J. Phys. A: Math. Gen. **21** (1988) 271-284. Printed in the UK

Optimal storage properties of neural network models

E Gardner† and B Derrida‡

† Department of Physics, Edinburgh University, Mayfield Road, Edinburgh, EH9 3JZ, UK
‡ Service de Physique Theorique, CEN Saclay, F 91191 Gif sur Yvette, France

Received 29 May 1987

Abstract. We calculate the number, $p = \alpha N$ of random N-bit patterns that an optimal neural network can store allowing a given fraction f of bit errors and with the condition that each right bit is stabilised by a local field at least equal to a parameter K. For each value of α and K, there is a minimum fraction $f_{\min}$ of wrong bits. We find a critical line, $\alpha_c(K)$ with $\alpha_c(0) = 2$. The minimum fraction of wrong bits vanishes for $\alpha < \alpha_c(K)$ and increases from zero for $\alpha > \alpha_c(K)$. The calculations are done using a saddle-point method and the order parameters at the saddle point are assumed to be replica symmetric. This solution is locally stable in a finite region of the K,α plane including the line, $\alpha_c(K)$ but there is a line above which the solution becomes unstable and replica symmetry must be broken.

1. Introduction

It is well known in the theory of spin glasses that when one considers an Ising Hamiltonian,

$$\mathscr{H} = -\sum_{i,j} J_{ij} S_i S_j \tag{1}$$

with random interactions J_{ij}, that there are a lot of metastable states (Bray and Moore 1980, 1981, De Dominicis *et al* 1980, Ettelaie and Moore 1985, Derrida and Gardner 1986, Gardner 1986) (at least at zero temperature). By definition, a metastable state at zero temperature is a spin configuration $\{S_i\}$ such that,

$$S_i = \operatorname{sgn}\left(\sum_j J_{ij} S_j\right). \tag{2}$$

For randomly chosen interactions, J_{ij}, one can then try to calculate the number of these metastable states, the distribution of their energies, magnetisations, overlaps . . . .

In the present paper, we address a different question; given p spin configurations (which we call patterns), $\{S_i^\mu\}$ for $1 \leqslant i \leqslant N$ and $1 \leqslant \mu \leqslant p$, what is the probability that randomly chosen interactions J_{ij} have these patterns as metastable states? This probability is related in a simple way to the volume in the space of all J_{ij} in which equation (2) is satisfied for all sites i and all patterns μ. This question is an important one in the theory of neural networks where one usually represents the patterns one wants to store by spin configurations $\{S_i^\mu\}$ and the synapses by pair interactions, J_{ij}.

Recent work has dealt with cases where is a simple storage rule giving each J_{ij} as a function of the patterns $\{S_i^\mu\}$. For example, $J_{ij} = N^{-1} \sum_\mu S_i^\mu S_j^\mu$ (Hebb rule) (Hebb 1949, Little 1974, Hopfield 1982, Amit *et al* 1985a, b, 1987a, b), $J_{ij} = \operatorname{sgn}(N^{-1} \sum_\mu S_i^\mu S_j^\mu)$

or more complicated rules (Kohonen 1984, Personnaz *et al* 1985, Kanter and Sompolinsky 1987, Toulouse *et al* 1986, Parisi 1986, Mézard *et al* 1986). In general, when the number, $p = \alpha N$, of patterns is increased, there is a critical value α_c above which the system cannot store the patterns. α_c depends on the rule used to calculate the J_{ij}.

The method we use here is similar to a maximum entropy approach since we consider that nothing is known about the J_{ij} except that they try to make the patterns as metastable as possible. We will now define the problem more precisely. Consider p patterns, $S_i^\mu = \pm 1$ ($1 \leqslant i \leqslant N$ and $1 \leqslant \mu \leqslant p$). These patterns are chosen at random and remain fixed. For each choice of the J_{ij}, we call $p_w(\{J_{ij}\})$ the number of patterns, μ, such that a given site i is wrong,

$$p_w(\{J_{ij}\}) = \sum_{\mu=1}^{p} \left[1 - \theta\left(S_i^\mu \sum_j J_{ij} S_j^\mu \left(\sum_j J_{ij}^2 \right)^{-1/2} - K \right) \right]. \tag{3}$$

We say here that a site i is right for pattern μ if S_i^μ is not only parallel to the local field, $h_i^\mu = \Sigma_j J_{ij} S_j^\mu / (\Sigma_j J_{ij}^2)^{1/2}$ but if also $h_i^\mu S_i^\mu$ is greater than a minimal value K. Of course, a wrong site is by definition a site which is not right. For a given number p of random patterns, one can ask what is the volume in the space of the J_{ij} which gives a certain fraction f of wrong patterns at site i. If one tries to minimise f for a given p, then one has to solve an optimisation problem whose cost function is given by (3). It is clear that if a choice of J_{ij} gives a certain f, the choice λJ_{ij} gives exactly the same f. So one needs to choose a constraint on the J_{ij} in order to have a finite volume. In the present paper, two constraints will be considered, a spherical one in §§ 2 and 3,

$$\sum_j J_{ij}^2 = N \tag{4}$$

and an Ising one,

$$J_{ij} = \pm 1 \tag{5}$$

in § 4. The basic quantity that we will consider is a partition function $Z(h)$ defined by

$$Z(h) = \langle \exp[-h p_w(\{J_{ij}\})] \rangle_{\{J_{ij}\}} \tag{6}$$

where p_w is given by (3). The J_{ij} here play the role of the dynamical variables (which are in 'thermal equilibrium') whereas the S_i^μ are the quenched variables. Expression (6) can be easily transformed into

$$Z(h) = \left\langle \prod_{\mu=1}^{p} \left[e^{-h} + (1 - e^{-h}) \theta\left(S_i^\mu \sum_j J_{ij} S_j^\mu N^{-1/2} - K \right) \right] \right\rangle_{\{J_{ij}\}}. \tag{7}$$

Several quantities of interest can be computed from a knowledge of $Z(h)$. For example, the average, $\langle p_w \rangle = pf$, the number of wrong patterns at site i is given by

$$pf = \langle p_w \rangle = -\frac{\mathrm{d}}{\mathrm{d}h} \ln Z(h). \tag{8}$$

In the limit, $h \to \infty$, the minimal fraction $f_{\min}$ of wrong patterns at site i (i.e. the fraction f obtained for the optimal choice of the J_{ij}) is given by

$$pf_{\min} = \lim_{h \to \infty} -\frac{d}{dh} \ln Z(h). \tag{9}$$

It is clear from the above expression that the meaningful quantity to study is $\ln Z(h)$. We have here restricted consideration to the single-site problem. However, since all quantities we are interested in are derived from derivatives of $\ln Z(h)$, this problem is equivalent to the complete problem where the constraints (3) are fixed at all sites provided that the variables J_{ij} at different sites i are independent. This need not of course be true if, for example, a symmetry constraint $J_{ij}=J_{ji}$ is imposed.

