



HAL
open science

Residual entropy and simulated annealing

R. Ettelaie, M.A. Moore

► **To cite this version:**

R. Ettelaie, M.A. Moore. Residual entropy and simulated annealing. Journal de Physique Lettres, 1985, 46 (19), pp.893-900. 10.1051/jphyslet:019850046019089300 . jpa-00232915

HAL Id: jpa-00232915

<https://hal.science/jpa-00232915>

Submitted on 4 Feb 2008

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.

LE JOURNAL DE PHYSIQUE - LETTRES

J. Physique Lett. **46** (1985) L-893 - L-900

1^{er} OCTOBRE 1985, PAGE L-893

Classification
Physics Abstracts
05.20

Residual entropy and simulated annealing

R. Ettelaie and M. A. Moore

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

(Reçu le 17 juin 1985, accepté le 16 août 1985)

Résumé. — On montre que la détermination de l'entropie résiduelle, dans la méthode d'optimisation par recuit simulé, fournit des informations utiles sur l'état fondamental vrai. Un verre d'Ising unidimensionnel est étudié pour montrer un exemple de ce procédé et, dans ce cas, l'entropie résiduelle est reliée au nombre d'états métastables qui restent stables par retournement d'un seul spin. L'entropie résiduelle ne décroît que logarithmiquement vers zéro avec le taux inverse de refroidissement.

Abstract. — Determining the residual entropy in the simulated annealing approach to optimization is shown to provide useful information on the true ground state energy. The one-dimensional Ising spin glass is studied to exemplify the procedure and in this case the residual entropy is related to the number of one-spin flip stable metastable states. The residual entropy decreases to zero only logarithmically slowly with the inverse cooling rate.

The method of simulated annealing as a general procedure of optimization in problems involving many independent degrees of freedom, was first proposed by S. Kirkpatrick *et al.* in 1983 [1], and since then has received a great deal of attention. The method is particularly suited to the more difficult optimization problems which are characterized not only by the many degrees of freedom but also by conflicting constraints which cannot be simultaneously satisfied. In such cases one would like to find a solution which is the best compromise between all the different constraints. In order to attach a quantitative meaning to the goodness of a particular solution a cost function is defined. Such a cost function reflects the nature of each constraint and its relative importance. Then, the problem of finding the best solution becomes that of finding the solution which minimizes the cost function. This is not a trivial task. The effect of having a lot of constraints usually introduces many local minima (or metastable states) (often of order $\exp(N)$, where N is the number of degrees of freedom). Hence, the amount of computational time required to pick the exact optimal solution increases exponentially with N [1]. (This is strictly time for an NP complete problem. Not all problems with an exponentially large number of local minima are NP complete).

Obviously in any case, for large values of N , one has to resort to an approximate solution algorithm or heuristic.

Situations of this nature occur in many fields of science and engineering. In the field of condensed matter physics the problem of determining the ground state of a spin glass is of this kind. The design of integrated circuits [2], wiring of computers [3], and the N -city travelling salesman problem [4], are other well known examples.

Simulated annealing is based upon the connection between statistical mechanics and optimization problems [1]. In this method one treats the cost-function as the energy of the system. Introducing a parameter T , which has the same dimensions as the cost-function, one can simulate the system using the Monte Carlo method as in statistical physics. The parameter T plays the rôle of the temperature. If the transition probabilities are defined according to the Metropolis algorithm [5] then, for any given initial state the system will converge towards an equilibrium state such that the probability of occurrence of a given configuration is proportional to $e^{-\beta E}$, where E is the value of the cost-function for that configuration and $\beta = 1/T$. At low temperatures relaxation times become very long and the system gets trapped in a metastable state. The temperature at which this trapping effect starts to occur is known as the freezing temperature.

The method of simulated annealing involves setting up a Monte Carlo algorithm at a temperature well above the freezing temperature of the system and then lowering the temperature slowly, ensuring that at each stage the system reaches equilibrium before the temperature is lowered any further. The process of slow cooling becomes particularly important close to the freezing temperature where the system takes a long time to equilibrate. Once the temperature gets below the freezing temperature one can cool rapidly to zero temperature, since below the freezing point the system would be trapped in a given metastable state and is unlikely to escape. The point about the simulated annealing method is that it makes sure that such a metastable state is one which is close to the ground state, i.e. it is a near optimal solution, and if the equilibration can be achieved for all T , then the ground state is guaranteed as $T \rightarrow 0$.

