Associative memory on a small-world neural network

L.G. Morelli[1,a], G. Abramson[2], and M.N. Kuperman[2]

[1] Abdus Salam International Center for Theoretical Physics, P.O. Box 586, 34100 Trieste, Italy
[2] Centro Atómico Bariloche, CONICET and Instituto Balseiro, 8400 S.C. de Bariloche, Argentina

Abstract. We study a model of associative memory based on a neural network with small-world structure. The efficacy of the network to retrieve one of the stored patterns exhibits a phase transition at a finite value of the disorder. The more ordered networks are unable to recover the patterns, and are always attracted to non-symmetric mixture states. Besides, for a range of the number of stored patterns, the efficacy has a maximum at an intermediate value of the disorder. We also give a statistical characterization of the spurious attractors for all values of the disorder of the network.

PACS. 84.35.+i Neural networks – 89.75.Hc Networks and genealogical trees – 87.18.Sn Neural networks

1 Small-world neural networks

Artificial neural networks have been used as a model for associative memory since the 80's, and a considerable amount of work has been made in the field [1–3]. Most of this work regards both the simulation and the theory of completely connected networks, as well as networks with a random dilution of the connectivity. It is known that particular prescriptions for the determination of the synaptic weights enable these systems to successfully retrieve a pattern out of a set of memorized ones. This behavior is observed in the system up to a certain value of the number of stored patterns, beyond which the network becomes unable to retrieve any of them. For reasons of simplicity of the models and their analytical tractability, complex architectures of the networks, more akin to those found in biological neural systems, have been largely left out of the theoretical analysis. Fortunately, since a few years ago, a class of models that has come to be known as "complex networks" began to be thoroughly studied. Complex networks seem more compatible with the geometrical properties of many biological and social phenomena than regular lattices, random networks, or completely connected systems [4–7]. Already in the seminal work of Watts and Strogatz [4], whose small-world model combines properties of regular and random networks, it was observed that the neural system of the nematode *C. elegans* shares topological properties with this model networks. Very recently, a number of works have begun to explore the relevance of a complex architecture in the behavior of neural networks as memory devices. The stability of the memorized patterns, as well as the capacity to retrieve a pattern from a

state affected by errors, has been studied by McGraw et al. in reference [8], for a variety of small-world and scale-free networks. The scale-free topology has been further studied by Torres et al. [9], who found that these networks perform better than random ones of the same dilution or connectivity. Other recent contributions can be found in [10–12]. In particular, reference [11] by Bohland et al. explores the same model we describe in he present work. Our results support theirs and provide a characterization of the transition and the nature of the attractors.

In this paper we study a neural network built upon the Watts-Strogatz model for small worlds. The model interpolates between regular and random networks by means of a parameter p, which characterizes the disorder of the network. The construction, as formulated in reference [4], begins with a one-dimensional regular lattice of N nodes, each one linked to its K nearest neighbors to the right and to the left, and with periodic boundary conditions. With probability p, each one of the right-pointing links, of every node, is rewired to a randomly chosen node in the network. Self connections and repeated connections are not allowed. The result is a disordered network, defined by the set N, K, p, that lies between a regular lattice ($p = 0$) and a random graph ($p = 1$). A wide range of these networks displays high local clusterization and short average distance between nodes, as many real complex networks. They can be defined by the *connectivity matrix* c_{ij}, where $c_{ij} = 1$ if there is a link between nodes i and j, and $c_{ij} = 0$ otherwise. This matrix establishes the synaptic connections between neurons, differing from the traditional Hopfield model, where the network is completely connected and the connectivity matrix is $c_{ij} = 1$, $\forall i, j$. It is also different from the traditional diluted disordered networks, which have also been considered in the literature, in which

[a] e-mail: `morelli@ictp.trieste.it`

randomly chosen elements in the connectivity matrix are set to zero.