In § 2, we calculate $\overline{\ln Z}$, the typical value of $\ln Z$, using the replica method (the bar means an average on the patterns) for the sphérical constraint (4). From this expression of $\ln Z(h)$, we obtain $f_{\min}$ as a function of $\alpha=p/N$ and K. We also give the expression $\alpha_c(K)$ such that $f_{\min}=0$ for $\alpha<\alpha_c$. In § 3, the stability of the replica symmetric solution of § 2 is analysed. The solution is always stable when $f=0$ (provided K is positive).

This seems reasonable because the space of solutions J_{ij} is connected; any two solutions can be continuously deformed into one another (i.e. there is a single valley). If, however, the mean fraction of wrong bits is positive, different solutions can correspond to errors in different bits and the solution space need not be connected (the set of solutions may be composed of many valleys). We find that the solution is stable in a finite region of the space, K, α, f. On the surface of this space where f is a minimum, the replica-symmetric solution is stable if α is small enough and there is a line above which it becomes unstable. This therefore provides a new example of a problem where replica symmetry has to be broken. In § 4, we will discuss the Ising constraint (5). Since the number of possible states of the J_{ij} is finite, the entropy can be calculated. We show that, in the replica symmetric approximation, the entropy or logarithm of the number of solutions at the minimum value of f is always $-\infty$ and the replica-symmetric solution is unstable. The calculation of the number of solutions is only valid in a finite region of the space K, f, α but not at values of α and K which give $f_{\min}(\geqslant 0)$.

2. The spherical model

In order to average $\ln Z(h)$, we use the replica method

$$\overline{\ln Z(h)}=\lim_{n\to 0}\frac{\overline{Z^n(h)}-1}{n}. \tag{10}$$

When one introduces replicas $(1\leqslant\alpha\leqslant n)$, each variable J_{ij} is replicated, J_{ij}^{α}, and with the spherical constraint (4), $\overline{Z^n(h)}$ is given by

$$\overline{Z^n(h)}=C\int\prod_{\alpha=1}^{n}\frac{\mathrm{d}\varepsilon_\alpha}{2\pi}\int\prod_j \mathrm{d}J_{ij}^{\alpha}\exp\mathrm{i}\varepsilon_\alpha\Big(\sum_j (J_{ij}^{\alpha})^2-N\Big)$$
$$\times\prod_\mu\Big[\mathrm{e}^{-h}+(1-\mathrm{e}^{-h})\theta\Big(N^{-1/2}S_i^\mu\sum_j J_{ij}S_j^\mu-K\Big)\Big] \tag{11}$$

where the constant C is just the normalisation of the volume of the space of J_{ij},

$$C=\Big[\frac{1}{2\pi}\int\mathrm{d}\varepsilon\int\prod_j\mathrm{d}J_{ij}\exp\mathrm{i}\varepsilon\Big(\sum_j J_{ij}^2-N\Big)\Big]^{-n}. \tag{12}$$

In the thermodynamic limit $(N\to\infty)$, $\overline{Z^n(h)}$ can be calculated using a saddle-point method since it can be written

$$\overline{Z^n(h)} = \int \sum_{\alpha<\beta} \mathrm{d}q_{\alpha\beta} \frac{\mathrm{d}\varphi_{\alpha\beta}}{(2\pi/N)} \prod_{\alpha} \frac{\mathrm{d}\varepsilon_\alpha}{2\pi} \exp N$$
$$\times \left(\frac{p}{N} G_0(\{q_{\alpha\beta}\}) + G_1(\{\varphi_{\alpha\beta}\}, \{\varepsilon_\alpha\}) + i \sum_{\alpha<\beta} \varphi_{\alpha\beta} q_{\alpha\beta} \right). \tag{13}$$

The derivation of (13) and the expressions for $G_0(\{q_{\alpha\beta}\})$ and $G_1(\{\varphi_{\alpha\beta}\}, \{\varepsilon_\alpha\})$ are given in appendix 1. If one assumes that $\ln Z(h)$ is given by a replica-symmetric saddle point in the limit $n \to 0$,

$$q_{\alpha\beta} = q \qquad \varphi_{\alpha\beta} = \mathrm{i}F \qquad \varepsilon_\alpha = \mathrm{i}E \tag{14}$$

then one gets (see (A1.17) and (A1.18)),

$$\frac{\ln Z(h)}{N} = \operatorname*{extr}_{E,F,q} \frac{1}{n} [\alpha G_0(q) + G_1(\mathrm{i}F, \mathrm{i}E) + \tfrac{1}{2} nFq] \tag{15}$$

where $\alpha = p/N$ and

$$G_0(q) = n \int_{-\infty}^{\infty} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}} \mathrm{d}t$$
$$\times \ln\left(1 + (e^{-h} - 1) \int_{-\infty}^{(K - tq^{1/2})(1-q)^{-1/2}} (2\pi)^{-1/2} \exp(-\lambda^2/2)\, \mathrm{d}\lambda \right) \tag{16}$$

$$G_1(\mathrm{i}F, \mathrm{i}E) = n[E + \tfrac{1}{2}\ln 2\pi - \tfrac{1}{2}\ln(2E+F) + \tfrac{1}{2}F/(2E+F)]. \tag{17}$$

