

Solvable models of working memories

M. Mézard, J.P. Nadal, G. Toulouse

▶ **To cite this version:**

M. Mézard, J.P. Nadal, G. Toulouse. Solvable models of working memories. Journal de Physique, 1986, 47 (9), pp.1457-1462. 10.1051/jphys:019860047090145700 . jpa-00210340

HAL Id: jpa-00210340
https://hal.science/jpa-00210340v1

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.



HAL Authorization

J. Physique **47** (1986) 1457-1462 SEPTEMBRE **1986**, PAGE 1457

Classification
Physics Abstracts
87.30G — 75.10H — 89.70 — 64.60C

Solvable models of working memories

M. Mézard (°), J. P. Nadal (*) and G. Toulouse (+)

Universita di Roma I, Dipartimento di Fisica, Piazzale Aldo Moro 2, 1-00185, Roma, Italy
(*) Groupe de Physique des Solides de l'Ecole Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France
(+) E.S.P.C.I., 10, rue Vauquelin, 75231 Paris Cedex 05, France

(*Reçu le 27 mars 1986, accepté le 7 mai 1986*)

Résumé. — Nous considérons une famille de modèles qui généralise le modèle de Hopfield, et qui peut s'étudier de façon analogue. Cette famille englobe des schémas de type palimpseste, dont les propriétés s'apparentent à celles d'une mémoire de travail (mémoire à court terme). En utilisant la méthode des répliques, nous obtenons un formalisme simple qui permet une comparaison détaillée de divers schémas d'apprentissage, et l'étude d'effets variés, tel l'apprentissage par répétition.

Abstract. — We consider a family of models, which generalizes the Hopfield model of neural networks, and can be solved likewise. This family contains palimpsestic schemes, which give memories that behave in a similar way as a working (short-term) memory. The replica method leads to a simple formalism that allows for a detailed comparison between various schemes, and the study of various effects, such as repetitive learning.

Introduction.

Networks of formal neurons provide models for associative memories [1-5] and much numerical and analytical progress has been made recently, especially on the Hopfield model [2, 3]. In particular Amit *et al.* [3] have solved for the thermodynamics of this model, using the replica method with the approximation of replica symmetry. Since then, alternative learning schemes have been proposed [4-6], which avoid the catastrophic deterioration of the memory when the number of stored patterns exceeds a critical value. In these schemes, new patterns may always be learned, at the expense of previously stored patterns which get progressively erased. For this reason, such memories have been called palimpsests, and may provide inspiration for the study of working (short term) memories.

In this paper, we define a family of models, which generalizes the Hopfield model and can be solved likewise. One of the palimpsestic schemes defined in [4] (the marginalist scheme) is a member of the family. Within the replica method, with the approximation of replica symmetry, we obtain analytical results that agree with previous numerical calculations, and allow for a detailed comparison between various schemes. The most interesting case is when one requests a very good retrieval quality of the learned patterns : remarkably, the formulation becomes very simple in this limit.

In section 1, we introduce the family of models. In section 2, we give the solution within the approximation of replica symmetry, and study the zero-temperature limit. In section 3, we consider various effects, such as repetitive learning of a pattern. The main results are summarized in the conclusion.

1. Palimpsestic schemes.

All the models we will consider have the same basic ingredients : the network is made of N interconnected formal neurons S_i. Each neuron can be either in the firing state ($S_i = +1$) or in the quiescent state ($S_i = -1$). The synaptic efficacies T_{ij} contain the information on a set of patterns $S^\mu = (S_i^\mu)_{i=1,N}$ which one wants to memorize. They can be either positive (excitatory) or negative (inhibitory). The network should work as an associative memory : setting the network in a pattern S^μ (or close to S^μ), it relaxes under a suitable dynamics towards a close stationary

(°) Permanent address : Laboratoire de Physique Théorique, Ecole Normale Supérieure, 24, rue Lhomond, 75231 Paris Cedex 05, France.