This feature is not true of the more common heuristic methods. It is not difficult to see the reason for this difference. For example, consider the spin glass problem with Ising spins. In this case the best known heuristic method involves setting up the system in a random configuration (which is equivalent to having T very high where each configuration has an equal chance of occurring), examining each spin in turn and reversing if the total energy is reduced. The process is continued until the system reaches a one spin-flip stable metastable state where no further improvement is possible. Since the method starts from a random configuration, any one of the metastable states has a certain chance of being picked. Of course, the probability would not be equal for all metastable states and would depend on the number of configurations about a particular metastable state. For simplicity let us suppose that all metastable states have « gathering bins » of configurations of the same size. Then the chance of the system ending up in a metastable state with energy $N\varepsilon$ will be proportional to $N_s(\varepsilon)$ where $N_s(\varepsilon)$ is the number of metastable states of energy ε per spin above the true ground state energy. The form of $N_s(\varepsilon)$ for a one-dimensional Ising spin glass model is shown by the solid curve in figure 1. One would expect quite generally that most problems would have similar $N_s(\varepsilon)$ functions. In particular, the maximum of $N_s(\varepsilon)$ will occur at an energy well above the ground state energy, and so this method of minimization will hardly ever pick the ground state or states close to it, but instead picks a state close to the maximum. This is an inherent limitation of these types of iterative improvement methods [6].

Now consider the same method, but this time assume that the initial configuration could always be chosen such that it would have an energy ε_m . Applying the iterative improvement method would then produce a metastable state with an energy less or at worse equal to ε_m . In this way one has eliminated the probability of getting trapped at metastable states above ε_m . In order to ensure that the initial configuration has an energy equal or close to ε_m , one could use the Monte Carlo method. At temperatures above the freezing temperature, the system will equilibrate after a finite number of Monte Carlo steps. At equilibrium the probability of the system

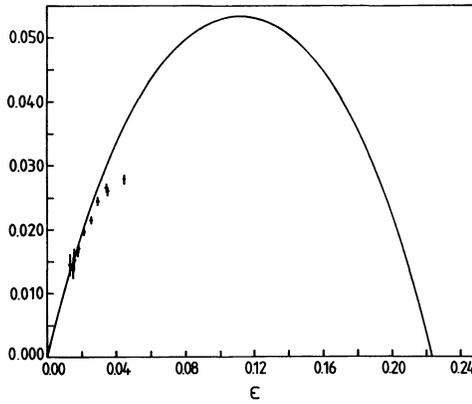


Fig. 1. — The square of the residual entropy per spin plotted against the residual energy/spin, ϵ . The solid curve represents the exact value of $[\ln N_s(\epsilon)/N]^2$ obtained from the analytic results, and is linear in ϵ for small ϵ .

having an energy ϵ would be proportional to $\Omega(\epsilon) e^{-\beta N \epsilon}$, where $\Omega(\epsilon)$ is the density of states and $e^{-\beta N \epsilon}$ the Boltzmann factor. From basic statistical physics, it is known that $\Omega(\epsilon) e^{-\beta N \epsilon}$ has a very sharp peak at a particular value of ϵ , say ϵ_m , which means that the system is almost always at a state with energy ϵ_m or very close to it. Thus, by equilibrating the system at a particular temperature using the Monte Carlo method and then applying the iterative improvement method we can ensure that the answers obtained would be better or at worse equal to ϵ_m . The value of ϵ_m is of course temperature dependent and can be reduced by lowering the temperature, as long as the temperature is above the freezing temperature. Below the freezing temperatures, due to the long relaxation times, it is not possible to achieve equilibrium without a very large number of Monte Carlo steps and enormous computing times, which obviously is not practical. In simulated annealing one slowly cools down the system to freezing temperatures making sure that the system is in equilibrium at every stage. By doing so one moves ϵ_m to the lowest practically obtainable value. Having done this, the system is rapidly cooled to $T = 0$ which is equivalent to applying an iterative improvement method. The application of the method to a number of optimization problems, including the design of integrated circuits [2], global wiring [3], the travelling salesman problem [4], quadratic sum assignment [7] and the minimum weighted matching problem [8], have all shown that simulated annealing can produce better results than those obtained by conventional methods.

