A statistical investigation of bidirectional associative memories (BAM)

J. Kurchan, L. Peliti, M. Saber

To cite this version:


HAL Id: jpa-00247018
https://hal.archives-ouvertes.fr/jpa-00247018
Submitted on 1 Jan 1994

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A statistical investigation of bidirectional associative memories (BAM)

J. Kurchan (1), L. Peliti (2 *) and M. Saber (3)

(1) Dipartimento di Fisica and Sezione INFN, Università di Roma « La Sapienza », P.le Aldo Moro 2, I-00185 Roma, Italy
(2) Institut Curie, Section de Physique et Chimie, Laboratoire Curie, 11 rue Pierre et Marie Curie, F-75231 Paris Cedex 05, France
(3) Département de Physique, Faculté des Sciences, Université « Moulay Ismaïl » B.P. 4010, Beni M’hamed, Meknès, Morocco

(Received 12 April 1994, received in final form 18 July 1994, accepted 21 July 1994)

Abstract. — We investigate by statistical mechanical methods a stochastic analogue of the bidirectional associative memories introduced by B. Kosko. We derive its storage capacity as a function of the total number of synapses and of the asymmetry of the network.

1. Introduction.

One of the main shortcomings in J. J. Hopfield’s model of associative memory [1, 2] is the fact that the stored patterns lack internal organization. Any bit is as important as any other bit. This runs contrary to the commonly experienced fact that the objects which we try to remember are often structured, so that some part of their description is more important than another. One possible way of taking this feature into account in Attractor Neural Networks [2] is to introduce correlations among the stored patterns. For example, Parga and Virasoro [3] have considered the way in which a hierarchical family of patterns can be stored in a Hopfield network, and Amit and Tsodyks [4] have shown that it is possible to introduce artificial correlations between temporally neighboring patterns, so that closeness in the Hamming distance gives a clue to temporal closeness.

One may choose a different approach, separating the information to be coded in different features: one possibility is categorization. This led quite naturally to the consideration of hierarchically structured networks, in which each layer is devoted to the treatment of a

(*) Bourgeois Henri de Rothschild. On leave of absence from Dipartimento di Scienze Fisiche and Unità INFN, Università « Federico II » Mostra d'Oltremare, Pad. 19, I-80125 Napoli, Italy. Associate INFN, Sezione di Napoli.
different feature. Several authors have considered hierarchical organizations of layers \([5-7]\). More general organizations have been considered \([8, 9]\). They have the common feature of considering layers of strongly interacting neurons, which are comparatively weakly connected among themselves. In the model considered by Brunel \([9]\), for example, a module (layer) codes for the category of the stored information, while another group of neurons codes for the remaining information. For example, one may separate a name into surname and given name, or the identity record of a person into a name and a photograph. Then incomplete information on either the name or the photograph may lead to the recall of the whole record.

Bart Kosko \([10]\) introduced in 1988 a completely different network architecture performing this task and named it a bidirectional associative memory (BAM). The task of the network is to retrieve a pair of associated activity patterns when a stimulus close to one or the other pattern is presented to the network. Kosko has shown how this task can be performed by a network with a simple two layer architecture. The interesting point he made is that the network works by « reverberation » even if there are no connections between the neurons on the same layer. He considers a network made up of two layers, A and B, of McCulloch-Pitts neurons. We denote by A and B the corresponding activity patterns :

\[
A = (a_1, a_2, \ldots, a_M); \quad B = (b_1, b_2, \ldots, b_N).
\]

Here the binary variables denoting the activities of the neurons are supposed to take up the values \(±1\).

The network evolves according to a deterministic process defined by the following scheme :

\[
(A, B) \rightarrow (A', B) \rightarrow (A', B') \rightarrow \ldots
\]

with successives parallel updatings of the two layers, one at the time. The updatings are defined by the following equations :

\[
a_i' = \text{sign} \left( \sum_{j=1}^{N} J_{ij} b_j \right),
\]

\[
b_j' = \text{sign} \left( \sum_{i=1}^{M} K_{ji} a_i' \right).
\]

It is important to remark that these equations provide no direct coupling between neurons belonging to the same layer at each time step.

