Preferential trapping on energy landscapes in regions containing deep-lying minima: The reason for the success of simulated annealing?

1997 J. Phys. A: Math. Gen. 30 2367

(http://iopscience.iop.org/0305-4470/30/7/018)

J. Phys. A: Math. Gen. **30** (1997) 2367–2389. Printed in the UK PII: S0305-4470(97)77815-6

Preferential trapping on energy landscapes in regions containing deep-lying minima: The reason for the success of simulated annealing?

J Christian Schön

Institut für Anorganische Chemie, Universität Bonn, Gerhard-Domagk-Str. 1, D-53121 Bonn, Germany

Received 13 September 1996, in final form 20 November 1996

Abstract. Evidence has been accumulating that many complex systems are characterized by energy landscapes that contain many pockets with exponentially growing densities of states. Such pockets possess characteristic trapping temperatures, E_i, such that random (Monte Carlo) walkers on such surfaces will be caught in the pockets, if $T < E_i$. It will be shown that the walkers are trapped preferentially in basins containing the deep-lying minima. This observation might serve as an explanation of simulated annealing's success to deliver good suboptima in spite of the use of optimization times far less than those times for which convergence of the algorithm has been proven. Furthermore, preferential trapping might also be involved in certain physical processes, e.g. the glass transition.

1. Introduction

Since its invention by Kirkpatrick *et al* [1] and Czerny [2] simulated annealing has indubitably become one of the most successful global optimization methods, with applications ranging from classical combinatorial optimization problems to the study of complex energy landscapes in physics, chemistry, biology, etc [3]. Nevertheless, this success becomes a major puzzle, when one tries to understand the actual mechanisms of simulated annealing. The heuristic argument usually employed starts with the observation that classical Monte Carlo (MC) simulations using the Metropolis criterion [4] sample the whole state space according to a Boltzmann distribution. Thus, the system will reach the global minimum, once the temperature has been slowly reduced to zero. Following this line of argument, Geman and Geman [5] have derived a temperature schedule that ensures this desired outcome of the optimization:

$$T(t) = A/\ln(t+1) \tag{1}$$

where A is a constant corresponding to the largest relevant energy barrier of the energy landscape. This result has been elaborated further by several authors [6–10] using the theory of (inhomogeneous) Markov chains to obtain necessary and sufficient conditions for the convergence to the Boltzmann distribution as T goes to zero.

However, practically no one uses this kind of schedule in actual optimization problems, since the time required is usually not available. Instead, empirical and semi-empirical schedules are employed [10–20], where the system tends to be quite far from equilibrium, especially at low temperatures. Thus, ergodicity (on time scales less than the available

0305-4470/97/072367+23$19.50

computation time) is broken below some critical temperature(s), and the system is trapped in some local basin of the energy landscape, which cannot be left within the remaining computation time. But in spite of this breakdown of the basic assumptions that underlie the standard view of why simulated annealing is successful, the optimizations result in very convincing suboptima. It would therefore be tremendously helpful for the design of efficient temperature schedules to gain information about how typical features of systems of interest are related to the performance of the algorithm.

Certain processes must clearly take place that channel the walker on the energy landscape of typical systems into the right basins, i.e. the ones containing the deepest minima. One possible attempt at a simple explanation might start with the assumption that there exists a single critical temperature, T_c, analogous to a glass transition temperature. Below this temperature, ergodicity is broken and the walker will be caught in the basin that possesses the highest number of states on the current 'energy level' of the walker. If now the basins with the largest rims (surfaces) were also the deepest ones ('large-rims-have-deep-wells'†), the walker would, with high probability, end up in the 'right' basin. While both the selection according to rim size and the 'large-rims-have-deep-wells' hypothesis are assumptions that are too severe to be fulfilled by general or even typical energy landscapes, they point to the need for further investigations of the mechanisms of trapping and breaking of ergodicity, together with the identification of characteristic features of complex energy landscapes.

Following this general line of argument, we propose in this paper that the cause of simulated annealing's success might be found in certain characteristic properties of the energy landscape itself. Evidence has been accumulating over the past few years that many complex systems showing broken ergodicity possess multiminima energy landscapes, where the local density of states near deep-lying minima grows exponentially [26–28]. As we will show, this exponential growth can lead to a selection towards very deep-lying minima during the optimization, i.e. the random walker gets preferentially trapped in basins that contain the deepest local minima, even though the number of surface states might favour some other basin.

In section 2, exponentially growing densities of states and their importance for trapping are discussed. Section 3 describes how the selection biased towards deep basins takes place and analyses the competition between two pockets. Finally, these results are applied to the annealing of multipocket landscapes, followed by a summary and a discussion of possible implications of preferential trapping in physical systems.

2. Exponentially growing densities of states

In the past four years, a number of multiminima systems have been investigated using the so-called lid [26] or threshold algorithms [29], which allow the complete or statistical enumeration of all states within a pocket of phase space close to a deep-lying minimum (only states that can be reached from this minimum without crossing a prescribed lid are considered). All those among the systems studied, which are classically expected to show

† This phrase is an inversion of the so-called 'deep-wells-have-large-rims' hypothesis [21], which was used to indicate that energy landscapes with singular holes (e.g. the (in)famous hole-on-a-golf-course problems [22–24]) should be excluded from the analysis. Called originally the 'inverse isoperimetric inequality', it states that the depth of a basin should be a monotonic function of the minimal 'diameter' and thus result in a lower bound on the size of the basin's rim as a function of the depth. It has been shown by Salamon and co-workers [25] that the inverse isoperimetric inequality holds true for $\mathbb{R}^n$ and some NP-complete problems (travelling-salesman problem, graph partitioning). The additional assumption that the deepest wells might actually have the largest rims has developed later in trying to understand the mechanism of simulated annealing.

broken ergodicity and some kind of 'glass transition'—the travelling salesman problem [26], spin glasses [27], and a lattice approximation to a real glass [28]—exhibited exponential growth of the local density of states near the deepest minima:

$$g_i(E) = C_i \exp((E - E_i^0)/E_i) \qquad \text{for } E \in [E_i^0, E_i^t] \tag{2}$$

where E_i^0 is the energy at the bottom of the basin and $D_i = E_i^t - E_i^0$ its depth. If one considers such a pocket in isolation, one can easily calculate its statistical and thermodynamic properties. Of special interest in the following are the expectation values of the energy and its square within basin i:

$$\langle E\rangle_i = \frac{\frac{\exp(\alpha E_i^t)}{\alpha}\left(E_i^t - \frac{1}{\alpha}\right) - \frac{\exp(\alpha E_i^0)}{\alpha}\left(E_i^0 - \frac{1}{\alpha}\right)}{\frac{\exp(\alpha E_i^t)}{\alpha} - \frac{\exp(\alpha E_i^0)}{\alpha}} \tag{3}$$

and

$$\langle E^2\rangle_i = \frac{\left[\frac{2}{\alpha^2} - \frac{2E_i^t}{\alpha} + (E_i^t)^2\right]\exp(\alpha E_i^t) - \left[\frac{2}{\alpha^2} - \frac{2E_i^0}{\alpha} + (E_i^0)^2\right]\exp(\alpha E_i^0)}{\frac{\exp(\alpha E_i^t)}{\alpha} - \frac{\exp(\alpha E_i^0)}{\alpha}}. \tag{4}$$

Here, $\alpha = (1/E_i - 1/T)$. We notice that for $T \gg E_i$,

$$\langle E\rangle_i \approx \frac{[E_i^t - E_i]\exp\left(\frac{E_i^t}{E_i}\right) - [E_i^0 - E_i]\exp\left(\frac{E_i^0}{E_i}\right)}{\exp\left(\frac{E_i^t}{E_i}\right) - \exp\left(\frac{E_i^0}{E_i}\right)} \tag{5a}$$

which for $D_i \gg E_i$ becomes $\langle E\rangle_i \approx E_i^t - E_i$. In the same limits,

$$\langle E^2\rangle_i \approx E_i^2 + (E_i^t - E_i)^2 \tag{5b}$$

and thus

$$\sigma_i = \sqrt{\langle E^2\rangle_i - \langle E\rangle_i^2} \approx E_i. \tag{5c}$$

Similarly, we find for $T \ll E_i$,

$$\langle E\rangle_i \approx E_i^0 + T \tag{6a}$$

and

$$\langle E^2\rangle_i \approx T^2 + (E_i^0 + T)^2 \Rightarrow \sigma_i \approx T. \tag{6b}$$

Finally, for $T = E_i (\Leftrightarrow \alpha = 0)$,

$$\langle E\rangle_i = \frac{E_i^t + E_i^0}{2} \tag{7a}$$

and

$$\langle E^2\rangle_i = \frac{(E_i^t)^2 + (E_i^0)^2 + E_i^t E_i^0}{3} \Rightarrow \sigma_i = \frac{1}{\sqrt{12}}(E_i^t - E_i^0). \tag{7b}$$

As an illustration, we have plotted $\langle E\rangle_i$ in figure 1 ($C_i = 1$, $E_i^t = 10$, $E_i^0 = 0$, $E_i = 1$).

At $T = E_i$, some phenomenon akin to a phase transition occurs [26]. For $T > E_i$, the pocket is nearly invisible (independent of its depth!), while for $T < E_i$, the barrier $D_i \gg T$ needs to be crossed in order to leave the pocket at all, where the time scale of escape is now controlled by an Arrhenius factor $\exp(-D_i/T)$†. Thus for temperatures below E_i, the

† If $D_i \approx E_i$, the pocket is usually irrelevant, since it contains only about C_i states. For very large values of $D_i (\approx E_i)$, we are dealing with a very steep but extremely narrow hole—a situation analogous to the golf-hole problem, which is a counter example to essentially all types of optimization techniques that aim to be an improvement over random search.

