Zero-temperature scaling and simulated annealing

R. Ettelaie, M.A. Moore

To cite this version:


HAL Id: jpa-00210551
https://hal.archives-ouvertes.fr/jpa-00210551
Submitted on 1 Jan 1987

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Zero-temperature scaling and simulated annealing

R. Ettelaie and M. A. Moore

Department of Theoretical Physics, The University, Manchester M13 9PL, U.K.

(Reçu le 2 mars 1987, accepté le 8 avril 1987)

Résumé. — Nous appliquons la technique du recuit simulé, avec des changements plus compliqués que le recouvrement d'un spin, au verre de spin d'Ising à une dimension. Nous montrons que la relation explicite entre l'entropie résiduelle à \( T = 0 \) dans le recuit simulé et le nombre d'états métastables, trouvée précédemment pour le renversement d'un seul spin, reste aussi vraie pour des changements plus compliqués. Combinant ce résultat avec un argument de loi d'échelle à température nulle, nous caractérisons l'amélioration produite par l'inclusion de changements plus compliqués. Ces résultats sont en accord avec des observations faites sur différents problèmes d'optimisation combinatoire.

Abstract. — Simulated annealing with a more complicated set of moves than single-spin flips is applied to the one-dimensional Ising spin glass. The explicit connection between the residual entropy at \( T = 0 \) in the simulated annealing and the number of metastable states, found previously for single-spin flips, is shown to be also true for these more complicated moves. This result together with a zero-temperature scaling argument is used to derive the degree of improvement that the inclusion of more complicated moves can produce. The results are in accord with the observations made for a number of problems in combinatorial optimization.

1. Introduction.

In a recent paper [1] we calculated the number of one spin-flip metastable states as a function of energy for a one-dimensional Ising spin glass. We found that for slow cooling rates the residual entropy was related to the number of such metastable states. The dynamic used in the cooling process was that of the Metropolis algorithm [2] and involved flipping one spin at a time. We also noticed that the residual entropy decreases to zero only logarithmically slowly with the inverse cooling rate. Hence, although the simulated annealing produced candidate ground states much lower in energy than those obtained by the quenched algorithm, it still requires an extremely slow cooling rate to find the true ground state.

An important part of any heuristic or simulated annealing method is the choice of its local rearrangements or local moves [3, 4]. It is known that in the case of some optimization problems more complicated sets of moves can give better answers. For example, in the travelling salesman problem, it has been found that a « 3-OPT » algorithm gives better answers than a « 2-OPT » one [5, 6]. Similarly a simulated annealing with three-bond rearrangement as its moves produced better solutions than that with two-bond moves [5]. (A tour is \( m \)-OPT if no shorter tour can be obtained by replacing \( m \) steps (bonds) in the tour with any other set of \( m \) bonds [18]). Although the use of more complicated rearrangements usually produces better answers, (albeit with more computer time), there are problems for which this is not the case. One such is the construction of binary sequences with low off-peak autocorrelations [7, 8]. For this problem, Bernasconi has found [8] that using multi-spin flip Monte Carlo did not produce any significant improvement when compared to the simpler one-spin flip Monte Carlo algorithm. Whether or not better results can be obtained depends on the distribution of the metastable states and the way this changes with a more complicated set of moves. In previous work, we had already found an explicit connection between the residual entropy at \( T = 0 \) in simulated annealing and the number of metastable states for one-dimensional Ising spin glasses [1]. The particular dynamic used in that case was single-spin flip. In this paper we show that the relation between the residual entropy and the number of metastable states is more general and applies also to more complicated dynamics, such as
two-spin moves. The results obtained here are for a linear Ising spin glass chain with Hamiltonian

$$H = -\sum_i J_i S_i S_{i+1}, \quad S_i = \pm 1, \quad i = 1, 2, \ldots, N.$$  \hspace{1cm} (1)

The exchange interactions $J_i$ are independent random variables with a gaussian distribution of width $\Delta J$,

$$P(J_i) = \left(\frac{2\pi (\Delta J)^2}{2}\right)^{-1/2} \exp\left(-\frac{J_i^2}{2(\Delta J)^2}\right).$$  \hspace{1cm} (2)

and $\Delta J$ will be taken to be one. Unfortunately the problem of finding the ground state energy of this system is not an NP complete problem. However, performing the following calculations for an NP complete case is not feasible. Nevertheless, we believe that when the simple 1-d Ising spin glass problem is treated by the method of simulated annealing it shows many of the features of more complex situations. This is because a one-dimensional Ising spin glass system has an exponentially large number of metastable states, and that there is nothing in the simulated annealing procedure which takes advantage of the features of the one-dimensional problem which makes it so easily solvable.

