Available at: http://www.ictp.it/~pub_off IC/2004/91

United Nations Educational Scientific and Cultural Organization
and
International Atomic Energy Agency
THE ABDUS SALAM INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS

SPATIAL ASYMMETRIC RETRIEVAL STATES IN SYMMETRIC HEBB NETWORK WITH UNIFORM CONNECTIVITY

Kostadin Koroutchev[1]
Depto. de Ingeniería Informática and Instituto de Ingeniería del Conocimiento, Universidad Autónoma de Madrid, 28049 Madrid, Spain
and
Institute for Computer Systems, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

and

Elka Korutcheva[2]
Depto. de Física Fundamental, Universidad Nacional de Educación a Distancia, c/Senda del Rey, No 9, 28080 Madrid, Spain,
Georgi Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences, 72 Tzarigradsko Chaussee Blvd., 1784 Sofia, Bulgaria
and
The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy.

Abstract

In this paper we show that during the retrieval process in a binary Hebb recursive neural network, spatial localized states can be observed when the connectivity of the network is distance-dependent. We point out that the minimal condition that leads to this type of behavior is the asymmetry between the retrieval and the learning states.

MIRAMARE – TRIESTE
September 2004

[1]kostadin.koruchev@ii.uam.es
[2]Regular Associate of ICTP. elka@fisfun.uned.es

1 Introduction

In a very recent publication [1] it was shown that using linear-threshold model neurons, the Hebb learning rule, sparse coding and distance-dependent asymmetric connectivity, spatial asymmetric retrieval states were observed and their biological relevance was pointed out. These asymmetric states are characterized by a spatial localization of the activity of the neurons, described by the formation of local bumps. For the biological relevance of the issue see the introductory part of Refs.[1] and [2].

The observation is intriguing, because all components of the network are intrinsically symmetric in respect to the positions of the neurons and the retrieved state is clearly asymmetric. However, until now, to our knowledge, spatial asymmetry states (SAS) have not been observed in more simple models like the Hebb neural networks (NN) models with binary neurons.

The aim of this article is to impose a minimal set of restriction on a Hebb network with binary neurons that can lead to SAS.

When the network is sufficiently diluted, say less then 5% of dilution, then the differences between asymmetric and symmetric connectivity are minimal [3]. Therefore we expect that the differences between symmetrically and asymmetrically connectivity is minimal in SAS. This can also be observed by simulations.

There are several factors that possibly contribute to SAS in model network.

Talking about spatial events in NN, one essentially introduces distance measures and topology between the neurons and also imposes some distribution on the connections' probability dependent on that topology. The major factor to observe spatial asymmetric activity is of course the spatially dependent connectivity of the network. Actually this is an essential condition, because by applying random permutation to the enumeration of the neurons of a network, one will obviously achieve states without SAS. Therefore, the topology of the connections must depend on the distance between the neurons.

Due to these arguments, a symmetric and distance-dependent connectivity for all neurons is chosen in this study.

We consider an attractor NN model of Hebbian type formed by N binary neurons $\{S_i\}, S_i \in \{-1, 1\}, i = 1, ..., N$, storing p patterns $\xi_i^\mu, \mu \in \{1...P\}$, and we assume a symmetric connectivity $c_{ij} = c_{ji} \in \{0, 1\}, c_{ii} = 0$ between the neurons. $c_{ij} = 1$ means that neuron i and j are connected. We regard only connectivities in which the fluctuations between the individual connectivity are small, e.g. $\forall_i \sum_j c_{ij} \approx cN$, where c is the mean connectivity.

The learned patterns are symmetrically distributed from the following distribution:

$$P(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1).$$

The Hamiltonian of the system is:

$$H_0 = \frac{1}{N}\sum_{ij\mu} S_i \xi_i^\mu c_{ij} \xi_j^\mu S_j$$

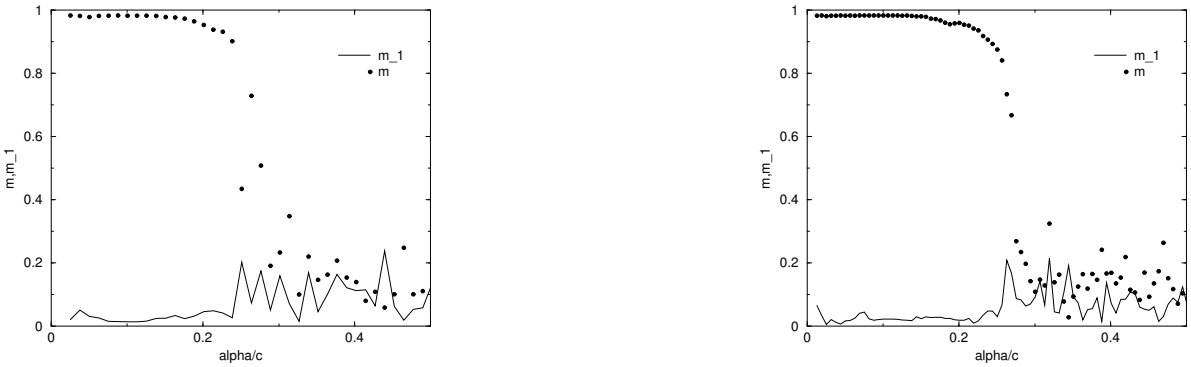

Figure 1: Binary attractor networks with the same distribution of the pattern and the retrieval activity. The overlap m and the power of its first Fourier transform are chosen as a measure of the existence of SAS. The left figure is for $N = 6400$, $c = 80/N$, $\sigma_x = 100$, the right one is for $N = 6400$, $c = 160/N$, $\sigma_x = 200$. None of the networks presents SAS. The same is true for sparse code, different dilution and other topologies.

and the retrieval states are supposed to obey

$$P(S_i) = \frac{1}{2}\delta(S_i - 1) + \frac{1}{2}\delta(S_i + 1).$$

However, as it will be shown later in the article, it is not possible to observe any SAS by these conditions, with the exception of the areas near the phase transition point between retrieval and non retrieval states.

In other words, imposing a symmetry between the retrieval and the learning states, i.e. equal probability distributions of the patterns and the network activities, no SAS exists. Spatial asymmetry can be observed only when asymmetry between both states is imposed.

Actually, by using binary network and symmetrically distributed patterns, the only asymmetry that can be imposed, independent on the position of the neurons, is the total number of the neurons in a given state. Having in mind that there are only two possible states, this condition leads to a condition on mean activity of the network.

To achieve this difference, we add an extra term H_a to the Hamiltonian

$$H_a = NR\left(\sum_i S_i/N - a\right)^2.$$

When $R \to \infty$, the H_a term tends to fix the sum of S_i to Na, yielding the proportion of neurons $S_i = 1$ of $1/2 + a/2$ and the proportion of the neurons with $S_i = -1$ equal to $1/2 - a/2$. If the goal is just to reduce the number of spins in high state $S_i = 1$ without fixing their number, the extra term in the Hamiltonian can be reduced to a linear one, that is easier to analyze theoretically. In this article it is shown that the last condition is sufficient to observe SAS.

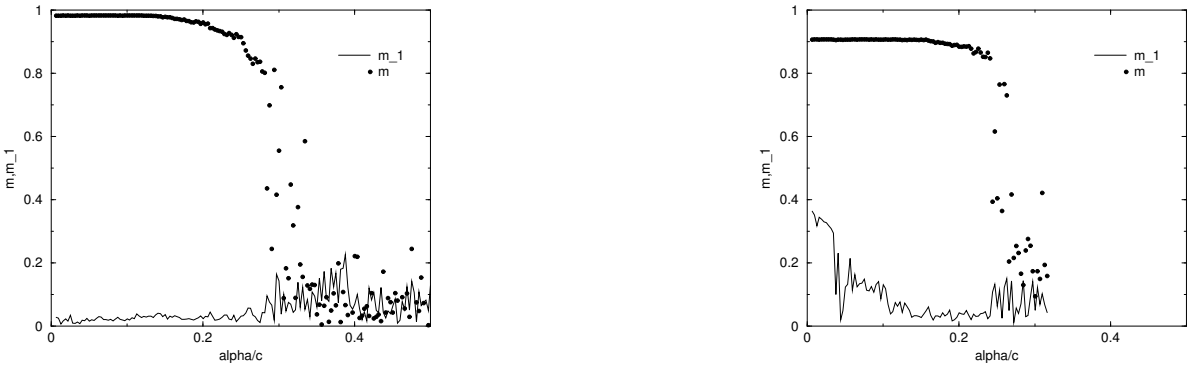

Figure 2: Spatial asymmetry observed by its first Fourier component power m_1. The left figure shows the network with $N = 6400$, $c = 0.05$, $\sigma_x = 500$ and $a = 0$. The right one shows the network with $N = 6400$, $c = 0.05$, $\sigma_x = 500$ and $a = 0.1$. It is clear that the SAS is observed only when $a \neq 0$. The probability to get a value of $m_1 \geq 0.1$ by chance is less then 10^{-5}.

Figure 3: Spatial asymmetry observed by measuring the smoothed local fields, by length 100 and 600. The left figure shows the network with $p/(cN) = 0.1$, the right one with $p/(cN) = 0.03$. $N = 6400$, $c = 0.05$, $\sigma_x = 500$, $a = 0.1$.