Article published online by EDP Sciences and available at http://dx.doi.org/10.1051/jphys:019860047090145700

state. Proximity is measured by the retrieval overlap

$$m = (1/N) \left\langle \sum_i S_i^\mu S_i \right\rangle, \qquad (1.1)$$

where the bracket is an average over stochastic noise (thermal averaging). When $m \sim 1$, retrieval is good. Assuming symmetric connections and relaxational dynamics, the (meta)stable states of the network are those of the Hamiltonian

$$H = -(1/2) \sum_{i \neq j} T_{ij} S_i S_j . \qquad (1.2)$$

As a first rough criterion of efficiency, a pattern S^μ is said to be memorized (or recognized) if its retrieval quality (1.1) is good.

The models we study differ in their learning schemes. These are the rules which fix the synaptic efficacies for the given set of patterns to be learned. The Hopfield scheme [1] is :

$$i \neq j \quad T_{ij} = (1/N) \sum_{\mu=1}^{p} S_i^\mu S_j^\mu . \qquad (1.3)$$

In this model, each pattern is learned with the same acquisition intensity $k = 1/N^2$. In the first palimpsestic scheme proposed in [4], named marginalist scheme, each pattern μ is learned with an intensity $k(\mu)$, which increases exponentially with μ. To avoid this undesirable exponential growth, we consider a slightly different version of this model : once t patterns have been stored, storing a new pattern S^{t+1} is done through the following modification of the synaptic efficacies :

$$T_{ij}(t+1) = \lambda[(\varepsilon/N) S_i^{t+1} S_j^{t+1} + T_{ij}(t)] \qquad (1.4)$$

where λ is such that the cumulated intensity (average squared synaptic efficacy) remains fixed :

$$\overline{T}_{ij}^2 - \overline{T}_{ij}^2 = 1/N . \qquad (1.5)$$

This normalization is convenient for the learning of a large number of patterns. The bar denotes the average over the quenched disorder $\{ S_i^\mu \}$. For independent random patterns, this implies

$$\lambda = (1 + \varepsilon^2/N)^{-1/2} . \qquad (1.6)$$

This scheme is stationary in the sense that the rule (1.4), (1.6) is time independent. This scheme will also be named marginalist since it is equivalent to the one of [4] through a rescaling of the temperature.

For simplicity of notation, we will use in the following a « time » such that the interval between two consecutive learning events is equal to one « time » unit. Emphasis is given first to the asymptotic regime, after an infinite number of patterns has been learned. Finite time effects will be considered later. Hence, if we denote $\xi^1 = (\xi_i^1)_{i=1,N}$ the last stored pattern, ξ^2 the previously stored one and so on, the model we consider is given by

$$T_{ij} = (\varepsilon/N) \sum_{\mu \geqslant 1} \lambda^\mu \xi_i^\mu \xi_j^\mu . \qquad (1.7)$$

In this scheme, the actual intensities are exponentially decreasing with storage ancestry. In the large N limit, $\lambda \sim \exp - \varepsilon^2/2N$. Thus, for the most recently learned patterns, that is for μ finite (relative to N), the intensity is constant. For $\mu \ll N$, the intensity is vanishingly small : this observation suggests that the network will act like a Hopfield system, with a capacity of order N, the memorized patterns being the most recently stored.

By a straightforward generalization of (1.7), we can actually introduce a whole family of models in which the strength of learning is time dependent :

$$T_{ij} = (1/N) \sum_\mu \Lambda(\mu/N) \xi_i^\mu \xi_j^\mu \qquad (1.8)$$

where $\Lambda(\mu)$ is any positive function such that

$$\int_0^\infty \mathrm{d}u \, \Lambda^2(u) = 1 . \qquad (1.9)$$

The marginalist scheme (1.7) is obtained for $\Lambda = \Lambda_m$:

$$\Lambda_m(\mu) = \varepsilon \exp - \mu\varepsilon^2/2 . \qquad (1.10)$$

It is particularly instructive to compare a given model with the Hopfield model. This model also belongs to the family (1.8)-(1.9) : it corresponds to learning with constant amplitude between some initial « time » $- N\tau$ up to the present moment. Hence, in our formulation it is given by the function $\Lambda = \Lambda_H$:

$$\begin{aligned} &0 \leqslant u \leqslant \tau \qquad \Lambda_H(u) = \varepsilon \\ &u > \tau \qquad\quad \Lambda_H(u) = 0 . \end{aligned} \qquad (1.11)$$