In the next section we shall first define what we mean by an $m$-spin flip metastable state, and then calculate the number and the distribution of such metastable states. These results indicate that the logarithm of the number of these metastable states, with energies close to the ground state, does not depend on $m$ and goes as $N^\sqrt{\epsilon}$, where $\epsilon$ is the difference between the energy (per spin) of a particular metastable state and the ground state energy. In § 3 we shall focus on the large $m$ limit and derive a scaling relation for the density of the metastable states as a function of energy. We show that such a relation can also be obtained using a zero-temperature scaling argument. This argument can be extended to higher dimensions $d$ and one finds that the density of metastable states is generally related to energy in the following way:

$$\frac{\ln N_s(\epsilon)}{N} = \frac{1}{m} f(\epsilon m^{1-\gamma/d}),$$  \hspace{1cm} (3)

where $\gamma$ is the zero-temperature scaling exponent [9], and $\ln N_s(\epsilon)$ is the logarithm of the number of metastable states with energy $\epsilon$ above the ground state energy, averaged over all possible bond configurations. We have not been able to determine the exact form of the function $f(x)$ for dimensions greater than one. Certain possibilities for $f(x)$ are suggested by the results for the one-dimensional case, and are discussed in sections § 3 and § 4. Whatever the detailed form of $f(x)$ it is likely that it will have a maximum at a certain value of $x$, say $x = x_0$. Then, an $m$-spin flip simulated annealing procedure, (or any other heuristic method using $m$-spin flip local moves for that matter), is expected at least to find a metastable state with an energy $\epsilon \approx x_0/m^{1-\gamma/d}$. This result clearly shows the important role that the exponent $\gamma$ plays in determining whether an increase in $m$ would give rise to better results. When $\gamma = 1-\gamma/d$ is zero or small, increasing $m$ will result in no significant improvement. Otherwise some improvement should be possible in principle, but it might not be enough to compensate for the longer computer times associated with the increased complexity of the algorithm (see § 4).

Finally, in section § 5, we have applied the simulated annealing method with $m = 2$ to a linear chain of $N$ Ising spins with Hamiltonian given by (1). The residual entropy is calculated for various cooling rates and system sizes. These results for the residual entropy are plotted against the residual energy in each case and compared to the analytically calculated curves for the density of two-spin flip metastable states. We find that the relation between the residual entropy and the density of metastable states obtained previously of $m = 1$ [1] also holds for $m = 2$; viz

$$S_{res}(e) = \ln N_s(e), \quad e \to 0,$$  \hspace{1cm} (4)

where $N_s(e)$ is the number of two-spin flip metastable states with energy per spin $\epsilon$ above the ground state energy and the bar denotes the average over the bonds $\{J_i\}$. A theoretical argument for equation (4) was given by Jackle [19]. We have also investigated the cooling rate dependence of the residual entropy.

2. The density of $m$-spin flip metastable states.

In order to define what is meant by an $m$-spin flip metastable state consider a linear chain of $N$ Ising spins with a Hamiltonian given by (2). An $m$-spin flip simulated annealing process applied to such a system will involve flipping a cluster of spins with sizes varying from one spin up to $m$ spins at any one step. An $m$-spin flip metastable state can be defined as a state for which the total energy of the system cannot be decreased by flipping any single spin or $n$ adjacent spins, where $n = 2, \ldots, m$. In order to calculate the number and the distribution of these states we shall work with the picture proposed by Fernandez and Medina [10]. In this picture the only contribution to the energy comes from the broken bonds each of which contribute $2|J_i|$. (A bond is « broken » if $J_i S_i S_{i+1} < 0$). Thus, the ground state, which has no broken bonds, will have zero energy. Since an $m$-spin flip metastable state is defined as a state for which the energy cannot be decreased by flipping a cluster of spins with any size from 1 up to $m$, it follows that when the system is in any such...
metastable state all the bonds which do not satisfy the condition

\[ |J_i| < |J_{i+k}|, \quad k = \pm 1, \ldots, \pm m \]  