2 Simulations

The dynamics of the network at time $t + 1$ and temperature $T = 0$ is

$$S_i(t+1) = \text{sign}\left(\sum_j \xi_i^\mu c_{ij} S_j(t) - T_h\right),$$

where T_h is the threshold of the system, which in general is nonzero, due to the extra energy term H_a. Taking the limit $R \to \infty$, H_a actually fix the number of neurons in state 1 to $(1 + a)N/2$, that can be implemented easily by a programme. That can easily be done by sorting the non-normalized internal fields $h_i = \sum_j \xi_i^\mu c_{ij} S_j(t)$ and choosing $h_{(1+a)N/2}$ as a threshold.

The topology of the network is chosen to be a circular ring, with distance

$$|i - j| \equiv \min(i - j + N \bmod N, j - i + N \bmod N).$$

The same connectivity as in Ref.[1] with typical connectivity distance $\sigma_x N$ is used, e.g.:

$$P(c_{ij} = 1) = c\left[\frac{1}{\sqrt{2\pi}\sigma_x N} e^{-(|i-j|/N)^2/2\sigma_x^2} + p_0\right].$$

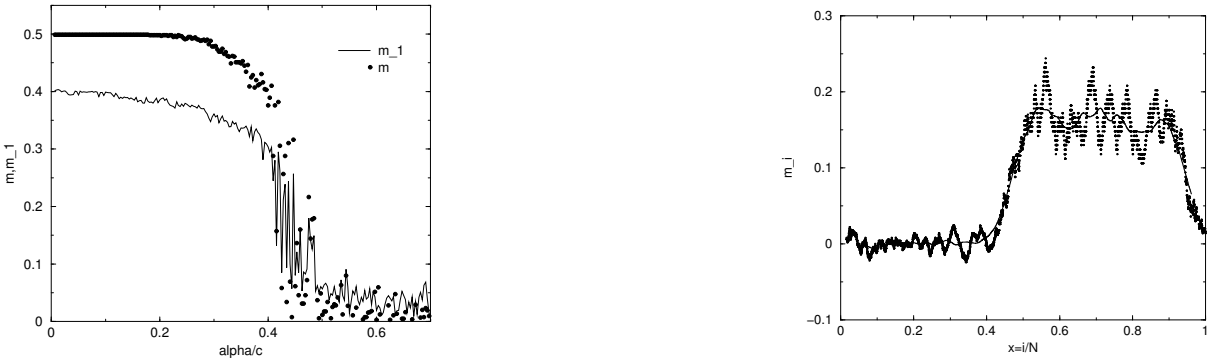

Figure 4: Sparse code influence on the spatial asymmetry observed by measuring the smoothed local fields. Smoothing by length 100 and 600. The sparsity of the code was chosen to be 0.2, the asymmetry factor $a = 0.5$. The load of the network in the right figure is $p/(cN) = 0.1$. $N = 6400$, $c = 0.05$, $\sigma_x = 500$, $a = 0.5$.

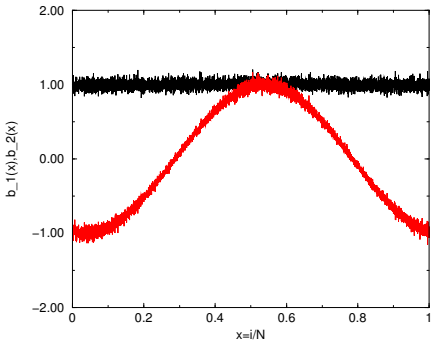

Figure 5: The components of the first eigenvectors of the connectivity matrix, normalized by the square root of their eigenvalue, in order to eliminate the effect of the size of the network. The first eigenvector has constant components, the second one - sine-like components. $N = 6400$, $c = 320/N$, $\lambda_1 = 319.8$, $\lambda_2 = 285.4$. Note that the differences between the first and the second eigenvectors have comparable magnitudes.

Here p_0 is chosen to normalize the expression in the brackets. When σ_x is small enough, then spatial asymmetry is expected.

If the retrieval state, corresponding to the pattern ξ_i^0 is S_i, the mean overlap is $m = \sum \xi_i^0 S_i/N$ and the local overlap at site i is

$$m_i = \xi_i^0 S_i.$$

These quantities, even smoothed, are N measures of the locality of the overlap, that are best for graphical representation, but serve only if the spatial asymmetry is evident. However we need a single numerical measure of SAS. In the case when S_i follow a single sine wave, the ideal measure of spatial asymmetry would be

$$m^{(1)} = \frac{1}{N} \left| \sum_k \xi_k^0 S_k e^{2\pi i k/N} \right|.$$

Because we are looking for a single-bump spatial activity, the ratio $m^{(1)}/m$ is a good measure of SAS. If the bump has a form of pure sine wave, then $m^{(1)}/m = 1/2$, and if it has a form

of a Gaussian with variance $\sigma_1 N$ with small σ_1, then $m^{(1)}/m \approx e^{-2\pi\sigma_1^2}$, that is about 0.8 for $\sigma_1 = 0.1$. Because the sine waves appear first, $m^{(1)}/m$ results also to be a sensitive asymmetric measure, at least compared with the visual inspection of m_i.

