Pattern retrieval in an asymmetric neural network with embedded limit cycles

1989 J. Phys. A: Math. Gen. 22 L525

(http://iopscience.iop.org/0305-4470/22/11/013)

J. Phys. A: Math. Gen. 22 (1989) L525-L532. Printed in the UK

LETTER TO THE EDITOR

Pattern retrieval in an asymmetric neural network with embedded limit cycles

Yuhei Mori, Peter Davis and Shigetoshi Nara
ATR Optical and Radio Communications Research Laboratories, Sanpeidani, Inuidani, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan

Received 9 January 1989

Abstract. A neural network model in which the connection matrix is formed by summing direct products of successive patterns in cyclic sequences is studied. It is found that the noise elimination performance of this model can be better than that of an autocorrelation model due to a reduced tendency for trapping in spurious attractors.

The problem of recall of memorised patterns from noisy input in neural network models has a long history (Nakano 1972, Amari 1972, Anderson 1972, Kohonen 1972, Hopfield 1982). In particular, memory capacity and statistical properties of symmetric autoassociative networks have been extensively investigated (Amit 1987 and references therein). Hopfield (1982) introduced an energy-like Lyapunov function and a statistical physics methodology to describe the relaxation of symmetric networks. The performance of symmetric networks as memory is impaired by trapping at local energy minima. In memory models there is the problem of distinguishing between spurious minima and minima corresponding to stored memories. Much work has been done on the origins and nature of spurious minima (Amit 1987) and their temperature dependence (Amit *et al* 1985, 1987, Feigelman and Ioffe 1986). A small amount of noise can assist escape from shallow local minima. Annealing in which the thermal noise temperature is reduced according to some cooling schedule has been proposed to help convergence to deep minima (Ackley *et al* 1985). In asymmetric networks there is no known Lyapunov function guaranteeing convergence to a fixed pattern attractor. More complicated dynamical behaviour is possible. Some studies have considered the effect of a percentage of asymmetric connections on autocorrelation memory retrieval performance (Hopfield 1982). Adding a weak random asymmetry to a symmetric network can have an effect similar to increasing the level of noise (Feigelmann and Ioffe 1987). Other studies have used non-symmetric networks from a different point of view. Shinomoto (1987) used an assumption that each synapse must be either excitatory or inhibitory to form an asymmetric memory matrix for which the multiplicity of spurious attractors are removed—the system either converges to one of the memorised patterns or to a homogeneous 'don't know' state. Parisi (1986) proposed that temporal instability due to strong asymmetry could be useful in the learning process, with the metastability of memory states serving to distinguish them from other 'chaotic' states. Non-symmetric matrices, in which a symmetric contribution causes convergence toward a stationary pattern and an asymmetric part causes transitions between patterns, have been used

0305-4470/89/110525+08$02.50

in models of temporal pattern sequence generation. There are a number of models which either use non-symmetric synaptic interactions with temporal features such as dynamic synaptic strength (Peretto 1986), or time-delayed transmission (Kleinfield 1986, Sompolinsky and Kanter 1986, Gutfreund and Mézard 1988), or use thermal noise to induce transitions (Buhmann and Schulten 1987). These models and others (Personnaz *et al* 1986, Dehaene *et al* 1987, Guyon *et al* 1988) address the interesting issue of the possible roles and properties of pattern transitions, cycles and more complicated dynamical behaviour in neural networks.

In this letter, we address the question of whether retrieval performance in synchronous associative memories can be improved by storing patterns in the form of heterocorrelation cycles without an explicit autocorrelation part. In particular, will the tendency for trapping in spurious attractors increase or decrease?