(5)

have to be satisfied. Bonds for which the above condition is true may or may not be satisfied. To see this, consider a bond \( J_i \) for which the condition (5) is not true. Then, at least one of its 2 \( m \)-nearest neighbours, say \( J_{i+n} (n = \pm 1, \ldots, \pm m) \), has a magnitude less than \( |J_i| \). If \( J_i \) is broken, one can flip \( n \) spins between \( J_i \) and \( J_{i+n} \) simultaneously which will make \( J_i \) unbroken and reverse the condition of \( J_{i+n} \) but otherwise leave all the other bonds in the system unchanged. As a result the total energy of the system will be decreased by \( 2|J_i| + 2|J_{i+n}| \) if \( J_{i+n} \) was originally broken and \( 2|J_{i+n}| - 2|J_i| \) otherwise. Since \( |J_{i+n}| < |J_i| \) the decrease in energy is always positive. It follows that a state with \( J_i \) broken cannot be an \( m \)-spin flip metastable state. In other words an \( m \)-spin flip metastable state is entirely specified by the condition of the bonds for which equation (5) applies. Using (2), the probability for such bonds occurring is given by

\[
\int_{-\infty}^{+\infty} \frac{\exp(-J_i^2/2)}{\sqrt{2\pi}} \left( 2 \int_{|J_i|}^{+\infty} \frac{\exp(-J_j^2/2)}{\sqrt{2\pi}} dJ_j \right)^m dJ_i.
\]  

(6)

Evaluating the integral gives a probability of \( 1/(2m+1) \) (as has been also obtained by Li [11]).

Hence, for a system of size \( N \), on average \( N/(2m+1) \) bonds will satisfy relation (5). Since any \( m \)-spin flip metastable state can be specified by the condition of these \( N/(2m+1) \) bonds (which can be broken or unbroken) then there will be \( 2^{N/(2m+1)} \) \( m \)-spin flip metastable states for a system of size \( N \).

The distribution of these \( m \)-spin flip metastable states as a function of energy \( \varepsilon \) can be calculated in a similar way to that of the one-spin flip case [1]. First we note that the residual energy of a particular metastable state can be written

\[
N\varepsilon = \sum_{j=1}^{N/(2m+1)} (1 + \sigma_j)|J_j|,
\]

(7)

where the summation is over all \( N/(2m+1) \) bonds for which condition (5) is true and \( \sigma_j \) is +1 if the bond is broken and −1 otherwise. Using (7), the logarithm of the number of metastable states with total energy \( N\varepsilon \) above the ground state energy becomes

\[
\ln N_s(\varepsilon) = \ln \sum_{j=1}^{N/(2m+1)} \delta(\varepsilon - \sum_{j=1}^{N/(2m+1)} (1 + \sigma_j)|J_j|),
\]

(8)

which of course has to be averaged over all possible values of \( J_j \) of \( N/(2m+1) \) bonds. Representing the \( \delta \) function by its integral form one gets

\[
\ln N_s(\varepsilon) = \left( \ln \sum_{j=1}^{N/(2m+1)} \int_{-\infty}^{+\infty} \frac{d\mu}{2\pi} \int_{-\infty}^{+\infty} \frac{d\mu}{2\pi} \prod_{j=1}^{N/(2m+1)} (1 + e^{-2i\mu |J_j|}) \right) \langle J \rangle,
\]

(9)

The sum \( \sum_{j=1}^{N/(2m+1)} \ln (1 + e^{-2i\mu |J_j|}) \) is self averaging and can be written as

\[
\langle N/(2m+1) \rangle \langle \ln (1 + e^{-2i\mu |J|}) \rangle_f
\]

where \( \langle \rangle \) is the average taken with respect to the distribution

\[
P_{\text{min}}(J) = \frac{2m+1}{\sqrt{2\pi}} e^{-J^2/2} \left( 2 \int_{|J|}^{+\infty} \frac{\exp(-J_j^2/2)}{\sqrt{2\pi}} dJ_j \right)^m
\]

(10)

of the bonds satisfying relation (5). Changing the variable \( i\mu \) to \( \mu \) puts the integral (9) into a form suitable for evaluation by the method of steepest descent. Carrying out the integral gives the following equation for \( \mu \):

\[
\varepsilon = \frac{1}{2m+1} \left( \frac{2|J| e^{-2\mu |J|}}{1 + e^{-2\mu |J|}} \right)_f
\]

(11)

and

\[
\ln N_s(\varepsilon) = \varepsilon\mu + \frac{1}{2m+1} \langle \ln(1 + e^{-2\mu |J|}) \rangle_f
\]

(12)

The maximum value of (12) is \( \ln 2/(2m+1) \) and occurs for \( \mu = 0 \). For any value of \( m \) the \( \ln N_s(\varepsilon)/N \) versus \( \varepsilon \) curves are symmetric about their corresponding maxima. It can be shown that
for small values of $\epsilon$, that is $\mu \to +\infty$, we have

