

## Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks

Daniel J. Amit and Hanoch Gutfreund

*Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel*

and

H. Sompolinsky

*Department of Physics, Bar Ilan University, Ramat Gan, Israel*

(Received 11 July 1985)

The Hopfield model for a neural network is studied in the limit when the number  $p$  of stored patterns increases with the size  $N$  of the network, as  $p = \alpha N$ . It is shown that, despite its spin-glass features, the model exhibits associative memory for  $\alpha < \alpha_c$ ,  $\alpha_c \geq 0.14$ . This is a result of the existence at low temperature of  $2p$  dynamically stable degenerate states, each of which is almost fully correlated with one of the patterns. These states become ground states at  $\alpha < 0.05$ . The phase diagram of this rich spin-glass is described.

PACS numbers: 87.30.Gy, 64.60.Cn, 75.10.Hk, 89.70.+c

Spin-glass models which exhibit features of learning, memory, and pattern recognition have become the focus of exciting numerical and analytical studies.<sup>1-4</sup> Of particular interest is Hopfield's model of associative memory.<sup>1</sup> In this model of neural network, a given set of patterns is embedded in the "synaptic" interactions between the neurons so as to make these patterns dynamically stable. A crucial issue is the storage capacity of the network. In a previous work,<sup>4</sup> we have established the stability of the embedded patterns, under the severe restriction that the number of stored patterns,  $p$ , remains finite as the size of the network,  $N$ , approaches infinity. On the other hand, there have been apparently conflicting statements regarding the capacity of the system as  $p$  increases to infinity with  $N$ .

Hopfield<sup>1</sup> concluded, on the basis of simulations and Gaussian noise arguments, that the system continues to provide associative memory for  $p \leq \alpha_c N$ , at  $T=0$ , with  $\alpha_c \approx 0.1-0.2$ , but degrades rapidly when  $p \geq \alpha_c N$ . In apparent contrast, Weisbuch<sup>5</sup> and Posner<sup>6</sup> have proved that at  $T=0$ , the original patterns will be locally stable only if  $p < N/(2 \ln N)$ .

In this Letter we study the model in the limit that  $p$  increases to infinity as  $p = \alpha N$ . We have determined the properties of the stable and metastable states of the system, thereby resolving the issue of its storage capacity. Apart from their relevance to neural networks the results reveal rich statistical mechanical properties which emerge from an unusual intertwining of ferromagnetic (FM) and spin-glass (SG) symmetry breaking. These features disappear for large  $\alpha$  and the model approaches the spin-glass model of Kirkpatrick and Sherrington.<sup>7</sup>

The network of  $N$  neurons is modeled by  $N$  spins,  $S_j = \pm 1$ . A pattern of neural firings corresponds to a spin configuration. The dynamics is modeled by a serial heat-bath Monte Carlo process,<sup>4</sup> governed by an

energy

$$H = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j \quad (1)$$

at temperature  $T (= \beta^{-1})$ . The couplings (synaptic efficiencies) are constructed of  $p$  given spin configurations (patterns), according to

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \quad (2)$$

with  $\xi_j^\mu (= \pm 1)$  quenched, independent, random variables.<sup>8</sup> They represent the  $p$  patterns, which were embedded in the system by a "learning" process. *Retrieval of memory is diagnosed as the dynamical persistence of a pattern, provoked by an external stimulus.* The discussion centers, therefore, around the nature of persistent states of the underlying dynamical process, and hence about the stable states of the free energy associated with  $H$ . Of particular pertinence are questions about the correlation of the dynamically stable states with the  $\xi^\mu$  and the dependence of those correlations on  $p$ .

