Monte Carlo dynamics of optimization problems: A scaling description

Paolo Sibani
Fysisk Institut, Odense Universitet, Campusvej 55, DK-5230 Odense M, Denmark

Jacob Mørch Pedersen*
Fysisk Laboratorium, Københavns Universitet, Universitetsparken 5, DK-2100 København Ø, Denmark

Karl Heinz Hoffmann
Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, D-6900 Heidelberg, West Germany

Peter Salamon
Department of Mathematical Sciences, San Diego State University, San Diego, California 92182
(Received 26 May 1989; revised manuscript received 31 May 1990)

We show that some hard optimization problems studied by Monte Carlo methods, such as simulated annealing, have features that can be estimated by a statistical analysis of the data, well before being actually observed. This applies, for instance, to the estimation of the ground-state energy of the problem. We start by showing that the density of states and the distribution of extremes of energy seen in a given time interval in the Monte Carlo dynamics of combinatorial optimization problems are strongly related to each other through the first-passage-time distribution of the stochastic dynamics of the system. We then introduce a scaling ansatz for this last quantity, which allows an estimate of the ground-state energy. Finally, we demonstrate the method on a "traveling-salesman" problem with known ground-state energy and apply it to the simulated annealing of a graph-bipartitioning problem.

I. INTRODUCTION AND DISCUSSION

Many problems arising in statistical mechanics such as finding the ground state of a spin glass[1-3] are hard combinatorial problems. As such, they can be studied by simulated annealing, a technique introduced in recent years[4,5] in which the optimization is considered as a statistical-mechanical problem in its own right. In the annealing approach, the cost function of the problem at hand is viewed as the energy of a stochastic dynamical system, which relaxes according to some Monte Carlo scheme. By lowering the temperature, one confines the system to states of progressively lower energy, eventually obtaining a near-optimal solution.

The basic analogy between combinatorial problems studied by Monte Carlo methods and physics has already been investigated, with respect to the statistical properties of the local minima. In particular, the possibility that these minima might be clustered according to some ultrametric measure, similar to what happens for the Sherrington-Kirkpatrick (SK) model of spin glasses,[6,7] has attracted some attention.

The *dynamics* of physical random systems and of combinatorial problems has important common features as well: In both cases, the configuration spaces have very many local energy minima, and the systems perform random walks with Arrhenius-type transition rates. It is hopeful that a theoretical description of the relaxation of combinatorial optimization problems will make it possible to improve existing simulated annealing methods, as well as yield useful insights on complicated physical models that are too hard to simulate. Also, the study of the dynamics of simulated annealing can help in the design and control of Monte Carlo experiments on highly frustrated physical systems at low temperatures.

In an attempt to define a possible theoretical approach, some questions naturally arise: Why should a characterization of the dynamics of a combinatorial problem be possible at all, how should it be attempted, and how general can it be? In the rest of this paper, we give a tentative answer to the first two questions.

Some qualitative insight can be gained indirectly if we consider why simulated annealing works at all. If the correlation between the energy of configurations that are close to each other in the topology of the move class is totally absent or varies erratically in phase space, it will fail. Consider, for example, searching a configuration space where the energy is mostly constant, except for some deep minima with very small attraction basins; this is the classic "golf hole" problem. No useful information can be collected by sampling energies in a neighborhood of most points. Therefore, annealing and random search will be identical on this problem, and both give poor results. The opposite case would be an energy function shaped as a surface of revolution with one global minimum. Here, the structure of the energy function can be completely determined by looking at a small neighborhood, and a gradient descent algorithm will quickly provide the answer. Simulated annealing with a judicious choice of the move class is empirically known to work rather well on a large class of hard combinatorial prob-

lems, providing fast and accurate answers. The phase space of these problems must then be somehow intermediate between the two examples given above, i.e., it must have some degree of structure, and this structure should reveal itself in the properties of the dynamics. We therefore think that the answer to our first question should be affirmative for a rather large set of problems.

In order to proceed further, we state and discuss the assumptions on which our approach is based, find some of their implications for a measurable quantity in Monte Carlo experiments, and check the results on two different combinatorial problems: an instance of the traveling salesman problem (TSP), for which the ground state is known (for any problem size) and a graph partitioning problem.