$$\frac{1}{N} \ln N_s(\epsilon) \approx 1.1456 \sqrt{\epsilon}, \quad (13)$$

independent of the value of $m$. This is an important result. It means that the number of low energy metastable states does not decrease as $m$ is increased. For small values of $m$, however, as $m$ is increased, the total number of metastable states drops by a large factor. For example, if $m$ is changed from 1 to 2, the total number of metastable states changes from $2^{N/3}$ to $2^{N/5}$. In figure 1 we have plotted $\left[\frac{1}{N} \ln N_s(\epsilon)\right]^2$ versus $\epsilon$ both for one and two-spin flip metastable states. It is seen that for very small values of $\epsilon$ the two overlap. However, for $\epsilon > 0.01$ there is already a significant difference. Thus, one can clearly see that a two-spin flip algorithm should give the metastable states close to the ground state a much better chance of being picked. It is important to point out at this stage that, in the case of $m = 2$, if one restricts the local moves to two-spin moves only, not allowing any one-spin rearrangements, then there will be no reduction in the number of the metastable states. In this case, the density of the metastable states versus energy graph would be identical to that for $m = 1$. However, not every metastable state would be a one-spin flip metastable state.

The fact that the number of very low-energy metastable states is independent of $m$ is not an unexpected result. For example, consider a state close to the ground state where the two weakest bonds in the system are the only broken bonds. Let us assume there are $n$ spins between these two bonds. This state is a metastable state for an $m$-spin flip algorithm as long as $m < n$. In the thermodynamic limit we expect $n$ to be very large, of order $N$, so this state is a metastable state for any finite value of $m$. For a high energy metastable state, however, one has many broken bonds which would not be separated by a large number of spins. Thus, an increase in the value of $m$ will remove these metastable states.

3. Large-\(m\) limit.

For large values of $m$, the probability distribution of the weakest bonds, i.e. probability distribution (10), takes a rather simple form:

$$P_{\min}(J) = (2m + 1) P(J) \left[ 2 \int_{|J|}^{\infty} dJ' P(J') \right]^{2m} \approx 2mP(0) \left( 1 - 2P(0)|J| \right)^{2m} \approx 2mP(0) \exp \left\{ -4mP(0)|J| \right\} \quad (14)$$

$P(J)$ is given by the Gaussian distribution (2). This allows one to evaluate equations (11) and (12) analytically for the large $m$ limit. Using (11) together with (14) we have

$$\varepsilon = \int_{-\infty}^{0} 2mP(0) \exp \left\{ -4mP(0)|J| \right\} \times \frac{2|J| e^{-2\mu|J|}}{2m(1 + e^{-2\mu|J|})} \, dJ. \quad (15)$$

Making a change of variable

$$4mP(0)J = y$$

$$\alpha = \mu / 2mP(0),$$

equation (15) becomes

$$(2m)^2 \varepsilon = \frac{1}{P(0)} \int_{0}^{\infty} dy \frac{ye^{-y}}{1 + e^{\alpha y}}. \quad (17)$$

Similarly for (12) making the same change of variable

$$\frac{1}{N} \ln N_s(\epsilon) = 2mP(0)\varepsilon \alpha + \frac{1}{2m} \int_{0}^{\infty} dy e^{-y} \ln \left( 1 + e^{-\alpha y} \right). \quad (18)$$

Using integration by parts, the above integral can be written in a simpler form:

$$\frac{1}{N} \ln N_s(\epsilon) = 2mP(0)\varepsilon \alpha + \frac{\ln 2}{2m} - \frac{\alpha}{2m} \int_{0}^{\infty} dy e^{-y} \frac{e^{-\alpha y}}{e^{\alpha y} + 1}. \quad (19)$$

Substituting equation (17) for $\varepsilon$ and a further integration by parts finally gives

$$(2m) \frac{1}{N} \ln N_s(\epsilon) = \ln 2 - \alpha^2 \int_{0}^{\infty} dy \frac{e^{(\alpha - 1)y}}{(e^{\alpha y} + 1)^2}. \quad (20)$$
Equations (17) and (20) together imply a scaling relation between $\ln N_s(\epsilon)/N$ and $\epsilon$ of the form

$$\frac{1}{N} \ln N_s(\epsilon) = \frac{1}{m} g(\epsilon m^2). \quad (21)$$

In figure 2, we have plotted the above function together with the corresponding curves for $m = 1$ and $m = 2$ scaled appropriately. It is not hard to show that $g(x) \sim \sqrt{x}$ for small values of $x$. This implies that for $\epsilon \ll 1/m^2$, we have

$$\ln N_s(\epsilon)/N \sim \sqrt{\epsilon},$$

independent of $m$, as mentioned previously.

