Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The storage capacity of Potts models for semantic memory retrieval

Emilio Kropff[1] **and Alessandro Treves**

SISSA, Cognitive Neuroscience, via Beirut 4, 34014 Trieste, Italy
E-mail: kropff@sissa.it and ale@sissa.it

Received 15 June 2005
Accepted 12 July 2005
Published 23 August 2005

Online at stacks.iop.org/JSTAT/2005/P08010
doi:10.1088/1742-5468/2005/08/P08010

Abstract. We introduce and analyse a minimal network model of semantic memory in the human brain. The model is a global associative memory structured as a collection of N local modules, each coding a feature, which can take S possible values, with a global sparseness a (the average fraction of features describing a concept). We show that, under optimal conditions, the number c_M of modules connected on average to a module can range widely between very sparse connectivity (*high dilution*, $c_M/N \to 0$) and full connectivity ($c_M \to N$), maintaining a global network storage capacity (the maximum number p_c of stored and retrievable concepts) that scales like $p_c \sim c_M S^2/a$, with logarithmic corrections consistent with the constraint that each synapse may store up to a fraction of a bit.

Keywords: spin glasses (theory), neuronal networks (theory)

[1] http://www.sissa.it/∼ale/limbo.html

1742-5468/05/P08010+19$30.00

Contents

1. Introduction

Hebbian associative plasticity appears to be the major mechanism responsible for sculpting connections between pyramidal neurons in the cortex, for both short- and long-range systems of synapses. This and other lines of evidence [1] suggest that autoassociative memory retrieval is a general mechanism in the cortex, occurring not only at the level of local networks, but also in higher-order processes involving many cortical areas. These areas are often regarded both from the anatomical and from the functional point of view as distinct but interacting modules, indicating that in order to model higher-order processes we must first understand better how multi-modular autoassociative memories may operate. In a class of models conceived along these lines, neurons in local modules, interconnected through short-range synapses, are capable of retrieving local activity patterns which, combining across the cortex through long-range synapses, compose global states of activity [2]. Since long-range synapses are also modified by associative plasticity, these states can be driven by attractor dynamics, and such networks are capable of retrieving previously learned global patterns.

This could serve as a simple model of semantic memory retrieval. The semantic memory system, as opposed to episodic memory, stores composite concepts, e.g. objects, and their relationships. Although information about distinct features pertaining to a given object (e.g. its shape, smell, texture, function) may be processed in different areas of the cortex, a cue including only some of the features, e.g. the shape and colour, may suffice to elicit retrieval of the entire memory representation of the object. Imaging studies show that, though distributed across the cortex, this activity is sparse and selective, and might involve regions associated to the concept being retrieved, even if not directly activated by the cue [3]. This process could well fit a description in terms of autoassociative multi-modular memory retrieval. In this perspective, while a local module codes for diverse values of a given feature, a combination of features gives rise to a concept, which behaves as an attractor of the global network and is thus susceptible of retrieval. The two-level description that characterizes this view is the principal difference with other attempts to describe semantic memory in terms of featural representations [4].

In order to reduce the complexity of a full multi-modular model [5, 6] one can consider a minimal model of semantic memory, which can be thought of as a global autoassociative memory in which the units, instead of representing, as usual, individual neurons, represent local cortical networks retrieving one of various (S) possible states of activity. The combined activity of these units generates a global state, which follows a retrieval dynamics. The first question arising from this proposal is how the global storage capacity of such a network is related to the different local and global parameters.

In the following section of this paper we present the model in mathematical terms. In the third section we compare, through a simple signal-to-noise analysis, different model variants proposed in the literature and extract the minimum requirements for a network of this kind to perform efficiently in terms of storage capacity. In the fourth section we analyse with more sophisticated techniques the simplest model endowed with a large capacity (the sparse Potts model) and, in particular, interesting cases such as the very sparse and the high-S limits. Following this we study modifications to the model that make it more realistic in terms of connectivity. Finally, we relate the results from the previous sections to a simple information capacity analysis.

2. S-state fully connected networks

Autoassociative memories are networks of N units connected to one another by weighted synapses. These synapses are trained in such a way that the network presents, in the ideal case, a number p of preassigned attractor states, also called stored patterns, or memories, represented by the vectors $\vec{\xi}^{\mu}$, with $\mu = 1 \ldots p$. If the state of the network is forced into the vicinity of an attractor (e.g., by presenting a cue correlated with one of the stored patterns) the natural dynamics of the network converges toward the attractor, in state space, and the memory item is said to be retrieved. A substantial amount of the literature on attractor networks is devoted to study the relationship between the number and type of stored patterns and the quality of retrieval.