The stochastic dynamics of the system can be described by the distribution of the waiting time for reaching a target energy from any given starting energy. The average waiting time of this distribution should, for an NP hard problem, diverge very fast with the size of the problem, at least for some (low) range of the initial and final energies.[19] Otherwise, the average waiting time for reaching the ground state or global minimum of the problem would, in the average, remain reasonably small, which counters our intuition about the nature of hard combinatorial problems. We take, therefore, the point of view that the distribution has no average. As we only consider finite combinatorial problems, this description can only be valid over a finite (but, hopefully, very large) span of time. A further idealization is to consider the first-passage problem as infinitely divisible, which amounts to treating the energy as a continuous variable. This description will fail on some (hopefully, negligibly small) energy scale. Finally, we assume that the stochastic properties of the random walk are scale invariant, which means that the waiting-time distribution for one set of starting and target points is the same as for any other set, except for a scaling factor, which depends on the position of the intervals with respect to the ground state of the problem. As a plausibility argument in support of our assumption, we remark that scaling relations of this type have been previously derived for a class of random walks on trees,[8] and that it has been argued[9] that random walks on trees provide a good coarse-grained description of thermal relaxation in phase spaces with many nested minima. Nevertheless, the validity of the assumptions rests on the agreement between the predicted behavior and what is actually observed in Monte Carlo experiments.

The waiting-time distribution, introduced above, is not a readily available quantity. We show, however, that it is related to a simple and natural measure of how far the annealing has progressed, namely, the least energy seen in the simulation during the time interval $(0,t)$. This stochastic variable, which is easily monitored during a numerical experiment, was first introduced by Jakobsen and co-workers,[10,11] who argued that its average value is a good measure of performance for optimization algorithms, and dubbed it best-so-far energy (E_{BSF}).

The main observation of this paper is that the BSF distribution has, for a nonempty class of problems, quite specific scaling properties, *provided that the appropriate zero of the energy axis is known,* i.e., provided that the ground-state energy E_g is correctly guessed. This makes it possible to estimate the ground-state energy long before it is reached in a simulation, on the basis of run-time information collected during the run.

II. MATHEMATICAL FORMALISM

Consider a numerical simulation of a combinatorial-problem run with just one copy of the system. The dynamics is assumed to be given by the Metropolis algorithm, with a temperature that might depend on the time, thus including the important case of simulated annealing. The BSF distribution of the problem is defined by $F(t,e)=\mathrm{Prob}(E_{\mathrm{BSF}}<E \text{ at time } t)$. It can be empirically sampled by running the dynamics with n identical copies of the system.

In this section, it is shown that $F(t,E)$ can be expressed in terms of the density of states of the combinatorial problem at hand (a "static" quantity) and the previously mentioned density of first-passage times through a given energy (a "dynamic" quantity).

Let $R_{E,E_0}(t)$ be the probability density for the first-passage time through E, with the starting point at $E_0>E$, and $F(t,E|E_0)$ and E_{BSF} distribution conditional to the starting point being E_0. Clearly, one has

$$F(t,E|E_0)=\int_0^t R_{E,E_0}(y)dy=1-W_{E,E_0}(t)\ , \tag{1}$$

where $W_{E,E_0}(t)$ is, by definition, the probability of waiting at least time t before hitting the energy E, averaged over all the possible initial configurations with energy E_0.

The E_{BSF} distribution is obtained by averaging Eq. (1) over the distribution of initial energies. Usually, the initial state is chosen at random, in which case, this distribution is the normalized density of states of the problem $\mathcal{D}(E_0)$. Thus

$$F(t,E)=1-\int_E^\infty W_{E,E_0}(t)\mathcal{D}(E_0)dE_0\ , \tag{2}$$

where we have used the fact that $W_{E,E_0}(t)=0$ for $E<E_0$.

The *ab initio* calculation of the waiting-time probability $W_{E,E_0}(t)$ requires the solution of the full dynamical problem itself. However, some of the properties of W follow from simple probabilistic reasoning and hold for any dynamics. These properties are given below:

$$W_{E,E_0}(0)=1\ , \tag{3}$$

$$W_{E,E_0}(\infty)=0\ , \tag{4}$$

$$\lim_{E\downarrow E_0} W_{E,E_0}(t)=0 \text{ for } t\neq 0\ , \tag{5}$$

$$\lim_{E\downarrow E_g} W_{E,E_0}(t)=1 \text{ for } t\neq\infty\ , \tag{6}$$

$$\lim_{E_0\to\infty} W_{E,E_0}(t)=1 \text{ for } t\neq\infty\ . \tag{7}$$

Here E_g is the ground state, and we have explicitly assumed that there is zero probability of reaching it in finite time. For a finite system, this cannot be strictly true. In

practice, the characteristic time scale, on which $W_{E_g,E}$ approaches zero, will diverge very quickly with the size of the system, unless the problem is trivial. Mathematically, the assumption can be satisfied by taking the limit of an infinite system.