Let us note that $m^{(1)}$ can be regarded as the power of the first Fourier component and m can be regarded as a 0^{-th} Fourier component, that is the power of the direct-current component. More sophisticated SAS measures can be elaborated, counting higher frequencies, but for the purpose of these simulation $m^{(1)}/m$ results good enough. If $m \approx 1$ then $m^{(1)}/m$ is equivalent to $m^{(1)}$.

Simulations with "small-world" topology and more sharp localized connectivity

$$P(c_{ij} = 1) \propto \frac{1 - b\cos\varphi}{1 - 2b\cos\varphi + b^2},$$

with $\varphi \equiv 2\pi|i-j|/N$ and b being some parameter, show similar results. The last connectivity has the advantage that the eigenvector of the connectivity matrix are cosine waves and the eigenvalues are known.

The results of the simulations for different σ and a are shown in Figures 1,2 and 3.

If $a = 0$, no asymmetry can be observed at any σ_x, up to the level of the network fragmentation (Fig. 1). No difference between asymmetric and symmetric connectivity is observable for any connectivity $c < 0.05$ and any of the topologies tested.

The sparse code increases SAS effects Fig. 4, but SAS cannot be observed by any sparsity a_s if the proportion of the firing neurons is kept to be equal to a_s (not shown).

The hint from the simulations is that no asymmetric states can be observed if the retrieval and the memorized state have the same level of activity. On the other hand, SAS is observed when $a \neq 0$, as shown in Figs. 2,3,4.

3 Analytical analysis

For the analytical analysis of SAS states, we consider the decomposition of the connectivity matrix c_{ij} by its eigenvectors $a_i^{(k)}$:

$$c_{ij} = \sum_k \lambda_k a_i^{(k)} a_j^{(k)}, \sum_i a_i^{(k)} a_i^{(l)} = \delta_{kl},$$

where λ_k are the corresponding (positive) eigenvalues. For convenience we denote $b_i^k \equiv a_i^{(k)}\sqrt{\lambda_k}$, having

$$c_{ij} = \sum_k b_i^k b_j^k.$$

We will assume that $a_i^{(k)}$ are ordered by its eigenvalues in decreasing order, e.g.

$$\forall a_j^{(k)}, a_j^{(l)} k > l \Rightarrow \lambda_k \leq \lambda_l$$

To have some intuition of what $a_j^{(k)}$ look like, we plot in Fig. 5 the first two eigenvectors.

Moreover, for a wide variety of connectivities, the first three eigenvectors approximate:

$$a_k^{(1)} = \sqrt{1/N}, \tag{1}$$

$$a_k^{(2)} = \sqrt{2/N}\cos(2\pi|k-k_0|/N) \tag{2}$$

and

$$a_k^{(3)} = \sqrt{2/N}\sin(2\pi|k-k_0|/N). \tag{3}$$

Following the classical analysis of Amit et al [4], we use Bogolyubov's method of quasi averages [5] to have into account a finite number of overlaps that condense macroscopically. To this aim we introduce an external field, conjugate to a finite number of patterns $\{\xi_i^\nu\}, \nu = 1, 2, ..., s$, adding a term

$$H_h = \sum_{\nu=1}^{s} h^\nu \sum_i \xi_i^\nu S_i \tag{4}$$

to the Hamiltonian.

Finally, as we already mentioned in the Introduction, in order to impose some asymmetry in the neural network's states, we also add the term

$$H_a = NR\left(\sum_i S_i/N - a\right) \tag{5}$$

The whole Hamiltonian we are studying is now:

$$H = \frac{1}{N}\sum_{ij\mu} S_i\xi_i^\mu c_{ij}\xi_j^\mu S_j + \sum_{\nu=1}^{s} h^\nu \sum_i \xi_i^\nu S_i + NR(\sum_i S_i/N - a). \tag{6}$$

By using the "replica method" [6] for the averaged free energy per spin we get:

$$f = -\lim_{n\to 0}\lim_{N\to\infty}\frac{1}{\beta nN}(\langle\langle Z^n\rangle\rangle - 1), \tag{7}$$

where $\langle\langle ...\rangle\rangle$ stands for the average over the pattern distribution $P(\xi_i^\mu)$, n is the number of the replicas, which are later taken to zero and β is the inverse temperature.