If the two matrices \(J\) and \(K\) are the transpose of each other, it is easy to show that the following quantity acts as a Lyapunov function, since it decreases (or remains unchanged) at each time step :

\[
E = - \sum_{j=1}^{N} \sum_{i=1}^{M} J_{ij} a_i b_j = - \sum_{i=1}^{M} \sum_{j=1}^{N} K_{ji} b_j a_i.
\]

As a consequence, the attractors of the dynamics are fixed points, i.e., the local minima of \(E\), which satisfy the equation

\[
a_i = \text{sign} \left( \sum_{j=1}^{N} J_{ij} b_j \right); \quad b_j = \text{sign} \left( \sum_{i=1}^{M} a_i J_{ij} \right).
\]

We shall consider from now on only the matrix \(J = K^T\).
Storage is performed by choosing a form of the coupling matrix \( J \) reminiscent of the Hopfield synaptic matrix:

\[
J_{ij} = \frac{1}{\sqrt{MN}} \sum_{\mu=1}^{\mu} \xi_i^\mu \eta_j^\mu
\]  

(1.6)

Here \( \xi_i^\mu = (\xi_1^\mu, \xi_2^\mu, \ldots, \xi_M^\mu) \) and \( \eta_j^\mu = (\eta_1^\mu, \eta_2^\mu, \ldots, \eta_N^\mu) \), \( \mu = 1, 2, \ldots, \mu \), are activity patterns: \( \xi_i^\mu, \eta_j^\mu = \pm 1 \). \( \forall i, j, \mu \). The factor \( \sqrt{MN} \) is irrelevant at this stage, and has been chosen for later convenience. We shall assume here, and throughout the paper, that the components \( \xi_i^\mu, \eta_j^\mu \) are independent random variables, assuming the values \( +1, -1 \) with equal probability.

Retrieval is performed in a way analogous to Hopfield models: one of the layers is prepared in an initial condition close to one of the corresponding stored patterns (\( \xi \) for layer A, \( \eta \) for layer B). Then the system evolves, taking care to update first the other layer: the same reasonings as for the Hopfield models show that eventually the system reaches a fixed point close to the form \( (\xi_i^\mu, \eta_j^\mu) \), where the pattern \( \mu_0 \) is determined by the initial condition.

One can suggest several possible uses of this architecture as an association device. Kosko [11] has also suggested some interesting modifications of it, and in particular the introduction of a learning rule leading to what he calls Adaptive Bidirectional Associative Memories (ABAM). Here we shall stick to the simple «nonadaptive» architecture we have just described.

Our aim is to investigate the phase diagram of a stochastic version of the BAM, in particular in order to determine its storage capacity and its performance. The method is close to the classic work of Amit, Gutfreund and Sompolinsky [12, 13] on Hopfield networks. It is a first step towards the consideration of more complex architectures, in which single neural layers act as a unity. The point which makes this architecture worth investigating is the fact that it may become more efficient in the case of great asymmetry between the two layers, as it happens, e.g., for the case of the association of a name with an image. In this case one needs only \( NM \) links instead of \( (N + M)^2/2 \) to have essentially comparable retrieval features, and the former figure is much smaller when \( N \ll M \). The price to be paid is in the attraction basin, which may be considerably smaller for the smaller layer.

We now define the stochastic version of the BAM. A fictitious «inverse temperature» \( \beta = 1/T \) can be introduced as usual by considering a stochastic generalization of the updating equations (1.3); namely, the probability that the state of spin \( i \) is equal to \( a_i^j \) at the next time step is given by

\[
p(a_i^j) = \frac{e^{\beta h_i a_i}}{e^{\beta h_i a_i} + e^{-\beta h_i a_i}}
\]

(1.7)

The local field \( h_i \) is defined by

\[
h_i = \frac{1}{\sqrt{MN}} \sum_{i} J_{ij} b_j.
\]

(1.8)

The layer B is governed by completely analogous equations: the probability that unit \( j \) takes up the state \( b_j^l \) at the next time step is given by

\[
p(b_j^l) = \frac{e^{\beta h_j b_j}}{e^{\beta h_j b_j} + e^{-\beta h_j b_j}}
\]

(1.9)
where the local field \( k_j \) is given by
\[
k_j = \sum_{i=1}^{M} J_{ij} a_i .
\] (1.10)

Let us remark that \( b_j \) and \( a_i \) in equations (1.7) and (1.9) respectively stand for the current state of the respective layers. There is no memory in the dynamics.

