



HAL
open science

Storage Capacity of Generalized Palimpsests

D. Bonnaz

► **To cite this version:**

D. Bonnaz. Storage Capacity of Generalized Palimpsests. Journal de Physique I, 1997, 7 (12), pp.1709-1721. 10.1051/jp1:1997165 . jpa-00247482

HAL Id: jpa-00247482

<https://hal.science/jpa-00247482>

Submitted on 4 Feb 2008

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.

Storage Capacity of Generalized Palimpsests

D. Bonnaz (*)

Institut de Physique Expérimentale, Université de Lausanne, 1015 Lausanne, Switzerland

(Received 24 April 1997, revised 11 June 1997, accepted 22 August 1997)

PACS.87.10.+e – General, theoretical, and mathematical biophysics (including logic of biosystems, quantum biology, and relevant aspects of thermodynamics, information theory, cybernetics, and bionics)

PACS.89.70.+c – Information science

PACS.75.10.Nr – Spin-glass and other random models

Abstract. — A simple analytical study of a short term memory model is performed. This model consists of a symmetric p-neuron interaction between N neurons. Learning is achieved by a generalized Hebb rule. Saturation is prevented by the introduction of a bound A to the couplings. At each time step, an input pattern is drawn at random, independently of the previous ones. The determination of the life time T of a memorized pattern viewed as a function of A and N is accomplished by a statistical study of the dynamic of the learning process which has been made possible under the assumption that the couplings evolve independently. This simplification reduces the determination of T to a one-dimensional problem, by considering energies rather than couplings. The choice of the optimal value A_{opt} of A is a compromise between the success of the learning process and the maximization of T . The essential results are expressed by the formulae $T \propto A^2$ and $A_{\text{opt}} \propto N^{\frac{p-1}{2}}$

1. Introduction

The Hopfield model [1] suffers from the sudden appearance of confusion as soon as the number of input patterns crosses a threshold. This model is therefore not suited to describe the behavior of the memory during arbitrary long learning phases. Many elegant ideas have been considered to remedy to this problem and calculate the capacity of such systems, by modifying the acquisition intensity [2] or by considering transmission delays (see [3] for instance). As suggested by Hopfield [4], the Hebb rule can also be modified by introducing bounds to the couplings. This alternative is common in recent models, which include multilayer networks as in [5] for example. Numerical studies can be found in [6, 7]. Two different analytical methods have also been developed: Hemmen [8] considered the learning process as a discrete dynamical system whose asymptotics can be calculated exactly; Derrida and Nadal [9] reduced the investigation, in the case of asymmetric diluted neural networks, to the study of a one-dimensional random walk subject to constraints. For a presentation of the rich behaviour exhibited by such systems, see [10].

(*) Author for correspondence (e-mail: dbonnaz@pop-server.unil.ch)

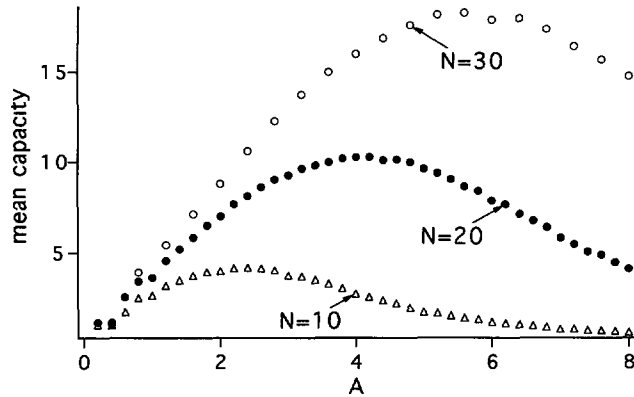


Fig. 1. — Plot of the mean capacity *versus* A for networks with $N = 10, 20$ and 30 neurons, obtained numerically in the case $p = 3$. No error is tolerated in the patterns retrieval.

The goal is to provide an elementary understanding of systems that are able to record every input patterns, inevitably to the detriment of good capacity. In order to improve storage capacity, a p -neuron interaction between N neurons is considered, following Abbott [11] (calculus of the storage capacity using the replica method) and Gardner [12] (among other things, determination of the storage capacity without the replica trick in the limit $p \rightarrow \infty$). The capacity is estimated by the mean life time T of one recorded pattern. Figure 1 shows the typical dependence of the capacity on the bound A for various sizes of the system.

We proceed to an analytical study of fully connected symmetric networks by assuming that the couplings are almost uncorrelated for $p \gg 2$. Our approach differs in that respect from [9]. The idea is then to consider energy models instead of interaction models, inspired by Derrida's random energy model of spin glasses [13].