For finite  $p$ , as  $N \rightarrow \infty$ , the situation is rather clear:<sup>4</sup> (1) At a critical temperature ( $T_c = 1$ ) the system undergoes a second-order phase transition, from a disordered phase to a phase of  $2p$  degenerate free-energy ground states—each one a Mattis<sup>9</sup> state, correlated with one of the embedded patterns  $\{\xi^\mu\}$ . With definition of the overlap of a state with the  $\nu$ th pattern as

$$m^\nu = N^{-1} \sum_i \langle S_i \rangle \xi_i^\nu, \quad (3)$$

where  $\langle . . . \rangle$  denotes a thermal average, the  $\mu$  Mattis state has  $m^\nu = m \delta^{\mu\nu}$ . The local magnetization is  $\langle S_i \rangle = \xi_i^\mu m$ , and  $m = \tanh \beta m$ . (2) Near  $T_c$ , the Mattis states are the only stable states. Below  $T = 0.46$  additional dynamically stable states appear, corresponding to well-defined mixtures of several patterns. The

number of these metastable states increases as  $T$  decreases. It also increases (at least exponentially) with increasing  $p$ . (3) The  $2p$  ground states as well as the metastable (mixture) states are separated by free-energy barriers of  $O(N)$ .

We now proceed to the case of a finite  $\alpha = p/N$ . We have evaluated the average free energy per spin,  $f = -\langle \langle \ln T \exp(-\beta H) \rangle \rangle / N\beta$ , of the Hamiltonian (1),(2) with  $p = \alpha N$  by the replica method. Here  $\langle \langle \dots \rangle \rangle$  denotes an average over the quenched disorder  $\{\xi_i^\nu\}$ . Since the system is fully connected,  $f$  can be calculated exactly in the  $N \rightarrow \infty$  limit by a mean-field theory. Most of our discussion will be within the replica-symmetric theory.<sup>7</sup> The occurrence and implications of replica-symmetry breaking<sup>10</sup> will be discussed at the end.

In the present model, the low-temperature phase will have weak random overlaps with most of the pat-

terns, each of which will be typically of  $O(1/\sqrt{N})$ . This can be realized from the fact that the zero-temperature energy per spin which is of order unity is just  $E = \frac{1}{2}\alpha - \sum_\mu (m^\mu)^2$ , where  $m^\mu$  are the overlaps defined in Eq. (3). However, it is possible that one or a finite number of overlaps condense macroscopically, i.e., retain a finite value as  $N \rightarrow \infty$ . We thus find three sets of order parameters: (i) the *macroscopic* overlaps  $m^\mu$  with the patterns  $\{\xi_i^\nu\}$ ,  $\nu = 1, 2, \dots, s$ . These will be denoted by an  $s$  component vector  $\mathbf{m}$ , where  $s$  remains finite as  $N \rightarrow \infty$ . (ii) The total mean square of the random overlaps with the other  $p - s$  patterns, denoted by  $r$ ,

$$r = \alpha^{-1} \sum_{\mu > s}^{N_\alpha} \langle \langle (m^\mu)^2 \rangle \rangle. \quad (4)$$

(iii) The Edwards-Anderson<sup>11</sup> order parameter  $q = \langle \langle \langle S_i \rangle^2 \rangle \rangle$ , which measures the local ordering. In terms of these quantities,  $f$  is given by

$$f = \frac{1}{2}\mathbf{m}^2 + \frac{1}{2}\alpha \left[ \beta^{-1} \ln[1 - \beta(1 - q)] + \frac{(1 - \beta)(1 - q)}{1 - \beta(1 - q)} + \beta r(1 - q) \right] - \beta^{-1} \langle \langle \ln 2 \cosh \beta [(\alpha r)^{1/2} z + \mathbf{m} \cdot \boldsymbol{\xi}] \rangle \rangle. \quad (5)$$