The state of a network at a given moment is given by the state of each of its units, σ_i for $i = 1 \ldots N$. The first quantitative analyses of autoassociative memories were of binary models [7], in which units could reach two possible states, $+1$ (active unit) and -1 (inactive unit), resembling Ising $\frac{1}{2}$ spins. In our case, in which units do not represent single neurons but rather local networks, we want active units to be able to reach one of S possible states, while inactive units remain in a 'zero' state. We thus choose the notation $\sigma_i = k$ for an active unit in state k and $\sigma_i = 0$ for an inactive unit. This particular choice has no effect on the results, since all quantities can be transformed to some other notation. On the other hand, the stored patterns $\vec{\xi}^{\mu}$ can be simply thought of as special states of the network. For this reason, it is natural to choose the same kind of representation for the activity of a unit i in pattern μ, ξ_i^{μ}.

Although in the first binary models of autoassociative memories patterns were constructed with a distribution of equally probable active and inactive units, the search for an accurate description of activity in the brain made it necessary to introduce sparse representations. This property of autoassociative memories is described by the sparseness a, defined as the average activity (the average fraction of active units) in the stored patterns. In our case, because we are assuming all S different activity states to be equally probable, we consider patterns defined by the following probability distribution

$$\begin{aligned} P(\xi_i^\mu = 0) &= 1 - a \\ P(\xi_i^\mu = k) &= \tilde{a} \equiv \frac{a}{S} \end{aligned} \tag{1}$$

for any active state k. In this way the probability to find a unit active in a pattern is the sparseness a. For *sparse* codes, this quantity is closer to 0 than to 1.

Following the assumption of Hebbian learning and, as is usual for a simplified analysis, symmetry in the weights ($J_{ij} = J_{ji}$), a general form for the weights is

$$J_{ij}^{kl} = \frac{1}{E} \sum_{\mu=1}^{p} v_{\xi_i^\mu k} v_{\xi_j^\mu l} \tag{2}$$

where E is some normalization constant and v_{mn} is an operator computing interactions between two states.

As one can notice, the long-range synaptic weights in equation (2) have different values for different pre- and post-synaptic states k and l. In this way we do not intend to model the actual distribution of synapses going from one cortical area to another (since they connect neurons and not abstract states), but rather the general mechanism of communication between these areas. In a recent study [8], the authors have raised the issue of finding the most suitable description of global cortical networks in terms of single long-range synapses connecting distant local areas. Applying statistical tools (dynamic causal modelling), they propose that MRI data can be described as being produced by networks with category-specific forward connections, roughly the kind of connections modelled by equation (2).

The state of generic unit i is determined by its local fields h_i^k, which sum the influences by other units in the network and are defined as

$$h_i^k = \sum_{j \neq i} \sum_{l} J_{ij}^{kl} u_{\sigma_j l} - U(1 - \delta_{k0}) \tag{3}$$

where we introduce the operators u_{mn}, analogous to v_{mn}, and a second (threshold) term, which has the function of regulating the activity level across the network [9, 10]. The unit i updates its state σ_i, with an asynchronous dynamics, in order to maximize the local field $h_i^{\sigma_i}$. In the general case, the probability to choose the state k is defined as

$$P(\sigma_i = k) = \frac{\exp(\beta h_i^k)}{\sum_{l=0}^{S} \exp(\beta h_i^l)}$$

where β is a parameter analogous to an inverse temperature.

Finally, we can include all of these elements, as is usual for the study of attractor networks, into a Hamiltonian framework. The Hamiltonian representation of binary networks can be extended to S-state models as

$$H = -\tfrac{1}{2}\sum_{i,j\neq i}^{N}\sum_{k,l}^{S} J_{ij}^{kl} u_{\sigma_i k} u_{\sigma_j l} + U\sum_{i}^{N}\sum_{k\neq 0}^{S} u_{\sigma_i k}. \tag{4}$$

Note that, for the case $S = 1$, equation (4) generalizes the Hamiltonians used in binary networks, given appropriate definitions of the weights J_{ij}^{kl} and of the operators u_{mn}.