2. Definition of the Model

The system considered here is composed of N neurons, with neuron i in state $S^i \in \{-1, 1\}$ (1 standing for firing, -1 for quiescent) interacting *via* the Hamiltonian:

$$H[\mathbf{S}] = - \sum_{i_1 < i_2 < \dots < i_p} J_{i_1 i_2 \dots i_p} S^{i_1} S^{i_2} \dots S^{i_p}, \quad 1 \leq p \leq N. \quad (1)$$

There are C_N^p distinct couplings $J_{i_1 i_2 \dots i_p}$. Each coupling evolves according to the discrete rule

$$J_{i_1 i_2 \dots i_p}(t) \rightarrow J_{i_1 i_2 \dots i_p}(t+1) = f_A \left(J_{i_1 i_2 \dots i_p}(t) + S_t^{i_1} S_t^{i_2} \dots S_t^{i_p} \right) \quad (2)$$

where

$$f_A(x) = \begin{cases} A, & x > A \\ x, & |x| \leq A \\ -A, & x < -A \end{cases}$$

is the limiting function and $\mathbf{S}_t = \{S_t^i \mid i = 1, 2, \dots, N\}$ is the random input pattern at time t . The memorization process is successful if \mathbf{S}_t corresponds to a strict minimum of energy in the configuration space. To understand how it works, suppose that the function f_A is removed.

The system registers more and more patterns but it becomes robust in the sense that the couplings increase, at the expense of flexibility (more precisely, the system cannot create minima of energy at the right places any more and spurious minima invade the configuration space). Furthermore, as a consequence of the learning process, the old patterns are progressively forgotten. These two facts contribute to a loss of ability of the system to record and remember at the same time. To avoid such a dramatic situation, the freedom of the couplings must be limited to insure that every input pattern is memorized. The system remains at the edge of robustness (the system is not critical in the sense that the life time distribution is not a power law for the values of A considered), to the detriment of capacity because less deep energy wells get faster obliterated. This problem can be viewed as an optimization process. The remaining part of the section is devoted to the study of the model by considering the asymptotic distribution of the couplings.

2.1. A TOY MODEL: THE CASE $p = 1$. — Consider first the case where each neuron interacts only with a site dependent external field. Since there is only one strict minimum of energy, the capacity is limited to one pattern. The quantity of interest is the probability that an input pattern is successfully registered. If P_A denotes the probability that a J_i has the correct sign (*i.e.* the same as S_i^c) right after the trial, then the probability of success is given by $(P_A)^N$, where

$$P_A = \begin{cases} \frac{A+1}{2A+1} & A \text{ integer} \\ \frac{1}{2} + \frac{1}{2A+1} & A \text{ half-integer} \\ \frac{1}{2} + \frac{2([2A] - [A]) + 1}{([2A] + 1)([2A] + 2)} & \text{otherwise} \end{cases} \quad (3)$$

where the brackets denote integer part. For further details, see appendix A. The probability of success goes exponentially to 0 as $N \rightarrow \infty$ unless $A < 1$, so any $A \in (0, 1)$ is optimal.

2.2. THE GENERAL CASE. — Consider now the case $p > 1$. Suppose for simplicity that A is a positive integer and that initial conditions are

$$J_{i_1 i_2 \dots i_p}(t=0) = 0 \quad \forall i_1 < i_2 < \dots < i_p \quad (4)$$

The equation (2) controlling the discrete evolution of each coupling can be rewritten as

$$J_{i_1 i_2 \dots i_p}(t) \rightarrow J_{i_1 i_2 \dots i_p}(t) + \Delta J_{i_1 i_2 \dots i_p}(t) \quad (5)$$

where $\Delta J_{i_1 i_2 \dots i_p}(t) = \pm 1$ with probability 1/2 for $J_{i_1 i_2 \dots i_p}(t) \neq \pm A$, with the sign depending on the choice of the input pattern. We can have also $\Delta J_{i_1 i_2 \dots i_p}(t) = 0$ or ∓ 1 with probability 1/2, the sign corresponding respectively to the cases $J_{i_1 i_2 \dots i_p}(t) = \pm A$.

Let $\rho_J(t)$ be the probability one coupling equals J , defined for $J \in \{-A, -A+1, \dots, A\}$, then they obey the equations

$$\begin{aligned} \rho_A(t+1) &= \frac{1}{2}\rho_A(t) + \frac{1}{2}\rho_{A-1}(t) \\ \rho_J(t+1) &= \frac{1}{2}\rho_{J+1}(t) + \frac{1}{2}\rho_{J-1}(t) \quad \text{for } J \neq \pm A \\ \rho_{-A}(t+1) &= \frac{1}{2}\rho_{-A}(t) + \frac{1}{2}\rho_{-A+1}(t) \end{aligned} \quad (6)$$

which have already been studied in [9]. When $t \rightarrow \infty$, the solution converges to the uniform distribution:

$$\rho_J(t \rightarrow \infty) = \frac{1}{2A+1} \quad \forall J \in \{-A, -A+1, \dots, A\}. \quad (7)$$