The average  $\langle \langle \dots \rangle \rangle$  in Eq. (5) and in the following stands for averaging over the discrete distribution of  $\boldsymbol{\xi}$  ( $\xi_i^\nu = \pm 1$ ,  $\nu = 1, \dots, s$ ) and over a Gaussian variable  $z$  with zero mean and unit variance. The saddle-point equations for the order parameters are

$$\mathbf{m} = \langle \langle \boldsymbol{\xi} \tanh \beta [(\alpha r)^{1/2} z + \mathbf{m} \cdot \boldsymbol{\xi}] \rangle \rangle, \quad (6)$$

$$q = \langle \langle \tanh^2 \beta [(\alpha r)^{1/2} z + \mathbf{m} \cdot \boldsymbol{\xi}] \rangle \rangle, \quad (7)$$

$$r = q[1 - \beta(1 - q)]^{-2}. \quad (8)$$

Note that the local field consists of two parts: A "ferromagnetic" part  $\mathbf{m}$  which results from the  $s$  condensed overlaps and a "spin-glass" part  $(r\alpha)^{1/2}z$ , generated by the sum of the overlaps with the rest of the patterns.

Equations (6)–(8) have two types of solutions which are locally stable to variations in  $q$ ,  $r$ , and  $\mathbf{m}$ : (1) A solution with  $\mathbf{m} = 0$ ,  $q, r \neq 0$ . It represents a SG state which does not have a macroscopic overlap with any of the patterns. It does not contribute to associative memory and is truly "spurious." (2) FM solutions with  $\mathbf{m} \neq 0$  in addition to  $q$  and  $r$ . These solutions, which exist for sufficiently small  $\alpha$ , make the system useful for associative memory.

The most important FM solutions are characterized by *macroscopic overlaps with a single pattern*,  $m^\nu = m\delta^{\nu p}$ . There are  $2N\alpha$  degenerate solutions of this type. As  $\alpha \rightarrow 0$  they approach the finite- $p$  Mattis states.

*Zero temperature.*—As  $\beta \rightarrow \infty$ , Eqs. (6)–(8) yield

$$m = \text{erf}[m/(2r\alpha)^{1/2}], \quad (9)$$

$q = 1 - CT$ , and  $r = (1 - C)^{-2}$ , with

$$C = (2/\pi r\alpha)^{1/2} \exp(-m^2/2r\alpha). \quad (10)$$

For  $\alpha > \alpha_c = 0.138$ , there is no solution with  $m \neq 0$ . As  $\alpha$  decreases below  $\alpha_c$ , two solutions with  $m \neq 0$  appear discontinuously. Out of the two, the one with the larger  $m$  is locally stable to variations in  $\mathbf{m}$ . This solution corresponds to a state which deviates only slightly from the precise stored pattern. The maximum deviation occurs at  $\alpha = \alpha_c$ , where  $m = 0.967$ , implying a 1.5% error. The percentage of errors decreases to zero very rapidly with decreasing  $\alpha$ , see Fig. 1. The energy spin of these states is (at  $T = 0$ )

$$E = \frac{1}{2}\alpha(1 - r) - \frac{1}{2}m^2. \quad (11)$$

At  $\alpha = \alpha_c$ ,  $E = -0.5014$  and it approaches  $-0.5$  as  $\alpha \rightarrow 0$ . For a fully correlated state ( $s_i = \xi_i^p$ ),  $m = 1$ ,  $r = 1$ , and  $E = -0.5$ . Thus, at finite  $\alpha$ , the system is able to slightly lower its energy by relaxing a small fraction of the spins, to accommodate for fluctuations in the overlaps of the other patterns.

As  $\alpha \rightarrow 0$ , Eq. (9) gives asymptotically,  $m \sim 1 - (2\alpha/\pi)^{1/2} \exp(-1/2\alpha)$ . The average number of errors is, therefore,

$$N_e \equiv \frac{1}{2}N(1 - m) = N(\alpha/2\pi)^{1/2} \exp(-1/2\alpha). \quad (12)$$

This result implies that the average fraction of errors

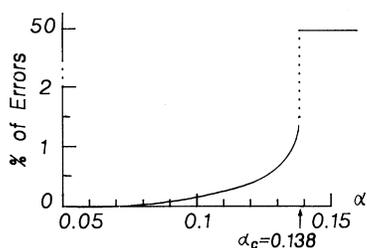


FIG. 1. Average percentage of errors in the FM states, as a function of  $\alpha$  at  $T=0$ .