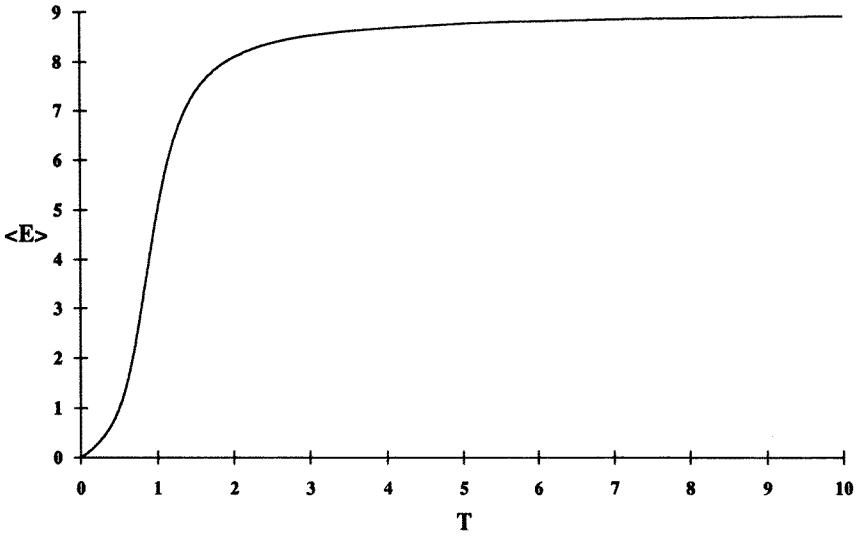

Figure 1. Plot of $\langle E\rangle_i$ as a function of temperature, for a pocket with an exponentially growing density of states ($C_i = 1$, $E_i^t = 10$, $E_i^0 = 0$, $E_i = 1$).

walker is essentially trapped in the pocket: the ergodicity of the system is broken on the time scale of the optimization for $T < E_i$. Note that in the limit $D_i \to \infty$, $\langle E\rangle_i$ exhibits a discontinuous jump at the trapping temperature E_i.

The concept of exponentially growing densities of states is a familiar one in the context of (doped) amorphous semiconductors [30]. There, the so-called band tails appear to exhibit an exponential growth of the electronic density of states within the band gap up to the conduction band for example [31, 32]. The dispersive charge transport in such semiconductors [33] has been modelled successfully by Scher and Montroll [34] as a trapping and emission process of the charge carriers, using a power-law distribution for the hopping times. It has been shown that such a distribution can be derived from the assumption of an exponential distribution of traps [35–37]. To avoid possible confusion with this kind of 'exponential trapping', we would like to point out certain differences to the trapping of random walkers on energy landscapes that we discuss in this paper.

(i) The exponential growth in the semiconductor refers to the global density of states. In general, these states will not be neighbours on the energy landscape (except through quantum mechanical tunnelling), since they tend to be spatially localized (in $\mathbb{R}^3$). Thus each such state is essentially a trap in its own right, with an associated local density of states, probably somewhat similar to the one of an isolated hydrogen-like atom, i.e. the local density of states does not grow exponentially†.

(ii) Consequently, the reason for trapping in semiconductors is *not* some critical temperature, T_c, connected to the exponential growth in the local density of states, and the trap is *not* invisible above T_c, independent of its depth. Instead, the depth, D, of the trap is large compared to the temperature of the system, leading to an Arrhenius factor $\exp(-D/T)$ that controls the escape out of the trap, i.e. the trap is always present and noticeable, but it will only capture the carriers for $T < D$.

† Often, these traps may be described as small localized polarons or hydrogen-like states with a large dielectric constant.

3. Competition between two exponential traps

3.1. Two traps joining at their rims without additional barriers

In order to analyse a system consisting of many pockets with different depths, E_i^0, and trapping temperatures, E_i, we first consider a double-basin landscape with two basins characterized by E_i^0, E_i^t, E_i, C_i, $i = 1, 2$ ($E_1 < E_2$, for definiteness). Let us assume that the prefactors are about equal, $C_1 \approx C_2$. This assumption is not critical, since we are dealing with exponentially growing densities of states. Furthermore, $E_1^t = E_2^t = E^t$ ($= 0$, for definiteness), and there should be no additional barriers separating the two pockets (figure 2). The case $E_1^t \neq E_2^t$ will be discussed in section 3.2.

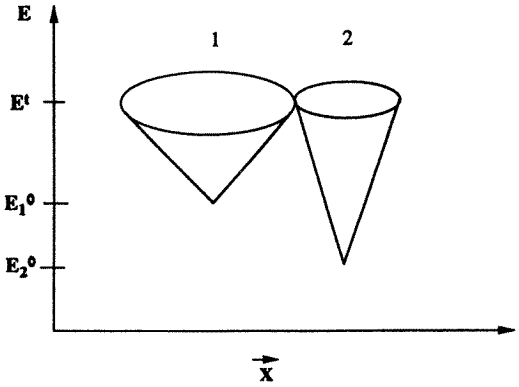

Figure 2. Sketch of the energy landscape, $E(\vec{x})$, of two pockets characterized by E_i^0, E_i^t, E_i, $C_i(= 1)$, $i = 1, 2$ with exponentially growing densities of states joined at their rim ($E_1^t = E_2^t$). The growth factors, $E_i(E_1 < E_2)$, are indicated by the slopes of the cones.

For $T > E_2 > E_1$, both basins will be invisible, and the system will approximately be in equilibrium, i.e. the walker samples the available phase space according to a Boltzmann distribution. This will change at $T = E_2$, since now the second basin can trap the walker. If the walker has not been trapped until T reaches E_1, however, it will end up in basin 1 instead. It appears reasonable to fix the 'critical' moment at $T = E_2$, since the trapping power of basin 1 will increase as T is lowered towards E_1 (clearly, some additional flow of probability will occur from basin 1 into basin 2 while $T \in [E_1, E_2]$, for a somewhat more detailed discussion of this issue, we refer to section 3.4). Thus, we will calculate the expected probabilities to be in basins 1 and 2, p_1 and p_2, respectively, assuming that the system is in thermodynamic equilibrium at $T = E_2$:

$$A = p_2(T = E_2) = \frac{D_2 C_2}{Z} \exp\left(\frac{D_2}{E_2}\right) = \frac{D_2}{Z} g_2^s \tag{8a}$$

and

$$B = p_1(T = E_2) = \frac{E_1 E_2}{E_2 - E_1} \frac{C_1}{Z} \exp\left(\frac{D_1}{E_1}\right) \left[1 - \exp\left(\frac{D_1(E_2 - E_1)}{E_1 E_2}\right)\right]$$
$$\approx \frac{E_1 C_1}{Z} \exp\left(\frac{D_1}{E_1}\right) = \frac{E_1}{Z} g_1^s \tag{8b}$$

where Z is the sum over states and acts as a normalization factor. Here g_i^s is the density of states at the surface of basin i (recall that $E_i^t = 0$, by definition). Clearly, if $A \gg B$, then

the walker is trapped in basin 2, while if $A \ll B$, the walker remains 'free' and will end up in basin 1 when $T = E_1$. We see that the ratio of surface states still enters, but it has to be compared to the ratio of the depth of the more slowly growing basin to the growth factor of the fast-growing one.

The condition $A \lessgtr B$ can be recast by taking the logarithm (for the purpose of the qualitative analysis, we set $C_1 = C_2$, thus $\ln(C_1/C_2) = 0$). Defining

$$f = D_1/E_1 - D_2/E_2 = ((1+a) - 1/(1+b))(D_2/E_1) = m(D_2/E_1) \quad (9a)$$

where $a = (D_1 - D_2)/D_2$ and $b = (E_2 - E_1)/E_1$, and

$$g = \ln(D_2/E_1) \quad (9b)$$

we get the condition $f \lessgtr g$ for trapping into basins 1 and 2, respectively. By assumption, $b > 0$ and $-1 < a < \infty$; and $a \lessgtr 0$ is equivalent to $D_1 \lessgtr D_2$. In figure 3, we show for three different values of D_2/E_1 the curve $f = g$ in the (a, b)-plane. The full curve is for a 'typical' value†, $D_2/E_1 = 10$, while the broken curves are the extreme locations of this curve, $D_2/E_1 = e$ (upper curve) and $D_2/E_1 = 1, \infty$ (lower curve).

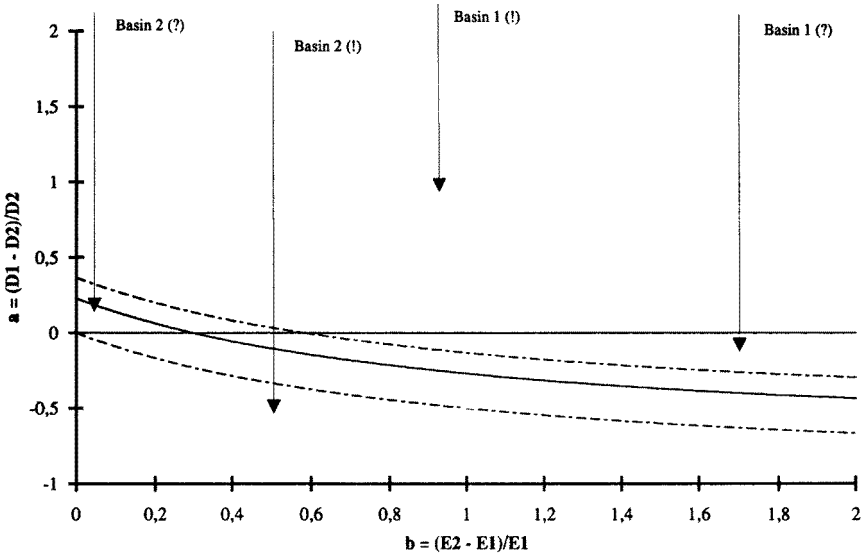

Figure 3. Plot of $f = g$ in the (a, b)-plane, with $f = D_1/E_1 - D_2/E_2 = ((1+a) - 1/(1+b))(D_2/E_1)$, equation (9*a*), and $g = \ln(D_2/E_1)$, equation (9*b*), for several different values of $D_2/E_1 = 1, \infty$ (lower broken curve), $D_2/E_1 = e$ (upper broken curve), and $D_2/E_1 = 10$ (full curve). For a given value of D_2/E_1, trapping into basins 1 and 2 occurs above and below the corresponding line, respectively. Exclamation (question) marks indicate whether this constitutes a positive (negative) outcome of the annealing.

We have marked the basins where the walker ends up with (!) and (?), in order to indicate whether this outcome is the desired one or not, respectively. We see that the system chooses the wrong basin for small values of b and small positive values of a, and larger values of b and small negative values of a. Note that if $b \to \infty$ and $D_1 \to 0$ ($a \to -1$), we are dealing with the 'golf-hole' problem.

† Judging by the growth factors in the TSP problem [26] and the network glasses [28] for example, typical values of D_2/E_1 appear to be in the range 10–100.

Finally, we consider the case $E_1 = E_2 = E'$, and $D_2 > D_1$. Now, we find for the expected probabilities:

$$A = p_2(T = E') = \frac{D_2 C_2}{Z} \exp\left(\frac{D_2}{E'}\right) = \frac{D_2}{Z} g_2^s \tag{10a}$$

and

$$B = p_1(T = E') = \frac{D_1 C_1}{Z} \exp\left(\frac{D_1}{E'}\right) = \frac{D_1}{Z} g_1^s. \tag{10b}$$

Clearly, the walker will be trapped in the deeper of the two basins (basin 2). Since only one trapping temperature, E', exists (and $E_1^t = E_2^t$), the trapping power of the basins is completely controlled by the number of surface states. Note that due to $E_1 = E_2$ and $E_1^t = E_2^t$, the 'large-rims-have-deep-wells' hypothesis holds in this situation.