The normalization (1.9) imposes

$$\varepsilon^2 = 1/\tau . \qquad (1.12)$$

Within this formulation, the Hopfield scheme is clearly not stationary : as the number $N\tau$ of stored patterns increases, the parameter ε which measures the effective uniform acquisition amplitude decreases.

For the general problem (1.8)-(1.9), the replica method can be used in the very same way as for the Hopfield model, that is following exactly the calculations of reference [3]. In the following, we will discuss the properties of these models within the assumption of replica symmetry : for any function Λ, the entropy at zero temperature is negative, but very small as in the Hopfield case. Thus, replica symmetry breaking effects are expected to be small, at least for recognition properties. Furthermore, we will mainly discuss the zero temperature limit, since it is the low temperature behaviour which is the most interesting.

However, we emphasize that any calculations that have been, or can be, done for the Hopfield model can be simply generalized for a generic function Λ.

2. Replica symmetric solution.

We compute the average free energy per spin

$$f = \overline{(-\operatorname{Log}\operatorname{Tr}\exp -\beta H)}/N\beta \qquad (2.1)$$

with the replica method. As in [3], we are led to introduce three sets of order parameters : the states we consider are characterized by

i) the macroscopic overlaps m^{μ_i}, $i = 1, s$, s being finite as $N \to \infty$,

ii) the total mean square of the small random overlaps with the other patterns μ, weighted by $\Lambda(\mu/N)$:

$$r = \sum_{\mu \neq \mu_i} \Lambda^2(\mu/N)\, \overline{\langle (m^\mu)^2 \rangle} \qquad (2.2)$$

iii) the Edwards-Anderson order parameter,

$$q = \overline{\langle S_i \rangle^2}. \qquad (2.3)$$

One gets the following expression for the free energy

$$f = 1/2 \sum_{j=1}^{s} (m^{\mu_j})^2\, \Lambda(\mu_j/N) - (\beta/2)\, r(q-1) + 1/2 \int_0^\infty du\, \Lambda(u) +$$

$$+ (1/2\,\beta) \int_0^\infty du \left\{ \operatorname{Log}\left[1 - \beta\Lambda(u)\,(1-q)\right] - q\Lambda(u)/(1 - \beta(1-q)\,\Lambda(u)) \right\}$$

$$- (1/\beta) \left\langle\!\!\left\langle \operatorname{Log} 2\cosh \beta \left[z\sqrt{r} + \sum_j \xi^{\mu_j}\, \Lambda(\mu_j/N)\, m^{\mu_j} \right] \right\rangle\!\!\right\rangle \qquad (2.4)$$

where $\langle\!\langle . \rangle\!\rangle$ means averaging over the Gaussian variable z with zero mean and unit variance, and over the discrete distribution of $\{ \xi^{\mu_i} \}$. The values of the order-parameters are determined by the saddle-point equations :

$$m^{\mu_j} = \left\langle\!\!\left\langle \xi^{\mu_j} \tanh \beta \left[z\sqrt{r} + \sum_{i=1}^{s} \xi^{\mu_i}\, \Lambda(\mu_i/N)\, m^{\mu_i}) \right] \right\rangle\!\!\right\rangle \qquad (2.5)$$

$$q = \left\langle\!\!\left\langle \tanh^2 \beta \left[z\sqrt{r} + \sum_{i=1}^{s} \xi^{\mu_i}\, \Lambda(\mu_i/N)\, m^{\mu_i} \right] \right\rangle\!\!\right\rangle \qquad (2.6)$$

$$r = \int_0^\infty du\, q\Lambda^2(u)/[1 - \beta(1-q)\,\Lambda(u)]^2 . \qquad (2.7)$$

We shall now concentrate on the zero temperature limit.