We consider a model in which the synchronous time evolution of the firing pattern in a neural network is described by

$$S_i(t+1) = \operatorname{sgn}\left\{\sum_{j=1}^{N} J_{ij} S_j(t)\right\} \tag{1}$$

where

$$S_i = \begin{cases} 1 \\ -1 \end{cases} \qquad i = 1\text{-}N \tag{2}$$

and N is the number of neurons. The patterns to be memorised are used to form the matrix J by summing direct products of pairs of patterns $\xi^{\mu,\nu}$ and $\xi^{\mu,\nu+1}$:

$$J = \sum_{\mu=1}^{K} \sum_{\nu=1}^{L} \xi^{\mu,\nu+1} \otimes \xi^{\mu,\nu}. \tag{3}$$

The patterns are organised in cycles, as illustrated in figure 1. μ is the cycle label and ν is the pattern label understood to be taken mod L. L is the number of patterns in one cycle and K is the number of cycles. The total number of stored memory patterns is $M = KL$. The diagonal elements are set to zero, $J_{ii} = 0$, as in the biologically motivated autocorrelation models (Hopfield 1982). The synchronous autocorrelation model in which the matrix J is symmetric corresponds to the case $L = 1$. In the $L = 1$ case there are only fixed-point or, more rarely, period-two attractors (Goles-Chacc *et al* 1985). For $L > 1$, since the matrix J is typically not symmetric, there is the possibility of more complicated asymptotic orbits, such as cycles and chaos.

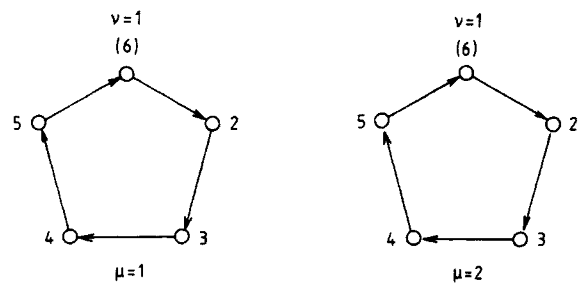

Figure 1. Example of cycles to be stored in a J matrix. μ is the cycle label and ν is the pattern label. Each direct product pair is shown by an arrow.

Starting from an initial pattern which is one of the memory patterns with added noise, the memory pattern or 'target' is retrived by iterating equation (1). Retrieval is successful if the noise is eliminated. Retrieval is unsuccessful when the iterated pattern converges to the same cycle as the target but with a phase slip, or to one of the other $L-1$ memorised cyles, or to some other 'spurious' attractor. Among the spurious attractors observed in numerical experiments are cycles with period L and, less often, cycles with longer period—typically small integral mutliples of L, but also, occasionally, very long cycles. Our object here is not to report details of the spurious attractors but to report a statistical evaluation of their effect on memory retrieval. In this sense we compare the noise elimination performance of the limit cycle model with the known performance of the autocorrelation model. When memory patterns are chosen randomly the retrieval performance is known to depend on the pattern ratio $\alpha = M/N$. In the autocorrelation case, at the intermediate value of $\alpha = \frac{3}{40}$ the existence of numerous spurious attractors significantly reduces the percentage of cases in which retrieval is successful (Kinzel 1985). We shall compre the performance of autocorrelation and cycle memory in this intermediate α case. We set $N = 400$ neurons and $M = 30$ patterns. The M stored patterns were chosen randomly, so their overlaps approximately obeyed a Gaussian distribution with zero average overlap.