3.2. The influence of barriers ($E_1^t \neq E_2^t$)

We now turn to the case that $E_1^t \neq E_2^t$, and to the possibility that there might exist an additional (effective) energy barrier between the pockets. It is important to realize that for $T > E_{1,2}$, these effective barriers refer to the energy differences between the top of the exponential pockets and the saddle connecting the two regions that contain these pockets, not to the differences between the saddle and the bottom of the basins. For simplicity, we set the energy of the saddle point(s) to zero ($\Rightarrow$ the heights of the two barriers become $-E_1^t$ and $-E_2^t$). Furthermore, we assume that the subexponential growth (e.g. a power-law growth) in the part of each of the two separate regions that lies between the surface of the actual basin and the saddle be such that the expectation value of the energy for each region (local equilibration) at $T \gg E_1, E_2$ is given by $\langle E \rangle_{1,2} \approx E_{1,2}^t + O(T)$. This assumption is fulfilled in many standard systems.

Clearly, as long as $T > -E_{1,2}^t$, the two regions will be in equilibrium, and we find for the ratio of the probabilities to be in the regions p_i, approximately:

$$p_1/p_2(T) \approx (N_1^s/N_2^s) \exp(D^t/T) \tag{11}$$

where $N_i^s \propto g_i^s \propto \exp(D_i/E_i)$ is the number of states at the surface of basin i, and $D^t = E_1^t - E_2^t$ is the difference between the two barriers.

If the temperature, T, falls substantially below the barrier heights, $-E_{1,2}^t$, the walker will be unable to cross the barrier in either direction, and will thus end up in the basin at the bottom of its current region. We are, therefore, dealing with a barrier crossing problem, and the choice of the walker will be determined by the ratio of the number of states, N_1^s/N_2^s, and some Boltzmann factor, $W = \exp(D^t/T_{\text{eff}})$, where T_{eff} is the freeze-in temperature relevant for a given situation. To estimate T_{eff}, we can define an 'Arrhenius' temperature $T_{\max} = \max(-E_1^t, -E_2^t)$, above which the equilibrium formula (equation (11)), can be assumed to hold. Furthermore, we can define a cut-off temperature $T_{\min} = \min(-E_1^t, -E_2^t)$, below which no communication between the pockets takes place. But for $T \in [T_{\min}, T_{\max}]$, there still exists some flow of probability from the high-lying basin to the lower one, possibly even a rather substantial one if $T_{\min}$ is very small and the number of relaxation steps per temperature, N_{steps}, is large. Thus, the exact value of T_{eff} is difficult to choose, similar to the case $T \in [E_1, E_2]$ in the previous subsection. In addition, the temperature, where ergodicity is broken, depends on the time available for equilibration, of course. In section 3.4, we will discuss this issue in somewhat more detail. But, as we will see in the following discussion, the actual choice of T_{eff} is not critical in a qualitative sense, since it should definitely lie

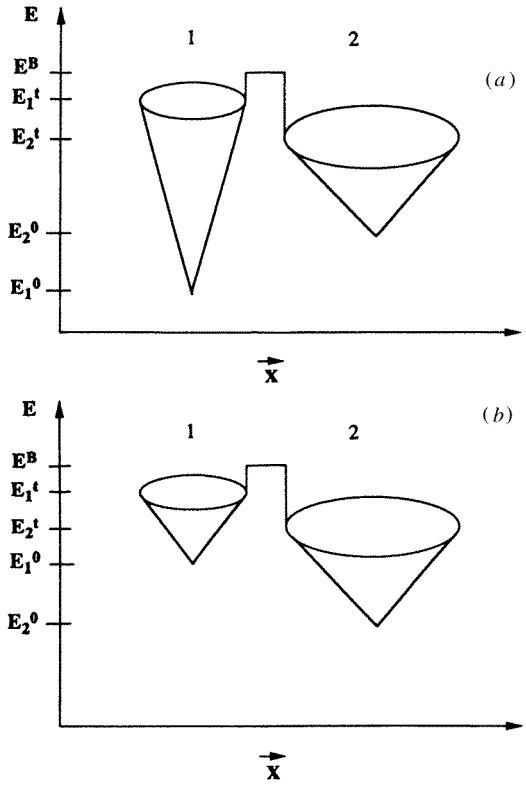

Figure 4. Sketch of the energy landscape, $E(\vec{x})$, of two pockets characterized by E_i^0, $E_i^t (D_i = E_i^t - E_i^0)$, E_i, $C_i (= 1)$, $i = 1, 2$ with exponentially growing densities of states joined at a saddle point at energy $E_B (E_1^t > E_2^t, D^t = E_1^t - E_2^t)$. The growth factors, E_i, are indicated by the slopes of the cones. (*a*) $D_1 > D_2 + D^t$; (*b*) $D_1 \leqslant D_2$.

between $T_{\max}$ and $T_{\min}$. We can distinguish three possible situations, (i) $T_{\text{eff}} < E_{1,2}$, (ii) $-E_{1,2}^t$, $T_{\text{eff}} \gg E_{1,2}$, and (iii) $-E_1^t \leqslant E_{1,2} \ll -E_2^t$.

(i) $T_{\text{eff}} < E_{1,2}$. Here, the barrier is irrelevant, since trapping occurs before the barrier can have an influence, and the results of the previous subsection apply.

(ii) $-E_{1,2}^t$, $T_{\text{eff}} \gg E_{1,2}$. The exponential trapping mechanism described in the previous subsection does not apply, and only the states on the surface of each basin are of importance, according to equation (11). But, because of the exponential growth within each pocket, the Boltzmann factor does not dominate the ratio p_1/p_2 in equation (11). For definiteness, let us take $E_1^t > E_2^t$.

First, we consider case (iia): $D_1 > D_2 + D^t$ (figure 4(*a*)). If now $E_1 < E_2$ or $E_1 \approx E_2$, then $D_1/E_1 - D_2/E_2 - D^t/T_{\text{eff}} > 0$. Thus it follows from equation (11), that the walker will end up in basin 1, as desired. If on the other hand $E_1 \gg E_2$, the combination of the larger number of surface states in basin 2 together with the Boltzmann factor wins, and the walker ends up in the 'wrong' basin. Note that the latter situation is similar to a deep golf hole on top of a mountain†, a situation extremely unfavourable to any optimization algorithm.

Next, there is case (iib): $D_1 \leqslant D_2$ (figure 4(*b*)). If now $E_1 > E_2$, $E_1 \approx E_2$, then

† Popularly called 'Stillinger's nightmare' [38].

$D_1/E_1 - D_2/E_2 - D^t/T_{\rm eff} < 0$. Therefore, the walker ends up in basin 2, which again is the desired one. If on the other hand $E_1 \ll E_2$, the right basin (basin 2) is not found, but the Boltzmann factor succeeds in negating some of the effect of the overwhelmingly large value for N_1^s. This is similar to the usual golf-hole problem discussed in section 3.1.

Finally, we have $D_1 \approx D_2 + D^t$. This case need not be considered, since both basins have approximately the same minimum value for the energy, and thus either choice is satisfactory for our purpose.

The third possibility is case (iii), $-E_1^t \leqslant E_{1,2} \ll -E_2^t$. Now, the exponential trapping of basin 1 becomes relevant, before the temperature $T_{\rm min}(= -E_1^t)$ has been reached. It appears reasonable to use $T_{\rm eff} \approx E_1$ in the expression for W. The discussion of the different cases is analogous to the one for case (ii). The results of this section are summarized in table 1.

Table 1. Most likely final locations of the simulated annealing walker for different combinations of effective energy barriers, $-E_{1,2}^t (D^t = -E_2^t + E_1^t > 0)$, and freeze-in temperatures, $T_{\rm eff}$, depths of the exponential part of the pockets, $D_{1,2}$, and growth factors (trapping temperatures), $E_{1,2}$. Exclamation (question) marks indicate a positive (negative) outcome of the annealing.

		$-E_{1,2}^t$, $T_{\rm eff} \gg E_{1,2}$	$E_1^t \leqslant E_{1,2} \ll -E_2^t$
$D_1 > D_2 + D^t$	$E_1 < E_2$	basin 1 (!)	basin 1 (!)
	$E_1 \approx E_2$	basin 1 (!)	basin 1 (!)
	$E_1 \gg E_2$	basin 2 (?)	basin 2 (?)
$D_1 \leqslant D_2$	$E_1 > E_2$	basin 2 (!)	basin 2 (!)
	$E_1 \approx E_2$	basin 2 (!)	basin 2 (!)
	$E_1 \ll E_2$	basin 1 (?)	basin 1 (?)

In conclusion, we see that the presence of high-energy barriers has the effect of mostly eliminating the exponential trapping: the decision whether the walker is caught in the basin with the deeper-lying surface takes place before the trap becomes active. However, the fact that a large part of each pocket exhibits exponential growth does strongly influence both the value of the effective freeze-in temperature and the frozen-in probability distributions at the moment $T_{\rm eff}$ is reached. Since $T_{\rm eff}$ is larger than the trapping temperatures ($T_{\rm eff} \geqslant E_{1,2}$), the exponential growth of the traps ($N_i^s \propto \exp(D_i/E_i)$) can therefore often compensate an unfavourable Boltzmann factor ($W = \exp(D^t/T_{\rm eff})$), in the sense that the probability of the walker to be in the region of the deeper pocket at the moment when the trapping temperature is finally reached is still very high (cf the discussion of case (iib) for $E_1 \approx E_2$).

Furthermore, the height of the barriers is measured from the top of the exponential pockets, leading to a reduction of the effective barrier height (and thus a decrease of $T_{\rm eff}$) compared with the full distance from the bottom of the basin to the saddle point. In many situations the resulting effective barrier might be even small enough to be made irrelevant by a computationally feasible increase in $N_{\rm steps}$ (cf figures 6 and 7 in section 3.3). Thus, due to the exponential growth of the pockets, the presence of energetic barriers does not completely dominate the final outcome of the annealing.