An indication of the capability of the network to recognize a given pattern of ancestry $p = \alpha N$, is given by looking at the solutions with a macroscopic overlap with the single pattern p :

$$m^{\nu_i} = m\delta_{\nu_i, p}. \qquad (2.8)$$

As $\beta \to \infty$, equations (2.5)-(2.7) yield

$$m = \sqrt{2/\pi} \int_0^x dz\, e^{-z^2/2} \qquad (2.9)$$

with

$$x = m\Lambda(\alpha)/\sqrt{r}, \qquad (2.10)$$

$$q = 1 - C/\beta,$$

with

$$C = (2/\pi r)^{1/2} \exp - x^2/2 \qquad (2.11)$$

$$r = \int_0^\infty du\, \Lambda^2(u)/[1 - C\Lambda(u)]^2 . \qquad (2.12)$$

Finally one obtains two coupled equations for the reduced variables x and C :

$$x\, e^{-x^2/2} = \Lambda(\alpha)\, C \int_0^x dz\, e^{-z^2/2} \qquad (2.13)$$

$$e^{-x^2} = \pi/2 \int_0^\infty du\, \Lambda^2(u)/[1 - C\Lambda(u)]^2 \qquad (2.14)$$

which, for a given function Λ, can be solved numerically. In particular, for the Hopfield scheme, $\Lambda = \Lambda_H$ (see (1.11), (1.12)), one recovers the equations of reference [3]. Numerical solution of these give the critical values $\alpha_c = 1/\varepsilon_c^2 = 0.138...$, $m_c = 0.97...$

For the marginalist scheme $\Lambda = \Lambda_m$, we find that there is no solution with $m \neq 0$ for

$$\varepsilon < \varepsilon_c = 2.465. \qquad (2.15)$$

For $\varepsilon = \varepsilon_c$, there is one (stable) solution with $\alpha = 0$,

and

$$m = 0.933 . \tag{2.16}$$

For $\varepsilon > \varepsilon_c$, there are two solutions with m non zero, for :

$$0 \leqslant \alpha \leqslant \alpha(\varepsilon)$$

and one finds that the (meta)stable one corresponds to the highest value of m. This state has a retrieval quality $m(\alpha, \varepsilon) > m_c$. The functions $\alpha(\varepsilon)$, $m(\alpha(\varepsilon), \varepsilon)$ and $m(0, \varepsilon)$ are shown in figure 1. $\alpha(\varepsilon)$ is the (stationary) capacity of the memory. It increases with ε until a maximal value at $\varepsilon = \varepsilon_{opt}$