Choosing one of the memory patterns as a target pattern, say $\xi^{R,r}$, pattern $\nu = r$ in cycle $\mu = R$, and choosing an overlap value q_0, we randomly generate an ensemble of initial states $S(0)$ with the specified overlap $q(0) = N^{-1}S(0) \cdot \xi^{R,r} = q_0$. Note that the overlap with the other patterns $(\mu, \nu) \neq (R, r)$ is $O(N^{-1/2})$. For each initial state we iterate equation (1) until the overlap defined as $q(t) = N^{-1}S(t) \cdot \xi^{R,r+t}$ converges to a cycle, or until the number of steps $t = 300$. We then average the overlap over L steps and call it the retrieval overlap. In defining the retrieval overlap we took into account two points. First, in the case of successful retrieval convergence is typically fast, occurring in $t = 3$-9 steps. Secondly, the measure of cases at this value of α with period greater than L is small. For example, in the case of initial overlap of 0.2, 76% go to memory cycles, 21% to spurious cycles of period L, 3% to longer cycles. The ensemble retrieval overlap obtained by averaging the retrieval overlap over the 5000 runs is plotted against intial overlap in figure 2 for the cases ($M = 30$, $L = 1$, $K = 30$), ($M = 30$, $L = 3$, $K = 10$) and ($M = 30$, $L = 10$, $K = 3$). The results in figure 2 show that the retrieval performance is better the longer the length L of the memory cycle.

Next we investigate this difference in performance by plotting in figure 3 a histogram of retrieval overlaps for the case of initial overlap $q_0 = 0.2$. In the autocorrelation case $L = 1$, 84% of trials resulted in convergence to patterns other than the target pattern. Most of these are spurious attractors with retrieval overlap about 0.4. The results for the cases of $L = 3$, $L = 10$ show that as cycle length increases the percentage of successful retrievals increases, and the percentage going to spurious attractors decreases. Moreover, the overlap of the spurious attractors with the target pattern decreases. In the case of $L = 10$, the retrieval overlap distribution of the 8% of trials which do not converge to the target is centred on zero overlap.

Why does retrieval performance improve with cycle length? Note that as the definition of retrieval overlap depends on the phase of the target cycle the improvement in performance is not just due to the decrease in number of attracting cycles. One way of thinking of the improved performance is in terms of a sort of pattern-shuffling effect. This is illustrated schematically in figure 4. The closed curves indicate the attracting basin boundary for a number of memorised patterns in the autocorrelation $L = 1$ case (Amari and Maginu 1988). The basins typically have complex shape and

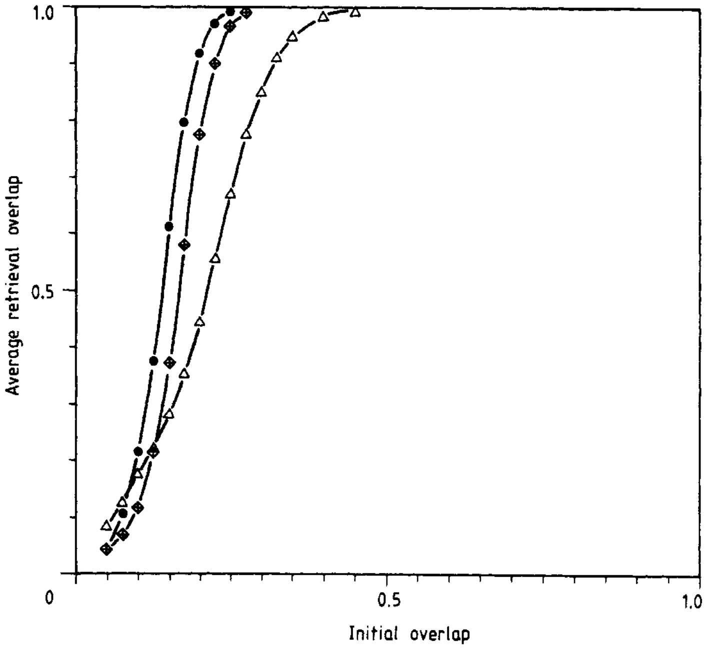

Figure 2. The ensemble retrieval overlap against initial overlap. The retrieval overlaps are averaged over the 5000 runs. △, autocorrelation; ◈, 3 patterns × 10 cycles; ●, 10 patterns × 3 cycles.