We now suggest that, for sufficiently low E and E_0, $W_{E,E_0}(t)$ depends on a combination of its three arguments, as given in the following scaling ansatz:

$$W_{E,E_0}(t)=W\left[\frac{t}{t_0}\left[\frac{E-E_g}{E_0-E}\right]^{1/\alpha}\right] \tag{8}$$

where W is any monotonically decreasing function, with values between 1 and 0, $0<\alpha<1$, and the constant t_0 is introduced in order to make the argument of W dimensionless. Note that W fulfills all the requirements on the waiting-time distribution stated in Eqs. (3)–(7).

Our formula is supported by the following heuristic considerations:

(1) Changing the distance between the initial and final states amounts to a rescaling of time. This happens also in a quite familiar example, i.e., for the first-passage time through the absorbing boundary at E of a simple one-dimensional diffusion process starting at E_0. In this case $W_{E,E_0}(t)$ just depends on $(E_0-E)/\sqrt{t}$. Here the exponent α is one half.

(2) While the medium is spatially homogeneous in the diffusion example, we have in the combinatorial problem, a privileged origin on the energy axis, namely, the ground-state energy. The factor $E-E_g$ removes the translational invariance and describes a behavior that becomes slower and slower as the ground state is approached. Finally, the idea of a scaling exponent less then one is based on the intuitive idea that the random exploration of the complicated phase space of a combinatorial problem can be coarse grained into a random walk on a tree.[9] The distributions for the tree models[12] are so-called stable distributions and have a long-time tail of the form $t^{-\alpha}$, with $0<\alpha<1$.[13,14]

For notational convenience, we now let $\tau=t/t_0$, and choose the origin of the energy axis as E_g until further notice. We then find the $E_{\rm BSF}$ distribution as

$$F(t,E)=1-\int_E^\infty W\left[\tau\left[\frac{E}{E_0-E}\right]^{1/\alpha}\right]\mathcal{D}(E_0)dE_0 \ . \tag{9}$$

The function W is expected to have an inverted sigmoid shape. We choose the scale of time t_0, such that the transition between the region where W is close to one and the region where it is close to zero occurs when the argument of W is unity.

We then split the above integral into separate integrals over these two regions:

$$F(t,E)=1-\int_0^{E\tau^\alpha} W\left[\tau\left[\frac{E}{y}\right]^{1/\alpha}\right]\mathcal{D}(E+y)dy - \int_{E\tau^\alpha}^\infty W\left[\tau\left[\frac{E}{y}\right]^{1/\alpha}\right]\mathcal{D}(E+y)dy \ . \tag{10}$$

The simplest approximation, and perhaps the only possible one at this level of description, is to neglect the first integral altogether and put $W=1$ in the second one. (This seems to be especially appropriate if the exponent α becomes small, in which case, the transition from the small argument to the large argument region takes place quite abruptly.) One then finds that

$$F(t,E)=\int_0^{E(1+\tau^\alpha)}\mathcal{D}(y)dy \ . \tag{11}$$

The nth moment of the $E_{\rm BSF}$ distribution should now fulfill the relation

$$\langle E^n\rangle(1+\tau^\alpha)^n=\int_0^\infty y^n\mathcal{D}(y)dy \ , \tag{12}$$

or, reintroducing the dependence on the ground-state energy,

$$\langle(E-E_g)^n\rangle(1+\tau^\alpha)^n=\text{const} \ . \tag{13}$$

The constant on the right-hand side of Eq. (13) is formally the nth moment of the normalized density of states. However, since the scaling ansatz used to derive Eq. (13) can, in general, only be expected to apply to the low-energy part of the spectrum, while the right-hand side of the same equation involves the high-energy part of the spectrum as well, especially for high and positive values of n, the formal identity cannot be trusted.

III. METHODS OF DATA ANALYSIS

Equation (13) is assumed to be strictly valid for averages performed over an infinite ensemble. Computational limits introduce therefore various possible sources of discrepancies between the data and the predictions of the theory, and different data can give slightly different answers. In this section, we therefore describe the methods we used to analyze the data presented in the following two sections.