For $d = 1$ and $y = -1$, equation (22) gives the correct result (21) as expected. For dimensions higher than one it is impossible to calculate the exact form of the scaling function $f(x)$. Two possibilities for $f(x)$ are likely in the light of the results for the one-dimensional case. If the spin glass can be regarded as a set of weakly interacting two-level systems at low temperatures [12, 13] then one would expect $f(x) \sim \sqrt{x}$ for small $x$. As seen above, this is certainly the case for the one-dimensional spin glass. In the next section we shall discuss this possibility in more detail. Another form for $f(x)$ might be that as $\epsilon \to 0$, $\ln N_s(\epsilon)$ becomes independent of the value of $m$, (as also happens for $d = 1$). One might expect this to be true in higher dimensions as well. This is especially likely when $y < 0$, although it is harder to envisage for problems with positive $y$. In both cases flipping a cluster of spins with a linear dimension $L$ produces an energy change in the order of $\Delta J L^y$. For a problem with negative $y$ this is a decreasing function of $L$. Therefore, for such problems, it is not unreasonable to expect the low energy metastable states to be very different from each other and from the ground state in structure. In other words, low energy metastable states do not become unstable until clusters with very large number of spins can be flipped. On the basis of this argument, for values of $m$ small compared to the size of the system, any change in $m$ should not alter the number of such low energy metastable states. Together with (22), this implies

$$\ln N_s(\epsilon)/N \sim \epsilon^{1/(1-y/d)} \text{ as } \epsilon \to 0. \quad (23)$$

Note that for very small values of the exponent $(1-y/d)$, $\ln N_s(\epsilon)/N$ is a very flat function of $\epsilon$, as $\epsilon \to 0$. For such problems, the energy surface consists of only a few valleys, one of which is deeper than any other. This deeper valley is the ground state. The energy landscape approximates that of a « golf-course » potential. This has important implications for the efficiency of multi-spin flip algorithms and we shall return to it later.

It must be stressed that the scaling argument given in this section only applies to systems which are homogeneous. Many problems, such as the 1-d Ising spin glass problem, are inhomogeneous if the number of spins $N$ is small. However, as the size of the system is made bigger any boundary effects become unimportant and the system can be treated as being effectively homogeneous. Other types of systems can exist which are only homogeneous above a certain length scale $l$. In that case a scaling expression similar to (22) can only be written for $m$-spin flip metastable states with $m \gg l^d$.

4. The scaling function $f(x)$ for two-level systems.

In previous sections we have repeatedly mentioned...
that for small values of \( \varepsilon \), \( \ln N_s(\varepsilon)/N \) for a one-dimensional Ising spin glass system varies as \( \sqrt{\varepsilon} \).

This result is by no means specific to the 1-d Ising spin glass but applies to any system consisting of a set of weakly interacting two-level systems. In particular there is some evidence that spin glasses at higher dimensions than one can also be regarded as such systems [12, 13]. In this section we shall look at a two-level system and derive explicitly the dependence of \( \ln N_s(\varepsilon) \) on \( \varepsilon \) for such a system. Consider a system made up of \( N \) weakly interacting, two-level sub-systems. Each sub-system can be in one of two possible states, the ground state or an excited state. Further, assume that the energy difference between the ground and excited level for the \( i \)-th sub-system is \( x_i \). Thus, the density of states for the total system is

\[
N_s(\varepsilon) = \sum_{\sigma_i = 0, 1} \delta \left( \varepsilon N - \sum_i \sigma_i x_i \right)
\]  

(24)

where \( N\varepsilon \) is the total energy above the ground state. Using (24) and following the same procedure as that in section § 2, we have

\[
\ln N_s(\varepsilon) = \ln \int_{-\infty}^{+\infty} \frac{\mu}{\pi^2} e^{-\mu^2} P(x) \alpha \frac{d}{d\alpha} \frac{1 + e^{-\mu x}}{1 + e^{\alpha}}.
\]

(25)

The average \( \langle \cdot \rangle \) is taken with respect to the probability distribution \( P(x) \) of \( x_i \)’s. Using the method of steepest descent to evaluate (25) gives

\[
\varepsilon = \int dx P(x) x e^{-\mu x} \frac{1}{1 + e^{-\mu x}}.
\]

(26)

Substituting \( \alpha = \mu x \), the integral (26) becomes

\[
\varepsilon = \frac{1}{\pi^2} \int P(\alpha/\mu) \frac{\alpha}{1 + e^\alpha} d\alpha.
\]

(27)

For small values of \( \varepsilon \), that is large values of \( \mu \), we have \( P(\alpha/\mu) \approx P(0) \) if \( \alpha \ll \mu \). For \( \alpha \) larger or comparable to \( \mu \), the exponential factor in the denominator of the integrand ensures that the contribution to the integral from this region is negligible. Hence, (27) can be written as

\[
\varepsilon = \frac{P(0)}{\mu^2} \int \frac{\alpha}{1 + e^\alpha} d\alpha.
\]

(28)

In other words \( \mu \sim \sqrt{1/P(0)/\varepsilon} \). Substituting this result into (25) gives

\[
\ln N_s(\varepsilon)/N \sim \sqrt{P(0) \varepsilon}, \quad \varepsilon \to 0.
\]

(29)

Note that this result is true for any \( P(x) \) as long as \( P(0) \) is finite. In the case of a \( d \)-dimensional spin glass system, (29) together with (22) give

\[
\ln N_s(\varepsilon)/N \sim \sqrt{\varepsilon/m^{1+y}/4}, \quad \varepsilon \to 0.
\]

(30)

At present it is not clear to us which of the two possibilities (30) or (23), if either, are correct.