Consider now only k different couplings among C_N^p . It is expected that they constitute, at least in the large A limit, a subset of independent variables if $k < p+1$. Indeed, if the effects of bounds are neglected, the variations of the couplings do not depend on the history. Thus it is sufficient to check relations such as for instance for $p=2$:

$$\langle (\Delta J_{12}(t))^{\alpha_1} (\Delta J_{13}(t))^{\alpha_2} \rangle = \langle (\Delta J_{12}(t))^{\alpha_1} \rangle \langle (\Delta J_{13}(t))^{\alpha_2} \rangle$$

for every positive integer α_1 and α_2 , but we have already that

$$\langle \Delta J_{12}(t) \Delta J_{13}(t) \Delta J_{23}(t) \rangle \neq \langle \Delta J_{12}(t) \rangle \langle \Delta J_{13}(t) \rangle \langle \Delta J_{23}(t) \rangle$$

where $\langle \dots \rangle$ stands for an average over the values of the components S_i^z of the input pattern. From now on, we assume that the couplings can be considered as independent variables which seems to be reasonable for $p \rightarrow N/2$ in the large N limit, but we expect it to be valid also for small values of p .

3. The Energetic Approach

Supposing that the couplings are independent, the central limit theorem (CLT) can be applied to calculate the distribution of energy of any pattern viewed as the sum of C_N^p independent stochastic variables. As a result of the symmetry of the coupling distributions, the distribution is independent of the pattern. So we calculate this distribution, denoted by $\rho(E)$, for the pattern $(1, 1, \dots, 1)$. The CLT states that $\rho(E)$ is a Gaussian with mean value $\mu = 0$ and variance $\sigma^2 = 1/3 A(A+1)C_N^p$. But the knowledge of $\rho(E)$ is not sufficient to calculate the life time of a memorized pattern. We have to introduce a way to describe the convergence of $\rho(E)$ towards a Gaussian distribution. Note that we already know that the energy of the input pattern decreases on average from E' to $E' - (1-Q)C_N^p$, where $Q = 1/(2A+1)$ is the probability that a coupling keeps the same value.

If the energy is E' at time t , it will be E at time $t+1$ with probability $P(E, E')$. In a general way, the distribution evolves according to

$$\rho(E; t+1) = \sum_{E'} P(E, E') \rho(E'; t). \quad (8)$$

Obviously, $\sum_E P(E, E') = 1$. Now we make the assumption that C_N^p is much greater than the distribution width. Then the approximation of the sum by an integral between $-\infty$ and $+\infty$ yields the integral transform

$$\rho(E; t+1) = \int_{-\infty}^{+\infty} dE' P(E, E') \rho(E'; t). \quad (9)$$

At this stage, an ansatz is made for $P(E, E')$ (to be justified below) depending on the parameters $\gamma \in (0, +\infty)$ and $\lambda \in (0, 1)$:

$$P(E, E') = \sqrt{\frac{\gamma}{\pi}} e^{-\gamma(E-\lambda E')^2} \quad (10)$$

Repeated applications of (9) lead asymptotically to a Gaussian distribution of mean value $\mu = 0$ and variance $\sigma^2 = \frac{1}{2}[\gamma(1 - \lambda^2)]^{-1}$. Appendix B presents these results in greater detail. The values of the two parameters will be obtained naturally by considering the justification of the relation (10).

For this purpose, let us define n'_i (and n_i) as the number of couplings with value $J = i$ at time t (and $t + 1$). The n_i are constraint by the relations:

$$\sum_{i=-A}^A n_i = M \quad \text{and} \quad -\sum_{i=-A}^A i n_i = E \tag{11}$$

with M standing for C_N^p . Let us also define $n_{j,i}$ as the number of couplings changing from $J = i$ at time t to j at time $t + 1$. For instance if $i \neq \pm A$, we can only have $j = i + 1$ or $i - 1$. The $n_{j,i}$ verify

$$\begin{aligned} n_{A,A} + n_{A-1,A} &= n'_A \\ n_{i+1,i} + n_{i-1,i} &= n'_i \quad \text{if } i \neq \pm A \\ n_{-A+1,-A} + n_{-A,-A} &= n'_{-A} \end{aligned} \tag{12}$$

and

$$\begin{aligned} n_{A,A} + n_{A,A-1} &= n_A \\ n_{i,i+1} + n_{i,i-1} &= n_i \quad \text{if } i \neq \pm A \\ n_{-A,-A+1} + n_{-A,-A} &= n_{-A}. \end{aligned} \tag{13}$$

