurate fit of the annealing schedule as seen in Fig. 2. Our model's reproduction of Ruppeiner's much slower and more accurate annealing of random graphs [18] is even more accurate and is presented in Fig. 3. The empirical heat capacities for random graphs shown in Fig. 4 have some scatter. However, it should be pointed out that the theoretical curve (calculated from eq. (2.3)) does not contain any free parameters. Furthermore, the numerical study is based on an ensemble of only 500 copies, with limited data points at medium to high temperatures.

Hence, the two-state model yields a good representation of the empirical data and is very cost-effective. Readers with a background in glasses might want to interpret the low-temperature part of the empirical data in Fig. 4 as linear in accordance with refs. [13,14]. Such an interpretation would indicate that the constant level separation J in the current model, eqs. (2.1)–(2.3), is inadequate and should be replaced by a continuous distribution. However, within the accuracy of the empirical data, the theoretical curve of Fig. 4, based on the simple assumption of constant J, provides the best overall fit without any fitting parameters.

Numerically differentiating Ruppeiner's constant thermodynamic speed annealing schedule, one may calculate the apparent relaxation time of the system. Fig. 5 shows an Arrhenius plot of $\ln \varepsilon$ vs. $1/x$. A straight line then indicates a constant barrier height. This is approximately the case, and there seems to be no need for additional barrier heights. The observed barrier height of $B = 1.2$ is, again considering the numerical scatter, relatively close to the barrier height of 0.86 that was found to afford a best fit to the annealing schedule and the heat capacities.

3 Conclusions

The type of thermal model and mathematical analysis used in Section 2 to derive the cooling strategy which minimizes entropy production for a process that is constrained to proceed in a fixed time has been used successfully for determination of optimal heating and cooling strategies for real physical heat transfer problems [20,21]. In simulated annealing problems one exploits the formal analogy between complex mathematical combinatorial problems and statistical mechanics to cast the problem in terms of thermodynamic variables such as temperature, energy, and heat capacity. For systems for which a concrete statistical mechanical model can be formulated, our simple thermodynamic model then permits determination of an *analytic* optimal annealing schedule. Lacking information on the detailed distribution of energy levels, this model cannot be expected to predict all the fundamental properties of all complicated systems such as real glasses and their spin glass models. But it has proven to yield a substantial computational saving for a number of mathematical optimization problems that have been successfully treated by simulated annealing.

It might be argued that, even though one has good physical reason to adopt a particular model, its parameters (J and B in the examples in Section 2) can only be determined *a posteriori*. This is only partly true, because the physical model of a particular optimization problem usually carries with it a good indication of the values of these parameters, or at least their bounds. Since the simple two-state model is used exclusively to determine the annealing schedule, and is not used in the annealing itself, the parameters are usually not critical, and one or two trial runs will quickly provide acceptable values.

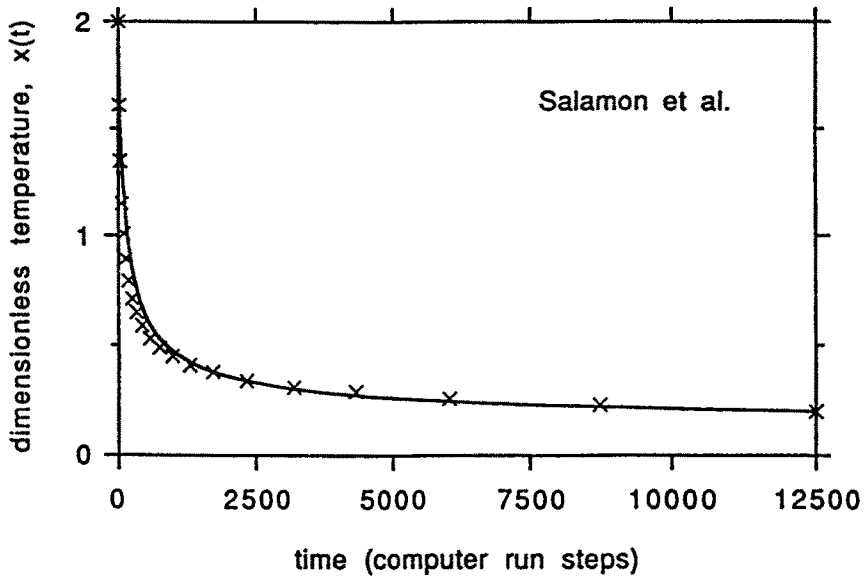

Fig. 2. Optimal annealing schedule $x(t)$ for graph partitioning used by Salamon *et al.* [4] (×) and as derived from the present two-state model with barrier height $B = 1.20$ (—).