This leads to the saddle-point equations

$$1 = \frac{2(E+F)}{(2E+F)^2} \tag{18}$$

$$q = \frac{F}{(2E+F)^2} \tag{19}$$

and

$$F + \alpha \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{n} G_0(q) \right) = 0. \tag{20}$$

From (18) and (19) it is easy to calculate E and F as functions of q, and one gets

$$F = \frac{q}{(1-q)^2} \qquad E = \frac{1-2q}{2(1-q)^2} \tag{21}$$

and therefore the replica-symmetric solution leads to

$$\frac{\overline{\ln Z(h)}}{N} = \operatorname*{extr}_{q} \left[\alpha \int_{-\infty}^{\infty} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}} \mathrm{dt} \right.$$
$$\times \ln\left(1 + (\mathrm{e}^{-h} - 1) \int_{-\infty}^{(K - tq^{1/2})(1-q)^{-1/2}} (2\pi)^{1/2} \exp(-\lambda^2/2)\, \mathrm{d}\lambda \right)$$
$$\left. + \tfrac{1}{2}\ln 2\pi + \tfrac{1}{2}\ln(1-q) + \tfrac{1}{2}(1-q)^{-1} \right]. \tag{22}$$

From this expression, one can deduce the minimal fraction f_{min} of wrong spins: in the limit $h \rightarrow \infty$ and $q \rightarrow 1$, one gets for the leading term

$$\frac{\overline{\ln Z(h)}}{N} = \operatorname*{extr}_{q}\left[\alpha\left(-\int_{K-[2h(1-q)]^{1/2}}^{K} \exp(-t^2/2)\frac{\mathrm{d}t}{(2\pi)^{1/2}}\frac{(t-K)^2}{2(1-q)} - h\int_{-\infty}^{K-[2h(1-q)]^{1/2}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\,\mathrm{d}t\right) + \tfrac{1}{2}(1-q)^{-1}\right] \tag{23}$$

which can be rewritten as

$$\frac{\overline{\ln Z(h)}}{N} = -h\alpha f_{\mathrm{min}} = h \operatorname*{extr}_{x}\left(-\alpha\int_{K-x}^{K}\frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}\frac{(t-K)^2}{x^2} - \alpha\int_{-\infty}^{K-x}\frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}} + \frac{1}{x^2}\right). \tag{24}$$

Setting the derivative of (24) with respect to x to zero, one finds that there is an optimal x only if $\alpha > \alpha_{\mathrm{c}}$ where

$$\alpha_{\mathrm{c}} = \left(\int_{-\infty}^{K}\frac{\mathrm{d}t}{(2\pi)^{1/2}}\exp(-t^2/2)(t-K)^2\right)^{-1}. \tag{25}$$

The function $\alpha_{\mathrm{c}}(K)$ is plotted in figure 1. $\alpha_{\mathrm{c}}(0)$ is equal to 2 in agreement with known results (Cover 1965, Venkatesh 1986a, b).

If $\alpha < \alpha_{\mathrm{c}}$, the extremum is given by

$$x = \infty \tag{26a}$$

and therefore

$$f_{\mathrm{min}} = 0 \tag{26b}$$

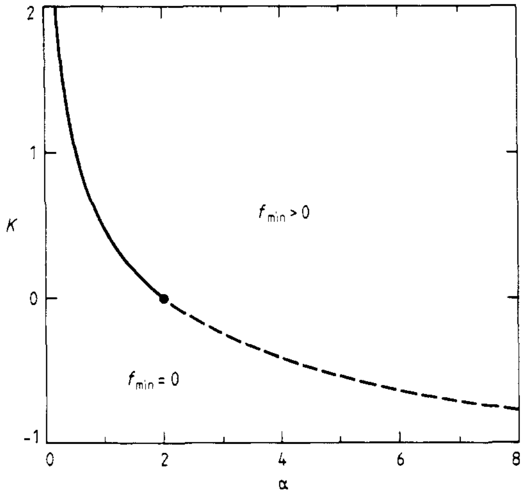

Figure 1. The critical line $\alpha_{\mathrm{c}}(K)$. For $\alpha < \alpha_{\mathrm{c}}(K)$, the minimum fraction of wrong bits $f_{\mathrm{min}} = 0$. The line is calculated assuming replica symmetry. For $K > 0$ (full curve) the replica-symmetric saddle point is stable around the line $\alpha_{\mathrm{c}}(K)$ whereas for $K < 0$ (broken curve) it is unstable.

whereas for $\alpha > \alpha_c$, x is a solution of

$$\alpha \int_{K-x}^{K} \frac{\mathrm{d}t}{(2\pi)^{1/2}} \exp(-t^2/2)(t-K)^2 = 1 \tag{27}$$

and $f_{\min}$ given by

$$f_{\min} = \int_{-x}^{K-x} \frac{\mathrm{d}t}{(2\pi)^{1/2}} \exp(-t^2/2) \tag{28}$$

$f_{\min}$ is plotted in figure 2 as a function of α for $K = 0, 0.5$ and 1.

These expressions of $\alpha_c(K)$ and $f_{\min}$ have been obtained assuming replica symmetry. We will see in the next section that the replica-symmetric solution becomes unstable when α increases. Nevertheless for $K > 0$, the whole line $\alpha_c(K)$ lies in a region where the replica-symmetric solution is stable. Recent numerical results (Krauth and Mézard 1987) seem to agree with expression (25).

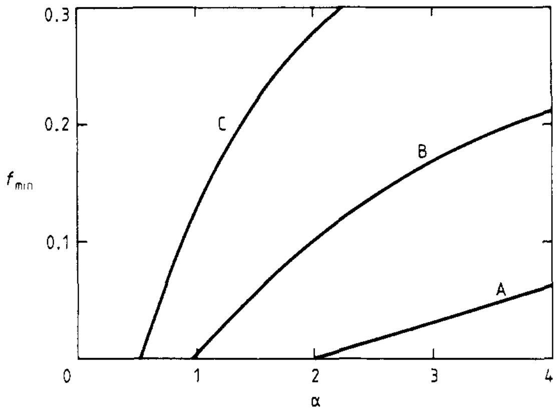

Figure 2. The minimum fraction, $f_{\min}$, of wrong bits as a function of α for $K = 0$ (curve A), 0.5 (B) and 1.0 (C).