These equations are not independent. In fact, the sum of the left-hand side terms of the equations (12) equals the sum of the left-hand side terms of the equations (13), so there are only $2(2A + 1) - 1$ relations for $2(2A + 1)$ variables. The probability sought is the number of ways to make each coupling vary for fixed $\{n_i\}$, $\{n'_i\}$ and $n_{i,j}$, divided by the total number of possibilities (2^M) and summed over all the choices of $\{n_i\}$, $\{n'_i\}$ and $n_{i,j}$ compatible with (12), (13) and (11):

$$\begin{aligned} P(E, E') &= 2^{-M} \sum_{\{n_i\}, \{n'_i\}, \{n_{i,j}\}} \frac{n'_A!}{n_{A,A}! n_{A-1,A}!} \cdot \frac{n'_i!}{n_{i+1,i}! n_{i-1,i}!} \cdot \frac{n'_{-A}!}{n_{-A+1,-A}! n_{-A,-A}!} \\ &= 2^{-M} \sum_{\{n_i\}, \{n'_i\}, \{n_{i,j}\}} \Pi n'_i! / \Pi n_{i,j}!. \end{aligned} \tag{14}$$

Every possible changes have the same probability because all the couplings are supposed to evolve independently and each coupling always takes a new possible value with the same probability $1/2$ (see (6)).

The combinatorial factors in the first sum are easy to understand. For instance, the first one represents the number of possibilities for $n_{A,A}$ couplings among n'_A to remain the same, and $n_{A,A-1} = n'_A - n_{A,A}$ to diminish. At this stage, using (12, 13), the $n_{i,j}$ in (14) are easily expressed as functions of $n_{A,A}$.

The calculus of (14) is performed in the large N limit by taking the leading order term in N . This term is obtained considering only the greatest term of the sum, which is calculated with optimal values of $\{n_i\}$, $\{n'_i\}$ and $n_{A,A}$ verifying the constraints. Nevertheless this is a difficult task. So we prefer to calculate $\{n_i\}$ and $\{n'_i\}$ first, and then deduce $P(E, E')$ by supposing it is, in the large N limit, equivalent to the probability to go from $\{n'_i\}$ to $\{n_i\}$. The price to pay is a slight discrepancy, for A small only, between the result for $\alpha(A) = 1/(2\sigma^2(A))$ and the expected value, as we shall see in equation (31).

The first step is to find the relation between E and $\{n_i\}_{i \in \{-A, -A+1, \dots, A\}}$. For each energy E correspond different choices of $\{n_i\}$, each choice corresponding in addition to different arrangements of the couplings. The probability of having the reduced state $\{n_{-A}, \dots, n_A\}$ is simply the number of acceptable arrangements of the couplings, which are equiprobable since all couplings are uniformly distributed, divided by the total number of arrangements $(2A+1)^M$.

$$(2A + 1)^{-M} \frac{M!}{\prod_{i=-A}^A n_i!} . \tag{15}$$

The energy distribution is then:

$$\rho(E) = \sum_{\{n_i\}} (2A + 1)^{-M} \frac{M!}{\prod_{j=-A}^A n_j!} \tag{16}$$

where the sum is performed under the constraints (11). Because of the difficulty to cope with this formula, a well-known method in statistical physics is used. The goal is to identify E with only one choice of the $\{n_i\}$, optimal in the sense that $1/\prod n_i!$ is a maximum. By applying the Lagrange multiplier method, we find:

$$n_i = M \frac{(X(E))^i}{\sum_{j=-A}^A (X(E))^j} , \quad i \in \{-A, \dots, A\} . \tag{17}$$

The Lagrange multiplier $X(E)$, is calculated by introducing the expression (17) for n_i in the relation $-\sum_{i=-A}^A i n_i = E$. Assuming that $\epsilon = E/M$ is small enough, the development of $X(E)$ to order ϵ is

$$X(E) = 1 - \frac{3}{A(A+1)} \epsilon . \tag{18}$$

From this we deduce that

$$n_i = \frac{M}{2A+1} (1 + a_i \epsilon) \quad \text{where} \quad a_i = -\frac{3}{A(A+1)} i . \tag{19}$$

At this stage we can recover the CLT result, calculating $\prod n_i!$ with the approximation $n! \approx e^{n[\ln(n)-1]}$. If we suppose that the n_i are of the form $\frac{M}{2A+1} (1 + a_i \epsilon + b_i \epsilon^2)$, then to order ϵ^2

$$\sum_{i=-A}^A n_i (\ln n_i - 1) = M \left[\ln \left(\frac{M}{2A+1} \right) - 1 \right] + \frac{M}{2(2A+1)} \sum_{i=-A}^A a_i^2 \epsilon^2 \tag{20}$$

This result to order ϵ^2 only provides from products of terms to order ϵ . The substitutions for every a_i from (19) lead to

$$\prod_{i=-A}^A n_i! \approx M! (2A+1)^{-M} e^{\frac{3}{2A(A+1)} \frac{E^2}{M}} \tag{21}$$

The distribution is then simply

$$\rho(E) \approx e^{-\frac{3}{2A(A+1)} \frac{E^2}{M}}, \tag{22}$$

in agreement with the CLT result.