Fortunately, as far as the efficiency of multi-spin flip algorithms are concerned, a detailed knowledge of \( f(x) \) is not required. Suppose it has a maximum at a certain value of \( x = x_0 \). This, together with equation (22) implies that the maximum value of \( \ln N_s(\varepsilon) \) occurs at \( \varepsilon = x_0/m^{1+y}/4 \). An \( m \) spin-flip simulated annealing or any other \( m \)-spin flip heuristic algorithm for that matter, ought to at least produce answers with \( \varepsilon \approx x_0/m^{1+y}/4 \). Let us assume that the average value of \( \varepsilon \) produced by a given algorithm is \( \delta x_0/m^{1+y}/4 \). \( \delta \) is a numerical factor between 0 and 1, the exact value of which depends on the details of the algorithm used. Clearly, for all those problems for which \( \lambda = 1 - y/d \approx 0 \), increasing the value of \( m \) should produce no significant decrease in the value of the residual energy produced by a given algorithm. Recently Bernasconi has found this to be the case for the problem of finding binary sequences with low off-peak autocorrelations [8]. He has proposed a cost function for this problem which is equivalent to the Hamiltonian of a one-dimensional spin system with long-range four-spin interactions. For such a system, Moore has found \( y = 1 \) [14]. This makes \( \lambda = 0 \). Thus, we predict that the use of multi-spin flip algorithms should not lead to significantly different results for this particular problem. This is in complete agreement with Bernasconi’s observations. Further, Bernasconi has pointed out that the ground state for this problem is extremely isolated in configuration space. Such a feature could be explained by equation (23), were it to be valid for this problem.

Next consider the travelling salesman problem. For this problem Karp [17] has suggested an algorithm which first divides the cities into groups of size \( m \). Next, the optimum tour for each group is found, and these are patched together to yield the final solution. It has been argued that for this problem the exponent \( y = 1 \), for any dimension greater than or equal to two [14]. However, in treating the travelling salesman problem it is more usual to apply the \( m \)-OPT algorithm without dividing the cities into groups. As a result it is possible to have spatially non-local moves. The derivation of the result \( y = 1 \), and the whole scaling approach, seem to require local moves. It is not clear to us what value the exponent \( y \) would have for non-local moves. We suggest, however, that \( y = 1 \) from the manner in which the tour evolves during the annealing process. Kirkpatrick has observed that for temperatures near zero the tour is optimal for very long distances with « mistakes » occurring in isolated local regions [5]. Thus, at least in the low temperature part of the simulated annealing process, moves are localized and « mistakes » move about until they collide and annihilate each other. Thus the argument
given in [14] should also apply in this case. With \( y = 1 \) and \( d = 2 \) we have \( \lambda = 1/2 > 0 \). Similarly for the one-dimensional Ising spin glass we have \( y = -1 \), hence \( \lambda > 0 \). Thus, for both problems some improvement is expected as \( m \) is increased. Again these predictions are in accord with Kirkpatrick's results for the TLS problem [5] and the results of the next section for the 1-d Ising spin glass problem.

A positive value of \( \lambda \) in itself does not justify the use of \( m \)-spin flip algorithms for a particular problem; in the case of many problems the increased computer time associated with the use of a more complex algorithm might be more than enough to offset any resulting improvements. To see this, first consider a quenching type algorithm based on single-spin moves only. The basic step for such an algorithm involves picking a single spin and moving this in such a way that no further reduction in the energy of the system is possible. For an Ising spin this means either flipping the spin or leaving the spin in its original state. In an analogous way the basic step for an \( m \)-spin flip quenched algorithm would involve choosing a cluster of \( m \)-spins and moving these spins until no further reduction in energy is possible either by flipping all the \( m \)-spins simultaneously or by flipping a sub-cluster of smaller size made up of these spins. This task can be performed by a transfer matrix method in a time bounded by \( \exp(m(d-1)/d) \).