3. Stability of the replica symmetric solution

An instability of the replica-symmetric solution (14) to the mean-field equations derived from equation (13) is determined by a sign change in (at least) one of the eigenvalues of quadratic fluctuations in the order parameters $q_{\alpha\beta}$, $\Phi_{\alpha\beta}$ and ε_α around the saddle point.

We will first consider the problem of diagonalising the matrices of second derivatives with respect to $q_{\alpha\beta}$ of G_0 and of second derivatives with respect to $\mathrm{i}\Phi_{\alpha\beta}$ and $\mathrm{i}\varepsilon_\alpha$ of G_1 separately. At the replica-symmetric saddle point, these matrices have the same symmetry properties as the matrix of quadratic fluctuations in the Edwards-Anderson (1975) order parameters in the Sherrington-Kirkpatrick (1975) model around the replica-symmetric solution. The eigenvectors (de Almeida and Thouless 1978) can be divided into two types: those which have the same eigenvalues as longitudinal fluctuations in the replica-symmetric space and those which are transverse to that space and have eigenvalues different from the longitudinal ones.

We will consider first the transverse eigenvectors and eigenvalues which are distinct from the longitudinal ones. The transverse eigenvectors of G_0 in terms of fluctuations in the variables $q_{\alpha\beta}$ are parallel to those of G_1 in terms of the variables $\mathrm{i}\Phi_{\alpha\beta}$ and have no component in the direction of fluctuations in the variables $\mathrm{i}\varepsilon_\alpha$. In each case there is a unique $\frac{1}{2}n(n-3)$-fold degenerate eigenvalue γ_1 and γ_2 respectively. These eigenvectors and eigenvalues are calculated in appendix 2.

The analysis of the longitudinal fluctuations could also be done. It turns out that there are three distinct eigenvalues in the longitudinal direction, and also a set of $3(n-1)$ transverse eigenvalues which are degenerate with the longitudinal ones. Since the replica-symmetric solution for E, F and q is unique, the solution should be stable with respect to fluctuations in this three-dimensional space, and therefore these eigenvectors should not lead to an instability.

The transverse eigenvalues of the quadratic fluctuations in the function

$$\alpha G_0(q_{\alpha\beta})+G_1(\varepsilon_\alpha,\varphi_{\alpha\beta})+\mathrm{i}q_{\alpha\beta}\varphi_{\alpha\beta} \tag{29}$$

on the right-hand side of equation (13) are therefore given by the two eigenvalues of the matrix,

$$\begin{pmatrix}\alpha\gamma_1 & 1\\ 1 & \gamma_2\end{pmatrix}. \tag{30}$$

In the limit $\alpha\to 0$, and hence $q\to 0$, the product of these eigenvalues is -1. The replica-symmetric result is correct in this limit since it is simply an integral over the phase space of couplings J_{ij}^α, and the result therefore must be stable. The sign is negative because of the change of variable from $\Phi_{\alpha\beta}$ to $\mathrm{i}\Phi_{\alpha\beta}$. The product can also be evaluated in the limit $\alpha\to\alpha_c$, $q\to 1$. A sign change in the product implies that one of the eigenvalues has changed sign for some value of α less than α_c. The replica-symmetric solution is unstable and the cost function calculations of the previous section therefore invalid if

$$\alpha_c\gamma_1,\ \gamma_2>1. \tag{31}$$

From appendix 2, (31) holds if

$$x\frac{1}{(2\pi)^{1/2}}\exp[(K-x)^2/2]>K\int_{K-x}^{K}\frac{\mathrm{d}t\exp(-t^2/2)}{(2\pi)^{1/2}}(K-t). \tag{32}$$

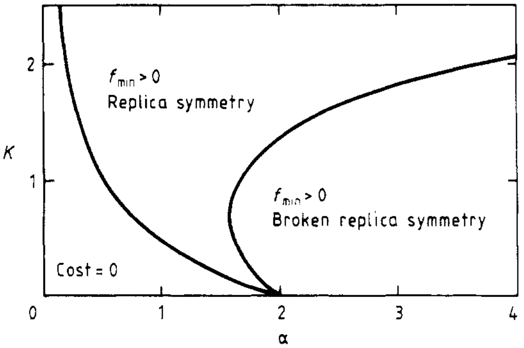

Figure 3. The critical line $\alpha_c(K)$ and the line where the replica-symmetric solution becomes unstable.

In the limit $x\to\infty$ or the fraction of errors $\to 0$, the product of eigenvectors has the same sign as at $\alpha=0$. However, as the fraction of errors increases, x decreases and the solution becomes unstable. For example, for small x ($x\ll K$) inequality (32) is always satisfied and the solution is always unstable. The instability line in the α,K plane is plotted in figure 3.

4. The Ising model

In this section, the calculation of § 2 will be repeated for the Ising constraint (equation (5)). Since the number of points in the phase space of interactions is finite for this constraint (for finite N), the total number Ω of solutions for a fraction f of wrong bits is given by

$$\overline{\ln\Omega}=\overline{\ln Z(h)}-h\frac{\overline{\mathrm{d}\ln Z(h)}}{\mathrm{d}h} \tag{33}$$

where h can be calculated as a function of f using equation (8). In the replica approach, we have again to consider $\overline{Z^n(h)}$,

$$\overline{Z^n(h)}=\int\prod_{\alpha<\beta}\mathrm{d}q_{\alpha\beta}\frac{\mathrm{d}\varphi_{\alpha\beta}}{2\pi}\exp N\left(\alpha G_0(\{q_{\alpha\beta}\})+G_1(\{\varphi_{\alpha\beta}\})+\mathrm{i}\sum_{\alpha<\beta}\varphi_{\alpha\beta}q_{\alpha\beta}\right) \tag{34}$$

where G_0 is given by equation (16) while G_1 is given assuming replica symmetry (14) by

$$G_1=n\left(\int_{-\infty}^{\infty}\frac{\mathrm{d}u\,\exp(-u^2/2)}{(2\pi)^{1/2}}\{\ln[2\cosh(u\sqrt{F})]-\tfrac{1}{2}F\}\right) \tag{35}$$