To continue the calculation of $P(E, E')$, we have to find the optimal value of n_{AA} to order ϵ by considering

$$\frac{d}{dn_{AA}} \left(\prod n_{ij}! \right) = 0. \tag{23}$$

This leads to a polynomial equation of degree $2A + 1$ in n_{AA} which admits only one real root. We use the value of this root and (19) to calculate the n_{ij} . This yields the next expression for the n_{ij} .

$$n_{\pm A \pm A} = \frac{M}{2(2A + 1)} \left(1 \pm \frac{3}{2A + 1} (\epsilon + \epsilon') \right) \tag{24}$$

and

$$n_{i-1, i} = \frac{M}{2(2A + 1)} (1 + \alpha_i \epsilon + \beta_i \epsilon') \tag{25}$$

where

$$\alpha_i = \begin{cases} -\frac{3}{2} \frac{(A + 1 - i)(A + i + 1)}{A(A + 1)} + \frac{3}{2A + 1}, & A - i \text{ odd} \\ -\frac{3}{2} \frac{(A - i)(A + i)}{A(A + 1)} - \frac{3}{2A + 1}, & A - i \text{ even} \end{cases} \tag{26}$$

$$\beta_i = \begin{cases} +\frac{3}{2} \frac{(A + 1 - i)(A + i - 1)}{A(A + 1)} + \frac{3}{2A + 1}, & A - i \text{ odd} \\ +\frac{3}{2} \frac{(A + 2 - i)(A + i)}{A(A + 1)} - \frac{3}{2A + 1}, & A - i \text{ even} \end{cases}$$

for $i \in \{-A + 1, \dots, A\}$. Switching α_i and β_i yields $n_{i, i-1}$. The calculation of $P(E, E')$ to order ϵ^2 gives an expression of the desired form (10) except for a factor $e^{-\kappa(E')}$ with $\kappa(E') \ll 1$. Fortunately the consequence is minor because the normalization is preserved with an appropriate rescaling. The parameters $\gamma(A)$ and $\lambda(A)$ are given by

$$\gamma(A) = \frac{1}{4(2A + 1)M} \left(2 \left(\frac{3}{2A + 1} \right)^2 + \sum_{i=-A+1}^A (\alpha_i^2 + \beta_i^2) \right) \tag{27}$$

and

$$\lambda(A) = -\frac{1}{\gamma(A)} \frac{1}{4(2A + 1)M} \left(2 \left(\frac{3}{2A + 1} \right)^2 + 2 \sum_{i=-A+1}^A \alpha_i \beta_i \right). \tag{28}$$

By substituting (26) into (27) and (28), we find

$$\gamma(A) = \frac{3}{10M} \frac{8A^4 + 16A^3 + 19A^2 + 11A + 6}{A(A + 1)(2A + 1)^2} \tag{29}$$

and

$$\lambda(A) = 1 - 10 \frac{A^2 + A + 1}{8A^4 + 16A^3 + 19A^2 + 11A + 6}. \tag{30}$$

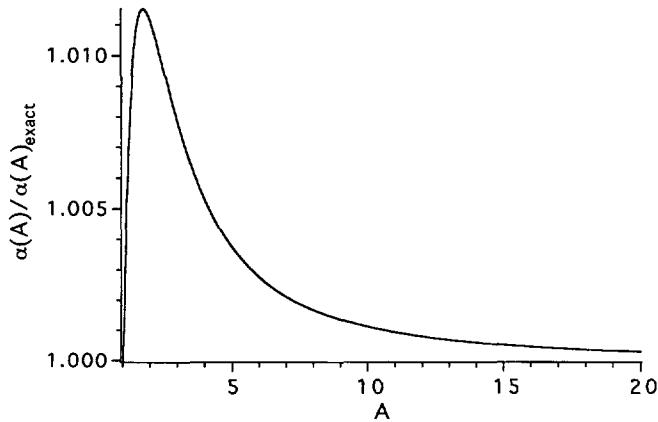


Fig. 2. — Approximate value of $\alpha(A)$ over exact $\alpha(A)$, versus A . The ratio does not depend on M .

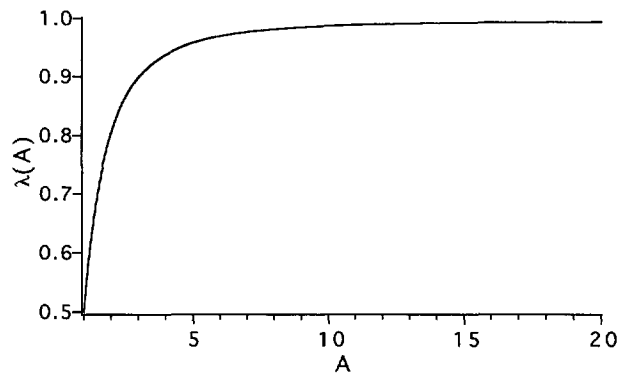


Fig. 3. — Plot of the function $\lambda(A)$.