In order to obtain the Boltzmann distribution of Hamiltonian \( E \) (Eq. (1.4)) the updating rule should be modified, by choosing at random, at each time step, the layer to be updated. It is easy to convince one self, in fact, that if the layers are updated in succession, as in the deterministic rule, the resulting dynamics does not obey detailed balance.

As a consequence, the presentation of the stimulus to the network cannot take place simply by setting one of the layers (say A) in a given initial state : since then, if the next layer to be updated turns out to be A, the stimulus is irretreivably lost. Hence one should assume that the stimulus is presented by applying to the chosen layer a sufficiently strong external field \( h_0 \) whose sign is determined by the pattern one wants to treat. We are however only concerned with the dynamics that follows once this external field is set to zero.

The paper is organized as follows : section 2 contains the analysis of the simple case in which the number \( p \) of stored patterns is finite in the thermodynamic limit Section 3 discusses the case of an infinite number of patterns. Section 4 contains a few conclusions.

2. Finite number of patterns.

We now derive the phase diagram of the stochastic BAM defined above when the number \( p \) of stored patterns is finite. In this situation the system is always self-averaging, and it is not necessary to introduce replicas. We consider therefore the partition function
\[
Z = \text{Tr} \exp(-\beta E) = \text{Tr} \exp \left\{ -\frac{\beta}{\sqrt{MN}} \sum_{\mu=1}^{\nu} (\xi^\mu \cdot a)(\eta^\mu \cdot b) \right\} .
\] (2.1)

We denote by \( \text{Tr} \) the sum over activity patterns :
\[
\text{Tr} = \sum_{a,b} \sum_{\{i \mid x_i = 1\}} \sum_{\{i \mid x_i = -1\}}
\] (2.2)

The scalar product \((\xi \cdot a)\) denotes the sum
\[
(\xi \cdot a) = \sum_{i=1}^{M} \xi_i a_i,
\] (2.3)

and analogously for B.

We now introduce two Hubbard-Stratonovich variables \( u_{\mu} \) and \( v_{\mu} \), exploiting the identity
\[
\chi \nu = \frac{1}{4} [(\chi + \nu)^2 - (\chi - \nu)^2]
\] (2.4)

We thus obtain
\[
Z = (\sqrt{MN} \beta)^{\nu} \int \prod_{\mu=1}^{\nu} \frac{du_{\mu} \cdot dv_{\mu}}{2 \pi i^{1/2}} \exp \left\{ -\frac{\beta}{2} \sqrt{MN} \sum_{\mu} (u_{\mu}^2 + v_{\mu}^2) \right\} \times \nonumber
\]
\[
\times \text{Tr} \exp \left\{ \frac{\beta}{\sqrt{2}} [u_{\mu} ((\xi^\mu \cdot a) + (\eta^\mu \cdot b)) + iv_{\mu} ((\xi^\mu \cdot a) - (\eta^\mu \cdot b))] \right\} .
\] (2.5)
We can now introduce the two auxiliary fields $m_\mu$ and $n_\mu$ by
\[
m_\mu = (u_\mu + iv_\mu)/\sqrt{2};
\]
\[
n_\mu = (u_\mu + iv_\mu)/\sqrt{2};
\]
so that
\[
m_\mu n_\mu = \frac{1}{2} (u_\mu^2 + v_\mu^2). \tag{2.7}
\]

We therefore obtain, up to a normalizing factor, the following expression for the partition function:
\[
Z = \int \prod_{\mu=1}^p \, dm_\mu \, dn_\mu \exp \left\{ -\beta \sqrt{MN} \sum_{\mu=1}^p m_\mu n_\mu \right\} \times \\
\times \text{Tr} \exp \left\{ \beta m_\mu (\xi^- a) + \beta n_\mu (\eta^- b) \right\}. \tag{2.8}
\]

The trace can be now simply evaluated, yielding
\[
\text{Tr} \exp \left\{ \beta \sum_{\mu=1}^p \left[ m_\mu (\xi^- a) + n_\mu (\eta^- b) \right] \right\} = \\
= \exp \left\{ \sum_{i=1}^M \ln 2 \cosh \left( \sum_{\mu=1}^p \beta \sqrt{M/N} m_\mu \xi_i^- \right) + \sum_{i=1}^N \ln 2 \cosh \left( \sum_{\mu=1}^p \beta \sqrt{N/M} n_\mu \eta_i^- \right) \right\}. \tag{2.9}
\]