The biological neuron carries out an operation on the inputs provided by other neurons, and it produces an output. A transformation of this continuous output into a binary variable makes it possible to formulate a simplified model in which the neurons are logical elements. In this binary representation, the state of each neuron is characterized by a single variable s_i. This variable can take two values representing the active and the inactive state,

$$s_i = \begin{cases} 1 & \text{if the neuron is active,} \\ -1 & \text{if the neuron is inactive.} \end{cases} \quad (1)$$

We assume that some patterns have been stored in the network by a learning process, which we do not deal with. The stored—or *memorized*—patterns are represented by network states ξ^μ, where $\mu = 1, \ldots, M$ labels the different patterns and M is their number. As usual, the patterns are generated at random, assigning with equal probability $1/2$ the values $\xi_i^\mu = \pm 1$. The patterns are uncorrelated and thus orthogonal in large networks:

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu = \delta_{\mu\nu}. \quad (2)$$

The state of the neurons is updated asynchronously, as in Glauber dynamics. At each simulation step, a neuron is chosen at random, and its new state is determined by the local field:

$$h_i = \sum_{j=1}^{N} \omega_{ij} s_j, \quad (3)$$

according to:

$$s_i = \text{sign}(h_i). \quad (4)$$

The synaptic weights ω_{ij} of the connections are given by Hebb's rule, restricted to the synapsis actually present in the network, as given by the connectivity matrix:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{M} c_{ij} \xi_i^\mu \xi_j^\mu \quad (5)$$

for $i, j = 1, \ldots, N$. Note that as the network model does not allow self connections the diagonal matrix elements are null. By definition, the synaptic matrix is symmetric.

In the model, "memory" is the capacity of the network to retrieve one of the stored patterns from an arbitrary initial condition. The retrieval of a stored or memorized pattern is the convergence of the system to that pattern by means of the dynamics.

Besides the dynamics prescribed by equations (3, 4), it is possible to consider a network is subject to thermal fluctuations [1,2]. In this work we will only consider the effects of a small amount of additive noise to verify the robustness of our results. A full discussion of the effect of a finite temperature in the dynamics will be left for future work. The stochastic asynchronous update, though,

prevents the system from having limit cycles, and the only attractors are fixed points.

In a fully connected network, the stored patterns ξ^μ would be, by construction of the synaptic weights (5), fixed points of the dynamics due to the orthogonality condition. Diluted networks, nevertheless, are also able to hold memorized patterns as long as, for a given number of patterns M, the connectivity is large enough. This holds for some classes of networks, such as random, symmetric and strongly diluted networks, as shown by Derrida et al. in [13], and also for symmetric ones with finite dilution, as demonstrated recently by Wemmenhove et al. [10]. Since the networks of our model have a topology which is not completely random we have verified, by direct inspection, that the memorized patterns are indeed fixed points of the dynamics for the dilution and the storage values used. Remarkably, the value of the disorder plays little or no role in the connectivity necessary to have the patterns as fixed points with statistical certainty: there is a sharp transition, at the same value of K for all p, above which the probability that a stored pattern is a fixed point is 1 (and below, it is 0). All our simulations are carried out in a region well beyond this critical value, where the stored patterns are fixed points. Of course, the fact that a pattern is a fixed point does not ensure that it is an attractor of the dynamical system, since it basin might be negligible or null. If the network is to perform well as a memory model, the basin of attraction of the memorized patterns has to occupy a significant volume in phase space, and we know that this is the case for completely random networks. It is our aim to explore how the topology of the network affects this matters, and eventually how the stored patterns lose there attraction as a function of the disorder parameter. As in the fully connected and the randomly diluted models, the reversed patterns $(-\xi_i)$, as well as a wealth of symmetric and asymmetric mixtures of patterns (also called spurious states), are also equilibria of the system and play a significant role in its behavior as a memory device.
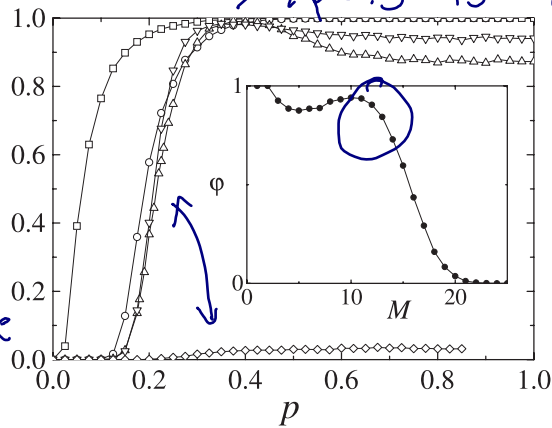
2 Effect of the disordered topology