3.3. Simple illustrative examples

To illustrate the trapping mechanism, let us consider several examples of two pockets joined at their rims, both with and without barriers. We use a discrete model where the (continuous) landscapes are represented as tree graphs, with all the states within one energy layer of a given pocket lumped together into a single node. For simplicity, we will choose $C_i = 1$ in all examples. Furthermore, we assume that the width of the layers is large enough that

the connections between the nodes can be restricted to those nodes that are neighbours in energy within each pocket. We can model the 'annealing' of such a system by constructing a Markov matrix, $\mathbf{M}(T)$, for the transition probabilities, using a procedure appropriate for such lumped tree graphs that has been described elsewhere [29]. Applying $\mathbf{M}(T)$ to the initial probability distribution (probability equals one in the top state, zero elsewhere), we get the probability distribution as a function of temperature during the annealing run.

In the first example, pocket 1 was represented by 12 nodes at energy levels $(0, -1, -2, \ldots, -11)$, and pocket 2 by six nodes at energies $(0, -1, \ldots, -5)$, as seen in figure 5(a). The top node, where the two basins joined, was given 200 000 states at the energy level $+0.1$. The density of states within each pocket was characterized by ($E_1^t = 0$, $E_1^0 = -11$, $E_1 = 2/\ln(5) \approx 1.24$) and ($E_2^t = 0$, $E_2^0 = -5$, $E_2 = 1/\ln(10) \approx 0.43$), respectively.

In order to be somewhat realistic we have chosen a relatively fast exponential temperature schedule, $T = T_{\text{init}}(f)^n$, where $f = 0.9$, $T_{\text{init}} = 5.0$ and $n = 30 (\Rightarrow T(30) \approx 0.21)$. At each temperature, $\mathbf{M}(T)$ was applied 2^7 times, in order to allow for equilibration

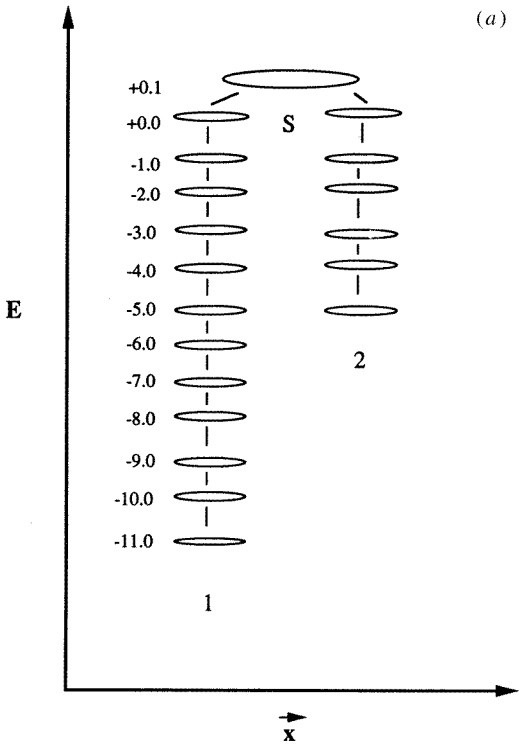

Figure 5. Fast annealing of two pockets with ($E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24$) and ($E_2^t = 0, E_2^0 = -5, E_2 = 1/\ln(10) \approx 0.43$), respectively. ($a$) Sketch of the tree graph used for the representation of the two pockets (1 and 2) and the saddle (S). Energies of the nodes are indicated; weights can be deduced from the definition of the pockets. (b) Plot of $T(n)$ (full squares), $E_{\text{av}}(n)$ (open squares), $E_{\text{av}}^{\text{eq}}(n)$ (full diamonds) and $z(n) = p_{\text{top}}(n)/p_{\text{top}}^{\text{eq}}(n)$ (open diamonds; $z = 1$: broken curve) as a function of temperature updates n. (c) Plot of $p_i(T(n))$ (diamonds), $p_i^{\text{eq}}(T(n))$ (squares), and the probability to end up in basin i upon quenching $p_i^{\text{qu}}(T(n))$ (triangles) as a function of temperature T (full symbols = minimum 1, open symbols = minimum 2).

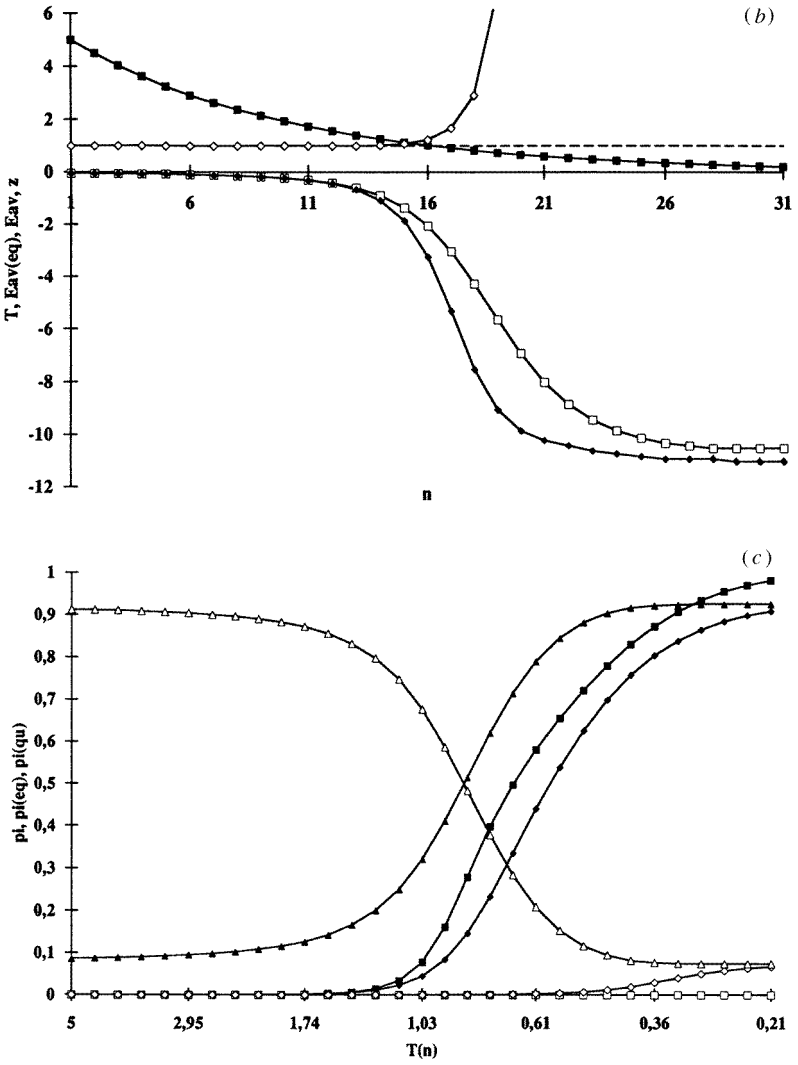

Figure 5. (Continued)

at $T > E_1 = 1.24$, i.e. the system could remain in equilibrium until trapping occurred.

The results are shown in figure 5. In figure 5(*b*), we have plotted $T(n)$, $E_{\mathrm{av}}(n)$, $E_{\mathrm{av}}^{eq}(n)$ and $z(n)$. Here, $z(n) = p_{\mathrm{top}}(n)/p_{\mathrm{top}}^{\mathrm{eq}}(n)$ is a measure for the degree to which the total system is in equilibrium [26, 29], where $p_{\mathrm{top}}(n)$ is the probability of occupation of the top node during the simulation, and $p_{\mathrm{top}}^{\mathrm{eq}}(n)$ is the expected (equilibrium) value of this probability. Furthermore, $E_{\mathrm{av}}^{\mathrm{eq}}(n)$ is the expected energy, supposing the system were in equilibrium at temperature $T(n)$, while $E_{\mathrm{av}}(n)$ is the actual value of the average energy during the annealing. Figure 5(*c*) shows the probability $p_i(T(n))$ to be in the minimum (lowest node) of basin i during the simulation and the expected (equilibrium) probability $p_i^{\mathrm{eq}}(T(n))$; in addition, the probability to end up in the minimum of basin i upon quenching at the moment we reach the temperature $T(n)$, $p_i^{\mathrm{qu}}(T(n))$, is shown for comparison.

First, we notice that upon instantaneous quenching, we would end up in basin 2 with

90% probability, while at the end of the simulated annealing run, this probability has essentially become reversed. Next, we see that at $T \approx 1 \ldots 1.5$, $z(n)$ begins to grow very rapidly, indicating that the system has dropped out of global equilibrium. Furthermore, p_1^{eq} increases dramatically at this point, reflecting the fact that the bottom of the first basin is no longer invisible, and the walker rapidly descends into this basin. A similar phenomenon occurs at $T \approx 0.3 \ldots 0.5$, where p_2 (not p_2^{eq}!) increases rapidly, since now the probability to be at the bottom of pocket 2 (given that one was somehow in basin 2 in the first place!) grows from zero to its maximal possible value ($= p_2^{\mathrm{qu}}$). We also note that at the same moment as the second trap opens, p_2^{qu} is stabilized at its current level $p_2^{\mathrm{qu}}(T \approx E_2)$. Thus the continuous flow of probability from basin 2 into basin 1 that took place for $E_1 > T > E_2$ has stopped.

Next, let us investigate the effect of introducing a substantial barrier. We will take the same two pockets as before, but place the nodes of pocket 2 at the energy levels $(-8, -7, \ldots, -3)$, i.e. the number of surface states still strongly favours basin 2, the trapping (if it occurs at all), takes place much earlier for basin 1 ($E_1 \approx 3E_2$), basin 1 remains deeper, and the height of the barrier is $-E_2^t = 3$. According to table 1, this would correspond to a situation, where the walker will most likely not find the 'right' basin.

The results are shown in figure 6. The expectation is borne out by the results: the annealing fails in finding the right basin, since E_2/E_1 is too large. The number of states at the surface of basin 2 combined with the barrier prevents the system from being present at the top of basin 1 with a high enough probability at the moment when trapping into basin 1 can occur ($T \approx E_1$).

However, if one lets the system equilibrate for a longer time at each temperature, the barrier can be climbed successfully, and the system remains in equilibrium even for some range of temperatures below E_1. Therefore, the barrier becomes irrelevant, and the analysis of section 3.1 applies: the system will end up trapped in basin 1, after all. This is depicted in figure 7, where we have increased the number of equilibration steps at each temperature to 2^{13}. The need for such a large increase is connected to the logarithmic temperature schedule discussed in the introduction (equation (1)) and section 3.4, which is necessary to equilibrate across a barrier. (Still, because of the exponential densities of states, the relevant barrier is only of size 3, not 8 or even 11.) We see that trapping in pocket 1 again sets in at about $T \approx 1$, while the trapping in basin 2 occurs at $T \approx 0.4$. The system breaks ergodicity at about $T \approx 0.8$, when p_1^{eq} separates from p_1. Still, it would require an increase of the number of equilibration steps to about 2^{18} to draw nearly all walkers across the barrier into basin 1.