When $M$ and $N$ go to infinity, the sum over the $M$ (respectively $N$) independent random variables $\xi_i, (\eta_i)$ yields, up to a factor, the average over the distribution of the patterns, which we denote by a bar:
\[
\sum_{i=1}^M \ln 2 \cosh \left( \sum_{\mu=1}^p \beta \sqrt{M/N} m_\mu \xi_i^- \right) \equiv M \ln 2 \cosh \left( \sum_{\mu=1}^p \beta \sqrt{M/N} m_\mu \xi_i^- \right), \tag{2.10}
\]
while an analogous relation holds for the second term. As in the analogous case of the Hopfield model, the system is self-averaging. It is therefore warranted to calculate the partition function by the saddle point method.

We consider the order parameters $\mathbf{m} = (m_1, m_2, \ldots, m_p)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_p)$ as vectors with $p$ components. The free energy $F$ is given as a function of their saddle point values by
\[
F = \sqrt{MN} (\mathbf{m} \cdot \mathbf{n}) - \beta^{-1} M \ln 2 \cosh \left( \beta \sqrt{N/M} (\xi \cdot \mathbf{m}) \right) - \\
- \beta^{-1} N \ln 2 \cosh \left( \beta \sqrt{M/N} (\eta \cdot \mathbf{n}) \right). 
\]

The saddle point equations are given by
\[
\mathbf{n} = \xi \tanh \left( \beta \sqrt{N/M} (\xi \cdot \mathbf{m}) \right); \\
\mathbf{m} = \eta \tanh \left( \beta \sqrt{M/N} (\eta \cdot \mathbf{n}) \right). \tag{2.11}
\]
The simplest case is the one of the Mattis (or retrieval) states, in which only one component of the order parameters is nonzero. In this case it is easy to see that the component must be the same for \( m \) and \( n \). The corresponding saddle point equations read:

\[
\begin{align*}
    n &= \tanh \left( \beta \sqrt{\frac{N}{M}} m \right); \\
    m &= \tanh \left( \beta \sqrt{\frac{M}{N}} n \right).
\end{align*}
\] (2.12)

We see that \( m \) and \( n \) must vanish or be nonzero together. In particular, \( m \) must be a solution of the equation

\[
    m = \tanh \left( \beta \sqrt{\frac{M}{N}} \tanh \left( \beta \sqrt{\frac{N}{M}} m \right) \right). \tag{2.13}
\]

which has the same qualitative behavior as the usual Ising mean field equation:

\[
    m = \tanh \left( \beta \tanh m \right). \tag{2.14}
\]

The discussion is then quite straightforward: if \( \beta \leq 1 \), only the zero solution exists (disordered state); for \( \beta > 1 \) there are two solutions of opposite sign, where \( m \) is one of the two nonzero solutions of equation (2.13) and \( n \) is given by

\[
    n = \tanh \left( \beta \sqrt{\frac{N}{M}} m \right). \tag{2.15}
\]

For \( \beta \to \infty \) both \( m \) and \( n \) tend to either \( +1 \) or \(-1 \). These thermodynamic states correspond to retrieval of a pair of associated patterns.

Other solutions of the mean field equations correspond to «mixed states», i.e., they have nonzero overlap with more than one pair of patterns. These solutions can be identified and discussed along the lines of references [12]. We have checked, in particular, that symmetric solutions (which overlap equally with more than one stored pattern) are at most metastable and lose their stability when \( \beta \) is slightly larger than one, i.e., immediately below the transition temperature.

3. Infinite number of patterns.

We now consider the expression for the partition function of \( n \) replicas of the system:

\[
    Z^n = \text{Tr} \exp \left\{ \frac{\beta}{\sqrt{MN}} \sum_{\mu=1}^{n} \sum_{t=1}^{e} (\xi^\mu \cdot \eta^\mu) (\xi^\mu \cdot \eta^\mu) \right\}. \tag{3.1}
\]