The replicated partition function is

$$\langle\langle Z^n\rangle\rangle = \left\langle\left\langle Tr_{S^\rho}\exp\left[\frac{\beta}{2N}\sum_{ij\mu\rho}(\xi_i^\mu S_i^\rho)c_{ij}(\xi_j^\mu S_j^\rho) - \frac{1}{2}\beta pn + \right.\right.\right.$$
$$\left.\left.\left.\beta\sum_\nu h^\nu\sum_{i,\rho}\xi_i^\nu S_i^\rho - \sum_\rho \beta NR(\sum_i S_i^\rho/N - a)\right]\right\rangle\right\rangle. \tag{8}$$

The term $\frac{1}{2}\beta pn$ comes from the $i = j$ term and therefore $c_{ii} = 1$. Following [4], we decouple the sites by using an expansion of the connectivity matrix c_{ij} over its eigenvalues $\lambda^l, l = 1, ..., M$ and eigenvectors a_i^l (eq.3).

We thus have:

$$\langle\langle Z^n\rangle\rangle = e^{-\beta pn/2}\left\langle\left\langle Tr_{S^\rho}\exp\left[\frac{\beta}{2N}\sum_{\mu\rho l}\sum_{ij}(\xi_i^\mu S_i^\rho b_i^l)(\xi_j^\mu S_j^\rho b_j^l) + \right.\right.\right.$$
$$\left.\left.\left.\beta\sum_\nu h^\nu\sum_{i\rho}\xi_i^\nu S_i^\rho - \sum_\rho \beta RN(\sum_i S_i^\rho/N - a)\right]\right\rangle\right\rangle. \tag{9}$$

Introducing variables $m^\mu_{\rho l}$ at each replica ρ, each configuration and each eigenvalue, we get:

$$\begin{aligned}\langle\langle Z^n\rangle\rangle &= e^{-\beta pn/2+\beta RaN} \\ &\left\langle\left\langle Tr_{S^\rho}\int\prod_{\mu l\rho}\frac{dm^\mu_l}{\sqrt{2\pi}}\exp\beta N\left(-\frac{1}{2}\sum_{\mu\rho l}(m^\mu_{\rho l})^2+\sum_{\mu\rho l}m^\mu_{\rho l}\frac{1}{cN}\sum_i\xi^\mu_i S^\rho_i b^l_i\right)\right.\right. \\ &\left.\left.\exp\beta N\left(-\frac{1}{2}\sum_{\nu\rho l}(m^\nu_{\rho l})^2+\sum_{\nu\rho l}m^\nu_{\rho l}\frac{1}{N}\sum_i\xi^\nu_i S^\rho_i b^l_i+h^\nu\frac{1}{N}\sum_i(\xi^\nu_i S^\rho_i+RS^\rho_i)\right)\right\rangle\right\rangle \quad (10)\end{aligned}$$

In the last expression we have split the sums over the first s-patterns and the remaining (infinite) $p-s$ ones.

After taking the averages over the patterns, supposing them equally distributed [3], the first term gives:

$$I=\exp\left(-\frac{\beta N}{2}\sum_{\mu\rho l}(m^\mu_{\rho l})^2+\sum_{i\mu}\ln\cosh\beta\sum_{\rho l}m^\mu_{\rho l}S^\rho_i b^l_i\right), \quad (11)$$

which expanded up to second order in m and its rescaling $m^\mu_\rho\to m^\mu_\rho/\sqrt{N}$, leads to:

$$I=\exp\beta\left(-\frac{1}{2}\sum_{\mu\rho l}(m^\mu_{\rho l})^2+\frac{\beta}{2N}\sum_{\rho\sigma lki\mu}m^\mu_{\rho l}m^\mu_{\sigma k}S^\rho_i S^\sigma_i b^l_i b^k_i\right). \quad (12)$$

We have:

$$\int\prod_{\mu\rho l}\frac{dm^\mu_{\rho l}}{\sqrt{2\pi}}I=\int\prod_{\rho\sigma lk}dq^{lk}_{\rho\sigma}\exp\left(-\frac{p}{2}Tr\ln[A^{lk}_{\rho\sigma}]\right)\prod_{\rho\sigma lk}\delta(q^{lk}_{\rho\sigma}-\frac{1}{N}\sum_i S^\rho_i S^\sigma_i b^l_i b^k_i). \quad (13)$$