Finally, in figure 8, we show the results of an annealing analogous to the previous example (figure 6), but with a more slowly growing density of states in basin 2: ($E_2^t = -3$, $E_2^0 = -8$, $E_2 = 1/\ln(3) \approx 0.91$). Thus $E_1 > E_2$, but not $E_1 \gg E_2$. Here, ergodicity is again broken at approximately the same temperature, at which the trapping into basin 1 occurs ($T \approx 1$). Although the barrier favours basin 2 somewhat (notice the dip in p_1^{qu} for $T > 1$), the much larger number of surface states of basin 1 compared with basin 2 ensures that most of the probability is still concentrated at the top of basin 1 when T drops below E_1.

3.4. Time-scale analysis

As we have seen in figures 6 and 7, the number of relaxation steps available per temperature interval, N_{steps}, can have an influence on the outcome of the annealing process. It is therefore appropriate to address this issue and to point out the choice of N_{steps} that is implied in our analysis in sections 3.1 and 3.2.

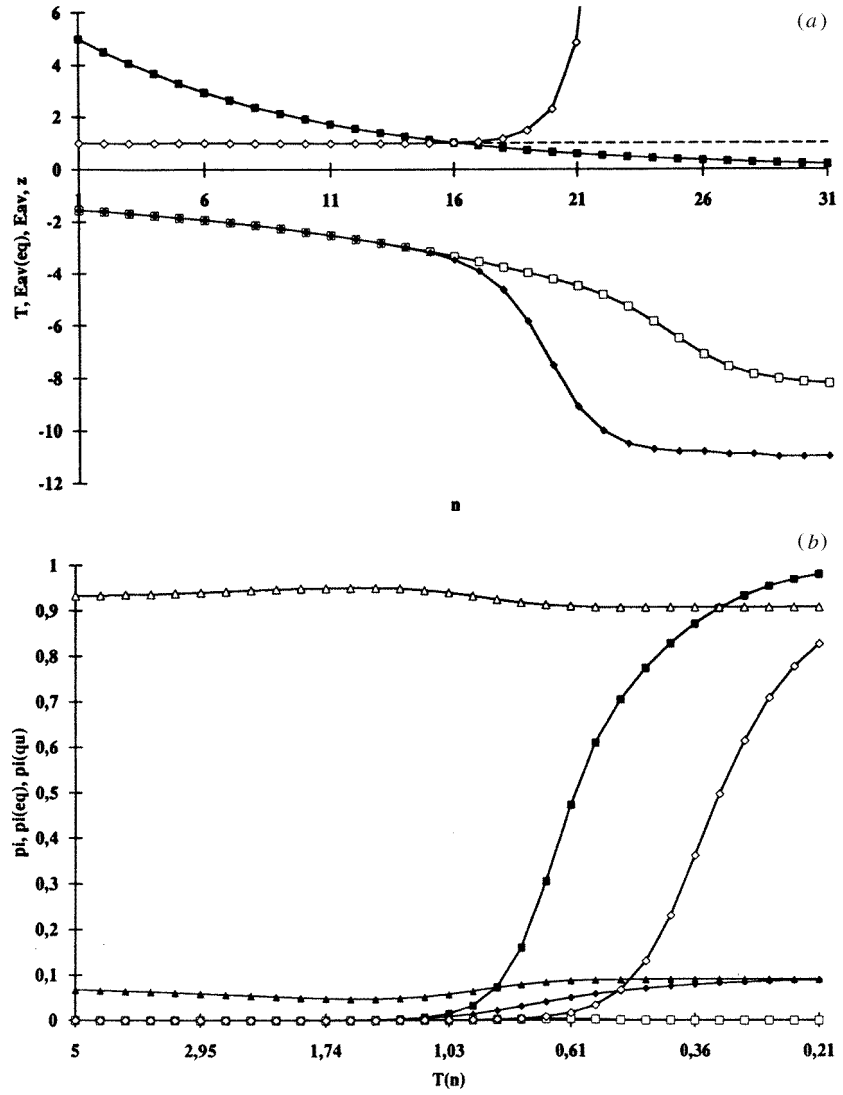

Figure 6. Fast annealing of two pockets with ($E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24$) and ($E_2^t = -3, E_2^0 = -8, E_2 = 1/\ln(10) \approx 0.43$), respectively. (*a*) Plot of $T(n)$ (full squares), $E_{\rm av}(n)$ (open squares), $E_{\rm av}^{\rm eq}(n)$ (full diamonds) and $z(n) = p_{\rm top}(n)/p_{\rm top}^{\rm eq}(n)$ (open diamonds; $z = 1$: broken curve) as a function of temperature updates n. (*b*) Plot of $p_i(T(n))$ (diamonds), $p_i^{\rm eq}(T(n))$ (squares), and the probability to end up in basin i upon quenching $p_i^{\rm qu}(T(n))$ (triangles) as a function of temperature T (full symbols = minimum 1, open symbols = minimum 2).

The 'easy' cases are of course the ones where either no time ($N_{\rm steps} = 0$) or practically infinite time ($N_{\rm steps} > \exp((D_i - E_i^t)/T)$ is available, i.e. a quench or a full equilibration across all barriers at every temperature, respectively. In sections 3.1 and 3.2, we have analysed the case that presumably is a common occurrence during simulated annealing for example, where the relaxation is slow enough to establish or maintain thermal equilibrium at the beginning, but so fast that trapping will play a decisive role.

Let us begin by considering a number of time scales that are relevant to the flow of

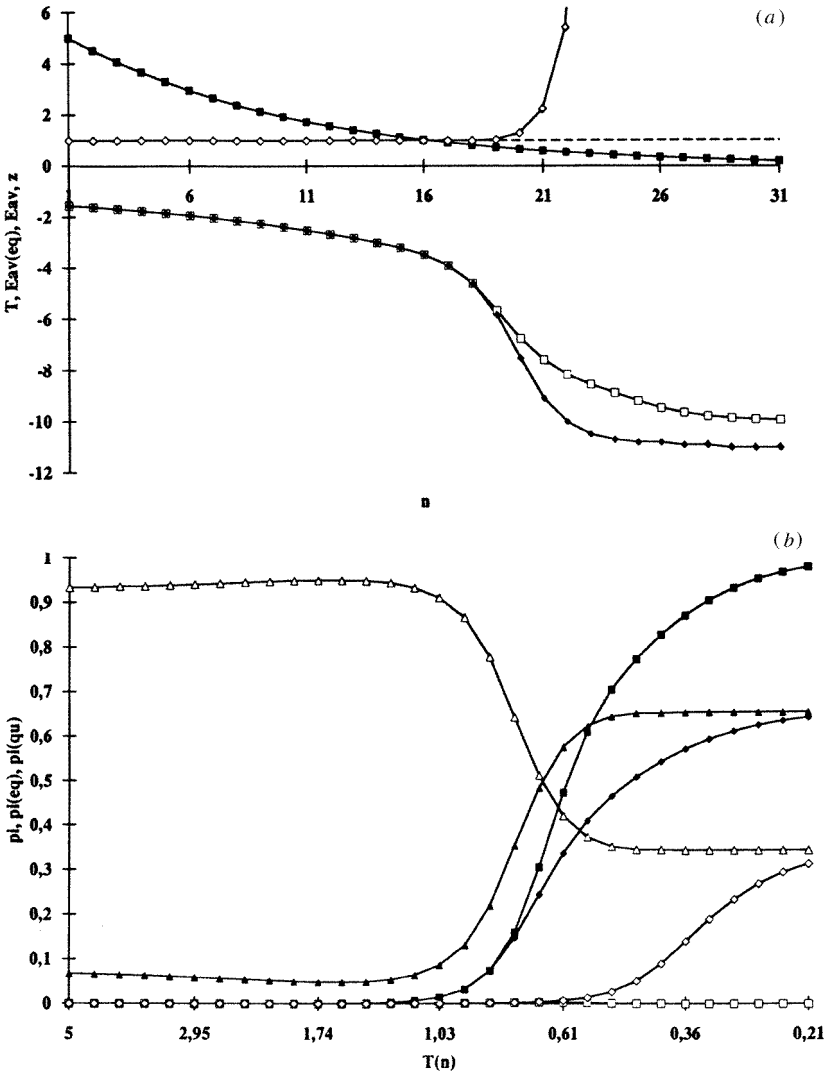

Figure 7. Slow annealing of two pockets with ($E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24$) and ($E_2^t = -3, E_2^0 = -8, E_2 = 1/\ln(10) \approx 0.43$), respectively. (*a*) Plot of $T(n)$ (full squares), $E_{\rm av}(n)$ (open squares), $E_{\rm av}^{\rm eq}(n)$ (full diamonds) and $z(n) = p_{\rm top}(n)/p_{\rm top}^{\rm eq}(n)$ (open diamonds; $z = 1$: broken curve) as a function of temperature updates n. (*b*) Plot of $p_i(T(n))$ (diamonds), $p_i^{\rm eq}(T(n))$ (squares), and the probability to end up in basin i upon quenching $p_i^{\rm qu}(T(n))$ (triangles) as a function of temperature T (full symbols = minimum 1, open symbols = minimum 2).

probability in the two-basin system. These reflect the energetic and entropic 'effective barriers' that are present in the system depending on the current temperature. For definiteness, we consider the flow of probability between the two basins for $T > E_{1,2}$. Here, the parts of the basins below the surface of the exponential region remain invisible. There exist two energetic effective barriers, $-E_{1,2}^t$. These enter the Boltzmann factors that describe the fraction of successful attempts to move from basin i to the saddle region 'S', located at $E_S = 0$. The entropic weight for each basin is given by the number of surface states $N_i^s \propto \exp(D_i/E_i)$.