In order to check our ansatz, we have with all our data plotted $\ln\langle(E-E_g)^n\rangle$ versus $\ln t$ for a time span of typically more than six orders of magnitude. In these plots, the data fall very well on straight lines for both the TSP and the graph-partitioning-problem run with constant temperature, except for a well-defined bend at short times (typically, two orders of magnitude), probably due to the fact that the appropriate time variable is $1+(t/t_0)^\alpha$ rather then t^α. In the case of simulated annealing on the graph-partitioning problem, the fit is somehow less satisfactory, and also less sensitive to the value of E_g.

Since scaling is a property of the distribution, rather than of a single moment, we combine the information from several moments in order to assess the quality of the fits. Given a *guessed* value of the ground-state energy E_g', the nth moment on the left-hand side of Eq. (13) can be calculated for any value of n for which the average is expected to exist.

The value of α can then be extracted by a least-squares fit in the $\ln\langle(E-E_g)^n\rangle$ versus $\ln t$ plot.

Let the experimentally determined and E_g'-dependent value of α be denoted by $\alpha_n(E_g')$. It follows from Eq. (13) that, for the true ground state, $n/\alpha_n(E_g)$ should be independent of n, i.e., one should have

$$n/\alpha_n(E_g)=1/\alpha_1(E_g)\ . \tag{14}$$

The best constant through the data $n/\alpha_n(E_g)$ is the arithmetic mean of the points. A relevant measure of the error is then the variance of the data points divided by the mean. We call this quantity $Q(E_g)$. Ideally, Q should vanish for perfect data, if E_g is chosen correctly. In practice, we estimate the ground-state energy as the E_g value that minimizes Q. In the procedure, we have used the first four positive and first four negative moments.

IV. NUMERICAL RESULTS FOR A TSP PROBLEM

As a check, we apply the formalism to a combinatorial problem with known ground-state energy. In this example, the temperature is kept fixed during each run in order to remove the additional complication introduced by the time dependence of the temperature in the annealing schedule. Also, walks at different temperatures probe different parts of the phase space. It is interesting to see explicitly whether they have similar statistical properties, as is assumed in our scaling ansatz, and to check how the predicted ground-state energy varies with the temperature.

Consider an $i\times j$ planar square grid, with unit distance between neighbor points. Place a city on each grid point. A configuration or "tour" is a permutation of the $m=ij$ cities, and its cost or energy function is simply given by the total Euclidean length of the tour, as found by summing the distance from one city to the next along the tour. It is easy to see that the lowest possible energy is $E_g=m$, if m is even, or $E_g=m-1+\sqrt{2}$, if m is odd.

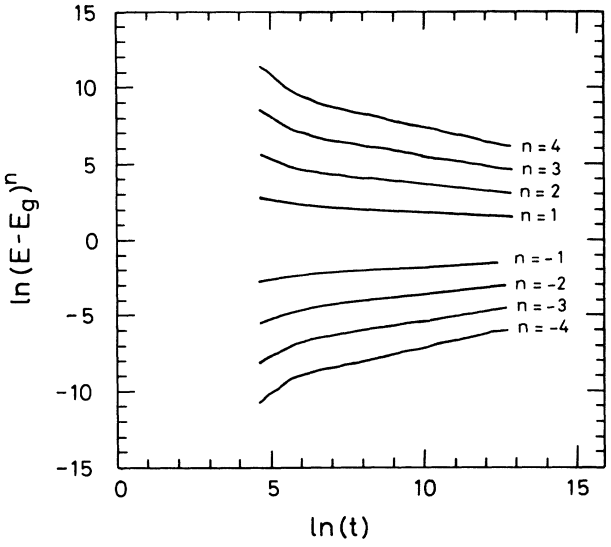

FIG. 1. Plot of $\ln\langle(E-E_g)^n\rangle$ vs $\ln t$ time for a TSP with 33 cities on a regular square lattice with lattice unit 1. The ensemble used has 100 elements, and the simulation was performed at a constant temperature $T=0.8$. The least energy seen in the whole ensemble is about 36, which is quite far from the ground-state energy $E_G=32+\sqrt{2}$. The data show that after a rather short initial transient, the asymptotic scaling form, implied by Eq. (13), is well satisfied for *all* time scales of interest.