$N_e/N$  vanishes as  $\alpha \rightarrow 0$  but the total number of errors vanishes only for  $\alpha < (2 \ln N)^{-1}$ , in agreement with the results of Refs. 5 and 6. Indeed, substituting  $\alpha = (2x \ln N)^{-1}$  in Eq. (12), one finds that  $N_e \sim N^{(1-x)}$ . Here we have shown that *with allowance for a small fraction of errors, very effectively controlled by small  $\alpha$ , the system provides a useful mechanism for associated memory*, at least up to  $\alpha_c \simeq 0.14$ , in agreement with the estimate of Hopfield.<sup>1</sup>

Note, that even for  $\alpha > \alpha_c$ , there may still exist spin states, with nonzero  $m$ , which are *stable to single spin flips*. These states, however, will decay at finite  $T$  much faster than the thermodynamically stable or metastable FM states below  $\alpha_c$ , which are surrounded by barriers of order  $N$ .

Equations (9) and (10) have a locally stable SG solution,  $m=0$ ,  $r = [1 + (2/\pi\alpha)^{1/2}]^2$  for all  $\alpha$ . Its energy, Eq. (11), equals  $E = -1/\pi - (\pi\alpha/2)^{1/2}$ . In the  $\alpha \rightarrow 0$  limit,  $E \rightarrow -1/\pi$ , and  $C = \beta(1-q) \sim 1$ . This limit coincides with the value of  $E$  of the finite- $p$  solutions, which mix  $n$  patterns, in the limit of  $n \rightarrow \infty$ .<sup>4</sup> This indicates that, as  $p \rightarrow \infty$ , the numerous states which mix infinitely many patterns merge to form the present SG phase. Comparing the energies of the SG and the FM states we find that the SG energy is lower for  $0.051 < \alpha < 0.138$ , whereas the FM state becomes the absolute minimum<sup>12</sup> below 0.051.

*Finite-temperature.*—The SG phase appears continuously as  $T$  decreases via a second-order transition. Expansion of Eqs. (9) and (10) for small  $q$  and zero  $m$  yields for the SG transition temperature the value of  $T_g = 1 + \sqrt{\alpha}$ . This phase is stable to the development of finite overlaps. The susceptibility of  $\mathbf{m}$  with respect to its conjugate field  $\mathbf{h}$ ,  $d\mathbf{m}/d\mathbf{h}$ , is positive, for all  $T < T_g$  in the SG phase. Indeed, above  $T=1$  other solutions do not exist for any  $\alpha$ . For  $\alpha < \alpha_c$ , FM states, with a single macroscopic overlap,  $m$ , appear discontinuously as  $T$  is decreased below  $T_M(\alpha)$ . The minimum value of  $T_M$  is 0.07, at  $\alpha = \alpha_c$ . As  $\alpha \rightarrow 0$ ,  $T_M$  increases to 1 and  $m(T_M)$  vanishes, thus approaching the finite- $p$  continuous transition; see Fig. 2.

Near  $T_M$ , the FM states are metastable. If

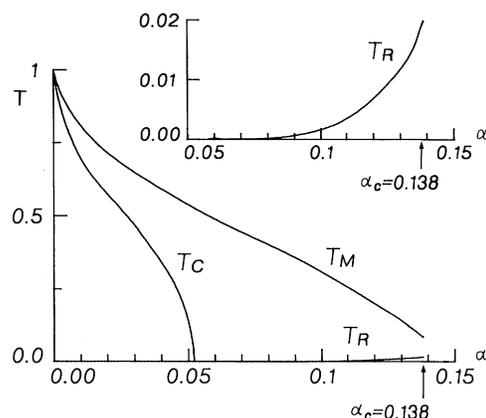


FIG. 2. Plots of critical temperatures of the FM states as a function of  $\alpha$ .  $T_M$  is the temperature at which FM states first appear.  $T_c$  is the first-order transition at which these states become global minima. Replica-symmetry breaking occurs below  $T_R$ , which is displayed on an expanded scale in the inset.