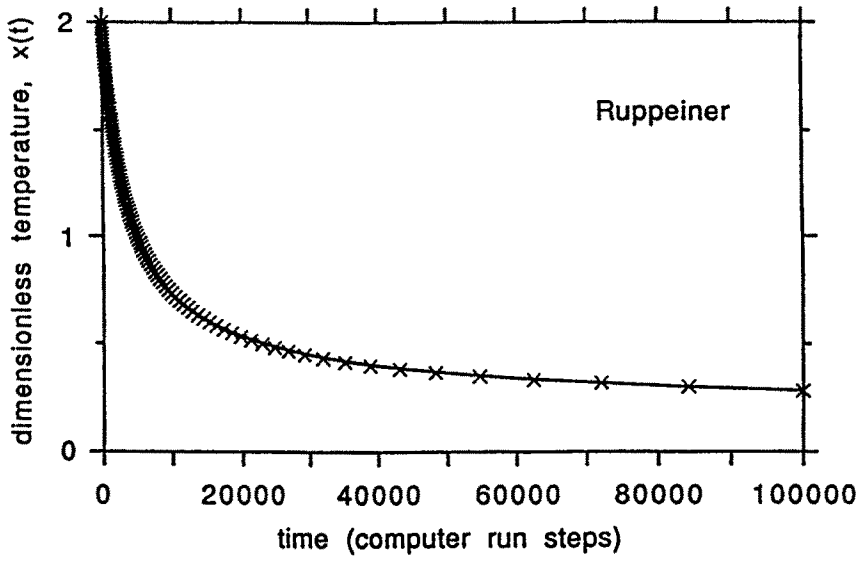

Fig. 3. Optimal annealing schedule $x(t)$ for graph partitioning used by Ruppeiner [18] (×) and as derived from the present two-state model with barrier height $B = 0.86$ (—).

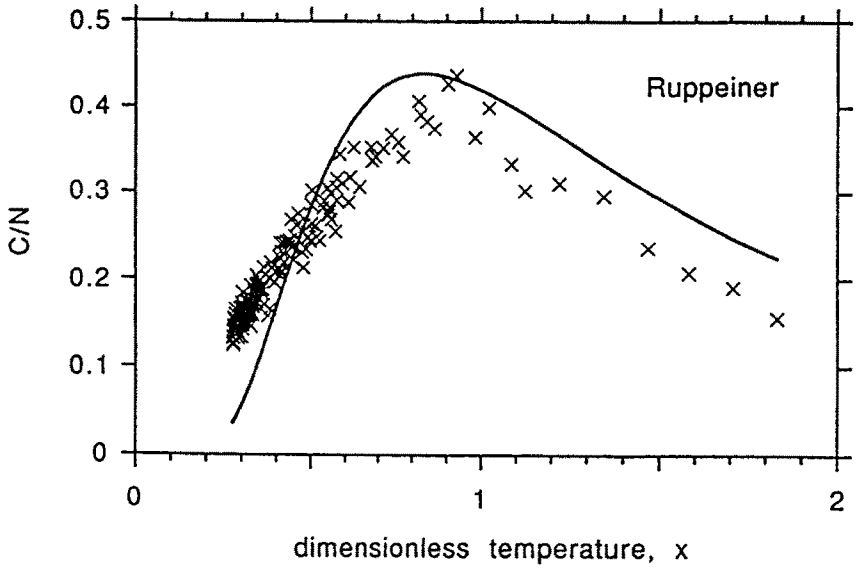

Fig. 4. Dimensionless heat capacity C/N vs. dimensionless temperature x for carefully annealed graph partitioning [18] (×) and in the two-state model without adjustable parameters (—).

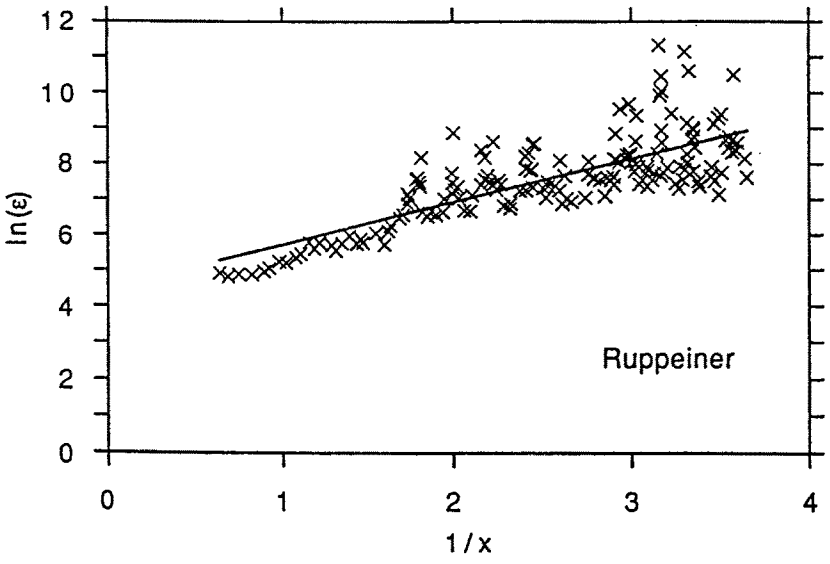

Fig. 5. Arrhenius plot of $\ln \epsilon$ vs. $1/x$ for carefully annealed graph partitioning [18] (×) with the best linear fit (—). A straight line indicates a constant barrier height.