With $\alpha = 1/(2\sigma^2)$ we get

$$\begin{aligned} \alpha(A) &= (1 - \lambda(A)^2) \gamma(A) \\ &= \frac{3}{2M} \frac{1}{A(A+1)} \left\{ 1 + \frac{1}{(2A+1)^2} \left(3 - 20 \frac{(A^2 + A + 1)^2}{8A^4 + 16A^3 + 19A^2 + 11A + 6} \right) \right\}. \end{aligned} \quad (31)$$

Note that λ does not depend on the size M of the system. λ can be seen as a quantitative measure of the robustness of the system, increasing if $\lambda \rightarrow 1$. The value of $\sigma^2(A)$ calculated from $\alpha(A)$ is compared with the previous result in Figure 2. These values do not differ more than about 1%. Furthermore they are identical for $A = 1$ and very close for $A \gg 1$. See also Figure 3 for $\lambda(A)$.

We are now able to estimate the storage capacity by calculating the life time of a recorded pattern. We assume that the energy distribution $\rho(E, t)$ is stationary (λ and α are fixed). The decrease of the energy of the input pattern, which equals $-M(1-Q) \approx -M$, and $-M(1-2p/N)$ for his neighbours, puts this state out of equilibrium. The effect of the noise of other records is to restore the equilibrium. More precisely, the Gaussian nature of $P(E, E')$ indicates that the mean energy \bar{E} and the variance Σ^2 obtained after one temporal step from energy E' are

respectively:

$$\bar{E} = \lambda E' \quad \text{and} \quad \Sigma^2 = \frac{1}{2\gamma}. \tag{32}$$

Each step corresponds on average to a simple multiplication of E' by λ . The fluctuations have the same amplitude for all E' .

In fact, we are interested in the difference of energy ΔE , between the input pattern and a neighbour pattern. The energies are the sum of M terms, $M(1 - p/N)$ of which are identical. Only Mp/N of them (always the same if the states are fixed as it is the case here) are opposite, so we drop the $M(1 - p/N)$ couplings which never play any role in the variations of ΔE . Therefore ΔE can be seen as a sum of only Mp/N terms. We conclude that $P(\Delta E, \Delta E')$ is given by

$$P(\Delta E, \Delta E') = \sqrt{\frac{\gamma_\Delta}{\pi}} e^{-\gamma_\Delta(\Delta E - \lambda_\Delta \Delta E')^2} \tag{33}$$

with the parameters γ_Δ , λ_Δ , Σ_Δ and σ_Δ calculated similarly as the parameters considered earlier, except for the substitution $M \rightarrow Mp/N$. Note that the correlations tends to soften the fluctuations of the difference of energies between the neighbours.

We are now in a position to calculate the mean life time T of a memory. Just after the memorization, the initial value of ΔE is $\Delta E_0 \approx 2Mp/N \pm \sigma_\Delta$. Then we deduce from (32) that ΔE is reduced by a factor λ_Δ^t after t new memorizations since the fluctuations, of typical value Σ_Δ , are negligible as long as $\Delta E \gg \Sigma_\Delta$. But when ΔE reaches the same order of magnitude as Σ_Δ , we expect that the hierarchy of patterns is rapidly modified. Thus we estimate naturally T by writing

$$\lambda_\Delta^T \frac{2p}{N} M \approx \Sigma_\Delta. \tag{34}$$

From now on, we assume that $1 \ll A$ (and $p \ll N$). Hence we obtain

$$\begin{aligned} T &\approx \frac{\ln\left(\sqrt{\frac{N}{2pM}}\right)}{\ln(\lambda_\Delta(A))} \\ &\approx (p-1) \ln(N) A^2 \end{aligned} \tag{35}$$

Figure 4 illustrates the evolution of the energies of an input pattern and one of his neighbours. This model also predicts how A_{opt} depends on N . If we want the memorization to be successful, ΔE_0 must be greater than the distribution width σ_Δ of ΔE , up to a numerical factor c , to insure a minimum of energy has been generated. That is, A verifies

$$c \sigma_\Delta(A) \leq \frac{2p}{N} M \quad \text{where} \quad \sigma_\Delta(A) \approx \sqrt{\frac{pM}{N}} A. \tag{36}$$

A_{opt} is thus the maximum admissible value of A :

$$A_{\text{opt}} \approx \sqrt{\frac{N^{p-1}}{(p-1)!}}. \tag{37}$$

4. Some Concluding Remarks

The main result is the strong dependence of T on A . The size of the system influences significantly the capacity, essentially because A_{opt} is increased. For $p = 2$, these results are in