(see A1.21). The mean-field equation for F implies

$$F=\frac{2}{\pi(1-q)^2} \tag{36}$$

as $q\to 1$ and

$$G_1+\frac{nFq}{2}\to\frac{n}{\pi(1-q)}. \tag{37}$$

The cost function, f, can therefore be expressed as an extremum over x as in equation (24) except that the last x^{-2} term is multiplied by a factor $2/\pi$. This is equivalent to multiplying α by a factor of $\pi/2$, and so equation (27) becomes

$$\frac{\alpha_c\pi}{2}\int_{K-x}^{K}\frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\mathrm{d}t(t-K)^2=1 \tag{38}$$

while $f_{\min}$ is given by equation (28). In particular, in the limit $f_{\min}\to 0$, $x\to\infty$ and $K=0$, the upper storage capacity α_c is given by

$$\alpha_c=4/\pi. \tag{39}$$

In contrast to the spherical model calculation where there is a region of the α,K plane where the replica-symmetric solution is stable, and $f_{\min}>0$, the Ising calculation is always incorrect as $q\to 1$. Clearly equation (39) is incorrect since the maximum amount of information which can be stored must be less than the number of bonds, and so α_c must be less than 1. The entropy or logarithm of the number of solutions can be calculated and as $q \to 1$ or $f_{\min} \to 0$ we find

$$\overline{\ln \Omega}/N \sim \ln(1-q) \to -\infty \tag{40}$$

implying an infinite negative entropy. The stability calculation of § 3 can also be repeated for this case and the product of the distinct transverse eigenvalues tends to infinity in this limit, implying that there is a sign change in one of the eigenvalues for some value of q less than 1.

The Ising calculation, of the number of solutions Ω of equation (33), is valid only for sufficiently small values of q. Specifically, there is a surface in the α, f, K space on one side of which replica symmetry must be broken. In particular, the calculation of $f_{\min}$ always requires replica-symmetry breaking.

5. Conclusions

In this paper, we have calculated the fractional volume of phase space with a given value of K and of the fraction of wrong bits f. In particular the minimum cost function, $f_{\min}$, has been calculated. All of these results, however, have been obtained assuming replica symmetry for the order parameters at the saddle point. For both constraints—the spherical one (4) and the Ising one (5)—there is a surface in the space K, α, f on one side of which this replica-symmetric solution is unstable and the calculations must therefore be incorrect. For the spherical constraint, the calculations are stable in a finite region of the $K, f_{\min}$ plane around $f_{\min} = 0$ and $K > 0$, whereas in the Ising case, the solution is unstable throughout this plane.

The calculational method of § 2 has been used in other situations, for example, to deal with correlated patterns (Gardner 1987, 1988a). It can also be extended in order to calculate the size of basins of attraction in a diluted version of this model (Gardner 1988b) where finite values of $K > 0$ do imply finite basins of attraction which increase with the magnitude of K.

Since explicit solutions for the optimal J_{ij} do not exist, it is important to have numerical methods for constructing them. For $f = 0$, extensions of the perceptron learning algorithm (Rosenblatt 1962, Minsky and Papert 1969) do exist which converge to solutions provided these solutions exist (Gardner 1988a, Krauth and Mézard 1987). These algorithms therefore generate exact storage with finite basins of attraction provided such solutions exist. Other algorithms which aim to give solutions with finite basins of attraction have also been discussed (Gardner *et al* 1987, Diederich and Opper 1987, Poppel and Krey 1987).

Acknowledgments

We would like to thank H Gutfreund, M Mézard and G Toulouse for useful discussions. EG thanks the Service de Physique Theorique for their hospitality whilst in Saclay. Financial support for EG was from a Dewar Fellowship from the University of Edinburgh.

Appendix 1

In this appendix, we show how equation (11) leads to expressions (13) and (14) and we calculate the explicit expressions for the functions G_0 and G_1 for the replica-symmetric solution. If the θ functions in (11) are replaced by

$$\theta\left(\frac{1}{\sqrt{N}}\sum_j S_i^\mu J_{ij}^\alpha S_j^\mu - K\right) = \frac{1}{2\pi}\int_K^\infty \mathrm{d}\lambda_\mu^\alpha \times \int_{-\infty}^{\infty} \mathrm{d}x_\mu^\alpha \exp\left[\mathrm{i}x_\mu^\alpha\left(\lambda_\mu^\alpha - \sum_j J_{ij}^\alpha N^{-1/2} S_i^\mu S_j^\mu\right)\right]. \qquad \text{(A1.1)}$$

Then $\overline{Z^n(h)}$ becomes

$$\overline{Z^n(h)} = C\int \prod_\alpha \frac{1}{2\pi}\mathrm{d}\varepsilon_\alpha \int \prod_j \mathrm{d}J_{ij}^\alpha \exp\left[\mathrm{i}\varepsilon_\alpha\left(\sum_j (J_{ij}^\alpha)^2 - N\right)\right] \times \prod_\mu \left(\mathrm{e}^{-h}\int_{-\infty}^{K}\frac{\mathrm{d}\lambda_\mu^\alpha}{2\pi} + \int_K^\infty \frac{\mathrm{d}\lambda_\mu^\alpha}{2\pi}\right)\int_{-\infty}^{\infty}\mathrm{d}x_\mu^\alpha \times \exp\left[\mathrm{i}x_\mu^\alpha\left(\lambda_\mu^\alpha - \sum_j J_{ij}^\alpha N^{-1/2} S_i^\mu S_j^\mu\right)\right]. \qquad \text{(A1.2)}$$

Averaging over the patterns, S_j^μ, we get a coupling between replicas

$$\exp\left(-\mathrm{i}\sum_\alpha x_\mu^\alpha N^{-1/2} J_{ij}^\alpha S_i^\mu S_j^\mu\right) = \cos\left(\sum_\alpha x_\mu^\alpha J_{ij}^\alpha N^{-1/2}\right). \qquad \text{(A1.3)}$$