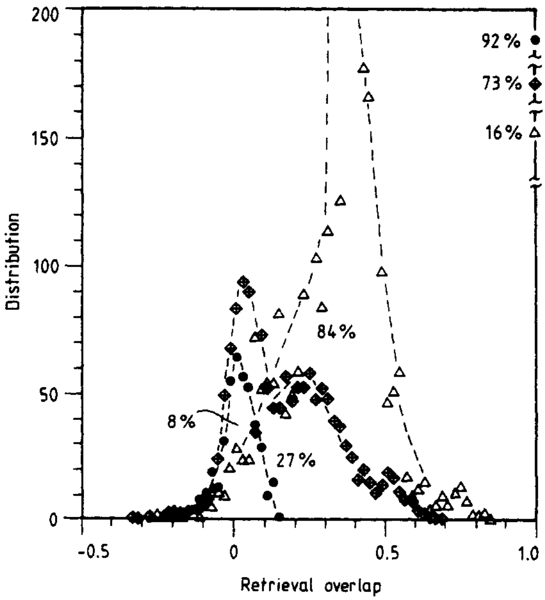

Figure 3. A histogram of retrieval overlaps for the case of initial overlap $q_0 = 0.2$. Symbols as in figure 2.

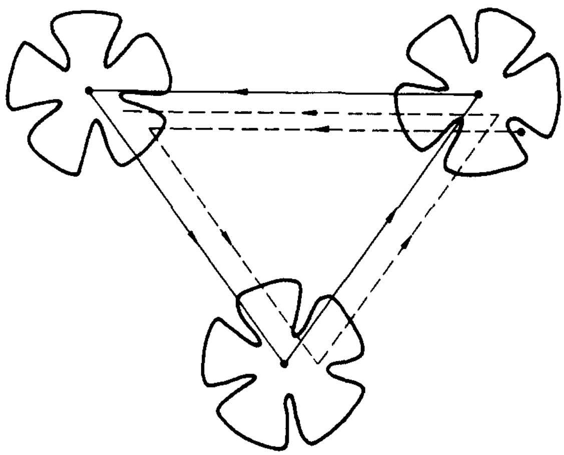

Figure 4. A schematic illustration of pattern shuffling in the cycle memory. The closed curves indicate the attracting basin boundary for each of the memorised patterns in the autocorrelation case. The full line indicates the converged orbital and the broken line the tangent orbital.

there are many orbits which are trapped close to a memorised pattern but outside its basin of attraction. If the memory patterns are instead stored in a cycle, orbits do not get trapped near the basin boundary of a single pattern, but visit the neighbourhoods of each memory pattern in the cycle, with a net increase in the number of orbits which converge to the target pattern. We suspect the shuffling of orbits due to cycling has a somewhat similar effect to that due to thermal noise in the $L=1$ case. We can express this idea a little more explicitly as follows. The evolution of the overlap with the target pattern is given by equations (4) and (5) for the autocorrelation and cycle cases respectively. For the autocorrelation case, assuming the target is pattern R and dropping the ν label,

$$q(R:t+1)=N^{-1}(\xi^R\cdot S(t+1))=N^{-1}\sum_{i=1}^{N}\operatorname{sgn}\left\{(\xi^R S(t)+\sum_{\mu\neq R}\xi_i^R\xi_i^\mu(\xi^\mu\cdot S(t)\right\}. \qquad (4)$$

For the cycle case, assuming the target pattern is pattern r in cycle R,

$$q(R,r:t+1)=N^{-1}(\xi^{R,k+1}\cdot S(t+1))$$

$$=N^{-1}\sum_{i=1}^{N}\operatorname{sgn}\left\{(\xi^{R,k}\cdot S(t)+\sum_{(\mu,\nu)\neq(R,k)}\xi_i^{R,k+1}\xi_i^{\mu,\nu+1}(\xi^{\mu,\nu}\cdot S(t)\right\}. \qquad (5)$$