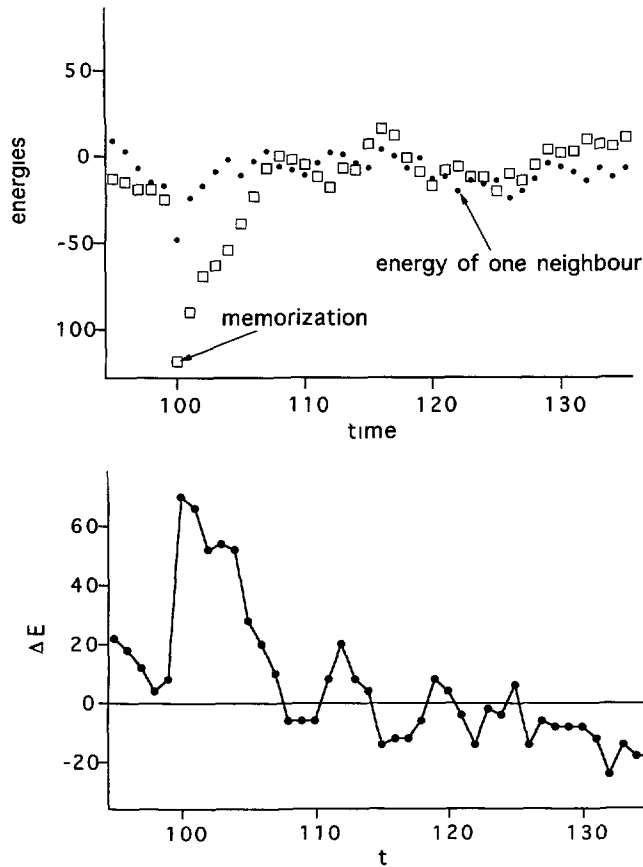


Fig. 4. — Time evolution of the energy of one pattern selected to be recorded at time $t = 100$. We have also represented the time evolution of the energy of one neighbour. The second diagram shows the energy difference between the two neighbours. With $N = 10$, $p = 3$, $A = 1.5$.

agreement with [6, 7, 9] where one can find $A_{\text{opt}} \propto \sqrt{N}$ and the capacity $\propto N$ (estimated here by T). They are also exact for the trivial case ($p = 1$). Furthermore, the capacity $\approx N^{p-1}$ is compatible with the total number of memory states given in [11, 12]. Of course, the capacity decreases for A smaller than the optimum and is exactly 1 when $A < 1$.

The typical number of bit stored per coupling is estimated by $NT/M \approx p(p-1) \ln(N)$. Thus it seems a good idea to consider values of $p > 2$, but there are also drawbacks, discussed for instance in [12].

In practice, the choice of p should be adapted to the purpose: it offers interesting possibilities to simulate systems with complex linkages where pair interaction is a poor, restrictive approximation. For instance, it allows to build a rugged landscape subject to changes in time, with significant number of extrema despite a small value of N .

Acknowledgments

I am indebted to A. Koch for discussions, and to O. Georg and M. Kunz for critical reading of the manuscript. I want also to thank the referees for their pertinent criticisms of the previous version and for the communication of some important references such as [9].

Appendix A

This appendix details the treatment of the $p = 1$ model. Since every couplings vary independently, it is only necessary to look at the time evolution of one of them (denoted by J). Assuming $J(t = 0) = A$, there are two cases: A is either an integer or half integer, or it is not. We begin with the easiest case: J can only take a value in the set $\{-A, -A + 1, \dots, A\}$. The evolution of J implies that the density ρ_J follows the equations (6), leading to the asymptotic solution

$$\rho_J(t \rightarrow \infty) = \frac{1}{2A + 1} \quad \forall J. \tag{38}$$

Now we calculate the probability P_A that $J (\neq 0)$ has the correct sign (+ for example), remembering that J makes a unit step in the desired direction. For integer values of A ,

$$P_A = \sum_{J=0}^A \rho_J = \frac{A + 1}{2A + 1} \tag{39}$$

while for half integer values of A

$$P_A = \sum_{J=-\frac{1}{2}}^A \rho_J = \frac{1}{2A + 1} + \frac{1}{2}. \tag{40}$$

In the other case J belongs to $\{-A, A - (L - 1), -A + 1, A - (L - 2), \dots, (L - 1) - A, A\}$, with $L = [2A] + 1$ (note that $-A + i = A - (L - 1) + i$ if and only if $2A$ is an integer). The equations are now:

$$\begin{aligned} \rho_A(t + 1) &= \frac{1}{2}\rho_A(t) + \frac{1}{2}\rho_{A-1}(t) + \frac{1}{2}\rho_{-A+(L-1)}(t) \\ \rho_J(t + 1) &= \frac{1}{2}\rho_{J+1}(t) + \frac{1}{2}\rho_{J-1}(t) \\ &\text{for } J \in \{A - (L - 1) + 1, \dots, A - 1\} \\ \rho_{A-(L-1)}(t + 1) &= \frac{1}{2}\rho_{A-(L-1)+1}(t). \end{aligned} \tag{41}$$

The values of ρ_J for the other J verify the same equations except for replacing every ρ_K by ρ_{-K} . The asymptotic solution is given by:

$$\rho_{A-i} = \rho_{i-A} = \frac{L - i}{L(L + 1)}, \quad i \in \{0, \dots, L - 1\} \tag{42}$$

and therefore

$$\begin{aligned} P_A &= \sum_{J>-1} \rho_J \\ &= \frac{1}{2} + \rho_{A-[A]-1} + \rho_{-A+[A]} \\ &= \frac{1}{2} + \frac{2L - 2[A] - 1}{L(L + 1)}. \end{aligned} \tag{43}$$

Substituting $L = [2A] + 1$ into the last equation gives (3).