We have performed extensive numerical simulations of the system, starting from a random unbiased initial condition[1]. After a transient, a fixed point is reached, whence no further changes occur to any neuron. In order to measure the efficacy of the network to recall a number M of stored random patterns, we define an *efficacy* φ as the fraction of realizations in which one of the stored patterns (or one of the reversed ones, $-\xi$) is perfectly retrieved. In Figure 1 we plot the order parameter φ as a function of the disorder parameter p. The different curves correspond to different numbers of stored patterns, $M = 1, 2, 5, 10,$ and 20. For this plot we have used $N = 5000$ and

[1] The reason to use a random initial condition is to achieve a good characterization of the whole phase space. In applications, it is usual to set the initial state at one of the stored patterns, affected by a variable amount of errors.

φ = prop. of times the pattern is recovered fully

[Handwritten annotations at top:] N=5000, K=100 → E = 250,000 $\frac{5000}{2^{12}} = 1.25$ $\frac{5000}{2^{13}} = .6$

M = 10: $\frac{5000}{2^9} = 9.7$, M=11: $\frac{5000}{2^{10}} = 4.88$, M=12: $\frac{5000}{2^{11}} = 2.44$

[Handwritten:] seems to be a phase trans. Performance at M=10

[Handwritten:] ~ p=.3 is a sweet spot

Fig. 1. Efficacy to retrieve a memorized pattern, φ, as a function of the disorder p. The curves correspond to different number of stored patterns: (squares) $M = 1$, (circles) $M = 2$, (up triangles) $M = 5$, (down triangles) $M = 10$, (diamonds) $M = 20$. Inset: The efficacy as a function of the number of stored patterns, at $p = 1$. Simulation parameters: $N = 5000$, $K = 100$, 10^4 realizations per point.

Fig. 2. Efficacy φ as a function of the disorder parameter p, for systems of different sizes (as shown in the legend), and $K = 100$. The number of stored patterns is $M = 5$, with 10^4 realizations per point. Inset: The same curves, scaled with the system size according to equation (6), collapse to a single curve Φ, with $p_c = 0.333$ and $\alpha = 0.2$.

[Handwritten right margin:] varying N has little impact on perf for fixed patterns

$K = 100$. Averages have been taken over 10^4 realizations. For each realization we use different patterns, as well as different initial conditions. Figure 1 shows that on highly ordered networks the system does not retrieve any stored pattern (as already observed in [11]). Then there is a transition as the disorder parameter p grows, and above some critical value of p, patterns are retrieved as fixed points yielding $\varphi > 0$. For $M = 1$ and $M = 2$, $\varphi \equiv 1$ above $p \approx 0.4$. But for $M > 2$ we find that φ does not grow monotonically with p. Instead, it decays as p grows after reaching a maximum value. This surprising non monotonic behavior with the disorder parameter p has been observed before in a problem of biased diffusion [14], and in an Ising model [15], both with asymmetric interactions.

[Handwritten left margin:] cool result about perf not monotonically increasing w/ N