We have introduced a vector notation for the summation over the «neural» indices. By means of the manipulations described in the previous section we obtain a form analogous to equation (2.8):

\[
    Z^n = \Xi \int \prod_{\mu,t} dm^t_\mu \, dn^t_\mu \, \text{Tr} \exp \left\{ \beta \sum_{\mu,t} m^t_\mu (\xi^\mu \cdot \eta^\mu) + \beta \sum_{\mu,t} n^t_\mu (\eta^\mu \cdot \eta^\mu) \right\} \times \\
    \times \exp \left\{ -\beta \sqrt{MN} \sum_{\mu,t} m^t_\mu n^t_\mu \right\}. \tag{3.2}
\]
where $\Xi$ is a constant. We can now distinguish between « low » ($\nu = 1, \ldots, r$) and « high » ($\mu = r + 1, \ldots, p$) components, according to reference [7]. Summing over the « high » patterns we obtain

$$\exp \left\{ \beta \sum_{\mu} \sum_{\ell} m^\ell_{\mu} (\xi \cdot \phi^\ell) \right\} = \exp \left\{ \sum_{\mu} \sum_{\ell} \ln \cosh \left( \beta \sum_{\nu} m^\ell_{\mu} a^\nu \right) \right\},$$

where we have denoted by the bar the average over patterns. An analogous equation holds for the B units. We now rescale the $m$ and $n$ fields according to the scheme

$$m^\ell_{\mu} \rightarrow \frac{m^\ell_{\mu}}{\sqrt{M}}, \quad n^\ell_{\mu} \rightarrow \frac{n^\ell_{\mu}}{\sqrt{N}}.$$ (3.4)

This yields the following contribution of the « high » patterns to the partition function:

$$\% \approx \exp \left\{ \frac{\beta^2}{2M} \sum_{\mu} \sum_{\ell} m^\ell_{\mu} m_{\mu}^\ell a^\ell a^\ell + \frac{\beta^2}{2N} \sum_{\mu} \sum_{\ell} n^\ell_{\mu} n_{\mu}^\ell b^\ell b^\ell \right\} \times \exp \left\{ -\beta \sum_{\mu} \sum_{\ell} m^\ell_{\mu} n_{\mu}^\ell \right\}.$$ (3.5)

We introduce the « spin glass parameters » $q_{\ell \ell}$ and $\bar{q}_{\ell \ell}$, by the definition

$$q_{\ell \ell} = \frac{1}{M} \sum_{\ell} a^\ell a^\ell, \quad \bar{q}_{\ell \ell} = \frac{1}{N} \sum_{\ell} b^\ell b^\ell; \quad k \neq \ell.$$ (3.6)

Integrating over the $m, n$ fields yields

$$\% = \text{const.} \exp \left\{ -\frac{D}{2} \ln \text{Det} [\beta^2 (Q + I)(\bar{Q} + I) - I] \right\}.$$ (3.7)

Here $I$ is the identity matrix in the space of replicas, while the matrices $Q$ and $\bar{Q}$ are defined by

$$[Q]_{\ell \ell} = q_{\ell \ell}, \quad [\bar{Q}]_{\ell \ell} = \bar{q}_{\ell \ell} \quad k \neq \ell.$$ (3.8)

We finally obtain the following expression for the partition function, up to irrelevant multiplicative constants:

$$Z' = \int \prod_{\ell} dm^\ell_{\mu} \prod_{\mu} d\tilde{m}^\ell_{\mu} \int \prod_{\ell} dq_{\ell \ell} \prod_{\ell} d\tilde{q}_{\ell \ell} \prod_{\ell} d\bar{q}_{\ell \ell} d\bar{\bar{q}}_{\ell \ell} \times$$

$$\times \exp \left\{ -\frac{D}{2} \text{Tr} \ln \left[ \beta^2 (Q + I)(\bar{Q} + I) - I \right] - \beta F_A - \beta F_B \right\}$$

$$-\frac{R}{2} \beta^2 \sum_{\ell \ell} r_{\ell \ell} q_{\ell \ell} - \frac{R}{2} \beta^2 \sum_{\ell \ell} \bar{r}_{\ell \ell} \bar{q}_{\ell \ell} - \beta \sqrt{MN} \sum_{\ell} m^\ell_{\mu} n^\ell_{\mu}.$$ (3.9)