Where N is large, it is only necessary to keep the first term in the expansion of the cosine

$$\ln\cos\left(\sum_\alpha x_\mu^\alpha J_{ij}^\alpha N^{-1/2}\right) = -\frac{1}{2N}\left(\sum_\alpha x_\mu^\alpha J_{ij}^\alpha\right)^2 \qquad \text{(A1.4)}$$

and $\overline{Z^n(h)}$ therefore becomes

$$\overline{Z^n(h)} = C\int \prod_\alpha \frac{\mathrm{d}\varepsilon_\alpha}{2\pi}\int \mathrm{d}J_{ij}^\alpha \prod_\mu \left(\mathrm{e}^{-h}\int_{-\infty}^{K}\frac{\mathrm{d}\lambda_\mu^\alpha}{2\pi} + \int_K^\infty \frac{\mathrm{d}\lambda_\mu^\alpha}{2\pi}\right) \times \prod_\mu \int_{-\infty}^{\infty}\mathrm{d}x_\mu^\alpha \exp[\Phi(\{\varepsilon_\alpha\},\{J_{ij}^\alpha\},\{x_\mu^\alpha\},\{\lambda_\mu^\alpha\})] \qquad \text{(A1.5)}$$

where

$$\Phi = \mathrm{i}\sum_\alpha \varepsilon_\alpha\left(\sum_j (J_{ij}^\alpha)^2 - N\right) + \mathrm{i}\sum_\alpha\sum_\mu x_\mu^\alpha\lambda_\mu^\alpha - \frac{1}{2N}\sum_\mu\sum_j\left(\sum_\alpha x_\mu^\alpha J_{ij}^\alpha\right)^2 \qquad \text{(A1.6)}$$

which can be easily transformed into

$$\Phi = \mathrm{i}\sum_\alpha \varepsilon_\alpha\left(\sum_j (J_{ij}^\alpha)^2 - N\right) + \mathrm{i}\sum_\alpha\sum_\mu x_\mu^\alpha\lambda_\mu^\alpha - \tfrac{1}{2}\sum_\mu\sum_\alpha (x_\mu^\alpha)^2 - \sum_\mu\sum_{\alpha<\beta} x_\mu^\alpha x_\mu^\beta\left(\frac{1}{N}\sum_j J_{ij}^\alpha J_{ij}^\beta\right) \qquad \text{(A1.7)}$$

where we have used the fact that $\Sigma_j\,(J_{ij}^\alpha)^2 = N$. For each pair of replicas, one can then introduce

$$1 = \int \frac{\mathrm{d}q_{\alpha\beta},\,\mathrm{d}\varphi_{\alpha\beta}}{(2\pi/N)}\exp\left[\mathrm{i}\varphi_{\alpha\beta}\left(Nq_{\alpha\beta} - \sum_j J_{ij}^\alpha J_{ij}^\beta\right)\right]. \qquad \text{(A1.8)}$$

Then, one obtains

$$\overline{Z^n(h)} = \int \prod_{\alpha<\beta} \frac{\mathrm{d}q_{\alpha\beta}\,\mathrm{d}\varphi_{\alpha\beta}}{2\pi} \prod_{\alpha} \frac{\mathrm{d}\varepsilon_\alpha}{2\pi}$$

$$\times \exp N\left(\alpha G_0(\{q_{\alpha\beta}\}) + \mathrm{i} \sum_{\alpha<\beta} \varphi_{\alpha\beta} q_{\alpha\beta} + G_1(\{\varphi_{\alpha\beta}\}, \{\varepsilon_\alpha\}) \right) \tag{A1.9}$$

where

$$\exp G_0(\{q_{\alpha\beta}\}) = \prod_{\alpha} \left(e^{-h} \int_{-\infty}^{K} \frac{\mathrm{d}\lambda^\alpha}{2\pi} + \int_{K}^{\infty} \frac{\mathrm{d}\lambda^\alpha}{2\pi} \right)$$

$$\times \int_{-\infty}^{\infty} \mathrm{d}x^\alpha \exp\left(\mathrm{i}x^\alpha\lambda^\alpha - \tfrac{1}{2}(x^\alpha)^2 - \sum_{\alpha<\beta} x^\alpha x^\beta q_{\alpha\beta} \right) \tag{A1.10}$$

where p is the number of patterns and

$$\exp G_1(\{\varphi_{\alpha\beta}\}, \{\varepsilon_\alpha\}) = \prod_{\alpha} \mathrm{d}J_i^\alpha \exp\left(\mathrm{i} \sum_{\alpha} \varepsilon_\alpha [(J_i^\alpha)^2 - 1] - \mathrm{i} \sum \varphi_{\alpha\beta} J_i^\alpha J_i^\beta \right). \tag{A1.11}$$

Expressions (A1.10) and (A1.11) can be greatly simplified when the calculation is limited to the replica-symmetric subspace

$$q_{\alpha\beta} = q \qquad \alpha \neq \beta \tag{A1.12}$$

$$\varphi_{\alpha\beta} = \varphi = \mathrm{i}F \qquad \alpha \neq \beta \tag{A1.13}$$

$$\varepsilon_\alpha = \varepsilon = \mathrm{i}E \qquad \text{for all } \alpha \tag{A1.14}$$

and one obtains

$$G_0(q) = \ln\left\{ \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2/2}\,\mathrm{d}t}{(2\pi)^{1/2}} \left[1 + (\mathrm{e}^{-h} - 1) \right.\right.$$

$$\left.\left. \times \int_{-\infty}^{K} \frac{\mathrm{d}\lambda}{[2\pi(1-q)]^{1/2}} \exp\left(-\frac{(\lambda + t\sqrt{q})^2}{2(1-q)} \right) \right]^n \right\} \tag{A1.15}$$

and

$$G_1(\mathrm{i}F, \mathrm{i}E) = nE + \tfrac{1}{2}n \ln(2\pi) - \tfrac{1}{2}n \ln(2E + F) - \tfrac{1}{2} \ln[1 - Fn/(2E + F)]. \tag{A1.16}$$