We now specify a form for the u_{mn} and v_{mn} operators. In the simplest and most symmetric case these operators have two alternative values, depending on whether m and n are equal or different states

$$\begin{aligned} u_{mn} &= (\kappa_u \delta_{mn} + \lambda_u) \\ v_{mn} &= (\kappa_v \delta_{mn} + \lambda_v)(1 - \delta_{n0}) \end{aligned} \tag{5}$$

where we have introduced four parameters. Particular choices for these parameters define the different models in which we are interested, including some proposed in the literature. In the v operators, which define the value of the weights, we have included a factor which ensures $J_{ij}^{kl} = 0$ if either k or l are the zero state, to implement the idea that Hebbian learning occurs only with active states. As we will see below, this appears to be a crucial element in the model.

3. Signal-to-noise analyses

We now show that, within the group of models defined in the previous section, there is a family (which we call 'well behaved') that exploits multiple states and sparseness in an optimal way in terms of storage capacity or, as usual, of $\alpha \equiv p/N$. We begin by applying an adjusted version of the arguments developed in [9].

A signal-to-noise analysis is a simplified way to estimate the stability of stored patterns by studying the field acting on a generic unit i during the perfect retrieval of a given pattern, assessing whether the state of this unit is likely to be stable or not. We can choose this retrieved pattern to be $\vec{\xi}^1$ without loss of generality. Equation (3) can then be rewritten as

$$h_i^k = \frac{1}{E} v_{\xi_i^1 k} \sum_{j\neq i}\sum_{l} u_{\sigma_j l} v_{\xi_j^1 l} + \frac{1}{E}\sum_{\mu>1} v_{\xi_i^\mu k}\sum_{j\neq i}\sum_{l} u_{\sigma_j l} v_{\xi_j^\mu l} - U(1-\delta_{k0}) \tag{6}$$

where the terms in the RHS stand for signal (ς), noise (ρ) and threshold respectively. Generally speaking, if the field had only the signal part then the state would be stable, but the noise can destabilize it.

As usual in this kind of analysis, we consider the contribution of the noise term in equation (6) as if it were a normally distributed random variable, i.e. through its average and its standard deviation. In general both quantities scale like p, but in some special cases the average noise is zero and the standard deviation scales only like $\sqrt{p}$, which means that one can store more patterns, as the noise level is reduced. It is clear that the well behaved family of models which we are looking for must fit into this favourable situation. As we said, a necessary but not sufficient condition is the average of the noise to be zero. There are two ways of imposing this into the model. The first way is to make $\lambda_u = -\tilde{a}\kappa_u$, but in this case the standard deviation still scales like p. The second way is to use

$$\lambda_v = -\tilde{a}\kappa_v \tag{7}$$

J. Stat. Mech. (2005) P08010

which makes the standard deviation scale like $\sqrt{p}$. Including this condition, the average signal and the standard deviation of the noise are

$$\varsigma = \frac{N\kappa_v^2}{E}\kappa_u\tilde{a}(1-\tilde{a})S(\delta_{\xi_i^1 k} - \tilde{a})(1-\delta_{k0})$$
$$\rho = \frac{N\kappa_v^2}{E}\kappa_u\tilde{a}(1-\tilde{a})\sqrt{\alpha a\left\{1-\tilde{a}\left[1-\left(1-\frac{\lambda_u}{\tilde{a}\kappa_u}\right)^2\right]\left[\frac{1-a}{1-\tilde{a}}\right]\right\}}(1-\delta_{k0})$$

where terms of low order N have been discarded.

The storage capacity α_c can be estimated as the largest value of α for which $h_i^{\xi_i^1}$ is still likely to be the largest among all $S+1$ local fields. The situation is quite different depending on whether ξ_i^1 is in an active state or not, so one needs to analyse both cases. Note first that $h_i^0 = 0$, so if $\xi_i^1 = 0$ the rest of the local fields must be negative. For this to hold true at least within one standard deviation of the noise distribution we require $\varsigma - U \pm \rho < 0$, or in other words

$$a + \frac{UE}{N\kappa_v^2\kappa_u\tilde{a}(1-\tilde{a})} > \sqrt{\alpha a\left\{1-\tilde{a}\left[1-\left(1-\frac{\lambda_u}{\tilde{a}\kappa_u}\right)^2\right]\left[\frac{1-a}{1-\tilde{a}}\right]\right\}}$$

where we have adopted a positive κ_u.