In the inset of Figure 1 we plot φ vs. M for a disordered network with $p = 1$. As the number of stored patterns M grows, the network is not able to retrieve them. The curve also shows a non monotonic behavior with M. The transition as the number of stored patterns grows has already been studied in diluted disordered networks (Ref. [1], Chap. 7). It is known that random dilution reduces capacity of a neural network (approximately $0.138N$ for a fully connected network [1]) in a way which is proportional to the fraction of available connections. For our system (which is very diluted) the transition, then, takes place at $M_c \approx 0.138(K/N)N = 13.8$, roughly as observed. Nevertheless, we are mostly interested in the behavior of the system regarding the different topologies characterized by p. The fact that the transition between the memory phase and the spin-glass phase occurs at a finite value of the disorder parameter is very interesting, since a few dynamical systems based on small-world architectures show it [16–18]. This occurs in spite of the fact that the average distance between nodes, the main geometrical property of the Watts-Strogatz model, has a transition at $p = 0$ [19]. Indeed, for several Ising-like systems, which bear some
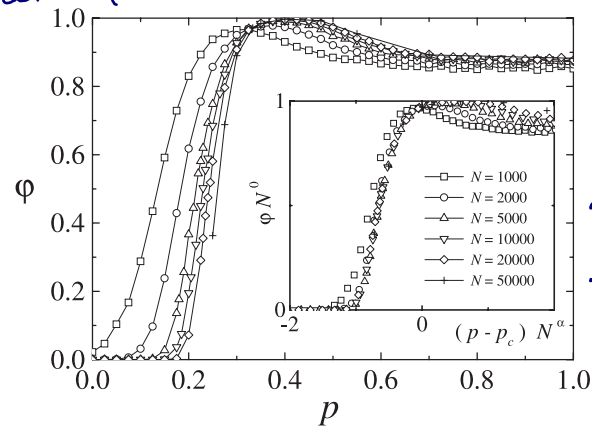
[Handwritten left margin:] what does this mean?

similarities with artificial neural networks, a phase transition occurs at $p = 0$ [19–22].

In order to understand the finite size effects in the system, and the behavior of the transition in the limit of an infinite system, we have made simulations on systems of different sizes. We have chosen to keep the connectivity parameter of the model constant through all the results we show, $K = 100$. In this regard, our results correspond to a neural network characterized by certain properties at the local level, for example the average connectivity of each neuron ($2K$ in our systems). Our finite size analysis shows the behavior of these networks in systems of increasing size N and in the limit $N \to \infty$.

The plot of φ vs. p for different values of N is shown in Figure 2. For this curves we have set $K = 100$ and $M = 5$, averaging over 10^4 independent realizations. As seen in the figure, all the curves seem to cross for the same value of the disorder parameter $p = p_c \approx 0.333$. Observe that the small-world transition, in a finite network, occurs at $p \approx 1/N$ (and at $p = 0$ for an infinite network), so the present transition occurs clearly at a finite value of the disorder parameter.

Based on numerical evidence, we find that the dependence of the efficacy on the system size can be built into a scaling function:

$$\varphi(p, N) = \Phi\left[(p - p_c)N^\alpha\right]. \qquad (6)$$

At the point of crossing of the curves, φ becomes independent of N.

Since the order parameter is not singular at the transition, we can expand Φ as a Taylor series around the critical control parameter p_c:

$$\varphi(p, N) = \Phi(0) + \Phi'(0)(p - p_c)N^\alpha, \qquad (7)$$

[Handwritten at bottom:] So phase transition in # of stored patterns occurs when we don't expect a node $\frac{N}{2^{M-1}} < 1$, ie when in each cluster.

to first order in $(p - p_c)$. Defining $\tilde{\varphi} = \varphi - \varphi(p_c)$ and $\tilde{p} = p - p_c$ we can write:

$$\left.\frac{\partial \tilde{\varphi}}{\partial \tilde{p}}(N)\right|_{\tilde{p}=0} = \Phi'(0) N^{\alpha}. \tag{8}$$

Plotting on a log-log scale the derivative $\partial \tilde{\varphi}/\partial \tilde{p}|_0$ vs. N, we obtain the exponent α as the slope of the line. Using data from $N = 2 \times 10^3$ to $N = 10^5$, we find $\alpha = 0.23 \pm 0.04$, and $\Phi'(0) = 0.096 \pm 0.016$. In the inset of Figure 2, we plot the re-scaled curves for different N. The best data collapse is obtained with $\alpha = 0.2$, compatible with the above result.