In order to investigate the effects of the cooling rate on the method, we have applied the method of simulated annealing to a linear chain of N Ising spins with Hamiltonian

$$\mathcal{H} = - \sum_i J_i S_i S_{i+1}, \quad S_i = \pm 1, \tag{1}$$

where the exchange interactions J_i are independent random variables and are assumed to have a Gaussian distribution.

$$p(J_i) = (2 \pi (\Delta J)^2)^{-1/2} \exp \left[- \frac{J_i^2}{2(\Delta J)^2} \right]. \tag{2}$$

The problem of finding the ground state energy $E(0)$ of this system is in fact trivial. $E(0) = -\sum_{i=1}^N |J_i|$ as one chooses the orientation of the Ising spins across each bond to maximize each bond's contribution to $-E$. Thus the problem is not NP complete. However, this system does possess an exponentially large number of metastable states and when treated by the method of simulated annealing shows all the features of more complex situations — such as freezing and the existence of residual entropy. Residual entropy is the value of the entropy calculated at $T = 0$ from say, the behaviour of the energy as a function of temperature in the course of the annealing (see below for a precise definition). If one were able to equilibrate throughout the annealing process, the residual entropy would be zero (for all systems with a non-macroscopic ground state degeneracy). However, one finds a non-zero residual entropy for any finite rate of cooling. One also finds that the energy of the typical metastable state which one reaches for a given cooling rate (the « residual energy ») lies typically on energy $N\varepsilon$ above the true ground state energy $E(0)$. A slower cooling rate makes both ε and the residual entropy $S_{\text{res}}(\varepsilon)$ smaller. For this simple model problem we are able to show that $S_{\text{res}}(\varepsilon)/N \sim \sqrt{\varepsilon}$ as $\varepsilon \rightarrow 0$. While this particular relation is specific to one-dimensional Ising spin glasses one would expect a similar relation in other problems. Thus an estimate of the true ground state energy should be generally obtainable for any model by extrapolating $S_{\text{res}}(\varepsilon)$ to zero.

As predicted by Jäckle [9] we find that for our one-dimensional model

$$S_{\text{res}}(\varepsilon) = \overline{\ln N_s(\varepsilon)}, \quad \varepsilon \rightarrow 0, \quad (3)$$

where a bar denotes the average over the bonds $\{J_i\}$. At rapid cooling rates and thus large values of ε there are deviations from equation (3). We also observe that to reduce the residual entropy by a factor of k requires an increase of computer time by a factor of order at least $\exp(k)$.

We shall first present the calculation of $N_s(\varepsilon)$ — the number of metastable states of energy ε whose energy cannot be reduced by flipping a single spin. It is more convenient to work with the picture proposed by Fernandez and Medina [10] who focus on the bonds instead of the spins. When a bond is satisfied it does not contribute to the energy and when it is not satisfied it contributes $2|J_i|$. Three types of bonds can be identified, step bonds, top bonds and bottom (weak) bonds. In a metastable state all the bonds except the weak bonds are satisfied; weak bonds may or may not be satisfied. A weak bond is defined as one which is smaller in magnitude than both its neighbouring bonds. The probability for such a bond occurring is

$$\int_{-\infty}^{+\infty} \frac{\exp\left[-\frac{J_i^2}{2}(\Delta J)^2\right]}{\sqrt{2\pi}\Delta J} \left\{ 2 \int_{|J_i|}^{\infty} \frac{\exp\left[-\frac{J_i'^2}{2}(\Delta J)^2\right]}{\sqrt{2\pi}\Delta J} dJ_i' \right\}^2 dJ_i = \frac{1}{3}. \quad (4)$$

For a system of size N we would have $N/3$ bottom bonds on average and hence $2^{N/3}$ metastable states. The number of metastable states of energy $N\varepsilon$ above the ground can be written as

$$N_s(\varepsilon) = \sum_{s_j=\pm 1} \delta\left(N\varepsilon - \sum_{j=1}^{N/3} (1 + s_j) |J_j|\right). \quad (5)$$