In the case in which ξ_i^1 is not the zero state two conditions must be fulfilled, namely $h_i^{\xi_i^1} > h_i^0$ and $h_i^{\xi_i^1} > h_i^{k\neq\xi_i^1}$. The first of them can be written as

$$S(1-\tilde{a}) - \frac{UE}{N\kappa_v^2\kappa_u\tilde{a}(1-\tilde{a})} > \sqrt{\alpha a\left\{1-\tilde{a}\left[1-\left(1-\frac{\lambda_u}{\tilde{a}\kappa_u}\right)^2\right]\left[\frac{1-a}{1-\tilde{a}}\right]\right\}}.$$

The most stringent of these three conditions determines α_c. By choosing a suitable threshold $U = (N/E)\kappa_v^2\kappa_u\tilde{a}(1-\tilde{a})[(S/2)-a]$ all conditions are made equivalent, thus optimizing the storage capacity. This choice determines a storage capacity of

$$\alpha_c \simeq \frac{S^2}{4a}\left\{1-\tilde{a}\left[1-\left(1-\frac{\lambda_u}{\tilde{a}\kappa_u}\right)^2\right]\left[\frac{1-a}{1-\tilde{a}}\right]\right\}^{-1}. \tag{8}$$

Note that the expression between curly brackets is equal to or greater than $1-\tilde{a}$. As a consequence, the system remains optimal as long as this expression remains of order 1, which, considering always a to be closer to 0 than to 1, occurs when the expression $(1-(\lambda_u/\tilde{a}\kappa_u))^2$ remains of order 1. For this to be true we must impose

$$|\lambda_u| \lesssim \tilde{a}\kappa_u. \tag{9}$$

We thus define the well behaved models as those which fulfil the conditions given by equations (7) and (9). This simple analysis indicates that the storage capacity of models in the well behaved family scales like S^2/a.

In the following subsections we examine different models proposed in literature, both within and outside the well behaved family.

J. Stat. Mech. (2005) P08010

3.1. Symmetric Potts model

The symmetric Potts model was the first S-state neural network to be proposed [11]. Its units can reach S equivalent states but no zero state. Though simple, a model constructed with these elements is enough to show the S^2 behaviour of the storage capacity, as we will see. It is defined by setting

$$a = 1$$
$$U = 0,$$

two conditions related to each other (if there is no zero state, the selectivity mechanism provided by the threshold is not necessary). Moreover $E = S^2N$, which is just a normalization, and

$$\kappa_u = \kappa_v = S$$
$$\lambda_u = \lambda_v = -1.$$

The conditions given by equations (7) and (9) are fulfilled, and the storage capacity in equation (8) is approximately

$$\alpha_c \approx \frac{S^2}{4}$$

provided S is large enough. The symmetric Potts model is then a well behaved model of sparseness $a = 1$.

This model is studied analytically with replica tools in [11], where the author finds an $S(S-1)$ behaviour of the storage capacity for low values of S. Unfortunately, the cited work lacks an analysis for high values of S, which is the interesting limit for modelling multi-modular networks. It is not too difficult, however, to clarify the behaviour in this limit.

The replica storage capacity is defined as the highest value of α for which there is a solution to the equation

$$y = \frac{-1 + S\int \mathrm{D}z\,[\phi(z+y)]^{S-1}}{\sqrt{(\alpha(S-1)/S)} + \int z\,\mathrm{D}z\,\{[\phi(z+y)]^{S-1} + (S-1)\phi(z-y)[\phi(z)]^{S-2}\}} \tag{10}$$

where

$$\phi(z) \equiv \frac{1 + \mathrm{erf}(z/\sqrt{2})}{2}. \tag{11}$$

Throughout this paper we use the Gaussian differential $\mathrm{D}z \equiv (\mathrm{e}^{-z^2/2}/\sqrt{2\pi})\,\mathrm{d}z$, and the integration limits, if not specified, are $-\infty$ and ∞.