$\alpha < 0.051$ , they become the global minima of  $f$  below a temperature  $T_c(\alpha)$ . At  $T_c(\alpha)$ , there is a thermodynamic first-order transition from a SG to a FM state, accompanied by a jump in  $m$ ,  $q$ , and  $r$ , and by a latent heat. This transition is similar to that which occurs in a Landau theory of a scalar  $\phi^6$  model.<sup>13</sup>  $T_c(\alpha)$  increases from 0 at  $\alpha=0.051$  to unity at  $\alpha=0$ ; see Fig. 2.

Additional FM solutions appear discontinuously for sufficiently small  $\alpha$  and  $T$ . These solutions are characterized by  $\mathbf{m}$  with more than one nonzero component. Some of these “mixture” states are metastable but none is an absolute minimum at any  $T$ . This happens first below  $\alpha=0.03$ , where a locally stable solution with three symmetric overlaps [ $\mathbf{m} = m(1, 1, 1)$ ] appears. For example, at  $\alpha=0.02$  they appear below  $T=0.14$ . There are  $-(1/3!)(2N\alpha)^3$  degenerate metastable states of this type. In general, the critical values of  $T$  and  $\alpha$  below which mixture solutions exist decrease with increasing dimensionality  $s$  of  $\mathbf{m}$ , as  $\alpha, T_M^2 \propto 1/s$ . As  $\alpha \rightarrow 0$ , these solutions coincide with the finite- $p$  mixture states, described in Ref. 4. Note the similar role played here by  $T$  and  $\alpha$ . The presence of either a thermal noise ( $T$ ) or a static internal one ( $\alpha$ ) smooths the free-energy surface in  $\mathbf{m}$  space, leading to the successive elimination of FM states as either  $T$  or  $\alpha$  increases.

So far we have assumed that replica symmetry is unbroken. This is clearly incorrect at  $T=0$ , where the entropy per spin is  $S = -\frac{1}{2}\alpha[\ln(1-C) + C/(1-C)]$  with  $C \equiv \beta(1-q)$ . It is negative for all replica-symmetric solutions, and hence is unphysical. The condition for stability of the replica-symmetric solu-

tions is that the "eigenvalue"

$$\lambda = [1 - \beta(1 - q)]^2 - \alpha\beta^2 \langle (\text{sech}^4 \beta [(r\alpha)^{1/2} z + \mathbf{m} \cdot \boldsymbol{\xi}]) \rangle \quad (13)$$

be positive. We find that  $\lambda < 0$  in the SG solution, for all  $T < T_g$ . On the other hand, the FM solutions are stable in a finite temperature regime,  $T_R < T < T_M$ . The instability temperature  $T_R(\alpha)$  is shown in Fig. 2, for the states with single overlaps. It decreases very rapidly to zero with  $\alpha$ , as

$$T_R(\alpha) \sim (8\alpha/9\pi)^{1/2} \exp(-1/2\alpha).$$

This symmetry breaking is not expected to modify most of our results. The existence, at small  $\alpha$ , of SG and FM metastable, or stable, states still holds. Furthermore, the properties of the FM states, e.g., the average number of errors and the energy, are hardly affected, since their replica-symmetry breaking occurs (continuously) only at very low temperatures  $T_R$ , where the system is already almost completely ordered ( $q \sim 1$ ). This is also exemplified by the fact that the  $T=0$  negative entropy of the replica-symmetric FM states is extremely small. Already at  $\alpha_c$ ,  $S \sim -1.4 \times 10^{-3}$  in the FM state, compared with  $-7 \times 10^{-2}$  in the SG state and  $-1.6 \times 10^{-1}$  in the model of Kirkpatrick and Sherrington.<sup>7</sup>

The free energy of the solution with broken replica symmetry is expected to be higher than that of the replica-symmetric one. Hence, the values of  $T_c(\alpha)$  and the maximum value of  $\alpha$  for which the FM states become the absolute minima will be higher than the above estimates, Fig. 2. Also, with replica-symmetry breaking, the value of  $\alpha_c$ , where metastable FM states first appear, may be slightly higher than 0.138.