Introducing the parameter $r^{lk}_{\rho\sigma}$, conjugate to $q^{lk}_{\rho\sigma}$, for the last expression we get:

$$\begin{aligned}\int\prod_{\mu\rho l}\frac{dm^\mu_{\rho l}}{\sqrt{2\pi}}I&=\int\prod_{\rho\sigma lk}dq^{lk}_{\rho\sigma}\prod_{\rho\sigma lk}dr^{lk}_{\rho\sigma}\exp\left(-\frac{p}{2}Tr\ln[A^{lk}_{\rho\sigma}]\right) \\ \exp N&\left(-\frac{1}{2}\alpha\beta^2\sum_{\rho\sigma lk}r^{lk}_{\rho\sigma}q^{lk}_{\rho\sigma}+\frac{1}{2}\alpha c\beta^2N^{-1}\sum_{i\rho\sigma lk}r^{lk}_{\rho\sigma}S^\rho_i S^\sigma_i b^l_i b^k_i\right), \quad (14)\end{aligned}$$

where the parameter $\alpha\equiv p/N$ is the storage capacity of the network and the matrix $A^{lk}_{\rho\sigma}$ is

$$A^{lk}_{\rho\sigma}=\delta_{\rho\sigma}\delta_{lk}(1-\beta\lambda^k)+\beta\delta_{\rho\sigma}q^{lk}-\beta q^{lk}. \quad (15)$$

The first term of $A^{lk}_{\rho\sigma}$ comes form the Gaussian integration over $m^\mu_{\rho l}m^\mu_{\sigma l}$, the second one is the value of $q^{lk}_{\rho\sigma}$ when $\rho\equiv\sigma$, i.e. $q^{lk}_{\rho\sigma}=\frac{1}{N}\sum_i(S^\rho_i)^2\sqrt{\lambda^l}\sqrt{\lambda^k}a^k_i a^l_i\equiv\sqrt{\lambda^l}\sqrt{\lambda^k}\delta_{lk}$, because of the fact that $S^2_i=1$ and the orthogonality of the eigenvectors $\sum_i a^k_i a^l_i=\delta_{lk}$. The third and the fourth term correspond to the situation when $\rho\neq\sigma$. For $\langle\langle Z^n\rangle\rangle$, after taking the limit $h^\nu\to 0$, we

[3] The sparse code distribution $P(\xi^\mu_i)=a\delta(\xi^\mu_i-1)+(1-a)\delta(\xi^\mu_i+1)$ leads to a renormalization of the temperature $\beta\to\beta(1-a^2)$

have:

$$\begin{aligned}\langle\langle Z^n\rangle\rangle &= e^{-\beta pn/2+\beta RanN}\int\prod_{\nu}dm^{\nu}_{\rho l}\int\prod_{\rho\sigma lk}dq^{lk}_{\rho\sigma}dr^{lk}_{\rho\sigma}\\ &\exp N\left(-\frac{\beta}{2}\sum_{\nu\rho l}(m^{\nu}_{\rho l})^2-\frac{1}{2}Tr\ln[A^{lk}_{\rho\sigma}]-\frac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma,lk}r^{lk}_{\rho\sigma}q^{lk}_{\rho\sigma}\right)\\ &\left\langle\left\langle Tr_{S^\rho}\exp N\left[\frac{1}{2}\alpha\beta^2N^{-1}\sum_{i\rho\sigma lk}r^{lk}_{\rho\sigma}S^{\rho}_iS^{\sigma}_ib^l_ib^k_i+\right.\right.\right.\\ &\left.\left.\left.\beta\sum_{\nu\rho l}m^{\nu}_{\rho l}\frac{1}{N}\sum_i\xi^{\nu}_iS^{\rho}_ib^l_i+\beta R\sum_iS^{\rho}_i\right]\right\rangle\right\rangle. \qquad (16)\end{aligned}$$

Supposing Replica Symmetry (RS) ansatz, we can write $m^{\nu}_{\rho}=m^{\nu}$, $q^{lk}_{\rho,\sigma}=q^{lk}$ for $\rho\neq\sigma$ and $r^{lk}_{\rho,\sigma}=r_{lk}$ for $\rho\neq\sigma$. The free energy (eq. 7) then reads:

$$\begin{aligned}f &= \frac{\alpha}{2}+Ra+\frac{\alpha}{2\beta n}Tr\ln[A^{lk}_{\rho\sigma}]+\frac{1}{2}\sum_{\nu l}(m^{\nu}_l)^2-\frac{\alpha\beta}{2}\sum_{lk}r_{lk}q^{lk}-\\ &\frac{1}{n\beta}\left\langle\left\langle\ln Tr_{S^\rho}\exp\left[\frac{1}{2}\alpha\beta^2N^{-1}\sum_{i\rho\sigma lk}r^{lk}_{\rho\sigma}S^{\rho}_iS^{\sigma}_ib^l_ib^k_i-\frac{1}{2}n\alpha\beta^2\sum_lr_{ll}\lambda^l+\right.\right.\right.\\ &\left.\left.\left.\beta\sum_{\nu\rho l}m^{\nu}_{\rho l}\frac{1}{N}\sum_i\xi^{\nu}_iS^{\rho}_ib^l_i+\beta R\sum_iS^{\rho}_i\right]\right\rangle\right\rangle. \qquad (17)\end{aligned}$$