We note that in equation (10) expressions of the form $[\phi(z)]^S$ can be approximated by displaced Heaviside functions for high values of S. Using this approximation we obtain an expression for the storage capacity in the high S limit

$$\alpha_c = \left[\frac{\phi(\sqrt{\pi/2})}{\sqrt{\pi/2} + \sqrt{2}\,\mathrm{erf}^{-1}(1 - (\ln(2)/S))}\right]^2 S^2. \tag{12}$$

The factor between brackets in this equation behaves like $\ln(S)^{-1/2}$ for high values of S, which means that the correction for high S to Kanter's low-S approximation is a factor of order $\ln(S)^{-1}$.

J. Stat. Mech. (2005) P08010

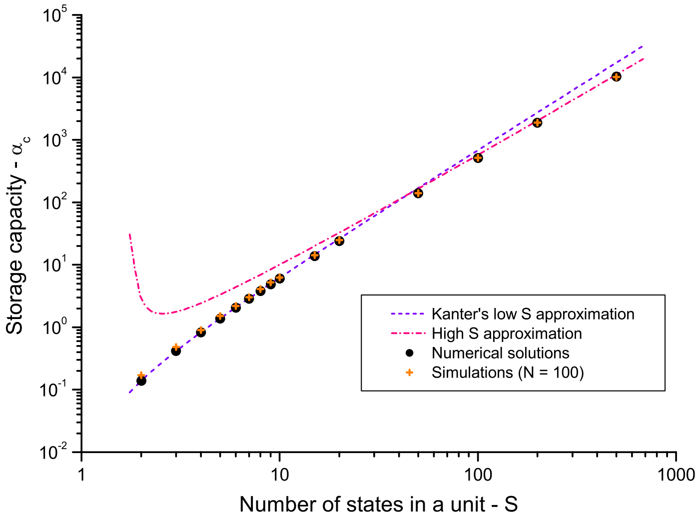

Figure 1. Storage capacity of a symmetric Potts network of $N = 100$ units for increasing S. Both axes are logarithmic. Black dots show numerical solutions for equation (10), which overlap almost perfectly with the simulations (plus signs). For low values of S ($S \lesssim 50$) Kanter's low-S approximation fits well, while the high values of S are well fitted by equation (12).

We show in figure 1 the results of simulations of a symmetric Potts network ($N = 100$) contrasted with Kanter's low-S approximation and our own high-S approximation of equation (12). The analytical predictions fit tightly the results of the simulations, both for low and high S.

3.2. Biased Potts model

This model is proposed and studied in [12]. The authors extend the symmetric Potts model to an S-state network with arbitrary probability distribution for the states of the units in stored patterns. We adapt their formalism to the case of S equivalent states, a zero state and sparseness a. The parameters are then

$$\begin{aligned} &U = 0 \\ &E = N \\ &u_{mn} = ((S+1)\delta_{mn} - 1) \\ &v_{mn} = (\delta_{mn} - P_n) \end{aligned} \tag{13}$$

where P_k is the probability of a unit in the stored patterns to be in state k. This model does not exactly fit our description because the v operators generate weights J_{ij}^{kl} that are not necessarily zero when k or l are zero. The signal-to-noise analysis for this situation shows a very poor storage capacity, scaling like a^2. If one adds a non-zero threshold ($U \sim a\,S$ in the optimal case) the storage capacity grows but remains of order 1. These two results show that allowing for non-zero weights to connect zero states is a drawback for the system. The poor performance can, however, be improved by multiplying the v operators by the corresponding $(1 - \delta_{n0})$ factors, and by adding the threshold. In this

J. Stat. Mech. (2005) P08010

way, instead of equation (13) we introduce our definition, equation (5), for the v operators, with the values for κs and λs arising naturally from the model as

$$\begin{aligned}
&\kappa_u = S+1\\
&\lambda_u = -1\\
&\kappa_v = 1\\
&\lambda_v = -\tilde{a}\\
&U \sim aS.
\end{aligned}$$

As in the symmetric Potts model, the condition given by equation (7) is fulfilled. However, the second condition (equation (9)) can be approximated for high S by

$$a \gtrsim 1/(1+1/S) \sim 1$$

which does not stand true for sparse coding. If, instead, $a \ll 1$, the critical value of α in equation (8) can be approximated as

$$\alpha_c \approx \frac{S^2}{4a}\left\{1+\frac{1}{a\,S}\right\}^{-1}.$$

Hence the storage capacity of the biased Potts model can be preserved close to optimal by imposing an ad hoc relation between two parameters that are *a priori* independent, to assure $1 \ll a\,S$. In this particular situation the model is well behaved. In the opposite limit, when $a\,S \ll 1$, the storage capacity scales like S^3, which is inferior to the S^2/a behaviour of the well behaved family.