This procedure is *not* proposed as a replacement for the far more general, powerful algorithm known as constant thermodynamic speed annealing which in principle is applicable to any problem, even when simple models for internal system structure cannot be devised. Rather, when such simple models are possible — as appears to be the case for a number of central problems in simulated annealing analyzed to date — the value of the analytic optimal cooling strategy is a substantial reduction in computer time, because the cooling strategy algorithm does not have to constantly update itself via evaluation of the heat capacity and relaxation time as they are changing along the annealing process. Our model thus removes a great deal of the complexity from the simulated annealing procedure.

For simulated annealing problems in graph partitioning, seismic inversion, and spin glasses, for which the simple two-state model appears to be appropriate, we find that both stationary and dynamic properties for the optimal solutions, previously arrived at via lengthy numerical simulations, can be predicted accurately with our approach. This includes analytic determination of the optimal cooling strategy itself.

Our predictions are not exact, because the combinatorial problems themselves may be very similar although not identical to the one-dimensional Ising model. For example, for the graph partitioning [4,7,9] the necklaces considered have either one or three nearest neighbours; the average is two as for the one-dimensional Ising model. Nonetheless our analytic predictions are in excellent agreement with prior numerical simulation results which would indicate that the basic physics of the model has been captured.

The approach presented here could be of general value in simulated annealing problems when some type of relatively simple statistical mechanical model can be crafted for a macroscopically complex system. Optimal cooling strategies can be calculated analytically (although one often must resort to a numerical solution), and stationary and dynamic properties of system energy and heat capacity can be evaluated.

Acknowledgements

We thank professors George Ruppeiner, Peter Salamon, and Paolo Sibani for kindly providing us with their raw data in soft form. B.A. gratefully acknowledges travel and hospitality grants from the Danish Ministry of Education and the Israeli Ministry of Science and Technology. We would also like to acknowledge the stimulating environment of the Telluride Summer Research Center, where part of this work was done.

References

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

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

[3] Lin S., B. W. Kernighan, Oper. Res. **21**, 498 (1973).

[4] Salamon P., J. Nulton, J. R. Harland, J. M. Pedersen, G. Ruppeiner, L. Liao, Comp. Phys. Comm. **49**, 423 (1988).

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

[6] Nulton J., P. Salamon, Phys. Rev. A **31**, 2529 (1985).

[7] Andresen B., K. H. Hoffmann, K. Mosegaard, J. Nulton, J. M. Pedersen, P. Salamon, J. Phys. (France) **49**, 1485 (1988).

[8] Kubo R.: *Statistical Mechanics*, 2nd Ed., North-Holland, Amsterdam, 1967.

[9] Ruppeiner G., J. M. Pedersen, P. Salamon, J. Phys. I **1**, 455 (1991).

[10] Fu Y., P. W. Anderson, J. Phys. A **19**, 1605 (1986).

[11] Grest G. S., C. M. Soukoulis, K. Levin, Phys. Rev. Lett. **56**, 1148 (1986).

[12] Ettelaie R., M. A. Moore, J. Phys. (France) **48**, 1255 (1987).

[13] Anderson P. W., B. I. Halperin, C. M. Varma, Phil. Mag. **25**, 1 (1972).

[14] Philips W. A., J. Low Temp. Phys. **7**, 351 (1972).

[15] Rees S., R. C. Ball, J. Phys. A **20**, 1239 (1987).

[16] Nulton J., B. Andresen, P. Salamon, Combinatorics of necklaces, in preparation.

[17] Geman S., D. Geman, IEEE Trans., PAMI **6**, 721 (1984).

[18] Ruppeiner G, Nucl. Phys. B (Proc. Suppl.) **5A**, 116 (1988).

[19] Sibani P., J. M. Pedersen, K. H. Hoffmann, P. Salamon, Phys. Rev. A **42**, 7080 (1990).

[20] Andresen B., J. M. Gordon, J. Appl. Phys. **71**, 76 (1992).

[21] Andresen B., J. M. Gordon, Int. J. Heat Fluid Flow **13**, 294 (1992).