Using the integral representation of the δ function, and calculating as usual $\overline{\ln N_s(\varepsilon)}$ rather than

$\overline{N_s(\epsilon)}$, we obtain

$$\begin{aligned} \overline{\ln N_s(\epsilon)} &= \left\langle \ln \left\{ \sum_{s_j = \pm 1} \int_{-\infty}^{+\infty} \frac{e^{iuN\epsilon}}{2\pi i} \prod_{j=1}^{N/3} e^{-iu(1+s_j)|J_j|} du \right\} \right\rangle_{\{J_j\}} \\ &= \left\langle \ln \left\{ \int_{-\infty}^{+\infty} \frac{e^{iuN\epsilon}}{2\pi i} \prod_{j=1}^{N/3} (1 + e^{-2iu|J_j|}) du \right\} \right\rangle_{\{J_j\}}, \end{aligned} \tag{6}$$

where the average is taken over all possible values of $N/3$ weak bonds. In order to perform this average, we note that

$$\prod_{j=1}^{N/3} (1 + e^{-2iu|J_j|}) = \exp \left\{ \sum_{j=1}^{N/3} \ln (1 + e^{-2iu|J_j|}) \right\}.$$

The sum is self-averaging, so we can write the above as

$$\exp \left\{ \frac{N}{3} \langle \ln (1 + e^{-2iu|J|}) \rangle_J \right\}. \tag{7}$$

The averaging $\langle \rangle_J$ is taken with respect to the probability distribution of the bottom (weak) bonds :

$$p_{\min}(J') = \frac{\exp \left[-\frac{J'^2}{2} (\Delta J)^2 \right]}{\sqrt{2\pi} \Delta J} \left\{ 2 \int_{|J'|}^{\infty} \frac{\exp \left[-\frac{J''^2}{2} (\Delta J)^2 \right]}{\sqrt{2\pi} \Delta J} dJ'' \right\}^2. \tag{8}$$

From now on for convenience we shall take ΔJ as unity. Substituting (7) into (6) and changing iu to u

$$\overline{\ln N_s(\epsilon)} = \ln \left\{ \int_{-i\infty}^{+i\infty} \frac{du}{2\pi} \exp \left\{ N\epsilon u + \frac{N}{3} \langle \ln (1 + e^{-2uJ}) \rangle_J \right\} \right\}. \tag{9}$$

The integral over u can now be evaluated for large N by the method of steepest descent. The results are shown in figure 1.

The number of metastable states has a maximum at an energy of about 0.112 above the ground state energy. This value agrees closely with that calculated by Fernandez and Medina [10] for the remanent energy after a rapid cooling. This result is not unexpected since after a rapid cooling the system would be trapped in one of the many metastable states which are symmetrically distributed near the maximum (Fig. 1). For this one-dimensional Ising spin glass it is easy to see using the arguments of [10] that each metastable state has the same *a priori* chance of being reached in a rapid quench from a random initial state.

With a simulated annealing method one hopes to obtain energies which are significantly closer to the ground state energy than 0.112. In order to do so an annealing schedule has to be defined. It is the normal practice to choose this annealing schedule such that, during the cooling process, the temperature is decreased by a constant ratio at each step [6]. However, for the purpose of this work we have found it more convenient to decrease the temperature by a fixed amount at each stage, all the way down to zero. This enables us to have a well defined cooling rate to which one can associate a residual entropy. According to equilibrium statistical mechanics the entropy

at temperature T , $S(T)$ is related to the heat capacity $C(T)$ by the relation

$$S(T) = S(\infty) - \int_T^\infty dT' \frac{C(T')}{T'}.$$

For our spin system $S(\infty)$ is readily evaluated and is just $N \ln 2$. Using $C(T) = \frac{\partial E}{\partial T}$ and integrating by parts gives

$$S(T) = S(\infty) - \int_T^\infty \frac{dT'}{T'^2} (E(T') - E(T)). \quad (10)$$

Following Jäckle [9] and Jäckle and Kinzel [11] we define an experimental or simulation entropy by performing the integral using the data for $E(T')$ obtained during the Monte Carlo annealing procedure. $S(T \rightarrow 0)$ should be zero for systems in equilibrium which obey the third law of thermodynamics i.e. for systems with a non-macroscopically degenerate ground state. However, as the Monte Carlo determined values of $E(T)$ lie above the equilibrium values of the energy, given by