After taking the Tr_{S^ρ} in the last expression, for the last term of it, we get:

$$-\frac{1}{n\beta}\frac{1}{N}\sum_i\int\frac{dz^{(i)}}{\sqrt{2\pi}}e^{-(z^{(i)})^2/2}[1+n\ln 2\cosh\beta(\sqrt{\alpha b^l_ir_{lk}b^k_i}z^{(i)}+m^{\nu}_l\xi^{\nu}_ib^l_i+R)]. \qquad (18)$$

Changing the integration variables $z^{(i)}\rightarrow z_ma^m_i$ and having in mind that because of ortonormality of a^m_i the Jacobian of the transformation is 1, one obtains:

$$-\frac{1}{nN}\sum_i\int\frac{dz_m}{\sqrt{2\pi}}e^{-z^2_m/2}[1+n\ln 2\cosh\beta(\sqrt{\alpha r_{lk}}w^{lk}_{m,i}z_m+m^{\nu}_l\xi^{\nu}_ib^l_i+R)], \qquad (19)$$

where $w^{lk}_{m,i}\equiv\sqrt{b^l_ib^k_i}(a^m_i)$.

The sum over i can be taken, because it depends only on the topology. The coefficient before z_m can be complex.

After taking the limit $n\rightarrow 0$, we end up with the following expression for the free energy:

$$\begin{aligned}f &= \frac{\alpha}{2}+Ra+\frac{\alpha}{2\beta n}Tr\ln[A^{lk}_{\rho\sigma}]+\frac{1}{2}\sum_{\nu l}(m^{\nu}_l)^2-\frac{\alpha\beta}{2}\sum_{lk}r_{lk}q^{lk}+\frac{\alpha\beta}{2}\sum_lr_{ll}\lambda^l-\\ &\left\langle\left\langle\frac{1}{\beta}\frac{1}{N}\sum_i\int\frac{dz_m}{\sqrt{2\pi}}\exp(-{z_m}^2/2)\ln 2\cosh\beta(\sqrt{\alpha r_{lk}}w^{lk}_{m,i}z_m+m^{\nu}_l\xi^{\nu}_ib^l_i+R)\right\rangle\right\rangle. \qquad (20)\end{aligned}$$

The calculation of the term $Tr\ln[A^{lk}_{\rho\sigma}]$ is done in analogy with the case of Ref. [4].

$$
\begin{aligned}
Tr\ln[A^{lk}_{\rho\sigma}] &= Tr_{\sigma\rho lk}\ln[\delta_{\rho\sigma}\delta_{lk}(1-\beta\lambda^k)+\beta\delta_{\rho\sigma}q^{lk}-\beta q^{lk}] = \\
&= Tr_{\sigma\rho lk}\ln\Big[(\delta_{\rho\sigma}-1/n)(\delta_{lk}(1-\beta\lambda^k)+\beta q^{lk}) \\
&\quad +1/n\left(\delta_{lk}(1-\beta\lambda^k)+\beta q^{lk}-\beta n q^{lk}\right)\Big] = \\
&= (n-1)Tr_{lk}\ln(\delta_{lk}(1-\beta\lambda^k) \\
&\quad +\beta q^{lk})+Tr_{lk}\ln(\delta_{lk}(1-\beta\lambda^k)+\beta q^{lk}(1-n))
\end{aligned}
$$

Let us denote $\delta_{lk}(1-\beta\lambda^k)+\beta q^{lk}$ as $\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q}$. Regrouping the expression and taking the limit $n\to 0$, we obtain:

$$
\begin{aligned}
\lim_{n\to 0}\frac{Tr\ln[A]}{n} &= Tr\ln[\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q}] \\
&\quad +\frac{d}{dn}Tr_{ml}\ln[\delta_{lk}(1-\beta\lambda^k)+\beta q^{lk}(1-n)]|_{n=0} = \qquad (21)\\
&= Tr\ln[\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q}]-\beta Tr[(\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q})^{-1}.\mathbf{Q}]
\end{aligned}
$$