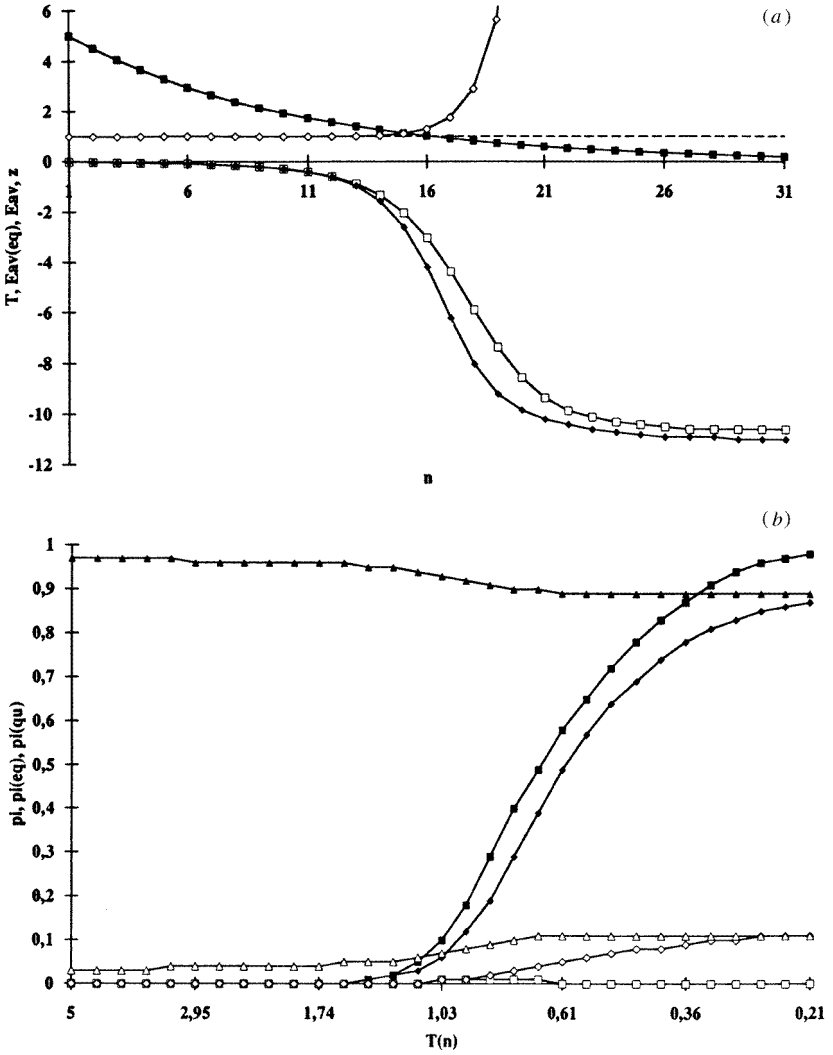

Figure 8. Fast annealing of two pockets with ($E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24$) and ($E_2^t = -3, E_2^0 = -8, E_2 = 1/\ln(3) \approx 0.91$), respectively. (*a*) Plot of $T(n)$ (full squares), $E_{\rm av}(n)$ (open squares), $E_{\rm av}^{\rm eq}(n)$ (full diamonds) and $z(n) = p_{\rm top}(n)/p_{\rm top}^{\rm eq}(n)$ (open diamonds; $z = 1$: broken curve) as a function of temperature updates n. (*b*) Plot of $p_i(T(n))$ (diamonds), $p_i^{\rm eq}(T(n))$ (squares), and the probability to end up in basin i upon quenching $p_i^{\rm qu}(T(n))$ (triangles) as a function of temperature T (full symbols = minimum 1, open symbols = minimum 2).

The flow of probability from, say, basin 1 to basin 2 is best analysed by splitting it into two effective steps, $f(1 \to 2) = f(1 \to s) f(s \to 2)$, where

$$f(1 \to s) = \exp(E_1^t/T) \tag{12a}$$

is the fraction of attempted moves that reach the saddle starting from basin 1, and

$$f(s \to 2) = N_2^s/(N_1^s + N_2^s) \tag{12b}$$

is the fraction that reaches basin 2 from the saddle.

Analogously, we have

$$f(2 \rightarrow s) = \exp(E_2^t/T) \tag{12c}$$

and

$$f(s \rightarrow 1) = N_1^s/(N_1^s + N_2^s). \tag{12d}$$

In equilibrium, the probabilities of occupation of the two basins, p_i, is found from the equation

$$p_1 f(1 \rightarrow 2) = p_2 f(2 \rightarrow 1) \tag{13}$$

which yields equation (11).

One should note that the real computation times associated with the effective MC steps $f(1 \rightarrow 2)$ and $f(2 \rightarrow 1)$, as observed during the use of a stochastic algorithm implementing simulated annealing, for example, in addition incorporate the ratio of possible moves within a basin or the saddle region to those possible between the basins and the saddle. Of course, these reflect the internal structure of the pockets and the saddle region. For our analysis, we will ignore this additional complication.

Measured in these effective MC steps, the time scale necessary to achieve equilibrium can be deduced by the following thought experiment. Assume all probability to be placed initially into basin 1 (worst case). Furthermore, assume that basin 2 is an absorbing state of the effective two-state system. Then the flow of probability per time step can be described by a (2×2) Markov matrix:

$$\{\mathbf{M}\} = \begin{Bmatrix} p(1 \rightarrow 1) = 1 - p(1 \rightarrow 2) & 0 \\ p(1 \rightarrow 2) & 1 \end{Bmatrix}, \boldsymbol{p}_n = \{\mathbf{M}\}\boldsymbol{p}_{n-1} = \{\mathbf{M}\}^n \boldsymbol{p}_0. \tag{14}$$

Now

$$\{\mathbf{M}\}^n = \begin{Bmatrix} (p(1 \rightarrow 1))^n & 0 \\ 1 - (p(1 \rightarrow 1))^n & 1 \end{Bmatrix} \tag{15a}$$

and for $n \rightarrow \infty$, we get

$$\lim_{n \rightarrow \infty} \{\mathbf{M}\}^n = \begin{Bmatrix} 0 & 0 \\ 1 & 1 \end{Bmatrix}. \tag{15b}$$

The number of steps, $N_{(1 \rightarrow 2)}$, necessary to draw a substantial amount of probability from basin 1 to basin 2 is found by setting

$$p(1 \rightarrow 1)^n = (1 - p(1 \rightarrow 2))^n \approx 1 - np(1 \rightarrow 2) = 0 \Leftrightarrow N_{(1 \rightarrow 2)} = 1/p(1 \rightarrow 2). \tag{16}$$

This is an analogue to the exponential decay factor $\lambda = -p(1 \rightarrow 2)$ one finds, if one transforms the discrete Markov-process into a continuous one

$$\{\mathbf{A}\} = \{\mathbf{M}\} - \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}, \dot{\boldsymbol{p}} = \{\mathbf{A}\}\boldsymbol{p} \tag{17}$$

and determines the eigenvalues of $\{\mathbf{A}\}$.

From this we find

$$N_{(1 \rightarrow 2)} = \exp(-E_1^t/T)(1 + N_1^s/N_2^s) \tag{18a}$$

and analogously

$$N_{(2 \rightarrow 1)} = \exp(-E_2^t/T)(1 + N_2^s/N_1^s). \tag{18b}$$

The number of steps required for equilibration at temperature T is then given by the larger of these two numbers, $N_{\max}$. This is an upper bound, of course, since we have deduced this value from a worst-case analysis.

We observe the following:

(i) If $N_{\rm steps} > N_{\rm max}$ ($= N_{(1\to 2)}$ for definiteness), the occupation of the two basins corresponds to the Boltzmann distribution. If $N_{(1\to 2)} > N_{\rm steps} > N_{(2\to 1)}$, it is possible to draw substantial amounts of additional probability from basin 2 into basin 1, as long as this moves the whole system closer to thermodynamic equilibrium. Finally, if $N_{(1\to 2)}$, $N_{(2\to 1)} > N_{\rm steps}$, essentially no change takes place with regard to the occupation of the two basins. (Depending on whether $N_{\rm max}$ or $N_{\rm min}$ are relevant for the continuation of the probability flow, we can define a 'freeze-in temperature' of the system, $T_{\rm eff}$.)

(ii) The analysis above can be carried over to the cases $E_1 < T < E_2$ and $T < E_{1,2}$. For example, we just need to replace $-E_2^t$ by $(D_2 + (-E_2^t))$ and N_2^s by the total number of states in the basin, $N_2^v \propto \exp(D_2/E_2)$, that are now involved (recall that the effective MC steps took redistribution moves within the basins into account).

This qualitative analysis yields table 2 for $N_{(1\to 2)}$ and $N_{(2\to 1)}$, where we assume for definiteness $E_2 > E_1$ (for $N_1^{v,s} \gg N_2^{v,s}$ or $N_1^{v,s} \ll N_2^{v,s}$, one could simplify the expressions for $N_{(1\to 2)}$, $N_{(2\to 1)}$, of course). With the help of table 2, we can roughly analyse the time evolution of the occupation of the two basins, for a given value for $N_{\rm steps}$.

Table 2. Number of steps necessary to draw a substantial amount of probability from basin 1 (2) into basin 2 (1), $N_{(1\to 2)}$ and $N_{(2\to 1)}$, respectively, for different ranges in temperature and relative sizes of the basins, $N_{1,2}^{v,s}$. See text for further discussion and notation.

	$N_{(1\to 2)}$	$N_{(2\to 1)}$
$T > E_{1,2}$	$\exp(-E_1^t/T) \times (1 + N_1^s/N_2^s)$	$\exp(-E_2^t/T) \times (1 + N_2^s/N_1^s)$
$E_2 > T > E_1$	$\exp(-E_1^t/T) \times (1 + N_1^s/N_2^v)$	$\exp((-E_2^t + D_2)/T) \times (1 + N_2^v/N_1^s)$
$T < E_{1,2}$	$\exp((-E_1^t + D_1)/T) \times (1 + N_1^v/N_2^v)$	$\exp((-E_2^t + D_2)/T) \times (1 + N_2^v/N_1^v)$

How do these considerations connect to the analysis in sections 3.1 and 3.2, i.e. which value of $N_{\rm steps}$ is considered there?

We begin with section 3.1. Here $E_2^t = E_1^t = 0$, and we assume that $N_{\rm steps}$ is large enough such that the system is in equilibrium for $T > E_{1,2}$. In order for trapping to occur for $T < E_2$, we need to assume that $N_{\rm steps} < \exp(D_2/T) \leqslant N_{\rm max}$, or else, the system would remain in thermodynamic equilibrium. The amount of probability trapped will depend on the time available in the temperature region between E_2 and E_1. There exist three possibilities:

(i) $N_1^s < N_2^v$. Thus $N_{\rm min} = 1$, and there will occur a large transfer of probability into basin 2, which is both steeper and deeper than basin 1.

(ii) $N_1^s > N_2^v$. If now $(N_1^s/N_2^v) > \exp(D_2/T) > N_{\rm steps}$, there is never enough time available to draw the probability from basin 1 into basin 2. Furthermore, the equilibrium would favour basin 1. Depending on the value of E_1, this may be a positive outcome or represent a golf-hole situation. This situation resembles a quench through the critical region in temperature from $T > E_2$ to $T < E_1$.