Except in the relatively narrow range of p where $\varphi \approx 1$, the system fails to retrieve any stored pattern in a significant fraction of the realizations: almost always when the network is very ordered, down to $p = 0$, and about 12% of the times when the network is very disordered, up to $p = 1$. What happens in the phase space as the network architecture changes? What happens to the trajectories, and why are the patterns missed? It seems natural to expect that the energy landscape is different for $p = 0$ than for $p = 1$. To address this problem we turn our attention to the properties of the overlaps of the equilibrium state with the memorized patterns. Suppose that after a transient the network has reached a fixed point ζ. We define the overlap of this fixed point with the patterns as

$$\theta^{\mu} = \frac{1}{N}\left|\sum_{i=1}^{N} \xi_i^{\mu} \zeta_i\right|. \tag{9}$$

Note that if the fixed point is a stored pattern, $\zeta = \xi^{\nu}$, then $\theta^{\nu} = 1$. In order to determine the type of fixed points that are reached when the network misses the patterns, we measure the overlap θ^{μ} of the fixed point with the stored patterns ξ^{μ}. The probability distribution $P(\theta)$ of these overlaps gives information on the kind of mixture that the fixed point is. Figure 3 shows the overlap distributions for several levels of disorder in the network. In this plots, $N = 2000$, $K = 100$, $M = 5$ and 10^6 realizations are used per curve. For the three higher values of p, the distributions have a high peak at $\theta = 1$, which is not shown for reasons of scale. This peak corresponds to the realizations that end up in a pattern, which happens frequently whenever $p > p_c$, as seen in Figure 2. The somewhat broader peak that these distributions have at low values of θ has the same origin, since the overlaps with the other $M-1$ patterns have a low value whenever a pattern is reached. Indeed, the overlap of two uncorrelated states has a mean value $\theta_0 = 0.022$. In the intermediate range of θ, the distribution presents a broad bump around $\theta = 1/2$. This corresponds to symmetric mixtures of the patterns, although the width of this bump suggests that asymmetric mixtures are present as well. In particular, the smaller peak present around $\theta \approx 0.35$ for the completely random network, corresponds to asymmetric mixtures. In contrast with these three cases—at and above the critical point—for ordered networks with $p = 0$ the overlap distribution is broad and does not have peak at $\theta = 1$. It
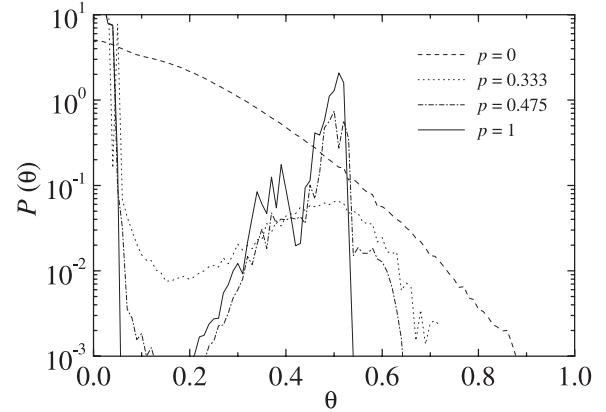

Fig. 3. Distribution of overlaps $P(\theta)$ after a fixed point has been achieved, between the state of the system and all stored patterns. Each curve corresponds to a value of p, as shown in the legend, typical of the different memory behaviors observed. A large peak at $\theta = 1$ (perfect retrieval of a pattern) is not shown for reasons of scale (see discussion in the text).

has a maximum at $\theta = 0$ and decays as θ grows, but large overlaps are observed in some realizations as the distribution shows. This is the only curve for which the complete distribution is shown. As the distribution suggests, the fixed points of these systems consist of very asymmetric mixtures.