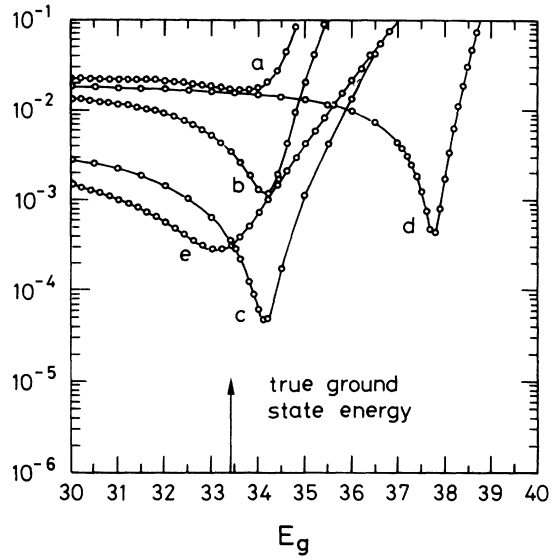

FIG. 2. Plot of figure of merit $Q(E_g)$ as a function of E_g for the TSP system with 33 cities. The actual ground state is $E_g=32+\sqrt{2}$, while the predicted ground-state energy is the minimum of Q. The curves are all calculated for an ensemble of 100 elements, with a constant-temperature simulation. The corresponding temperatures are $T=0.6$, (curve *a*), 0.8 (curve *b*), 1 (curve *c*), 1.2 (curve *d*), and 1.4 (curve *e*). The least energy seen in the ensemble is for curves *a*–*c*, very close to the energy value at which the curve stops. For the other two curves (*d* and *e*), least energy seen was in excess of 40. All of the simulations lasted 400 000 ensemble updates, except for case *e*, which lasted twice as long. Note that the minima are quite sharp, since the vertical scale is logarithmic. The worst deviation between the predicted and the actual ground state is 10%. Also noteworthy is that the quality of the minimum is not monotonously improved as the temperature is lowered. This might be because of the fact that when the temperature is much lower then $T=1$, which is the distance between the cities, then the system is practically quenched, at least on the time scales which are realistic in a simulation.

As a move class, we choose the Lin two bond algorithm,[15] in which two randomly chosen bonds are removed and replaced by two others in any attempted move. The moves are accepted or rejected according to the usual Metropolis algorithm.

We analyzed a small problem with 23 cities, in order to assess the importance of finite-size effects, and a larger problem with 33 cities. In the former case, a sizeable part of the ensemble accumulated very close to the ground state of the system, while, in the latter case, the simulation never came close to the ground state. Nevertheless, the scaling form assumed in Eq. (13) is very well fulfilled in both cases. In Fig. 1 we show the natural logarithm of the first four positive and first four negative moments of the $E_{\rm BSF}$ distribution, as a function of $\ln t$, for the 33-city problem. As anticipated, the curves are perfectly straight lines after a short transient. Very similar curves are obtained for the 23-city case. In both cases, the ensembles contained 100 systems.

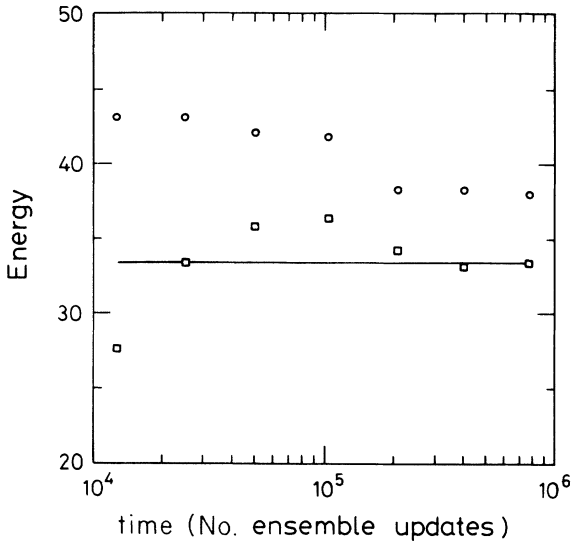

FIG. 3. The TSP ground-state estimates (squares), obtained from the minimum of $Q(E_g)$, plotted as a function of the duration of the simulation, t. The ensemble size was 100, and the temperature $T=1.4$. The estimate converges to the actual ground state of the system, which is indicated by the horizontal line at $E_g=32+\sqrt{2}$. The estimate is already quite good at an early stage in the simulation. However, for very short times, the algorithm fails to produce a minimum in Q. This happens for $t=6022$ (not shown in the plot). The circles show the least energy actually seen in the ensemble during the time t.

While we used the true ground-state energy values to obtain the plots of Fig. 1, we now estimate the ground-state energy following the technique described in Sec. III. We therefore extract the scaling parameter $\alpha_n(E_g')$ for each n in a range of guessed values of E_g in the vicinity of the true ground state, and calculate the figure of merit $Q(E_g')$ as a function of the guessed ground-state energy E_g'. The estimated ground-state energy is then found as the one minimizing Q.