Here $F_A$ and $F_B$ are the spin traces defined by

$$\exp(-\beta F_A) = \text{Tr} \exp \left\{ \frac{D}{2M} \beta^2 \sum_{\ell \ell} r_{\ell \ell} a^\ell a^\ell + \beta \sqrt{\frac{N}{M}} \sum_{\ell} m^\ell_{\mu} \xi_{\ell \ell} a^\ell \right\};$$ (3.10)
and the analogue for B:

$$\exp(-\beta F_B) = \operatorname{Tr} \left\{ \frac{p}{2N} \beta^2 \sum_{i \neq \ell} \bar{r}_{i\ell} b^i b^\ell + \beta \sqrt{\frac{M}{N}} \sum_i n_i^\ell \eta^\ell b^i \right\}. \quad (3.11)$$

We define:

$$\gamma_N = \sqrt{\frac{N}{M}}; \quad \gamma_M = \sqrt{\frac{M}{N}} = \frac{1}{\gamma_N}; \quad \alpha_q = \frac{p}{N}; \quad \alpha_M = \frac{p}{M}; \quad \alpha = \frac{p}{\sqrt{MN}}. \quad (3.12)$$

We now assume replica symmetry, implying that $r_{i\ell} = r_{j\ell} = q_j$, $m_i^\ell = m_i$, (and analogous relations for $\bar{r}$, $\bar{q}$ and $n_i^\ell$). This is expected to be a good approximation, just as in the Hopfield case [14]. The spin traces can be then explicitly computed to yield

$$\exp(-\beta F_A - \beta F_B) = \int \frac{dz}{\sqrt{2\pi}} \int \frac{dz}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} z^2 - \frac{1}{2} z^2 \right) \times$$

$$\times \exp \left\{ n \ln 2 \cosh \left[ \beta \left( z \sqrt{\alpha_M r_N} + \gamma_N \sum m_i \xi^i \right) \right] \right\} \times \exp \left\{ n \ln 2 \cosh \left[ \beta \left( z \sqrt{\alpha_N \bar{r}} + \gamma_M \sum n_i \eta^i \right) \right] \right\} - \frac{n}{2} \beta^2 (\alpha_M r + \alpha_N \bar{r}). \quad (3.13)$$

One the other hand, an explicit evaluation yields:

$$\epsilon_i = \operatorname{Det} \left[ \beta^2 (Q + I)(\bar{Q} + I) - I \right] = \left[ \beta^2 (1 - q)(1 - \bar{q}) - 1 \right]^{n-1} \times$$

$$\times \left\{ \left[ \beta^2 (1 - q)(1 - \bar{q}) - 1 \right] + n \beta^2 (q + \bar{q} - 2 q\bar{q}) \right\}. \quad (3.14)$$

We thus obtain the limit as $n \to 0$ of $\Gamma$:

$$\Gamma \equiv n \left\{ \ln \left[ \beta^2 (1 - q)(1 - \bar{q}) - 1 \right] + \frac{\beta^2 (q + \bar{q} - 2 q\bar{q})}{\beta^2 (1 - q)(1 - \bar{q}) - 1} \right\}. \quad (3.15)$$

By considering the straightforward limits

$$\sum_{i\ell} r_{i\ell} q_{i\ell} = n(n-1)rq \rightarrow nrq;$$

$$\sum_{i\ell} \bar{r}_{i\ell} \bar{q}_{i\ell} \rightarrow -n\bar{r}\bar{q}; \quad (3.16)$$

$$-\beta \sqrt{MN} \sum_{i} m_i^\ell n_i^\ell \rightarrow -n\beta \sqrt{MN} m_i n_i.$
we obtain the following expression for the average free energy \( G = \ln Z \):

\[
\frac{G}{n \times MN} \approx - \gamma M \int_{-\infty}^{\infty} \ln 2 \cosh \left[ \beta \left( z \sqrt{\alpha_M r} + \gamma_N \sum m_i \xi_i \right) \right] - \\
- \gamma N \int_{-\infty}^{\infty} \ln 2 \cosh \left[ \beta \left( \bar{z} \sqrt{\alpha_N \bar{r}} + \gamma_M \sum n_i \eta_i \right) \right] \\
+ \beta \sum m_i n_j + \frac{\alpha}{2} \beta^2 r(1 - q) + \frac{\alpha}{2} \beta^2 \bar{r}(1 - \bar{q}) \\
+ \alpha \left\{ \ln [\beta^2 (1 - q)(1 - \bar{q}) - 1] + \frac{\beta^2 (q + \bar{q} - 2 q \bar{q})}{\beta^2 (1 - q)(1 - \bar{q}) - 1} \right\}
\]

(3.17)