Appendix B

We study here the evolution of $\rho(E)$ towards a Gaussian distribution. The basic equation is

$$\rho(E; t+1) = \sqrt{\frac{\gamma}{\pi}} \int dE' e^{-\gamma(E-\lambda E')^2} \rho(E'; t) \quad (44)$$

where $\gamma \in (0, +\infty)$ and $\lambda \in \mathbb{R}$. The normalization is easily seen to be preserved. Suppose now that $\rho(E)$ is a Gaussian distribution:

$$\rho(E; t) = \sqrt{\frac{\alpha(t)}{\pi}} e^{-\alpha(t)[E-\mu(t)]^2} \quad (45)$$

Applying (44) yields

$$\rho(E; t+1) = \sqrt{\frac{\alpha(t)\gamma}{\pi[\gamma\lambda^2 + \alpha(t)]}} e^{-\frac{\alpha(t)\gamma}{\gamma\lambda^2 + \alpha(t)} [E-\lambda\mu(t)]^2} \quad (46)$$

Therefore (44) maps a Gaussian distribution into another Gaussian distribution. The parameters change according to the relations

$$\alpha(t+1) = \frac{\alpha(t)\gamma}{\gamma\lambda^2 + \alpha(t)} \quad \text{and} \quad \mu(t+1) = \lambda\mu(t). \quad (47)$$

The study of the iterations of

$$f(\alpha) = \frac{\alpha\gamma}{\gamma\lambda^2 + \alpha} \quad \text{and} \quad g(\mu) = \lambda\mu \quad (48)$$

leads to the results:

$$\alpha = \gamma(1 - \lambda^2) \quad \text{is an attractive fixed point} \left(\left. \frac{df}{d\alpha} \right|_{\gamma(1-\lambda^2)} = \lambda^2 < 1 \right) \quad (49)$$

and

$$\alpha = 0 \quad \text{is a repelling fixed point} \left(\left. \frac{df}{d\alpha} \right|_0 = \frac{1}{\lambda^2} > 1 \right) \quad (50)$$

for $0 < |\lambda| < 1$. $\mu = 0$ is an attractive fixed point for the same condition on λ too.

Assume now that $\rho(E)$ can be written as the linear combination of Gaussian distributions:

$$\rho(E) = \sum_i a_i \rho_i(E) \quad (51)$$

where $\rho_i(E)$ are Gaussian distributions with arbitrary means and variances. The coefficients a_i verify $\sum_i a_i = 1$ (normalization condition). Then each Gaussian distribution converges independently to the same asymptotic Gaussian distribution (because (44) is linear) thus $\rho(E)$ converges to this Gaussian distribution.

References

- [1] Hopfield J.J., *Proc. Nat. Acad. Sci. USA* **79** (1982) 2554.
- [2] Mézard M., Nadal J.P. and Toulouse G., *J. Phys. France* **47** (1986) 1457.
- [3] Herz A.V., Li Z. and van Hemmen J.L., *Phys. Rev. Lett.* **66** (1991) 1370.
- [4] Hopfield J.J., in *Modelling in Analysis and Biomedicine*, C. Nicolini, Ed. (World Scientific, Singapore, 1984) esp. p. 381.
- [5] Alstrøm P. and Stassinopoulos D. *Phys. Rev. E* **51** (1995) 5027.
- [6] Parisi G., *J. Phys. A: Math. Gen.* **19** (1986) L617.
- [7] Nadal J.P., Toulouse G., Changeux J.P. and Dehaene S., *Europhys. Lett.* **1** (1986) 535.
- [8] van Hemmen J.L., Keller G. and Kühn R., *Europhys. Lett.* **5** (1988) 663.
- [9] Derrida B. and Nadal J.P., *J. Stat. Phys.* **49** (1987) 993.
- [10] Wong K.Y.M., Kahn P.E. and Sherrington D., *Physica A* **185** (1992) 461.
- [11] Abbott L.F. and Arian Y., *Phys. Rev. A* **36** (1987) 5091.
- [12] Gardner E., *J. Phys. A: Math. Gen.* **20** (1987) 3453.
- [13] Derrida B., *Phys. Rev. B* **24** (1981) 2613.