Figure 2 shows how the method works for different simulation temperatures. It is seen that the outcome of the estimation has only a weak temperature dependence. In Fig. 3 we analyze the dependence of the estimate on the simulation length. We see that the estimate has a damped oscillatory behavior, and that it clearly converges to the correct value as the run time is increased. It is also interesting that rather good estimates can be obtained early in the simulation. In the same plot, we also show the very least energy actually seen in the simulation for the whole ensemble. This quantity is consistently much higher then the estimated ground-state energy, showing that the prediction of the ground state works long before the ground state is approached.

V. NUMERICAL RESULTS FOR GRAPH BIPARTITIONING

In this section, we apply our scaling technique to a different combinatorial problem. We consider the partitioning of the nodes of a random graph into two equal-size subsets. The cost function, or energy of the problem, is the number of cut edges. Our choice was motivated by the fact that this is a standard NP-complete problem with important industrial applications.[16]

We consider a variety of graphs, differing in the connectivity and the number of vertices. In all of the cases, we could make an estimate of the ground state according to our method. Here we report a detailed analysis for two instances: one graph, henceforth called A, has 500 vertices and edges that were generated independently with probability 0.004. The second graph, henceforth called B, has 400 vertices and connection probability 0.007.

For graph A we performed both a constant-temperature and a simulated-annealing simulation, while for graph B we only considered the simulated-annealing behavior.

We choose the constant thermodynamic speed annealing schedule described in Refs. 17 and 18. This schedule requires that the annealing be performed with an ensemble of copies of the system, each copy running independently of the others but sharing a common temperature. The temperature schedule is calculated adaptively, on the basis of run-time information extracted from various ensemble averages.[17] In our runs, we had to compromise between ensemble size and length of the simulation. After trying ensemble sizes of 200 and 100, and checking that the results were not affected, we did all the simulations with the smaller ensemble size. The quality of the scaling plots was as good as for the TSP for the constant temperature run, while it was of varying quality for the various annealing runs.

In Fig. 4 we show one scaling plot for graph B (simulated annealing). We used a ground-state energy of 50 in

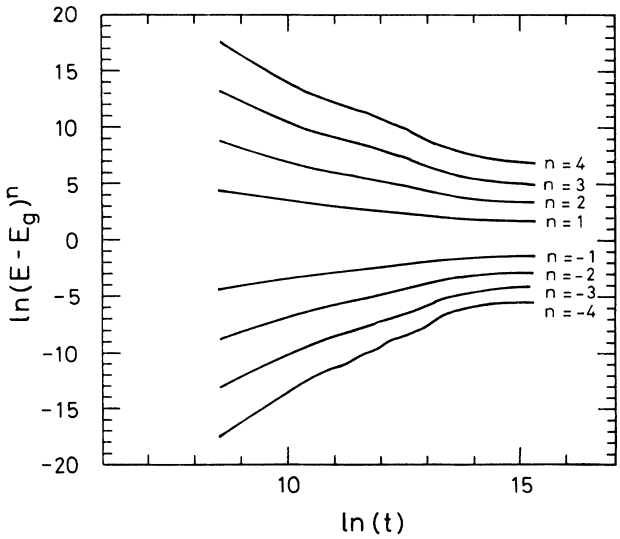

FIG. 4. Plot of $\ln\langle(E-E_g)^n\rangle$ vs $\ln t$ time for a graph-bipartitioning problem. The data stem from a simulated-annealing run. The graph had 400 nodes and a connection probability of 0.007. The curves flatten in the last part of the simulation, showing that there is no further progress in the optimization. Apart from this final regime, the scaling assumption seems to be satisfied over almost four time decades.

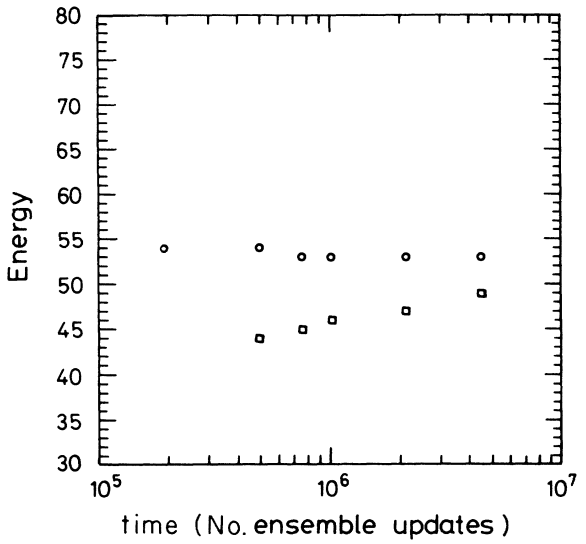

FIG. 5. Corresponds to Fig. 3, but for the graph-partitioning problem described in Fig. 4. The ground-state estimates (squares) and the actually seen least energy are seen to converge nicely toward each other.