The first term inside the curly bracket is the molecular field for the target pattern. The bigger the overlap the more dominant this term. The second term inside the curly bracket has the potential to detract from the first term. In the autocorrelation case $L=1$, it is possible that for some i in equation (4), $\xi_i^R\xi_i^\mu$ and $(\xi^{\mu,\nu}\cdot S(t))$ have opposite sign for a number of μ and thus that the second term inverts the effect of the first term even if the overlap $(\xi^{\mu,\nu}\cdot S(t))$ is small for $\mu\neq R$. Near a spurious attractor, this happens every iteration for particular i. However, in the cycle case, $L\neq 1$, since iteration replaces $\xi_i^{R,k+1}$ by $\xi_i^{R,k+2}$, it becomes less likely that, for successive iterations, $\xi_i^{R,k+1}\xi_i^{\mu\nu+1}$ and $(\xi^{\mu,\nu}\cdot S(t))$ have opposite sign for enough μ to result in the inversion of the first term. Thus we expect the reduction of the tendency for trapping in spurious attractors to be due to the dynamic self-averaging effect of the interchange of patterns in the second term. We believe a similar effect is seen in the multilayer feedforward network of Meir and Domany (1988). Use of memory cycles is a simple way of realising this effect in a single-layer recursive network. The modulation of the second term by pattern cycling has a somewhat similar effect to that of thermal noise aiding the escape from spurious local energy minima in the autocorrelation ($L = 1$) case. Since the significance of the modulation of the second term decreases as the pattern converges to the target, this effect can be thought of as 'self-annealing'.

Description of overlap dynamics using equations of type (4) and (5) have used various approximations to express this effect via the second term of the overlap with other patterns (Amari 1977, Kinzel 1985, Amari and Maginu 1988, Meir and Domany 1988). Amari (1977) and Kinzel (1985) derived the following equation for the case when the second term in equations (4) and (5) is assumed to be a random term obeying a Gaussian distribution:

$$q(t+1) = \mathrm{erf}(q(t)/\sigma) \qquad \sigma = \sqrt{2M/N}. \tag{6}$$

At the level of equation (6), the difference between the autocorrelation and cycle cases cannot be seen. Amari (1988), approximately in the autocorrelation case, and Meir and Donany (1988), exactly in the multilayer case, have improved on (6) by deriving step-dependent $\sigma(t)$. To examine the dynamical effect of the second term in equations (4) and (5) we look at the overlap dynamics in our numerical example with $N = 400$, $M = 30$ using a Lorentz plot of overlap at time t against overlap at time $(t-1)$. This is shown in figure 5, with overlap at t expressed as $\theta = \cos^{-1}(q(t))$. Figures 5(a, b) correspond to the cycle case $L = 10$, and 5(c, d) correspond to the autocorrelation case. The results for initial overlaps $q_0 = 0.2$ and 0.4 are represented. In each case the orbits for 100 different initial patterns are plotted. For initial overlap 0.4 most orbits monotonically converge to the target pattern, where $\theta = 0$. For initial overlap 0.2, the convergence is not necessarily monotonic, and a proportion of orbits converge to spurious attractors. The full line in the figures corresponds to equation (6). In each of the four examples in figure 5, the first iteration is consistent with the formula in equation (6), as the initial patterns are chosen randomly. However, from the second iteration the convergence is slower (i.e. closer to the line $\theta_{t+1} = \theta_t$) than the full line due to the appearance of non-zero overlaps with patterns other than the target pattern, and the corresponding increase in the value of the second term in equations (4) and (5). Near the target pattern, $\theta \simeq 0$, where the value of the second term is small, the Lorentz orbit again approaches that for the full curve. Comparing the plots for limit cycle and autocorrelation cases, we see that when the overlap is large the orbits in the two cases in figures 5(a, c) are similar, but when the initial overlap is small as in 5(b, d) and the effect of the second term is more significant the difference in orbits is considerable. The convergence is slower for the autocorrelation case. In many cases the orbit gets trapped in a spurious attractor near $\theta = 50°$. For the cycle case, however, over 90% of orbits go to the target at $\theta = 0$. The remainder go to spurious attractors near $\theta = 90°$.