The previous analysis unveiled the structure of the phase space and the difference between the low and the high p regimes. Still, what is the reason for the catastrophic loss of memory below the critical value of disorder? We have found that, for low values of disorder, the fixed points retrieve scattered pieces of several stored patterns. These fixed points consist of localized regions that overlap with different patterns. Indeed, at $p = 0$, the network is topologically very clusterized, and there exist local neighborhoods relatively isolated from each other. These neighborhoods begin to disappear by the action of the shortcuts provided by the random rewiring at higher values of p, until the whole system becomes essentially a single neighborhood. Then, at $p = 0$, from an arbitrary initial condition, different regions of the network eventually align themselves with different patterns. The final result is a completely asymmetric mixture, impossible to classify due to the arbitrariness of its origin and nature. These are the states that the broad distribution of overlaps describes, in Figure 3, for $p = 0$. The existence of asymmetric mixtures as attractors in this kind of associative memory model have been observed before (see for example [1], Chap. 4). But since they are very rare in the completely random or in the completely connected networks, they are very difficult to observe. In the present context, however, they play an essential role in the destruction of the ability of the system to retrieve the patterns.

In order to quantify this, we proceed to define a correlation measure that provides a clear picture of the situation. We introduce the difference of the fixed point ζ with
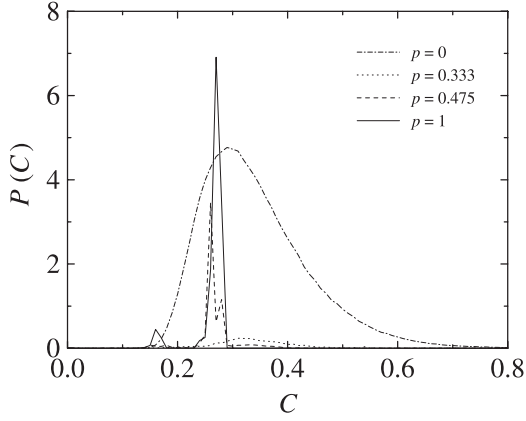
Fig. 4. Distribution of the local correlation that characterizes the level of alignment with a stored pattern (Eq. (13)). System parameters: $N = 2000$, $K = 100$, 10^6 realizations per curve. A peak at $C = 1$, shared by the three curves with the higher values of p, is not shown for reasons of scale (see discussion in the text).

a given pattern:

$$d_i^\mu = \xi_i^\mu \zeta_i = \begin{cases} 1 & \text{if} \quad \xi_i^\mu = \zeta_i, \\ -1 & \text{if} \quad \xi_i^\mu \neq \zeta_i. \end{cases} \quad (10)$$

Then we define a local magnetization for the difference vector d^μ, for every node i:

$$m_i^\mu = \frac{1}{1 + k_i} \left| d_i^\mu + \sum_{j \in \mathcal{V}_i} d_j^\mu \right|, \quad (11)$$

where $\mathcal{V}_i$ is the set of neighbors of node i. The local magnetization m_i^μ measures the local alignment with the pattern μ or its reversed companion. The maximum value $m_i^\mu = 1$ arises when $d_j^\mu = d_i^\mu \ \forall j \in \mathcal{V}_i$. The presence of connected domains where the fixed point ζ overlaps with the pattern ξ^μ should be detected as short range correlations between the local magnetizations. The correlation between the local magnetizations of the difference vector with the pattern μ is then defined as:

$$C^\mu = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{k_i} \sum_{j \in \mathcal{V}_i} m_i^\mu m_j^\mu. \quad (12)$$

As we intend to capture the existence of correlations in the difference with patterns that appear in the mixture that makes up the fixed point ζ, we define the maximum correlation

$$C = \max_\mu \{C^\mu\}. \quad (13)$$

Figure 4 presents the probability distribution $P(C)$ for different levels of network disorder. Each distribution is constructed over 10^6 realizations of $N = 2000$ networks, with connectivity $K = 100$. For $p = 0$ we observe a broad peak centered around $C \approx 0.3$. This is a quantitative measure of the occurrence of correlations on ordered networks,

as we pointed out. For the other values of p considered in the figure, the distribution has a sharp peak at $C = 1$ which we have not shown for reasons of scale, corresponding to the fixed points that coincide with a pattern, and consequently give the highest possible value of the correlation. Besides this peak, the most disordered systems show a narrow peak at $C \approx 0.25$, and the curve for $p = 1$ also a smaller one at $C \approx 0.15$. These two peaks correspond to symmetric and asymmetric mixtures, respectively. For $p = 0.333$, very close to the critical point, the distribution presents a very small bump at $C \approx 0.3$. It is easy to see, form the extended region of $P(C)$ in the curve for $p = 0$, that the mixtures are characterized by higher local correlation in the ordered system than in the disordered ones.