3.3. Sparse Potts model

The simplest version of a well behaved model is perhaps the one introduced as a model for semantic memory [13], with the parameter values

$$\begin{aligned}
&E = Na(1-\tilde{a})\\
&\kappa_u = \kappa_v = 1\\
&\lambda_u = 0\\
&\lambda_v = -\tilde{a}\\
&U \sim 1/2.
\end{aligned}$$

With these parameters, the sparse Potts model is clearly well behaved, and the storage capacity in equation (8) becomes

$$\alpha_c \simeq \frac{S^2}{4a}.$$

4. Replica analysis

Having introduced a simple model with optimal storage capacity, we can proceed to analyse the corrections to the signal-to-noise estimation by treating the problem in a

J. Stat. Mech. (2005) P08010

more refined way with the classical replica method. The Hamiltonian in equation (4) can be rewritten for the sparse Potts model as

$$H = -\tfrac{1}{2} \sum_{i,j\neq i}^{N} \sum_{k,l}^{S} J_{ij}^{kl} \delta_{\sigma_i k} \delta_{\sigma_j l} + U \sum_{i}^{N} (1 - \delta_{\sigma_i 0})$$

with

$$J_{ij}^{kl} = \frac{1}{Na(1-\tilde{a})} \sum_{\mu=1}^{p} (\delta_{\xi_i^\mu k} - \tilde{a})(\delta_{\xi_j^\mu l} - \tilde{a})(1 - \delta_{k0})(1 - \delta_{l0})$$

constructed using

$$v_{mn} = (\delta_{mn} - \tilde{a})(1 - \delta_{n0}).$$

We consider the limit $p \to \infty$ and $N \to \infty$ with the ratio $\alpha \equiv p/N$ fixed. Patterns with index ν (μ) are condensed (not condensed). Following the replica analysis [7] the free energy can be calculated as

$$\begin{aligned} f = \lim_{n\to 0} \frac{a(1-\tilde{a})}{2n} \sum_{\rho=1}^{n} \sum_{\nu} (m_\rho^\nu)^2 \\ &+ \frac{\alpha}{2n\beta} \mathrm{Tr}\left(\ln[a(1-\tilde{a})(\mathbb{I} - \beta\tilde{a}\mathbf{q})]\right) + \frac{\alpha\beta\tilde{a}^2}{2n} \sum_{\rho,\lambda=1}^{n} q_{\rho\lambda} r_{\rho\lambda} \\ &+ \frac{\tilde{a}}{n}\left(\frac{\alpha}{2} + U\,S\right) \sum_{\rho=1}^{n} q_{\rho\rho} \\ &- \frac{1}{n\beta} \left\langle\left\langle \ln \mathrm{Tr}_{\sigma_\rho} \exp\left\{ \beta \sum_{\rho=1}^{n} \sum_{\nu} m_\rho^\nu v_{\xi^\nu \sigma_\rho} \right.\right.\right. \\ &\left.\left.\left. + \frac{\alpha\beta^2}{2S(1-\tilde{a})} \sum_{\rho,\lambda=1}^{n} r_{\rho\lambda} \sum_{k} P_k v_{k\sigma_\rho} v_{k\sigma_\lambda} \right\} \right\rangle\right\rangle \end{aligned}$$

where P_k is the probability of a neuron to be in state k in a stored pattern, as defined in equation (1). The order parameters m stand for the overlaps of the states with different patterns, and $q_{\rho\lambda}$ is analogous to the Edwards–Anderson parameter [14], with the following definitions

$$\begin{aligned} m_\rho^\nu &= \frac{1}{N\,a(1-\tilde{a})} \left\langle\left\langle \sum_{i=1}^{N} \langle v_{\xi_i^\nu \sigma_i^\rho} \rangle \right\rangle\right\rangle \\ q_{\rho\lambda} &= \frac{1}{N\,\tilde{a}\,a(1-\tilde{a})} \sum_{i=1}^{N} \left\langle\left\langle \sum_{k} P_k \langle v_{k\sigma_i^\rho} v_{k\sigma_i^\lambda} \rangle \right\rangle\right\rangle \\ r_{\rho\lambda} &= \frac{S(1-\tilde{a})}{\alpha} \sum_{\mu} \langle\langle m_\rho^\mu m_\lambda^\mu \rangle\rangle - \left(\frac{2S\,U}{\alpha} + 1\right) \frac{\delta_{\rho\lambda}}{\beta\tilde{a}} \end{aligned}$$

so that they are all of order 1. Consider, for example, that if $\sigma_i^\rho = \xi_i^\nu$ for all i then $m_\rho^\nu = 1$ on average, while $m_\rho^\nu = 0$ on average if both quantities are independent variables.