$$\varepsilon_{opt} = 4.108 \tag{2.17}$$

with a capacity

$$\alpha_{opt} = 0.04895 \tag{2.18}$$

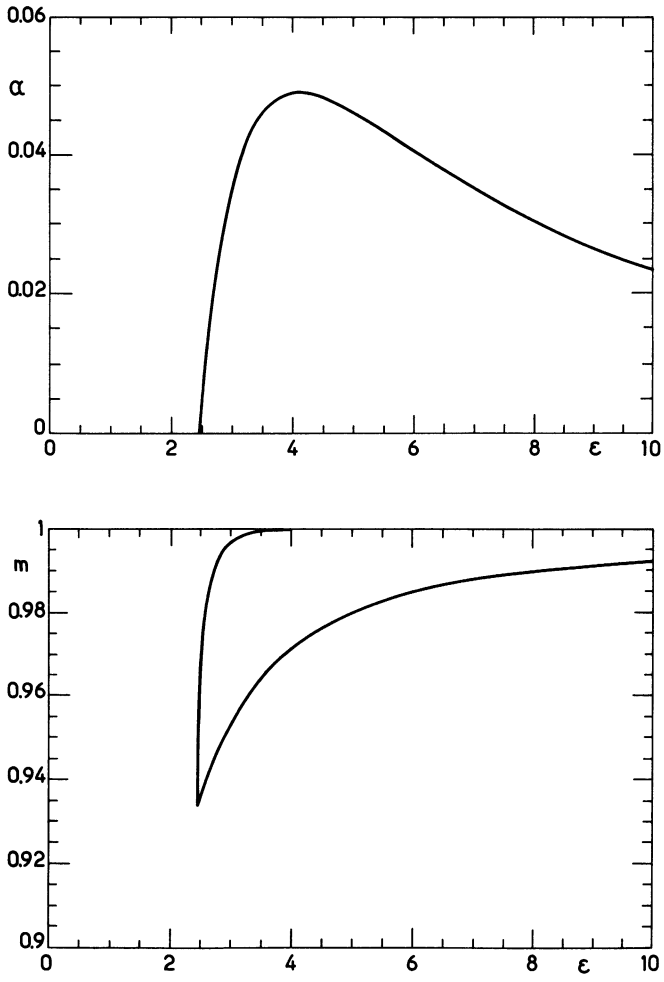


Fig. 1. — a) Capacity $\alpha = p/N$ of the marginalist scheme as a function of ε. This capacity is zero for $\varepsilon < \varepsilon_c = 2.4648...$, and goes through a maximum at $\varepsilon = \varepsilon_{opt} = 4.108$. b) Retrieval quality as a function of ε. The upper curve is the retrieval quality $m(0, \varepsilon)$ of the most recently stored pattern. The lower curve is the retrieval quality $m(\alpha(\varepsilon), \varepsilon)$ of the pattern $p = \alpha(\varepsilon) N$, that is of the most anciently stored, but still memorized, pattern.

and decreases for $\varepsilon > \varepsilon_{opt}$ — the number αN of memorized patterns being 1 in the large ε limit.

It is also of interest to request a retrieval quality at least equal to a given value of m — hence of x, see (2.9). The interesting limit consists in requiring a very good retrieval quality, that is m close to one. In this limit, it is easy to compute the capacity for any function Λ. In particular for models with a tunable parameter, such as the marginalist scheme, one finds explicitly the optimal value of the parameter and the corresponding optimal capacity.

When $m \to 1$, $x \to \infty$ and $C \to 0$. Then one has

$$m = 1 - \sqrt{2/\pi}(1/x)\, e^{-x^2/2} \tag{2.19}$$

$$\Lambda(\alpha) = x . \tag{2.20}$$

If Λ is a decreasing function, the capacity α_x is the largest value of α for which (2.20) holds. For the Hopfield scheme Λ_H one recovers

$$\alpha_x = 1/\varepsilon^2 \tag{2.21}$$

for

$$\varepsilon \geqslant x . \tag{2.22}$$

The maximal capacity is reached at $\varepsilon = \varepsilon_H$, with the capacity α_H :

$$\varepsilon_H = x, \quad \alpha_H = 1/x^2 . \tag{2.23}$$

In fact, this capacity α_H is the maximal possible capacity whatever Λ is. This is easily seen in this limit of good retrieval : due to the normalization condition (1.9) on Λ, if Λ is piece-wise continuous, α_x can be equal to $\alpha_H = 1/x^2$ if and only if Λ is constant, equal to x, on an interval of length $1/x^2$, and zero everywhere else. This result confirms the fact, observed in previous numerical simulations [4], that there is a price to pay, viz. a decrease of the capacity, to obtain a working memory, robust to new learning.