The last expression reproduces the limit of Ref.[4] when the index $k=1,\lambda_k=1$ and $\mathbf{Q}=q$. The expression for the free energy thus reads:

$$
\begin{aligned}
f &= \frac{\alpha}{2}+Ra+\frac{\alpha}{2\beta}\left(Tr\ln[\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q}]-\beta Tr[(\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q})^{-1}.\mathbf{Q}]\right) \\
&\quad +\frac{1}{2}\sum_{\nu l}(m^\nu_l)^2-\frac{\alpha\beta}{2}\sum_{lk}r_{lk}q^{lk}+\frac{\alpha\beta}{2}\sum_l r_{ll}\lambda^l- \\
&\quad \left\langle\left\langle\frac{1}{\beta}\frac{1}{N}\sum_i\int\frac{dz_m}{\sqrt{2\pi}}\exp(-{z_m}^2/2)\ln 2\cosh\beta(\sqrt{\alpha r_{lk}}w^{lk}_{m,i}z_m+m^\nu_l\xi^\nu_i b^l_i+R)\right\rangle\right\rangle. \qquad (22)
\end{aligned}
$$

The saddle point method applied to the free energy eq.(22) gives the following equations for the order parameters:

$$
m^\nu_s=\left\langle\left\langle\frac{1}{N}\sum_i\xi^\nu_i b^s_i\tanh\beta(\sqrt{\alpha r_{lk}}w^{lk}_{m,i}z_m+m^\nu_l\xi^\nu_i b^l_i+R)\right\rangle\right\rangle, \qquad (23)
$$

$$
q^{st}=\left\langle\left\langle\frac{1}{N}\sum_i b^s_i b^t_i\int\frac{dz_m}{\sqrt{2\pi}}e^{-{z_m}^2/2}\tanh^2\beta(\sqrt{\alpha r_{lk}}w^{lk}_{m,i}z_m+m^\nu_l\xi^\nu_i b^l_i+R)\right\rangle\right\rangle \qquad (24)
$$

and

$$
r_{kl}=[(\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q})^{-1}\mathbf{Q}(\mathbf{E}-\beta\mathbf{\Lambda}+\beta\mathbf{Q})^{-1}]_{kl}. \qquad (25)
$$

For zero temperature $T=0$, the equations can be transformed in the same way as in [4], having in mind that $C_{st}\equiv\beta(\delta_{st}\lambda_s-q^{st})$ are finite.

Now we assume that only few order parameters, namely those that correspond to the several largest eigenvalues, are different from zero.

In order to perform numerical estimations of the parameters we keep only three order parameters m_k and assume that the eigenvectors have the form of eqs. (1-3). Then the components of ξ, projected over a^0, a^1, a^2 are independent and we can assume that a self-averaging occurs with distribution of the self-averaged components proportional to the original distribution of ξ, in the case of a^0, and proportional to $\arcsin(x)$, with uniformly distributed x in the interval $[0, 1]$, in the case of a^1 and a^2. Having this in mind for $R = 0$ and solving numerically the corresponding equations at $T = 0$, we obtain $m_2 = m_3 = 0, m_1 \neq 0$ as only solution.

For $R > 0.5$ the numerical simulation shows that the solution with $m_1 \neq 0$ is not stable. The components m_2 and m_3 are different from zero, with $m_2^2 + m_3^2$ converging to some limit. Note that the components m_2 and m_3 are degenerated with the same energy, thus any fluctuation of the connectivity will break this symmetry and will fix their values.

Further analysis of the system, subject to ongoing work, will present a more detailed solution of the problem.

4 Conclusions

In this paper we have studied the minimal conditions for the appearance of spatial dependent activity in a binary neural network model. This analysis has been done analytically and also confirmed by simulations, which have been possible due to the finite number of relevant eigenvalues of the connectivity matrix. The latter gives a closed form for the equations describing the different order parameters and permits their later analysis.

We have shown that the appearance of asymmetric states of the neurons is related to the fact that the probability distributions of the patterns and the neural states are different. We point out that this condition is the minimal one to observe the effect.

Further analysis of the problem will be presented in a forthcoming publication.

Acknowledgments

The authors thank A. Treves and Y. Roudi for stimulating discussions. They warmly acknowledge the generous financial support from the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy, where this investigation was done. They also acknowledge the financial support from the Spanish Grants CICyT, TIC 01-572 and DGI.M.CyT.BFM2001-291-C02-01, respectively. E.K. is also supported by Grant "Promoción de la Investigación UNED'02". This work was done within the framework of the Associateship Scheme of ICTP.

References

[1] Y.Roudi and A.Treves, JSTAT (2004) P07010.

[2] J.Rubin and A.Bose, Network: Comput.Neural Syst. **15**(2004)133.

[3] J.Hertz, *Introduction to the theory of neural computation*, Perseus Publishing Group 1991.

[4] D.Amit, H.Gutfreund and H.Sompolinsky, Ann.Phys.**173**(1987)30.

[5] N.N.Bogolyubov, Physica(Suppl.) **26**(1960)S1.

[6] M.Mézard, G.Parisi and M.A.Virasoro, *Spin-glass theory and beyond*, World Scientific, 1987.