J. Stat. Mech. (2005) P08010

We now make two assumptions. First, we consider for simplicity that there is only one condensed pattern, making the index ν superfluous. Second, we assume that there is replica symmetry, and substitute

$$m_\rho^\nu = m$$
$$q_{\rho\lambda} = \begin{cases} q & \text{if } \rho \neq \lambda \\ \tilde{q} & \text{if } \rho = \lambda \end{cases}$$
$$r_{\rho\lambda} = \begin{cases} r & \text{if } \rho \neq \lambda \\ \tilde{r} & \text{if } \rho = \lambda. \end{cases}$$

Taking this into account, we arrive at the final expression for the free energy

$$f = a(1-\tilde{a})\frac{m^2}{2} + \frac{\alpha}{2\beta}\left[\ln\left(a(1-\tilde{a})\right) + \ln(1-\tilde{a}C) - \frac{\beta q \tilde{a}}{(1-\tilde{a}C)}\right] + \frac{\beta\alpha\tilde{a}^2}{2}(\tilde{q}\tilde{r} - qr) + \left[\frac{\alpha}{2} + S\,U\right]\tilde{q}\tilde{a} - \frac{1}{\beta}\left\langle\!\left\langle \int \mathrm{D}\mathbf{z} \ln\left(1 + \sum_{\sigma\neq 0}\exp(\beta\mathcal{H}_\sigma^\xi)\right)\right\rangle\!\right\rangle$$

where the finite-valued variable C has been introduced

$$C \equiv \beta(\tilde{q} - q)$$

in such a way that it is of order 1 and

$$\mathcal{H}_\sigma^\xi \equiv m\, v_{\xi\sigma} - \frac{\alpha a}{S^2}\frac{\beta(r-\tilde{r})}{2}(1-\delta_{\sigma 0}) + \sum_k \sqrt{\frac{\alpha r\, P_k}{S(1-\tilde{a})}}\, z_k v_{k\sigma}. \tag{14}$$

Note that $\mathcal{H}_0^\xi = 0$.

We now derive the fixed-point equation for m as an example of how the limit $\beta \to \infty$ is taken. The equation for finite β is

$$m = \frac{1}{a(1-\tilde{a})}\left\langle\!\left\langle \int \mathrm{D}\mathbf{z} \sum_\sigma v_{\xi\sigma}\left[\frac{1}{1 + \sum_{\rho\neq\sigma}\exp\{\beta(\mathcal{H}_\rho^\xi - \mathcal{H}_\sigma^\xi)\}}\right]\right\rangle\!\right\rangle.$$

In the limit $\beta \to \infty$ the expression between brackets is 1 if $\mathcal{H}_\sigma^\xi > \mathcal{H}_\rho^\xi$ for every $\rho \neq \sigma$ and 0 otherwise. It can thus be expressed as a product of Heaviside functions. The equation for m at zero temperature is then

$$m = \frac{1}{a(1-\tilde{a})}\sum_{\sigma\neq 0}\left\langle\!\left\langle \int \mathrm{D}\mathbf{z}\, v_{\xi\sigma}\prod_{\rho\neq\sigma}\Theta[\mathcal{H}_\sigma^\xi - \mathcal{H}_\rho^\xi]\right\rangle\!\right\rangle.$$