Better methods than this may exist requiring times of order \( \exp(m^p) \), with \( p \approx (d - 1)/d \). This discussion is not restricted to the quenched algorithm and also applies to the simulated annealing method. Thus, the time per spin required by an \( m \)-spin flip simulated annealing process is expected to be \( \exp(m^p) \) times longer than a single-spin flip simulated annealing (with the same cooling rate). Therefore, one might ask whether it is more appropriate to use the single-spin flip algorithm with an increased cooling time rather than the \( m \)-spin flip algorithm. Huse and Fisher [15] have argued that for spin glasses the residual energy is given by \( \epsilon \sim 1/(\ln \tau)^{\xi} \) for long cooling times \( \tau \), where \( 0 < \xi \approx 2 \). (Note that this result is only confirmed by the numerical work of Grest et al., [20] for NP complete spin glass problems). Increasing the cooling rate by a factor \( \exp(m^p) \) gives an improvement of

\[
\frac{1}{\ln \{\tau \exp(m^p)^{\xi}\}} / \frac{1}{(\ln \tau)^{\xi}} \sim 1/m^{p\xi}
\]

for large values of \( m \). Alternatively using \( m \)-spin flip moves decreases the value of the residual energy by a factor \( 1/m^{1 - y/d} \). Hence, if \( \lambda = 1 - y/d > p \zeta \) the use of \( m \)-spin flip algorithms are justified. An example of such a problem is the 2-d Ising spin glass where \( p \zeta \approx 1 \) and \( \lambda \approx 1.15 \). It must be pointed out that these results are only valid in the limit of large \( m \) and long \( \tau \). For small values of \( m \) and practical values of \( \tau \), a small increase in \( m \), (for example changing \( m \) from 1 to 2), may still produce a worthwhile improvement in some cases, even when \( \lambda \) is not larger than \( p \zeta \).

5. Two spin-flip simulated annealing : results.

In this section we apply simulated annealing with two-spin flip dynamics to a one-dimensional Ising spin glass chain with a Hamiltonian given by (1). This allows us to test the relationship between the residual entropy and the density of the metastable states for a dynamic more complicated than single-spin flip. The algorithm we have used is essentially that of Metropolis [2]. At each stage a spin \( i \) is chosen randomly. Then a random number between 0 and 1 is picked. If the number is greater than \( p = 0.5 \) only spin \( i \) is flipped. Otherwise both spins \( i \) and \( i + 1 \) and flipped together. Next, the change in energy \( \Delta E \) is calculated and the move is accepted if \( \Delta E < 0 \). If \( \Delta E > 0 \) the move is only accepted with a probability \( \exp(-\beta \Delta E) \), where \( \beta = 1/T \). The ratio of the two-spin flip moves to the total number of moves can be altered by changing the value of \( p \). We found that the results did not depend strongly on the value of \( p \) when it lay between 0.25 and 0.75. For convenience and to make the number of one and two-spin flip moves equal, it was decided to keep \( p = 0.5 \) throughout the whole of the simulated annealing process for all system sizes. In order to ensure that the detailed balance equation is satisfied and that the algorithm does give sensible answers, two systems of sizes 2 500 and 200 were cooled from a high temperature of \( T = 2.0 \) down to \( T = 0.05 \) using the two-spin flip Monte Carlo algorithm described above. The values of the energy were sampled at temperature intervals of 0.05. These results were compared to the true equilibrium energy for each system, given by

\[
E_{eq}(T) = -\sum_{j=1}^{N} J_j \tanh \beta J_j .
\]

In both cases the Monte Carlo data agreed closely with the true values of energy down to temperatures of 0.25. Below this value, due to freezing effects, a difference was seen as expected.

A number of systems with sizes varying from 50 to 2 500 were cooled from a temperature of \( T = 1.0 \) down to \( T = 0 \) in intervals of 0.05 using a number of different cooling rates. In each case the value of the residual entropy was obtained using the relation

\[
S_{res}(T) = S(\infty) - \int_{0}^{\infty} \frac{E(T) - E(0)}{T^2} dT,
\]