this plot, corresponding to our best estimate. The curves flatten in the last part of the simulation, showing that there is no further progress in the optimization. Figure 5 shows how the predicted ground-state energy (squares) depends on the length of the run. As in Fig. 3, we also show the very lowest energy seen in the ensemble (circles). The two data sets are seen to converge toward each other, suggesting that the ground state for this problem should be close to 50. An analytical estimate for the ground-state energy of random graph-partitioning problems, based on statistical considerations only,[3] predicts a value of 27, based on the connectivity and the size of the graph. As in the TSP, the estimate changes slowly with time.

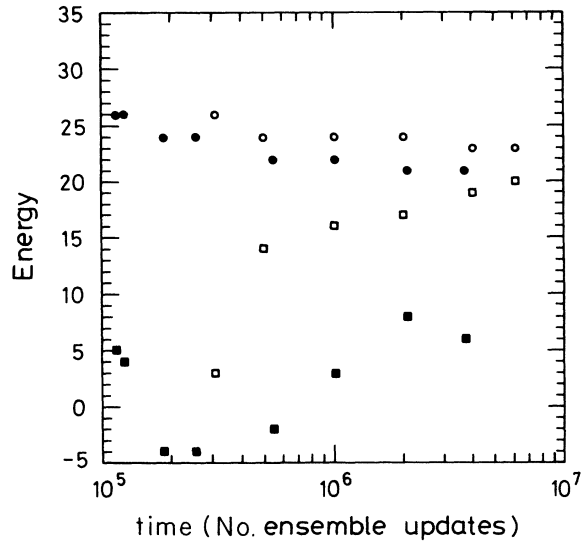

FIG. 6. Ground-state estimate (squares) and the least energy seen over the ensemble (circles) as a function of time. Three points are negative, which should be interpreted as a zero-energy guess. The open circles and squares are from a simulation where the temperature has been kept constant at $T=0.35$. The closed symbols refer to a simulated-annealing run. On the time scale of the simulation, the annealing can find lower-energy states than the constant-temperature run, but performs less well in terms of the ground-state energy prediction.

Figure 6 is analogous to Fig. 5, but shows the data for graph A. Both constant-temperature (open circles and squares) and simulated-annealing data (closed circles and squares) are shown. As one would expect, simulated annealing performs best in terms of the least energy seen. However, the constant-temperature run seems to be better in terms of the ground-state prediction, because the data seem to converge nicely to data for the least energy seen, while the simulated-annealing prediction for this instance is not as yet converging. As already mentioned, this might be due to the fact that the scaling property is better fulfilled for the constant temperature simulations. This particular graph was actually the least convincing from this point of view among the cases we considered.

From Fig. 6 we would expect the ground-state energy to be close to 20, which is much lower than in graph B, due to the fact that the connectivity is much less.

VI. SUMMARY AND CONCLUSIONS

In this paper we attempt a description of the dynamics of stochastic relaxation processes in systems with many local energy minima, such as hard combinatorial problems and frustrated physical systems.

We describe the dynamics in terms of the distribution of a stochastic quantity, the E_{BSF}. The E_{BSF} is the lowest energy seen in the simulation in the time span $(0,t)$, and one samples its distribution by running the Monte Carlo simulation with many identical copies of the system, rather than with a single one. This distribution relaxes toward a δ function at the ground-state energy of the system, as the time goes to infinity. In practice, the ground-state energy is never sampled, except for very small systems, and the relaxation of the distribution toward its theoretical equilibrium limit is very slow, reflecting the fact that the system is trapped in local energy minima.

We argue that, for a nonempty class of problems, the distributions of $E_{\mathrm{BSF}}-E_g$, where E_g is the ground-state energy of the system, behaves asymptotically as a power law with a small exponent α. By checking how well the relation is satisfied for different E_g values, it is possible to attempt an estimate of the ground-state energy of the system. We note that the actually observed scaling exponents have always been much smaller than unity, in agreement with the hypothesis.