Finally, let us comment on the models of sequential transition with delay. In this class of models (Kleinfield 1986, Sompolinsky and Kanter 1986, Gutfreund and Mézard 1988), in addition to a autocorrelation component associated with $q(t)$ there is a hetero component, relatively weighted by a factor λ, associated with $q(t-\tau)$, where τ is the delay. For large enough delay $\tau \gg 1$, the system settles at one target pattern for a time and then makes a transition to another. The behaviour on long timescales can be described by a strobe map relating the overlaps on successive patterns. The strobe

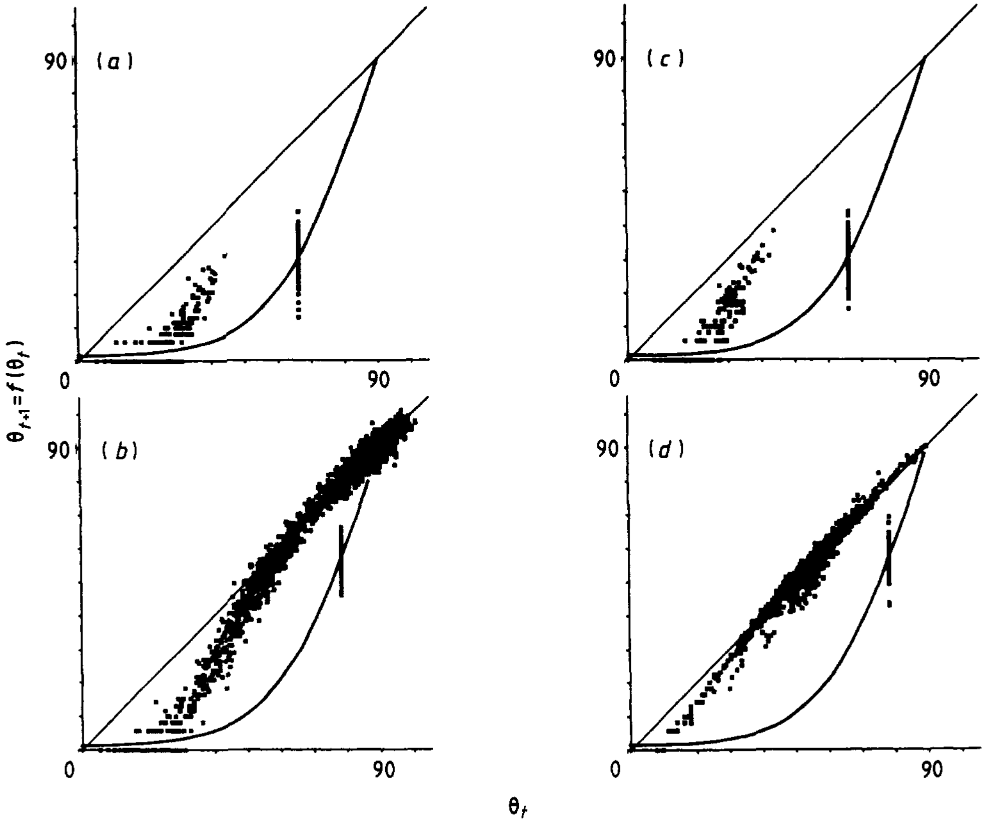

Figure 5. Lorentz plots of overlap at time t against overlap at time $(t-1)$. The overlap at t is expressed as $\theta = \cos^{-1}(q(t))$. Here, (a) and (b) correspond to the cycle case $L = 10$, and (c) and (d) correspond to the autocorrelation case. The results for initial overlaps $q_0 = 0.2$ are presented in (b) and (d) and for $q_0 = 0.4$ in (a) and (c).