For the marginalist scheme, in the limit $m \to 1$, one finds

$$\begin{aligned} &\varepsilon \geqslant \varepsilon_c = x \quad \alpha_x(\varepsilon) = (2/\varepsilon^2)\,\mathrm{Log}(\varepsilon/x) \\ &\varepsilon < \varepsilon_c \qquad \alpha_x = 0 . \end{aligned} \tag{2.24}$$

This gives a maximal value of α_x at a value $\varepsilon_{opt}(x)$:

$$\varepsilon_{opt}(x) = x\sqrt{e}, \tag{2.25}$$

with a capacity

$$\alpha_{opt}(x) = 1/ex^2 . \tag{2.26}$$

Note that $\alpha_{opt} = 1/\varepsilon_{opt}^2$, $\varepsilon_{opt} = \sqrt{e}\varepsilon_c$, whatever x is.

In order to compare with the numerical estimations of reference [4], let us suppose that $m > 0.97$. Numerical solution of the equations (2.13), (2.14) gives the exact values

$$\begin{aligned} \varepsilon_c &= 2.529 \\ \varepsilon_{opt} &= 4.088 \\ \alpha_{opt} &= 0.0489 . \end{aligned} \tag{2.27}$$

Note that these values are close to the values (2.15), (2.17) and (2.18), obtained without restriction on m. The values reported for $\varepsilon_c \sim 2.2$, $\varepsilon_{opt} \sim 4$, were in reasonable agreement with (2.27), but at $N = 100$, a capacity of 0.065 was estimated. The corresponding discrepancy can be attributed mainly to finite size effects : further numerical simulations at $\varepsilon = 4$, for N up to 320 give an estimate 0.059 ± 0.005 for the capacity.

3. Learning by reinforcement, and finite time effects.

Other interesting features can be obtained within our general formulation :

i) the repeated learning of a given pattern with a period L, within a background of random patterns; ii) the learning of a sequence of L patterns, repeated *ad infinitum*; iii) the learning of a number L of patterns starting from tabula rasa at « time » $- L$.

The last problem (iii) will be treated in this section because it appears to be formally very similar to problem (ii). The two first problems will illustrate the well known efficiency of repetitive learning (« to teach is to repeat »). That a learning scheme leads to such effects is not surprising. What is interesting is the possibility, within our formalism, to quantify these effects.

i) We illustrate here the repetitive priming effect. We consider the learning of patterns independently chosen at random, except one given pattern which is learned periodically, with a period $L = gN$.

If $p = \alpha N$ is the last « time » when this pattern has been learned, its actual weight is

$$\Lambda_g(\alpha) = \sum_{l \geqslant 0} \Lambda(\alpha + lg) . \tag{3.1}$$

Thus instead of the function Λ, we have to consider the fonction Λ_g

$$\begin{aligned} \Lambda_g(\alpha) &= \Lambda(u) \quad u \neq \alpha \bmod g \\ \Lambda_g(\alpha) &= \sum_{l \geqslant 0} \Lambda(\alpha + lg) . \end{aligned} \tag{3.2}$$

Since Λ_g differs from Λ only at a denumerable set of points, we still have

$$\int_0^\infty \Lambda_g^2(u)\, du = 1 . \tag{3.3}$$

Hence we can apply the preceeding formalism for Λ_g. If we look at the states having a macroscopic overlap with the periodically learned pattern, the equations are the same as (2.9)-(2.12), with $\Lambda_g(\alpha)$ instead of $\Lambda(\alpha)$ in (2.10).