On the other hand, if $\exp(D_2/T) > N_1^s/N_2^v \approx N_{\rm steps}$, there is sufficient time to shift probability from basin 1 to basin 2. This represents the case where exponential trapping is most likely to have an effect, i.e. where, for rather short relaxation times, the trapping into basin 2 can overcome the effect of the much larger number of surface states in basin 1.

Clearly, after $T < E_1$, $N_{\rm steps} < N_{\rm min} \leqslant \exp(D_2/T)$, $\exp(D_1/T)$, the relaxation thus is essentially a quench at this point.

In section 3.2, the situation is slightly more complicated, since here the effective barriers between the basins, measured from the surface, $-E_{1,2}^t$, come into play.

If we choose $N_{\rm steps}$ so large that the system remains in equilibrium until trapping occurs,

these barriers can be ignored and the discussion of section 3.1 applies. In section 3.2, it was thus assumed that the relaxation is so fast that the barriers are going to be relevant. First, both were non-negligible, i.e. below an effective temperature $T_{\mathrm{eff}} > E_{1,2}$, N_{steps} is not sufficient to maintain the system in equilibrium. Therefore, preferential trapping cannot occur, and the only effect of the exponential growth is seen in the combination of exponential growth factors and basin depths that influence the ratio of the probabilities in equation (11).

Next, we treated the case where one of the barriers was negligible ($-E_1^t = 0$). In this case $N_{(1\to 2)}$ continues to be small compared to N_{steps}, as long as $T > E_1$. Thus, there can exist a substantial flow of probability from basin 1 to basin 2, as long as the system would move towards equilibrium due to this flow. Again, we dealt with fast relaxation, i.e. $N_{(2\to 1)} \gg N_{\mathrm{steps}}$ (no flow from basin 2 to basin 1) for temperatures below the freeze-in temperature for the flow from basin 2 to basin 1, $T_{\max}$.

4. The success of simulated annealing

The results of the previous section can be summarized by saying that for a landscape consisting of two pockets with exponentially growing densities of states, which contain the important deep-lying minima, the random walker during a simulated annealing run, for example, is funnelled preferentially into the trap that exhibits the deepest minima. Of course, extreme counter examples can always be constructed (e.g. golf-hole type situations), but the net channelling effect of exponential traps compared with situations with only power-law growth in the density of states, for example, is quite remarkable.

Using these results, we can now understand why simulated annealing works for an energy landscape that consists of trapping pockets with exponentially growing densities of states.

(i) For $T > \max(E_i)$, the walker does not get trapped anywhere and statistical equilibrium is maintained, as long as the schedule allows for the crossing of possible barriers measured from the top of the exponential pockets.

(ii) At temperatures between maximum (E_i) and minimum (E_i), trapping occurs where a strong preference towards traps with deep-lying minima is shown.

(iii) For temperatures below the trapping temperature of the basin that has caught the walker, the walker will only explore this pocket.

As an illustration, we have considered a simple problem consisting of four basins with parameters ($E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24$), ($E_2^t = 0, E_2^0 = -5, E_2 = 1/\ln(10) \approx 0.43$), ($E_3^t = 0, E_3^0 = -5, E_3 = 1/\ln(2) \approx 1.44$), and ($E_4^t = 0, E_4^0 = -3, E_4 = 1/\ln(20) \approx 0.33$), respectively. Again, the nodes within the basins are spaced in units of one, and the four basins are joined at the top node containing 200 000 states at $E = +0.1$. We have again used the fast temperature schedule from section 3.3.

The results are plotted in figure 9. Figure 9(*a*) shows again $T(n)$, $E_{\mathrm{av}}(n)$, $E_{\mathrm{av}}^{\mathrm{eq}}(n)$ and $z(n)$, but we have only depicted $p_i^{\mathrm{qu}}(T(n))$, ($i = 1, \ldots, 4$) in figure 9(*b*) in order to avoid crowding. We notice that at the end of the annealing, the 'right' basin has the highest probability, as expected. Furthermore, ergodicity is broken at about $T \approx 1.3$, similar to figure 5. Although pocket 4 contains slightly more surface states than pocket 1, the former finishes last, even behind the tiny, but deep, pocket 3. We notice that the ratios of the final quenched probabilities $p_1^{\mathrm{qu}}(31)/p_2^{\mathrm{qu}}(31) \approx 11.8$, and $p_3^{\mathrm{qu}}(31)/p_4^{\mathrm{qu}}(31) \approx 9.0$ are approximately equal; a reasonable result, since pockets 3 and 4 resemble a scaled-down version of pockets 1 and 2.

Furthermore, figure 9(*b*) shows very nicely how for pockets of about equal depth (2 and 3) the one with the larger number of surface states wins, although only in a reduced

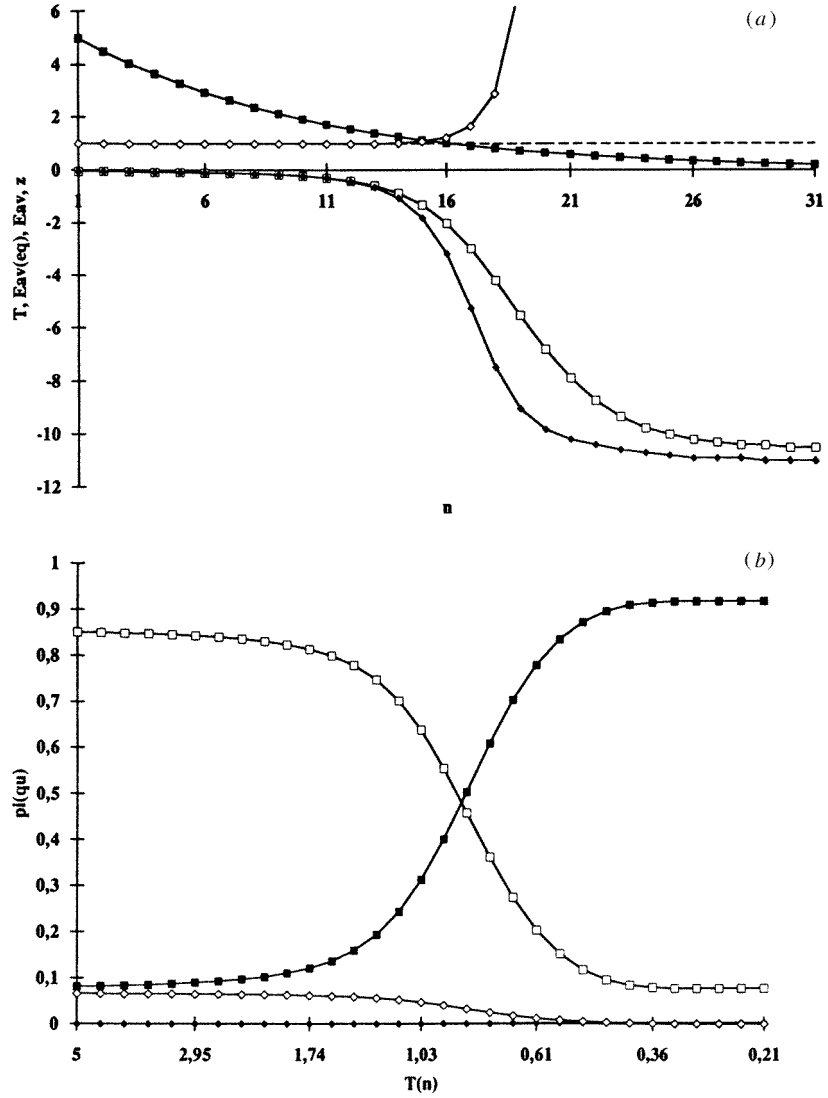

Figure 9. Fast annealing of four pockets with $(E_1^t = 0, E_1^0 = -11, E_1 = 2/\ln(5) \approx 1.24)$, $(E_2^t = 0, E_2^0 = -5, E_2 = 1/\ln(10) \approx 0.43)$, $(E_3^t = 0, E_3^0 = -5, E_3 = 1/\ln(2) \approx 1.44)$, and $(E_4^t = 0, E_4^0 = -3, E_4 = 1/\ln(20) \approx 0.33)$, respectively. (*a*) Plot of $T(n)$ (full squares), $E_{\rm av}(n)$ (open squares), $E_{\rm av}^{\rm eq}(n)$ (full diamonds) and $z(n) = p_{\rm top}(n)/p_{\rm top}^{\rm eq}(n)$ (open diamonds; $z = 1$: broken curve) as a function of temperature updates n. (*b*) Plot of the probability to end up in basin i upon quenching $p_i^{\rm qu}(T(n))$ as a function of temperature T (full squares = minimum 1, open squares = minimum 2, full diamonds = minimum 3, open diamonds = minimum 4).

fashion: $p_2^{\rm qu}(31)/p_3^{\rm qu}(31) \approx 31$ versus $g_2^s/g_3^s = 3125$. Finally, the comparison of pockets 1 and 3 shows that, for similar growth laws and thus similar trapping temperatures, the deeper pocket wins, because the 'large-rims-have-deep-wells-hypothesis' is fulfilled and relevant: $p_1^{\rm qu}(31)/p_3^{\rm qu}(31) \approx 367$ and $g_1^s/g_3^s \approx 219$.

Three important issues remain to be addressed with regard to this scenario:

(i) To what extent can the system be assumed to be in equilibrium at $T > \max(E_i)$?

(ii) To what extent does the analysis for two competing pockets apply when very many basins are present?

(iii) What happens with the walker inside a trap?

Let us begin with the second point. If many trapping basins are present, one can group them according to their E_i, E_i^0 and E_i^t values. Then the size of each group can be taken into account through an adjustment in the prefactor C_i in the density of states. However, since we are dealing with exponential growth laws, these prefactors will only be important if massive imbalances among the different classes of pockets exist. As long as statistical equilibrium can be maintained before trapping occurs, the arguments of the previous section can be transferred to the case of many competing minima. There now exists a sequence of trapping temperatures. If the class of traps with the highest value of E_i, e.g. E_1, does not succeed in catching the walker at $T = E_1$, the reason will lie in the overwhelming number of surface states of the remaining traps. Thus for temperatures below E_1, this first class of traps drops out of the competition. This process is repeated, until the walker is finally trapped.