Here we have introduced the shorthand \( \int_{z} = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2 \pi}} \). The saddle point equations are obtained by taking the derivatives of \( G \). The derivatives with respect to \( m \) and \( n \) are straightforward:

\[
\int_{z} \xi_i^{'} \tanh \left[ \beta \left( z \sqrt{\alpha_M r} + \gamma_N \sum m_i \xi_i^{'} \right) \right] = n_i, \\
\int_{z} \eta_i^{'} \tanh \left[ \beta \left( \bar{z} \sqrt{\alpha_N \bar{r}} + \gamma_M \sum n_i \eta_i^{'} \right) \right] = m_i.
\]

(3.18)

The derivatives with respect to the Lagrangian multipliers \( r \) and \( \bar{r} \) yield:

\[
\int_{z} \tanh^2 \left[ \beta \left( z \sqrt{\alpha_M r} + \gamma_N \sum m_i \xi_i^{'} \right) \right] = q; \\
\int_{z} \tanh^2 \left[ \beta \left( \bar{z} \sqrt{\alpha_N \bar{r}} + \gamma_M \sum n_i \eta_i^{'} \right) \right] = \bar{q}.
\]

(3.19)

On the other hand, the saddle points with respect to \( q \) and \( \bar{q} \) yield:

\[
\frac{r}{2} = \frac{\bar{q} + \beta^2 q(1 - \bar{q})^2}{[1 - \beta^2 (1 - q)(1 - \bar{q})]^2}, \\
\frac{\bar{r}}{2} = \frac{q + \beta^2 \bar{q}(1 - q)^2}{[1 - \beta^2 (1 - q)(1 - \bar{q})]^2}.
\]

(3.20)

It is convenient to define

\[
c = \beta (1 - q); \\
\bar{c} = \beta (1 - \bar{q}).
\]

(3.21)
We now discuss the $\beta \to \infty$ limit, where the saddle point equations become:

$$
\xi_{\nu} \text{erf} \left( \frac{\gamma_N \sum n_i \xi_i}{\sqrt{2} \alpha_M r} \right) = n_{\nu};
$$

$$
\eta_{\nu} \text{erf} \left( \frac{\gamma_M \sum n_i \eta_i}{\sqrt{2} \alpha_N \tilde{r}} \right) = m_{\nu}.
$$

$$
c = \sqrt{\frac{2}{\pi \alpha_M r}} \exp \left[ -\frac{\gamma_N^2 (\mathbf{m} \cdot \xi)^2}{2 \alpha_M r} \right];
$$

$$
\tilde{c} = \sqrt{\frac{2}{\pi \alpha_N \tilde{r}}} \exp \left[ -\frac{\gamma_M^2 (\mathbf{n} \cdot \eta)^2}{2 \alpha_N \tilde{r}} \right].
$$

We introduce the variables $\nu$ and $\tilde{\nu}$, defined by

$$
\nu = \frac{\gamma_N m}{\sqrt{2} \alpha_M r}, \quad \tilde{\nu} = \frac{\gamma_M n}{\sqrt{2} \alpha_N \tilde{r}}.
$$
All the variables can now be expressed in terms of \(y\) and \(\bar{y}\), yielding

\[
\begin{align*}
\alpha &= \frac{2\gamma e^{-\gamma^2}}{\sqrt{\pi \gamma} \text{erf}(\bar{y})} ; \\
\bar{\alpha} &= \frac{2\bar{\gamma} e^{-\bar{\gamma}^2}}{\sqrt{\pi \gamma} \text{erf}(\gamma)} ; \\
\frac{1 + \bar{c}^2}{(1 - \bar{c}\bar{c})^2} &= \frac{\text{erf}^2(\bar{y})}{4 \alpha_N \gamma^2} , \\
\frac{1 + c^2}{(1 - c\bar{c})^2} &= \frac{\text{erf}^2(y)}{4 \alpha_M \gamma^2} .
\end{align*}
\]  

(3.27) (3.28)

In this way the problem is reduced to solving the two self-consistency equations that determine \(y\) and \(\bar{y}\). These equations can be read off equations (3.27), (3.28). Once \(y\) and \(\bar{y}\) are known, the overlaps can be determined from

\[
m = \text{erf}(\bar{y}) ; \\
n = \text{erf}(y) .
\]  

(3.29)

We show in figure 1 the overlaps \(m\) (upper curve) and \(n\) as a function of \(\gamma = \gamma_N\), when \(N < M\). If \(N > M\) the situation is symmetrical, upon exchanging \(m\) with \(n\) (and \(\gamma_M\) with \(\gamma_N\)).