Since $\Lambda_g(\alpha) > \Lambda(\alpha)$, we see already that the retrieval quality will be enhanced by this periodic learning. Consider in particular the case $\alpha = g$, which means that we are looking at the retrieval quality just before re-learning. At zero temperature, the maximal value g^* of g, for which the retrieval quality is at least equal to a given value m, is given by

$$\Lambda_{g^*}(g^*) = \Lambda(\alpha_x) . \tag{3.4}$$

We recall that x is related to m *via* (2.19). For the marginalist scheme, this gives the relation

$$g^*(\varepsilon) = (2/\varepsilon^2) \operatorname{Log} [1 + \exp(\varepsilon^2\, \alpha_x(\varepsilon)/2)] . \tag{3.5}$$

This shows that for any ε

$$g^*(\varepsilon) > \alpha_x(\varepsilon) . \tag{3.6}$$

Thus, whereas a given pattern learned only once is retained during a time $N\alpha_x(\varepsilon)$, a pattern learned every N_g patterns is never forgotten provided $g \leqslant g^*$, where g^* is much greater than $\alpha_x(\varepsilon)$. Even for $\varepsilon < \varepsilon_c(x)$, where $\alpha_x(\varepsilon) = 0$, g^* is finite :

$$\varepsilon < \varepsilon_c(x) \quad g^* = (2/\varepsilon^2) \operatorname{Log} 2 . \tag{3.7}$$

In fact, through the diminution of the noise due to the other stored patterns, the smaller ε, the greater g^*.

ii) Consider now the repeated learning of the same sequence of $L = gN$ patterns. At « time » zero, the weight of the pth pattern is

$$1 \leqslant p \leqslant L \qquad \lambda_g(p/N) = \sum_{l \geqslant 0} \Lambda((p/N) + lg) .$$

To cast this problem into the general formulation, we have to consider the normalized weight Λ_g :

$$\begin{aligned} 0 \leqslant u \leqslant g \quad \Lambda_g(u) &= \lambda_g(u) \Big/ \left[\int du\, \lambda_g^2(u) \right]^{1/2} \\ u > g \quad \Lambda_g(u) &= 0 \end{aligned} \tag{3.8}$$

Now the model defined by Λ_g belongs to our general family : the new equations look the same as (2.9)-(2.14), with Λ replaced by Λ_g. For the marginalist scheme, one has

$$0 \leqslant u \leqslant g \quad \Lambda_g(u) = \varepsilon\, e^{-\varepsilon^2 u/2} / [1 - e^{-\varepsilon^2 g}]^{1/2} \tag{3.9}$$

In particular, for ε small, i.e.

$$\varepsilon^2\, g \ll 1$$

Λ_g is constant for $u \leqslant g$:

$$0 \leqslant u \leqslant g \quad \Lambda_g \simeq 1/\sqrt{g} . \tag{3.10}$$

This means that for ε small enough, the resulting memory is equivalent to the Hopfield limit. Within a palimpsestic scheme, one way to approach the optimal capacity, that is the Hopfield capacity, is to repeat again and again the sequence of patterns to be learned, but with a very low intensity.

iii) Finite « time » effects. The learning of $L = gN$ patterns (starting from tabula rasa) appears to be formally very similar to the learning of a periodic sequence. We want to recover the results of sections 1 and 2 in the infinite g limit : we consider the storing

of the pattern of ancestry $p = \alpha N$ with a weight $\Lambda(\alpha)$, for $\alpha \leqslant g$. But to cast this problem into our general formulation, we have to consider the normalized weight $\Lambda_g(\alpha)$, such that (1.9) remains true for any finite g :

$$1 \leqslant p \leqslant L \quad \Lambda_g(p/N) = \Lambda(p/N) \Big/ \left(\int_0^g \mathrm{d}u \, \Lambda^2(u) \right)^{1/2} \tag{3.11}$$

$$u > g \qquad \Lambda_g(u) = 0$$

In the limit $g \to \infty$, Λ_g becomes identical to Λ. For the marginalist scheme, Λ_g is again given by (3.9) : the two problems are thus equivalent.