$$E_{\text{eq}}(T) = - \sum_{i=1}^N J_i \tanh \beta J_i \quad (11)$$

for temperatures below the freezing temperature ($\simeq 0.3$) one finds that as $T \rightarrow 0$, $S(T) \rightarrow S_{\text{res}}(\varepsilon) \neq 0$. The residual entropy $S_{\text{res}}(\varepsilon)$ will only vanish if at the same time the residual energy $N\varepsilon$ is zero. The idea of determining the residual entropy as a test of the accuracy of the simulated annealing procedure was in fact suggested by its inventors [1].

Jäckle and Kinzel [11] attempted to verify equation (3) for the two-dimensional Ising spin glass but were handicapped by having only a rough knowledge of the form of $N_s(\varepsilon)$. We have tested the relation between the residual entropy and density of states for our one-dimensional Ising spin glass model. Systems of various sizes ranging from 50 to 10,000 spins were cooled from a starting temperature of $T = 1$ down to $T = 0$ at intervals of 0.05. The process was repeated for various cooling rates by changing the number of Monte Carlo steps per spin per temperature interval. Averaging the total energy over the Monte Carlo trajectory for each temperature interval, a value for $E(T)$ was obtained for each temperature. In each case the residual entropy was calculated using (10) from this data in the limit $T \rightarrow 0$. For $T > 1$, $E(T)$ was obtained from equation (11). The sample to sample errors were reduced by averaging the result over nine systems of the same size, but with a different set of bonds for each case. The residual entropies obtained in this way are plotted *versus* residual energies (the difference between the energy obtained for $T = 0$ at the end of cooling process and the exact ground state energy), in figure 1. The point with maximum residual entropy on the graph was obtained for a system of size 10,000 spins with a cooling rate as fast as 3 Monte Carlo steps per spin per each temperature interval and that with the lowest value was for a system of size 50 with 10,000 Monte Carlo steps per spin per temperature interval. For fast cooling rates the results show a clear deviation from the theoretically calculated curve. However, for slower cooling rates, the residual entropy seems to satisfy equation (3) remarkably well. These results and those of Jäckle and Kinzel [11] for the two-dimensional system suggest that perhaps such relations between the residual entropy and the number of metastable states of the residual energy might exist for any optimization problem with the same features. Since it is possible to calculate the residual entropy for any optimization problem for which the energy (or its equivalent) is well defined, graphs similar to that in figure 1 can always be obtained.

It might be that in some cases the extrapolation of such graphs down to zero residual entropy would provide a useful method for estimating the ground state energy. In the case of our data such extrapolation does not seem to be feasible unless a few more points with lower residual entropies could be obtained. This is a quite difficult task if one does not wish to use systems with sizes much less than 50, which are of less interest due to finite size errors (very few number metastable states). Slowing down the cooling process beyond a certain rate does not seem to produce any significant reduction in the residual entropy or energy. In order to make a more systematic study of this effect, we have plotted the residual entropy obtained for a variety of cooling rates, i.e. the number of Monte Carlo steps per spin per temperature interval (MCS/temperature interval) in figure 2.

All the results are obtained for a system of size 2,500 spins and temperature steps of 0.05 and were calculated in a similar way to that described before. The data suggest a logarithmic relation between the residual entropy and the number of MCS/temperature interval. The best line through the points gives $S_{\text{res}} = -0.016 \log n + 0.18$, « n » is the number of MCS/temperature interval [12]. Assuming that such relation holds down to zero residual entropy, the number of MCS/temperature interval required to obtain the ground state energy (i.e. $S_{\text{res}}(0) = 0$) would be $\sim 2 \times 10^7$ MCS/temperature interval. With temperature intervals spaced by 0.05, this is an inverse cooling rate of 4×10^{12} MCS/unit temperature. We believe that in practice the above relation does not hold down to zero entropy and that the straight line in figure 2 would eventually flatten out [11]. This means that even slower cooling rates than that estimated above would be required and that the method of simulated annealing is not in general able to produce the ground state energy. The more common annealing schedules, in which the cooling rate is not kept constant, do not change the situation as at no time, in the annealing schedule, cooling rates as slow as those above could be achieved. However, it must be said that the simulated annealing does produce results which are closer to the ground state energy than the normal quenched algorithm. This is specially true as the size of system becomes larger. For a system of size 10,000 a cooling rate of only 3 MCS/temperature interval, with temperature steps of 0.05, gives a residual energy of 0.04 which is still better than 0.1, the best result that was obtained by the quenched algorithm for the same amount of computer time.