map obtained in a fully diluted model (Gutfreund and Mézard 1988) corresponds to (6) in the limit of infinite λ. Also, a correspondence valid for all λ holds at the fixed points, with a redefined factor σ which includes λ. The threshold value of M/N below which there is a non-zero fixed point in equation (6) is one measure of capacity. It is found that this capacity improves for finite λ (Gutfreund and Mézard 1988), meaning that it is possible to store more patterns by storing them in sequences. This improvement in capacity seen already at the level of approximation of equation (6) is of different origin to the improvement in retrieval characteristics demonstrated in our model which is due to reduction of trapping in spurious attractors. The addition of thermal noise has been found to help avoid spurious attractors in a delay-type model (Gutfreund and Mézard 1988). It would be productive to investigate the retrieval performance and nature of spurious attractors in the delay-type models in the light of our results.

In conclusion we have shown on the basis of numerical experiment that retrieval performance in associative memories can be improved by storing patterns in the form of heterocorrelation cycles rather than in autocorrelation form and that this is due to the reduction of the tendency for trapping in spurious attractors. A qualitative explanation of this effect was given in terms of pattern cycling causing 'self-annealing'. A more rigorous analysis of this effect is left for future work. In related work we have studied a cycle model in which varying a system parameter (such as synaptic connection range) causes the destabilisation of the cycles and bifurcation to chaos (Nara and Davis 1989). We aim to apply complex dynamical memory structure, including multiple-period spurious attractors and chaos, in experiments with novel non-von Neumann information processing functions.

The authors are grateful to S Shinomoto and K Ikeda for stimulating discussions on related topics in neural nets and complex dynamics.

References

Ackley D H, Hinton G E and Sejnowski T J 1985 *Cog. Sci.* **9** 147
Amari S 1972 *IEEE Trans. Systems, Man, Cybern.* **SMC-2** 643
—— 1977 *Biol. Cybern.* **26** 175
Amari S and Maginu K 1988 *Neural Networks* **1** 63
Amit D J 1987 *Proc. Int. Symp. on the Physics of Structure Formation, Tubingen, 1986* (Berlin: Springer)
Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007
—— 1987 *Ann. Phys., NY* **173** 30
Anderson J A 1972 *Math. Biosci.* **14** 197
Buhmann J and Schulten K 1987 *Europhys. Lett.* **4** 1205
Dehaene S, Changeux J P and Nadal J P 1987 *Proc. Natl Acad. Sci. USA* **84** 2727
Feigelman M V and Ioffe L B 1986 *Europhys. Lett.* **1** 197
—— 1987 *Int. J. Mod. Phys.* B **1** 51
Goles-Chacc E, Fogelman-Soulie F and Pellegrin D 1985 *Discrete Appl. Math.* **12** 261
Gutfreund H and Mézard M 1988 *Phys. Rev. Lett.* **61** 235
Guyon I, Personnaz L, Nadal J P and Dreyfus G 1988 *Phys. Rev.* A **38** 6365
Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
—— 1985 *Biol. Cybern.* **52** 141
Kinzel W 1985 *Z. Phys.* B **60** 205
Kleinfeld D 1986 *Proc. Natl Acad. Sci. USA* **83** 9469
Kohonen T 1972 *IEEE Trans. Comput.* **C-21** 353
Meir R and Domany E 1988 *Phys. Rev.* A **37** 608
Nakano K 1972 *IEEE Trans. Systems, Man, Cybern.* **SMC-2** 380
Nara S and Davis P 1989 *Attractor transitions in a neural network with embedded limit cycles* in preparation
Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675
Peretto P and Niez J J 1986 *Disordered Systems and Biological Organization* ed E Bienenstock, F Fogelman-Soulié and G Weisbuch (Berlin: Springer)
Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev.* A **34** 4217
Shinomoto S 1987 *Biol. Cybern.* **57** 197
Sompolinsky H and Kanter I 1986 *Phys. Rev. Lett.* **57** 2861