It might be argued that the number of states belonging to the unsuccessful basins will begin to mount and in the end become much larger than the number of states in the last few competing classes of traps, leading to a contradiction. However, this argument does underestimate the power of exponential growth, and it will only apply if the number of different classes of basins is very large. Let us assume that N basins with different trapping temperatures $E_1 > E_2 \ldots > E_N$ exist, and that for each basin i the number of states on the surface of those pockets with $E_j < E_i$ exceeds the number of states of basin i (which is also essentially given by the number of surface states) by some large factor c. We want to consider the extreme case that the contradiction occurs for the last basin,

$$g_N^s < \sum_{i=1}^{N-1} g_i^s. \tag{19}$$

Then, the condition on c and N for this inequality to hold is given approximately by

$$(N-1) > (c+1)\ln(c) \tag{20}$$

i.e. the contradiction occurs, if the number of distinct competing basins is much larger than the already large factor c. If the trapping were to take place earlier, i.e. at some other value $i < N$, then already $(i-1)$ needs to be larger than $(c+1)\ln(c)$, in order to lead to a contradiction. Of course, one cannot exclude this possibility, but it requires that there exist at least $(c+1)\ln(c)$ classes of basins sufficiently different to require separate treatments.

The existence of many similar basins on the energy landscape might actually help in addressing the first question, and explain why it often appears that the system seems to remain in equilibrium until trapping, even though the phase space volume of the whole landscape is gigantic and important regions are separated by relatively large barriers. Instead of having to explore the whole energy landscape, all typical basins might be accessible 'locally' (without crossing large barriers), thus reducing the time necessary for accumulating a 'representative' sample. Such a similarity of different regions of the energy landscape might be a reasonable assumption in many systems, especially spin glasses and amorphous solids.

There remains the question of what happens to the walker within the traps. This will depend very much on the local energy landscape of the trapping basin. If this local landscape shows but little structure, with at most small energy barriers separating the local minima, the system will remain in thermodynamic equilibrium within the pocket and find the deepest minimum, with high probability.

On the other hand, one often encounters situations, where not only the number of states, but also the number of minima grows exponentially within the pocket [26–28]! Now the discussion of trapping in multiminima situations applies, as long as the trapping temperatures of these sub-basins are similar to the trapping temperature of the basin as a whole. The most problematic situations occur if these local trapping temperatures are for the most part higher than the trapping temperature of the whole pocket. Thus, once the walker is trapped, it falls into some local minimum, and it takes a very long time to find the global minimum walking on the very rugged landscape within the pocket. The success of the random walk within the pocket is then largely controlled by the barrier heights within the pocket, which can be much higher than E_i. However, we are often not interested in the global minimum, and a very good suboptimum is sufficient. If this is the case, then we note that the existence of exponential growth for both minima and states to a certain degree implies that those sub-basins with large trapping power will at the same time be rather deep. Thus, the walker often ends up in a good suboptimum, and only the final optimization necessary to achieve the last few per cent of improvement would require logarithmic temperature schedules for example [5], such as equation (1), if one were to use simulated annealing for this purpose. This is probably the reason that, for example, the lid method [39] or the related 'bouncing' algorithms [40, 41] appear to be more efficient in determining the global minimum of a given pocket or improving upon a very good suboptimum. But these algorithms usually start from an already very good suboptimum inside such a basin, which was found during an earlier optimization run using simulated annealing for example.

In this context, one should also make a final comment on those situations encountered in section 3 where the walker is likely to be caught in the wrong basin. In nearly all instances, we are dealing with some variant of the 'golf-hole' problem [22–24], which is specially designed to confound all optimization methods that claim to be an improvement over random search. Since simulated annealing is known to fail for such problems, it is reasonable to exclude these situations from our analysis, strengthening our claim that the trapping in exponentially growing pockets with different trapping temperatures constitutes an explanation for the success of simulated annealing on landscapes dominated by exponential traps.

5. Summary and discussion

To summarize, we have seen that random walkers exploring energy landscapes consisting of pockets with exponentially growing densities of states will be channelled preferentially into those basins that contain the deep-lying minima. This observation can help explain the success of the simulated annealing algorithm in the determination of good suboptima in complex multiminima systems.

We would like to point out, however, that the arguments presented here are not intended to be mathematically rigorous in the sense of the necessary and sufficient conditions on convergence mentioned in the introduction. Since we are trying to model Markov processes where breaking of ergodicity (on the time scale of the process) is important, and at best local equilibrium can be maintained, the standard methods to obtain results on convergence cannot be easily transferred. Nevertheless, as the qualitative discussion of relevant time scales in section 3.4 indicates, the derivation of such conditions is certainly possible for precisely defined energy landscapes and, of course, does remain an important aim.

Still, we feel that even the more qualitative approach taken here will be of use in understanding simulated annealing. The analysis presented in this paper clearly indicates, that a knowledge of the distribution of trapping temperatures, and the ratio of basin depth

and exponential growth factors will allow the design of more efficient temperature schedules for a given system, or even suggest modifications, e.g. in the moveclass (= neighbourhood structure) employed, of the global optimization algorithm itself.

Finally, the apparently common occurrence of exponentially growing local densities of states in models for physical systems that exhibit glass transitions provides food for thought: might there exist some deeper connections between the trapping in exponential pockets discussed in this paper and the (kinetically controlled) glass transition? After all, the time evolution of a physical system in contact with a heat bath is well represented by a random walk on an energy landscape, in a probabilistic sense.

In those systems where all important minima reside at the bottom of exponential pockets with similar growth laws (e.g. spin glasses?), a glass transition could be identified with the onset of trapping. The only requirement would be that the barriers above the surfaces of the traps should not be too large to avoid that the dynamics is barrier controlled instead of trap controlled.

Amorphous solids would be more complex, however, since the global minimum appears to be the crystalline phase. For the 'crystal' region of state space, one would expect that the growth in the local density of states would not follow an exponential function, but, for example, would be more likely to obey a power law†. Since the trapping temperature of a pocket with power-law growth is essentially infinite, $(1/E_i \to 0)$, crystallization will occur very quickly once the temperature drops below the melting temperature of the material and the walker (representing the system) is within or near the 'crystal' basin. But if either the trapping temperature of the exponential pockets representing amorphous configurations is comparable with the melting temperature, or the 'crystal' pocket is too small (a 'golf-hole' problem for nature!), the walker will be trapped in one of the amorphous regions. Of course, these considerations do not take the influence of the barrier structure above the traps into account—how are the exponential regions and the power-law basins welded together? A final answer to the question, whether trapping in pockets with exponentially growing densities of states is the cause of the glass transition, will surely require more detailed information about the energy landscape of amorphous solids.

Acknowledgments

I would like to thank the participants of the workshops at the Telluride Summer Research Center, on thermodynamics of trapping in 1995 and stochastic optimization in 1996, especially P Sibani, P Salamon, K H Hoffmann, I Morgenstern, B Andresen, J Nulton and R S Berry, for interesting and valuable discussions. This work has been supported by the DFG via the SFB408.

References

[1] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 *Science* **220** 671
[2] Czerny V 1985 *J. Opt. Theory Appl.* **45** 41–51
[3] van Laarhoven P J M and Aarts E H L 1987 *Simulated Annealing* (Dordrecht: Reidel)
[4] Metropolis N, Rosenbluth A, Rosenbluth M, Teller A and Teller E 1953 *J. Chem. Phys.* **21** 1087–92
[5] Geman S and Geman D 1984 *IEEE* **PAMI-6** 721–41
[6] Anily S and Federgruen A 1987 *Oper. Res.* **35** 867–74
[7] Anily S and Federgruen A 1987 *J. Appl. Prob.* **24** 657–67

† Some preliminary evidence for this assumption might be found in results gained from the analysis of some simple systems, using the threshold algorithm: a neon layer [29], and the system $MgF_2(z = 2)$ [42].

[8] Mitra D, Romeo F and Sangiovanni-Vincentelli A L 1986 *Adv. Appl. Prob.* **18** 747–71
[9] Gidas B 1985 *J. Stat. Phys.* **39** 73–131
[10] Hajek B 1988 *Math. Oper. R.* **13** 311–29
[11] Salamon P, Nulton J, Robinson J, Pedersen J, Ruppeiner G and Liao L 1988 *Comput. Phys. Commun.* **49** 423–8
[12] Flick J D, Salamon P and Andresen B 1987 *Inform. Sci.* **42** 239–53
[13] Goldstein L 1985 Mean square rates of convergence in the continuous time simulated annealing algorithm on Rd *Preprint*
[14] Hoffmann K H 1988 *Model Optimisation in Exploration Geophysics* ed A Vogel (Braunschweig: Vieweg Verlag) pp 65–70
[15] Lam J and Delosme J-M 1987 An adaptive annealing schedule *Preprint*
[16] Nulton J D and Salamon P 1988 *Phys. Rev.* A **37** 1351–6
[17] Pedersen J M, Mosegaard K and Jakobsen M O 1989 Optimized optimization *Preprint*
[18] Ruppeiner G, Pedersen J M and Salamon P 1991 *J. Physique* **1** 455–70
[19] Ettelaie R and Moore M A 1987 *J. Phys. France* **48** 1255–63
[20] Rees S and Ball R C 1987 *J. Phys. A: Math. Gen.* **20** 1239–49
[21] Salamon P 1986 unpublished
[22] Baum E 1986 *Phys. Rev. Lett.* **57** 2764–7
[23] Baskaran G and Stein D L 1987 *Phys. Rev. Lett.* **59** 373
[24] Wille L T 1987 *Phys. Rev. Lett.* **59** 372
[25] Hoffmann K H, Harland J, Nulton J and Salamon P 1988 An information theoretic analysis of simulated annealing *Preprint*
[26] Sibani P, Schön J C, Salamon P and Andersson J-O 1993 *Europhys. Lett.* **22** 479–85
[27] Sibani P and Andersson J-O 1994 *Physica* **206A** 1–12
[28] Schön J C 1997 in preparation
[29] Schön J C, Putz H and Jansen M 1996 *J. Phys.: Condens. Matter* **8** 143–56
[30] Wood D L and Tauc J 1972 *Phys. Rev.* B **5** 3144
[31] Soukoulis C M and Cohen M H 1984 *J. Noncryst. Solids* **66** 279
[32] Amer N M and Jackson W B 1984 *Semiconductors and Semimetals* ed R K Willardson and A C Beer (New York: Academic) pp 83–112
[33] Pfister G and Scher H 1978 *Adv. Phys.* **27** 747
[34] Scher H and Montroll E W 1975 *Phys. Rev.* B **12** 2455
[35] Silver M and Cohen L 1977 *Phys. Rev.* B **15** 3276
[36] Tiedje T and Rose A 1981 *Solid State Commun.* **37** 49–52
[37] Schmidlin F W 1977 *Phys. Rev.* B **16** 2362
[38] Berry R S 1995 personal communication
[39] Sibani P 1996 personal communication
[40] Müller-Krumbhaar H 1988 *Europhys. Lett.* **7** 479–84
[41] Morgenstern I 1996 personal communication
[42] Wevers M, Schön J C and Jansen M 1997 in preparation