3 Discussion

We have studied a model of associative memory based on neural networks with a complex topology. This kind of connectivity can be considered as more similar to the biological networks than the completely connected or randomly diluted networks. Many of the general features of these systems are preserved: the network is able to retrieve a memorized pattern, up to a saturation. Besides, we have found a critical dependence of the efficacy of retrieval on the disorder parameter of the network: a collapse of the memory capability takes place at a finite value of the disorder parameter. The optimal performance of the system occurs at an intermediate value of the disorder, just above the critical value. This enhanced performance occurs far away from the region of $p = 1$, which is equivalent to the well known models of completely connected or randomly connected neural networks. We have characterized the different phases by the properties of the mixture states, that prevent the system to reach one of the memorized states.

We have understood the failure of the more ordered networks to retrieve a stored pattern due to the partition of the system into arbitrary neighborhoods aligned with more than one pattern. This is something that the disordered networks cannot do, and in fact the distributions of the overlaps and of the correlations quantify this effect. It does not escape us that we cannot, at this stage, provide an explanation of the enhanced performance of the intermediate region.

We have checked the robustness of our results with respect to a small amount of noise in the dynamics. This has been implemented by flipping, with probability ϵ, one neuron at random after each deterministic step. For values of ϵ up to 0.01, the results are indistinguishable from the noiseless system. For greater values of ϵ the system becomes more and more ineffective to retrieve a pattern, but the general form of the curves $\varphi(p)$ is preserved for the whole range of p. A systematic analysis of the problem of a truly noisy network, characterized by a temperature, remains to be done.

References

1. D.J. Amit, *Modeling brain function: the world of attractor neural networks* (Cambridge University Press, Cambridge, 1989)
2. J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the theory of neural computation* (Addison-Wesley Publishing Company, 1991)
3. P. Peretto, *An introduction to the modeling of neural networks* (Cambridge University Press, Cambridge, 1992)
4. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
5. A-L. Barabási, R. Albert, Science **286**, 509 (1999)
6. M.E.J. Newman, J. Stat. Phys. **101**, 819 (2000)
7. D.J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999)
8. P.N. McGraw, M. Menzinger, Phys. Rev. E **68**, 047102 (2003)
9. J.J. Torres, M.A. Muñoz, J. Marro, P.L. Garrido, arXiv:cond-mat/0310205 (2003)
10. B. Wemmenhove, A.C.C. Coolen, J. Phys. A **36**, 9617 (2003)
11. J.W. Bohland, A.A. Minai, Neurocomputing **38-40**, 489 (2001)
12. C.L. Labiouse, A.A. Salah, I. Starikova, *Proceedings of the Santa Fe Complex Systems Summer School, Santa Fe, NM. (Santa Fe Institute, 2002)*
13. B. Derrida, E. Gardner, A. Zippelius, Europhys. Lett. **4**, 167 (1987)
14. D.H. Zanette, Europhys. Lett. **60**, 945 (2002)
15. A. Sánchez, J.M. López, M.A. Rodríguez, Phys. Rev. Lett. **88**, 048701 (2002)
16. M. Kuperman, G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001)
17. D.H. Zanette, Phys. Rev. E **65**, 041908 (2002)
18. G. Szabo, A. Szolnoki, R. Izsak, J. Phys. A **37**, 2599 (2004)
19. A. Barrat, M. Weigt, Eur. Phys. J. B **13**, 547 (2000)
20. J.Y. Zhu, H. Zhu, Phys. Rev. E **67**, 026125 (2003)
21. C.P. Herrero, Phys. Rev. E **65**, 066110 (2002)
22. B.J. Kim, H. Hong, P. Holme, G.S. Jeong, P. Minnhagen, M.Y. Choi, Phys. Rev. E **64**, 056135 (2001)