where \( \epsilon_{res} = (E(0) - E_g)/N \); and \( E_g \) is the true
ground state energy, \(- \sum |J_i|, [1, 16]\). For temperatures below \(T = 1.0\) we used the values of \(E(T)\) obtained from the Monte Carlo simulation. Above \(T = 1.0\) we used the exact values of \(E(T)\) calculated from equation (31). For each given cooling rate, the values of \(S_{res}^2\) and \(e_{res}\) obtained were averaged over nine systems of equal size, each having a different set of bonds. This was done in order to reduce the sample to sample fluctuations. In figure 3 we have plotted \(S_{res}^2\) versus \(e_{res}\) for each case. The minimum value of the residual entropy was obtained for a system of size 200 with a cooling rate of 10,000 Monte Carlo steps per spin per temperature interval (which was fixed at \(\Delta T = 0.05\)). The maximum value was obtained for a system of size 2500 with a cooling rate of 30 MCS per spin per temperature interval. The theoretically calculated curve for \([\ln N_s(e)/N]\) versus energy, for the two-spin flip case, is also included in figure 3 for comparison. As can be seen the Monte Carlo results lie very close to the theoretically predicted curve, especially for slow cooling rates. Thus, it seems that the relationship equation (4) between the residual entropy and the number of metastable states is just as true for more complicated moves as well as for the single-spin flip case. Furthermore, comparing these results with those obtained by a simulated annealing using only one-spin flip moves [1] shows that the inclusion of two-spin flip moves produces a significant improvement. For example, for a cooling rate of 7 000 MCS/spin/temperature interval, the residual entropy of a system of size 2500 was found to be 0.076 with \(m = 2\). We estimate that one needs a cooling rate slower than \(10^6\) MCS/spin/temperature interval in order to obtain the same result with a one-spin flip simulated annealing. For a given cooling rate and a fixed system size, the average computer time required for a simulated annealing with \(m = 2\) was found to be about three times longer than that needed by a simulated annealing which only used single-spin flip moves.

We have also used our data to investigate the dependence of the residual entropy on the cooling rate. Unfortunately the theoretical predictions for this dependence are only valid in the limit of very long cooling times (slow cooling rates). On the other hand, it is only possible to use relatively fast cooling rates in simulation work. For spin glasses, Huse and Fisher have suggested \(\varepsilon \sim 1/(\ln \tau)^{4/3}\) as \(\tau \to \infty\) [15]. The exponent \(\zeta\) is predicted to have an upper bound of 2. Since \(S \sim \sqrt{\varepsilon}\) for small values of \(\varepsilon\) then: \(S_{res} \sim 1/(\ln \tau)^{4/3}\) as \(\varepsilon \to 0\). In the light of this result we have plotted \(1/S_{res}\) versus \((\log_{10} \tau)\) for a system of size 2500. Values of \(S_{res}\) for each cooling rate were calculated in the same way as described previously. It was found that a relation of the form \(S_{res} = 0.593/(\log_{10} \tau + 4.51)\) satisfies the data quite well. For \(\tau \gg 10^{4.51}\) this result is consistent with the predictions of Huse and Fisher with \(\zeta = 2\). However, it must be pointed out that our data are obtained over a limited range of cooling rates. To make a proper study of the cooling rate dependence of the residual entropy one needs a cooling rate far slower than those used here in order to ensure that the asymptotic regime has been reached.

6. Summary and conclusion.

A simulated annealing process involving two-spin flip moves as well as one-spin ones was applied to an Ising spin glass chain with nearest-neighbour interaction. The energy, \(E(T)\), was sampled at intervals of 0.05 starting from a temperature of \(T = 1.0\) down

![Fig. 3. — The square of the residual entropy per spin, obtained from the two-spin flip simulated annealing is plotted against the residual energy/spin, \(\varepsilon\). The solid curve represents the exact value of \([\ln N_s(\varepsilon)/N]\)² calculated for two-spin flip metastable states.

![Fig. 4. — Inverse of the residual entropy per spin plotted against the cooling rate for a system of size 2500. The cooling rate is defined as the number of MCS/spin per temperature interval, where the temperature intervals were equal to 0.05 for all points on the graph. The line through the points is that of the best fit.](image-url)
to $T = 0.0$. Using this data the residual entropy was calculated. The process was repeated for various cooling rates and system sizes. The residual entropies obtained in this way were plotted against the corresponding residual energies (energy at $T = 0$). The points lay closely on the analytically calculated number of two-spin flip metastable state versus energy curve. This explicit connection between the residual entropy and the number of local minima for a more complicated set of moves than the simple one-spin flip, suggests that the success of any $m$-spin flip algorithm depends on the way that it alters the number of metastable states. The determination of the density of metastable states is a very difficult task. However, it is possible to write a scaling expression, equation (3) for its behaviour using a zero-temperature scaling argument. Calculation of $\ln N_s(e)$ for the one-dimensional spin glass problem in the limits of large $m$, (where $m$ is the maximum number of spins flipped simultaneously), confirms the scaling expression. It turns out that the zero-temperature scaling exponent $y$ for a particular problem determines whether an increase in $m$ would produce any significant improvement in the results obtained for that problem. If $\lambda = 1 - \frac{y}{d}$ (where $d$ is the dimension of the problem) is zero or very small no improvement is expected; otherwise some improvement should be possible. This prediction seems to be in agreement with the various studies made on other optimization problems.

References