For a quantitative discussion, we introduce a figure of merit Q which, for noiseless data, is zero if the scaling relation is perfectly satisfied. We then analyze Q as a function of a putative ground-state energy, and show that Q is small and has a very sharp minimum in two different combinatorial problems, i.e., the previously mentioned TSP and the annealing of a graph-partitioning problem.

For the test problem, the minimum coincides very well with the actual ground-state energy. We conjecture that the minimum of Q could, in general, be taken as a good estimate of the ground-state energy, or at least for the least energy that would be seen in an actual simulation if run for an infinitely long time.

In our first example, the temperature was always kept constant. In the second example, we tried both a constant temperature and an *adaptive* annealing schedule for the temperature decrease. In this schedule, the temperature turns out to have itself an algebraic dependence on time, and the relaxation of the E_{BSF} distribution is still quite well described by an exponent, although the quality of the fits in some examples is less satisfactory than in the constant-temperature case.

In both our examples, the theoretical scaling predictions are found to be rather well fulfilled. In the TSP, the predicted ground-state energy is surprisingly precise. In the other example, we do not have an independent way of checking the correctness of the ground-state estimate, because small examples, where the ground state can actually be found, tend to have trivial behavior. Nevertheless, we can observe that the predicted ground-state energy and the observed E_{BSF} converge toward each other as the running time is increased. We also note that the predicted ground-state energy increases with the connectivity of the graph.

In conclusion we have shown that, in some instances of hard combinatorial problems, it is possible to describe the slow relaxation dynamics by analyzing the statistics of an ensemble of systems. We believe that, neglecting finite-size effects that will never appear in problems of realistic size, the algebraic decay of the E_{BSF} distribution should be a common feature of very many different problems, where the dynamics is dominated by local energy minima.

Our method has some possible practical implications: Since the form of the energy landscape depends on the chosen move class, it is possible to rate different move classes according to how fast a decay (how large an α) they produce. Finally, the predicted value of the ground-state energy can be used for estimating the marginal benefit gained by continuing the optimization. Since the scaling law gives the time needed to approach the goal, one has, in fact, an adaptive stopping criterion. Previous criteria[3] do not take into account the features of the specific instance of the problem being annealed.

ACKNOWLEDGMENTS

We would like to acknowledge the stimulating environment of the Teluride Summer Research Center, where part of this work was done. Part of the code used for the simulations was written in collaboration with George Ruppeiner, New College of the University of South Florida. P. Sibani and J. M. Pedersen would like to thank Det Naturvidenskabelige Forskningsråd for partial travel and computer support.

---

*Present address: Ødegaard & Danneskiold-Samsøe ApS Consulting Engineers, Kroghsgade 1, DK-2100 København Ø, Denmark.

[1]D. Sherrington, and K. Y. M. Wong, J. Phys. A **20**, L785 (1987).

[2]K. Y. M. Wong and D. Sherrington J. Phys. A **20**, L793 (1987).

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

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

[5]V. Černy, J. Opt. Theory Appl. **45**, 41 (1985).

[6]S. Kirkpatrick and G. Toulouse, J. Phys. (Paris) **46**, 1227 (1985).

[7]S. Bacci and N. Parga, J. Phys. A **22**, 3023 (1989).

[8]P. Sibani, Phys. Rev. B **35**, 8572 (1987).

[9]K. H. Hoffmann and P. Sibani, Phys. Rev. A **38**, 4261 (1988).

[10]M. O Jakobson, K. Mosegaard, and J. M. Pedersen, in *Model Optimization in Exploration Geophysics 2,* edited by A. Vogel (Vieweg, Braunschweig, 1988), p. 361.

[11]J. M. Pedersen, K. Mosegaard, M. O. Jakobsen, and P. Salamon (unpublished).

[12]P. Sibani, Phys. Rev. B **34**, 3555 (1986).

[13]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1971), Vol. 2.

[14]W. R. Schneider, in *Stochastic Processes in Classical and Quantum Physics,* Vol. 262 of *Lecture Notes in Physics* (Springer-Verlag, Berlin, 1986), pp. 491–511.

[15]S. Lin, Bell Sys. Tech. J. **44**, 2245 (1965).

[16]D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon (unpublished).

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

[18]G. Ruppeiner, J. M. Pedersen, and P. Salamon (unpublished).

[19]NP hard refers to a category of problems which are believed to be insoluble by algorithms whose run time grows as a polynomial in the size of the problem. See, for example, M. R. Garey and D. S. Johnson, *Computers and Intractability* (Freeman, San Francisco, 1979).