If one works at a given value of ε, this gives immediately that, for short times, the complete set of patterns is memorized. This is the case until a critical value $g_c(\varepsilon)$. For $g > g_c$, only a fraction of the patterns are memorized, and the capacity decreases from g_c toward its asymptotic value $\alpha(\varepsilon)$. This is illustrated in figure 2 for $\varepsilon = 4$. Solving numerically the equations for Λ_g one finds g_c slightly greater than $1/\varepsilon^2$:

$$g_c(\varepsilon = 4.0) = 0.0672 \tag{3.12}$$

whereas

$$\alpha(\varepsilon = 4.0) = 0.0488 . \tag{3.13}$$

While the capacity decreases from $g_c(\varepsilon)$ to $\alpha(\varepsilon)$, the retrieval quality decreases slightly from a little better than 98 % to about 97 %.

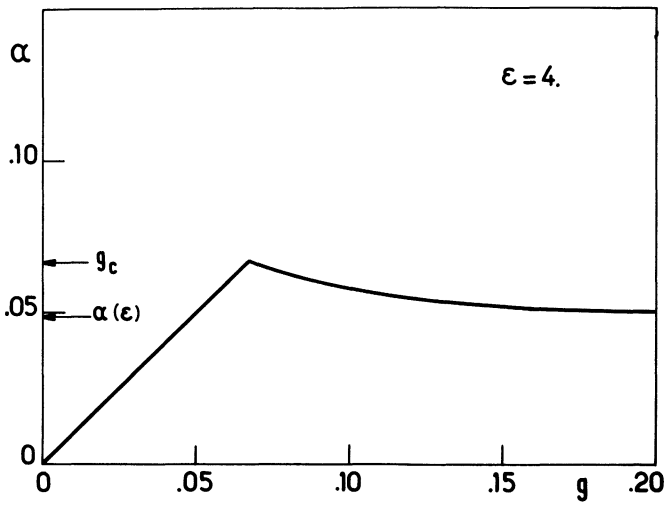


Fig. 2. — Capacity of the marginalist scheme for $\varepsilon = 4.0$, as a function of $g = L/N$, where L is the total number of stored patterns. All the patterns are memorized for g up to $g_c = 0.06719...$ For $g > g_c$, the capacity decreases toward its asymptotic value $\alpha(\varepsilon)$.

3. Conclusion.

In this paper we have shown how one can analyse very simply a whole family of models for working memory. Using the replica symmetry approximation, the resulting formalism is quite simple, and allows one to answer many questions about the behaviour of such memories. In particular we confirm that the capacity of these working memories are lower than the capacity of the Hopfield model. Typical phenomena of human short-term memory, such as the repetitive priming effect, are simply modelized and quantified within our formalism.

The incentive for considering non uniform acquisition intensities came from biology. From the view point of statistical physics, it is clear that everything that has been computed [2, 3] in the case of uniform intensities can be generalized. This paper has only made a first step in this promising direction.

Acknowledgments.

Fruitful discussions with J. P. Changeux and S. Dehaene are gratefully acknowledged.

References

[1] Hopfield, J. J., *Proc. Natl. Acad. Sci.* **79** (1982) 2554 ; Peretto, P., *Biol. Cybern.* **50** (1984) 51.

[2] Amit, D. J., Gutfreund, H., Sompolinsky, H., *Phys.* Toulouse, G., Dehaene, S., Changeux, J. P., *Proc. Natl. Acad. Sci.*, **83** (1986) 1695 ; Personnaz, L., Guyon, I., Dreyfus, G., Toulouse, G., *J. Stat. Phys.* **43** (1986) 411 ; Feigelman, M. V., Ioffe, L. B., *Europhys. Lett.* **1** (1986) 197 ; Sompolinsky, H., preprint.

[3] Amit, D. J., Gutfreund, H., Sompolinsky, H., *Phys. Rev. Lett.* **55** (1985) 1530 ; Amit, D. J., Gutfreund, H., Sompolinsky, H., preprint (1986).

[4] Nadal, J. P., Toulouse, G., Changeux, J. P., Dehaene, S., *Europhys. Lett.*, **1** (1986) 535.

[5] Parisi, G., preprint.

[6] A suggestion for such learning schemes was already present in a paragraph of the original paper of Hopfield [1] and should have been underlined in [4].