In summary we have studied the dependence of the residual entropy on the number of metastable states of given residual energy for a one-dimensional Ising spin glass [13]. Good agreement exists between the data and the analytically calculated curve for the number of metastable states, at slow cooling rates. It was also found that the residual entropy has a logarithmic dependence on the inverse cooling rate which implies that an extremely long cooling process is required to obtain zero residual entropy. Since such slow cooling rates cannot be achieved in practice it

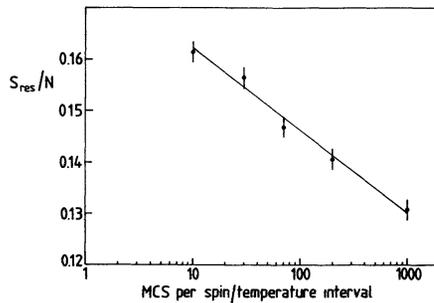


Fig. 2. — Dependence of the residual entropy on the cooling rate, defined as the number of the Monte Carlo steps/spin per temperature interval, for a system of size 2,500 spins. The temperature intervals were spaced by 0.05 for all the points on the graph. The line through the points is that of the best fit.

follows that a simulated annealing method is not really capable of producing the ground state energy. Even so, for large systems, the method of simulated annealing does produce results which are much better than those obtained by the quenched algorithm.

Acknowledgments.

We are indebted to Dr. A. J. Bray for helping us to calculate $\overline{\ln N_s(\varepsilon)}$.

References

- [1] KIRKPATRICK, S., GELATT, C. D. Jr. and VECCHI, M. P., *Science* **220** (1983) 671.
- [2] SIARRY, P. and DREYFUS, G., *J. Physique Lett.* **45** (1983) L-39.
- [3] VECCHI, M. P. and KIRKPATRICK, S., *Research Report RC 9555* (IBM Thomas J. Watson Research Center, Yorktown Heights, New York) 1982.
- [4] BONOMI, E. and LUTTON, J. L., *SIAM Rev.* **26** (1984) 551.
- [5] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E., *J. Chem. Phys.* **21** (1953) 1087.
- [6] KIRKPATRICK, S., *J. Stat. Phys.* **34** (1984) 975.
- [7] BONOMI, E. and LUTTON, J. L., to be published.
- [8] LUTTON, J. L. and BONOMI, E., to be published.
- [9] JÄCKLE, J., *Philos. Mag. B* **44** (1981) 533.
- [10] FERNANDEZ, J. F. and MEDINA, R., *Phys. Rev. B* **19** (1979) 3561.
- [11] JÄCKLE, J. and KINZEL, W., *J. Phys. A* **16** (1983) L-163.
- [12] Similar logarithmic dependences on the cooling rate have been suggested in other contexts. See for instance : GEMAN, S., GEMAN, D., « Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images » in *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** (1984) 721.
GIDES, B., « Non-Stationary Markov Chains and Convergence of the Annealing Algorithm » Rutgers preprint (1984).
- [13] The $\sqrt{\varepsilon}$ dependence of $S_{\text{res}}(\varepsilon)$ while certainly not a universal result, may, however, be quite common, since it is similar to an entropy $S(E)$ which behaves as \sqrt{E} or to a heat capacity which is linear in T . Such behaviour is found in a variety of amorphous systems including spin glasses and is attributed usually to the existence of two-level states : see : ANDERSON, P. W., HALPERIN, B. I. and VARMA, C. M., *Philos. Mag.* **25** (1972) 1.
PHILLIPS, W. A., *J. Low Temp. Phys.* **7** (1972) 351.