In the limit, $n \to 0$, one gets,

$$G_0(q) = n \int_{-\infty}^{\infty} \frac{\exp(-t^2/2)\,\mathrm{d}t}{(2\pi)^{1/2}}$$

$$\times \ln\left(1 + (\mathrm{e}^{-h} - 1) \int_{-\infty}^{(K - tq^{1/2})(1-q)^{-1/2}} \exp(-\lambda^2/2)(2\pi)^{-1/2}\,\mathrm{d}\lambda \right) \tag{A1.17}$$

and

$$G_1(q) = n[E + \tfrac{1}{2} \ln 2\pi - \tfrac{1}{2} \ln(2E + F) + \tfrac{1}{2}F/(2E + F)]. \tag{A1.18}$$

This completes the calculation in the spherical case. In the Ising case, the calculation is almost the same. The differences are the following. The variables ε_α disappear everywhere (in (A1.2), (A1.5), (A1.6) and (A1.9)) so that G_1 in (A1.11) becomes a function of the $\varphi_{\alpha\beta}$ only. The integrals over the J_{ij}^{α} are replaced by sums over the two values +1 and −1. Therefore, (A1.11) becomes

$$\exp G_1(\{\varphi_{\alpha\beta}\}) = \sum_{J_\alpha=\pm 1} \exp\left(-\mathrm{i} \sum_{\alpha<\beta} \varphi_{\alpha\beta} J^\alpha J^\beta\right). \tag{A1.19}$$

Assuming the replica symmetry,

$$\varphi_{\alpha\beta} = \mathrm{i}F \qquad \alpha \neq \beta \tag{A1.20}$$

one gets

$$\exp G_1(F) = \exp(-\tfrac{1}{2}nF) \int \frac{\mathrm{d}u}{(2\pi)^{1/2}} \exp\{-u^2/2 + n \ln[2\cosh(u\sqrt{F})]\}. \tag{A1.21}$$

Everything else remains the same and, in particular, G_0 is unchanged.

Appendix 2

In this appendix, the transverse eigenvectors and eigenvalues of the matrices, $\partial^2 G_0/\partial q_{\alpha\beta}\,\partial q_{\gamma\delta}$ and $\partial^2 G_1/\partial(\mathrm{i}\varphi_{\alpha\beta})\partial(\mathrm{i}\varphi_{\gamma\delta})$ will be calculated. From equation (A1.10)

$$\frac{\partial^2 G_0}{\partial q_{\alpha\beta}\,\partial q_{\gamma\delta}} = \langle x^\alpha x^\beta x^\gamma x^\delta\rangle - \langle x^\alpha x^\beta\rangle\langle x^\gamma x^\delta\rangle \tag{A2.1}$$

where $\langle\ \rangle$ is defined

$$\begin{aligned}\langle f(x)\rangle \equiv \prod_\alpha &\left(\mathrm{e}^{-h}\int_{-\infty}^{K}\frac{\mathrm{d}\lambda^\alpha}{2\pi} + \int_K^\infty \frac{\mathrm{d}\lambda^\alpha}{2\pi}\right)\int_{-\infty}^{\infty} \mathrm{d}x^\alpha f(x)\\ &\times\exp\left(\mathrm{i}\lambda^\alpha x^\alpha - \tfrac{1}{2}\sum_\alpha (x^\alpha)^2 - \sum_{\alpha<\beta} q^{\alpha\beta}x^\alpha x^\beta\right)\\ &\times\left[\prod_\alpha\left(\mathrm{e}^{-h}\int_{-\infty}^{K}\frac{\mathrm{d}\lambda^\alpha}{2\pi} + \int_K^\infty \frac{d\lambda^\alpha}{2\pi}\right)\int_{-\infty}^{\infty}\mathrm{d}x^\alpha\right.\\ &\left.\times\ \exp\left(\mathrm{i}\lambda^\alpha x^\alpha - \tfrac{1}{2}\sum_\alpha (x^\alpha)^2 - \sum_{\alpha<\beta} q^{\alpha\beta}x^\alpha x^\beta\right)\right]^{-1}.\end{aligned} \tag{A2.2}$$

At the replica-symmetric saddle point, (A2.1) can take three possible values:

$$\begin{aligned}P &= \frac{\partial^2 G_0}{\partial q_{\alpha\beta}\,\partial q_{\alpha\beta}} = \int \frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}[(\overline{x^2}[t])^2] - \left(\int\frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}(\bar{x}[t])^2\right)^2\\ Q &= \frac{\partial^2 G_0}{\partial q_{\alpha\beta}\partial q_{\alpha\gamma}} = \int \frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}[\overline{x^2}(t)(\bar{x}[t])^2]\\ &\qquad -\left(\int\frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}(\bar{x}[t])^2\right)^2 \qquad \beta\neq\gamma\\ R &= \frac{\partial^2 G_0}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} = \int \frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}\bar{x}^4[t]\\ &\qquad -\left(\int\frac{\mathrm{d}t\,\exp(-t^2/2)}{(2\pi)^{1/2}}(\bar{x}[t])^2\right)^2 \qquad \alpha\neq\gamma,\ \beta\neq\delta\end{aligned} \tag{A2.3}$$

where $\overline{f(x)}[t]$ is defined

$$\overline{f(x)}[t]=\left(\mathrm{e}^{-h}\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{2\pi}+(1-\mathrm{e}^{-h})\int_{K-q^{1/2}t}^{\infty}\frac{\mathrm{d}\lambda}{2\pi}\right)\int_{-\infty}^{\infty}\mathrm{d}x\,f(x)\exp[\mathrm{i}x\lambda-\tfrac{1}{2}(1-q)x^2]$$
$$\times\left[\left(\mathrm{e}^{-h}\int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda}{2\pi}+(1-\mathrm{e}^{-h})\int_{K-q^{1/2}t}^{\infty}\frac{\mathrm{d}\lambda}{2\pi}\right)\right.$$
$$\left.\times\int_{-\infty}^{\infty}\mathrm{d}x\exp[\mathrm{i}x\lambda-\tfrac{1}{2}(1-q)x^2]\right]^{-1}\qquad\text{(A2.4)}$$

since the denominator of equation (A2.2) tends to one in the limit $n\to0$ and equation (A2.4) is obtained using the same method as in appendix 1. The matrix (A2.1) is thus of the same form as that in the de Almeida-Thouless calculation (1978) and the transverse eigenvectors $\eta^{\alpha\beta}$ have the form,