The most important implication of replica-symmetry breaking at low  $T$  concerns the structure of the states of the system. Calculations near  $T = T_g$  indicate that the symmetry-breaking scheme is of the same nature as that of the model of Kirkpatrick and Sherrington,<sup>7</sup> with continuous-order functions  $q(x)$  and  $r(x)$  ( $0 \leq x \leq 1$ ) replacing  $q$  and  $r$ .<sup>10</sup> According to the accepted interpretation of this scheme, the following remarkable organization of states, at small  $\alpha$  and  $T$ , emerges. States are first classified according to their macroscopic overlap  $\mathbf{m}$  with the embedded patterns. This already leads to enormous degeneracy, with  $2N\alpha$  degenerate states overlapping (almost completely) with a single pattern. States with different macroscopic  $\mathbf{m}$  are far apart from each other in phase space, and are separated by barriers of order  $N$ . Energy differences between nondegenerate states (e.g., SG and FM) are proportional to  $N$ . On a finer level, each of these macroscopically different stable and metastable phases actually represents (at  $T=0$ ) many (an infinity as  $N \rightarrow \infty$ ) degenerate states, organized in a hierarchi-

cal ultrametric structure.<sup>14</sup> These states have the same macroscopic properties but differ in the organization of their random components (i.e., the location of the "errors"). These macroscopically equivalent states are separated by barriers which are probably not higher than  $O(\sqrt{N})$ .

We have studied the effect of the application of external fields conjugate to  $\xi_i$ . Many new features appear. They will be described in an extended presentation of this work.

We have been stimulated by discussions with A. J. Bray, E. C. Posner, and G. Weisbuch. We thank J. J. Hopfield for helpful comments on the manuscript. The research of two of us (D.J.A. and H.S.) was supported in part by the Fund of Basic Research administered by the Israel Academy of Science and Humanities.

<sup>1</sup>J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982), and **81**, 3088 (1984); J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).

<sup>2</sup>W. A. Little, Math. Biosci. **19**, 101 (1974); W. A. Little and G. L. Shaw, Math. Biosci. **39**, 281 (1978).

<sup>3</sup>P. Peretto, Biol. Cybernet. **50**, 51 (1984), and in "Disordered Systems and Biological Organization," Proceedings of the Les Houches Meeting, February 1985, edited by E. Bienenstock *et al.* (to be published); M. Y. Choy and B. A. Huberman, Phys. Rev. A **28**, 1204 (1983).

<sup>4</sup>D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).

<sup>5</sup>G. Weisbuch, private communication, and in "Disordered Systems and Biological Organization," Proceedings of the Les Houches Meeting, February 1985, edited by E. Bienenstock *et al.* (to be published).

<sup>6</sup>E. C. Posner, private communication.

<sup>7</sup>S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).

<sup>8</sup>Most of our results depend on the discreteness of the  $\xi^\mu$ . The model with a Gaussian distribution of  $\xi^\mu$  was studied by J. P. Provost, to be published, and by A. J. Bray, private communication.

<sup>9</sup>D. C. Mattis, Phys. Lett. **56A**, 421 (1976).

<sup>10</sup>G. Parisi, Phys. Rev. Lett. **50**, 1946 (1983).

<sup>11</sup>S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).

<sup>12</sup>Note that unlike the SG order parameters,  $\mathbf{m}$  is determined by the *minimum* of the free energy.

<sup>13</sup>R. Bausch, Z. Phys. **254**, 81 (1972).

<sup>14</sup>M. Mezard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, J. Phys. (Paris) **45**, 843 (1984).