The critical capacity

\[
\alpha_c = \frac{p}{\sqrt{MN}} ,
\]  

(3.30)

above which there are no nonzero solutions of the above equations is plotted in figure 2 as a function of \(\gamma = \gamma_N\).

For small values of \(\gamma = \gamma_N\), the storage capacity becomes proportional to \(\gamma\), as one may argue from a simple noise-to-signal ratio argument [15]. We can obtain this result as follows. Let us define:

\[
d = \gamma c ; \\
\bar{d} = \bar{c}/\gamma ; \\
\bar{\alpha} = \alpha/\gamma .
\]  

(3.31)

It is now easy to see that \(d\), \(\bar{d}\) and \(\bar{\alpha}\) tend to finite limits as \(\gamma \to 0\). Indeed, if we let these definitions into equations (3.27), (3.28), we obtain, in the limit \(\gamma \to 0\):

\[
d = \frac{2\gamma e^{-\gamma^2}}{\sqrt{\pi} \text{erf}(\bar{y})} , \\
\bar{d} = \frac{2\bar{\gamma} e^{-\bar{\gamma}^2}}{\sqrt{\pi} \text{erf}(\gamma)} .
\]  

(3.32)

while the relation between \(y\) and \(\bar{y}\) becomes

\[
\bar{y} = \frac{\sqrt{\pi} \text{erf}(y)}{2 e^{-\gamma}} .
\]  

(3.33)

The critical capacity \(\bar{\alpha}_c\) is then obtained as

\[
\sqrt{\bar{\alpha}_c} = \max \left\{ \frac{\text{erf}(\bar{y})}{2 \sqrt{\gamma}} - \frac{e^{-\gamma}}{\sqrt{\pi}} \right\} ,
\]  

(3.34)

where \(y\) is expressed in terms of \(\bar{y}\) via equation (3.33). One obtains

\[
\alpha_c = 0.248677 \gamma .
\]  

(3.35)
Fig. 1 — Overlaps $m = (1/M) \langle \sum \xi_i, a_i \rangle$ (continuous curve) and $n = (1/N) \langle \sum \eta_i, b_i \rangle$ (broken curve) with the stored patterns in the Mattis states at $\beta = \infty$ and $\alpha = \alpha_{\perp}$ as a function of $\gamma = \sqrt{N/M}$.

Fig. 2. — Critical capacity $\alpha = \alpha_{\perp} = p_{\perp} \sqrt{MN}$ as a function of $\gamma = \sqrt{N/M}$. The ratio $\alpha/\gamma$ is also plotted (scale on the r.h.s.). The limit of this ratio for small $\gamma$ is equal to 0.24877.
The ratio $\alpha / \gamma$ is also plotted in figure 2, in order to make this behavior more evident. On the other hand, the number of stored patterns approaches $\alpha_c^H \sqrt{2}$ for $\gamma \to 1$. where $\alpha_c^H = 0.1398$ is the critical storage for a Hopfield network. One may explain this result by arguing that storing a pair of patterns corresponds to doubling the information stored per pattern: and therefore the total information stored in the $M^2/2$ synapses, at saturation, is the same as in an ordinary Little-Hopfield network.


This is, to our knowledge the first analysis of the behavior of BAMs by statistical mechanics methods (1). We have shown in this paper that the methods introduced by Amit and collaborators [12, 13] can be extended without unexpected difficulties to cover this architecture, and that the results one obtains for the storage capacity of the network can be interpreted by simple arguments in the limits of symmetrical and strongly asymmetrical networks. It is now possible to envisage the systematic study of related architectures, in which the number of related segments is higher than two.

Acknowledgments.

We are grateful to B. Tirozzi for having communicated to us reference [16] and to S. Franz for useful discussions. The visit of M. Saber to Paris, where part of this work was done, was supported by the Section de Physique et Chimie of the Institut Curie.

References

[15] We thank Silvio Franz for clarifying this point.

(1) Since completion of this work we have been informed that the model had been investigated, via a different method, by Englisich et al. [16]. To the extent of their overlap, their results agree with ours.