$$\begin{aligned}&\eta^{\alpha_0\beta_0}=c\\&\eta^{\alpha_0\alpha}=\eta^{\alpha\beta_0}=d\qquad\alpha\neq\alpha_0,\beta_0\\&\eta^{\alpha\beta}=e\qquad\alpha,\beta\neq\alpha_0,\beta_0\end{aligned}\qquad\text{(A2.5)}$$

where α_0 and β_0 are a particular pair of replicas. In order that the eigenvector is normal to the n eigenvectors which are degenerate with the longitudinal eigenvectors, the conditions

$$c=(2-n)d\qquad d=\tfrac{1}{2}(3-n)e\qquad\text{(A2.6)}$$

must be satisfied. Substitution of conditions (A2.6) into the eigenvalue equation gives an $n(n-3)/2$-fold degenerate eigenvalue

$$P-2Q+R=\int_{-\infty}^{\infty}\mathrm{d}t\exp(-t^2/2)(2\pi)^{-1/2}(\overline{x^2}(t)-(\bar{x}[t])^2)^2.\qquad\text{(A2.7)}$$

The evaluation of this integral in the limit $q\to1$ is similar to the method used in appendix 1. In this limit, the integral over t is dominated by values of t,

$$0<K-\sqrt{q}\,t<x\qquad\text{(A2.8)}$$

and

$$\bar{x}[t]=(1-\mathrm{e}^{-h})\mathrm{i}\exp[-(K-\sqrt{q}\,t)^2/2(1-q)]$$
$$\times[2\pi(1-q)]^{-1/2}\left(\mathrm{e}^{-h}+(1-\mathrm{e}^{-h})\int_{(K-q^{1/2}t)(1-q)^{-1/2}}^{\infty}\mathrm{d}\lambda(2\pi)^{-1/2}\mathrm{e}^{-\lambda^2/2}\right)^{-1}$$
$$\text{(A2.9)}$$

$$\overline{x^2}[t]=-(1-\mathrm{e}^{-h})\exp[-(K-\sqrt{q}\,t)^2/2(1-q)](K-\sqrt{q}\,t)[2\pi(1-q)^3]^{-1/2}$$
$$\times\left(\mathrm{e}^{-h}+(1-\mathrm{e}^{-h})\int_{(K-q^{1/2}t)(1-q)^{-1/2}}^{\infty}\mathrm{d}\lambda(2\pi)^{-1/2}\mathrm{e}^{-\lambda^2/2}\right)^{-1}\qquad\text{(A2.10)}$$

and so γ_1 diverges as $q\to1$,

$$\gamma_1(\alpha_c)=\frac{1}{(1-q)^2}\int_{K-x}^{K}\mathrm{d}t\exp(-t^2/2)(2\pi)^{-1/2}.\qquad\text{(A2.11)}$$

The calculation of the transverse eigenvalues of $\partial^2G_1/\partial\mathrm{i}\varphi_{\alpha\beta}\partial\mathrm{i}\varphi_{\gamma\delta}$ is similar. The eigenvectors are again given by (A2.5) and (A2.6) and the eigenvalue by $P'-2Q'+R'$

where

$$P' = \frac{\partial^2 G_1}{\partial(\mathrm{i}\varphi_{\alpha\beta})\partial(\mathrm{i}\varphi_{\alpha\beta})}$$

$$Q' = \frac{\partial^2 G_1}{\partial(\mathrm{i}\varphi_{\alpha\beta})\partial(\mathrm{i}\varphi_{\alpha\gamma})} \qquad \beta \neq \gamma \tag{A2.12}$$

$$R' = \frac{\partial^2 G_1}{\partial(\mathrm{i}\varphi_{\alpha\beta})\partial(\mathrm{i}\varphi_{\gamma\delta})} \qquad \alpha \neq \gamma, \beta \neq \delta$$

and γ_2 can be found easily

$$\gamma_2 = (1-q)^2. \tag{A2.13}$$

References

de Almeida J R and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
Amit D J, Gutfreund H and Sompolinsky H 1985a *Phys. Rev. Lett.* **55** 1530
—— 1985b *Phys. Rev.* A **32** 1007
—— 1987a *Ann. Phys., NY* **173** 30
—— 1987b *Phys. Rev.* A in press
Bray A J and Moore M A 1980 *J. Phys. C: Solid State Phys.* **13** L469
—— 1981 *J. Phys. C: Solid State Phys.* **14** 1313
Cover T M 1965 *IEEE Trans. Electron. Comput.* **EC-14** 326
De Dominicis C, Gabay M, Garel T and Orland H 1980 *J. Physique* **41** 933
Derrida B and Gardner E 1986 *J. Physique* **47** 959
Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
Ettelaie R and Moore M A 1985 *J. Physique Lett.* **46** 1893
Gardner E 1986 *J. Phys. A: Math. Gen.* **19** L1047
—— 1987 *Europhys. Lett.* **4** 481
—— 1988a *J. Phys. A: Math. Gen.* **21** 257
—— 1988b *Preprint* Edinburgh, in preparation
Gardner E, Stroud N and Wallace D J 1987 *Preprint* Edinburgh 87/394
Hebb D O 1949 *The Organisation of Behaviour* (New York: Wiley)
Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
Kohonen T 1984 *Self Organisation and Associative Memory* (Berlin: Springer)
Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745
Little W A 1964 *Math. Biosci.* **19** 101
Mézard M, Nadal J P and Toulouse G 1986 *J. Physique* **47** 1457
Minsky M and Papert S 1969 *Perceptrons* (Cambridge, MA: MIT Press)
Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L617
Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **16** L359
Poppel G and Krey U 1987 *Preprint* Regensburg
Rosenblatt F 1962 *Principles of Neurodynamics* (New York: Spartan Books)
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **32** 1792
Toulouse G, Dehaene S and Changeux J P 1986 *Proc. Natl Acad. Sci. USA* **83** 1695
Venkatesh S 1986a *Proc. Conf. on Neural Networks for Computing, Snowbird, UT (AIP Conf. Proc. 151)* ed J S Denker (New York: AIP)
—— 1986b *PhD thesis* California Institute of Technology