In the same way we derive the rest of the fixed-point equations at zero temperature

$$q \xrightarrow[\beta\to\infty]{} \tilde{q} = \frac{1}{a}\sum_{\sigma\neq 0}\left\langle\!\left\langle \int \mathrm{D}\mathbf{z} \prod_{\rho\neq\sigma}\Theta[\mathcal{H}_\sigma^\xi - \mathcal{H}_\rho^\xi]\right\rangle\!\right\rangle$$
$$C = \frac{1}{\tilde{a}^2\sqrt{\alpha r}}\sum_{\sigma\neq 0}\sum_k\left\langle\!\left\langle \int \mathrm{D}\mathbf{z} \sqrt{\frac{P_k}{S(1-\tilde{a})}}\, v_{k\sigma} z_k \prod_{\rho\neq\sigma}\Theta[\mathcal{H}_\sigma^\xi - \mathcal{H}_\rho^\xi]\right\rangle\!\right\rangle \tag{15}$$
$$\tilde{r} \xrightarrow[\beta\to\infty]{} r = \frac{q}{(1-\tilde{a}C)^2}$$
$$\beta(r - \tilde{r}) = 2U\frac{S^2}{a\alpha} - \frac{C}{1-\tilde{a}C}.$$

J. Stat. Mech. (2005) P08010

The differences between r and $\tilde{r}$, and between q and $\tilde{q}$, are of order $1/\beta$. From the last equation it can be seen that the threshold U has the effect of changing the sign of $(r-\tilde{r})$ and allowing α to scale like S^2/a, with the variables C, r and $\tilde{r}$, as we have said, of order 1 with respect to a and S.

4.1. Reduced saddle-point equations

It is possible to calculate the averages in equations (15) by reducing the problem to the following variables, which represent respectively signal and noise contributions

$$
\begin{aligned}
y &\equiv m\sqrt{\frac{S^2}{\alpha a}\frac{(1-\tilde{a})}{r}} \equiv m\sqrt{\frac{(1-\tilde{a})}{\tilde{\alpha} r}} \\
x &\equiv \frac{\tilde{\alpha}\beta(r-\tilde{r})}{2}\sqrt{\frac{(1-\tilde{a})}{\tilde{\alpha} r}}
\end{aligned}
$$

where we have introduced the normalized (order 1) storage capacity $\tilde{\alpha} \equiv \alpha a/S^2$, which clarifies that both variables x and y are also of order 1.

At the saddle point, using equations (15), we obtain

$$
\begin{aligned}
y &= \sqrt{\frac{1-\tilde{a}}{\tilde{\alpha}}}\left(\frac{m}{\sqrt{q}+C\sqrt{r}}\right) \\
x &= \sqrt{\frac{1-\tilde{a}}{\tilde{\alpha}}}\left[U-\tilde{\alpha}C\sqrt{\frac{r}{q}}\right]\left[\frac{1}{\sqrt{q}+\tilde{a}C\sqrt{r}}\right]
\end{aligned}
\tag{16}
$$

which shows that the relevant quantities to describe the system are m, q, and $C\sqrt{r}$. Following this we compute the averages and get from equation (15) the corresponding equations in terms of y and x

$$
\begin{aligned}
q &= \frac{(1-a)}{\tilde{a}}\int \mathrm{D}w \int_{y\tilde{a}+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,\phi(z)^S \\
&\quad + \int \mathrm{D}w \int_{-y(1-\tilde{a})+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,\phi(z+y)^S \\
&\quad + (S-1)\int \mathrm{D}w \int_{y\tilde{a}+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,\phi(z-y)\phi(z)^{S-1} \\
m &= \frac{1}{1-\tilde{a}}\int \mathrm{D}w \int_{-y(1-\tilde{a})+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,\phi(z+y)^S - q\frac{\tilde{a}}{1-\tilde{a}} \\
C\sqrt{r} &= \frac{1}{\sqrt{\tilde{\alpha}(1-\tilde{a})}}\left\{\frac{(1-a)}{\tilde{a}}\int \mathrm{D}w \int_{y\tilde{a}+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,(z+\mathrm{i}\sqrt{\tilde{a}}w)\phi(z)^S \right. \\
&\quad + \int \mathrm{D}w \int_{-y(1-\tilde{a})+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,(z+\mathrm{i}\sqrt{\tilde{a}}w)\phi(z+y)^S \\
&\quad \left. + (S-1)\int \mathrm{D}w \int_{y\tilde{a}+x-\mathrm{i}\sqrt{\tilde{a}}w}^{\infty} \mathrm{D}z\,(z+\mathrm{i}\sqrt{\tilde{a}}w)\phi(z-y)\phi(z)^{S-1}\right\}.
\end{aligned}
\tag{17}
$$

Putting together equations (16) and (17) one can construct the system of two equations that determine the storage capacity. We show an example of their solution

J. Stat. Mech. (2005) P08010

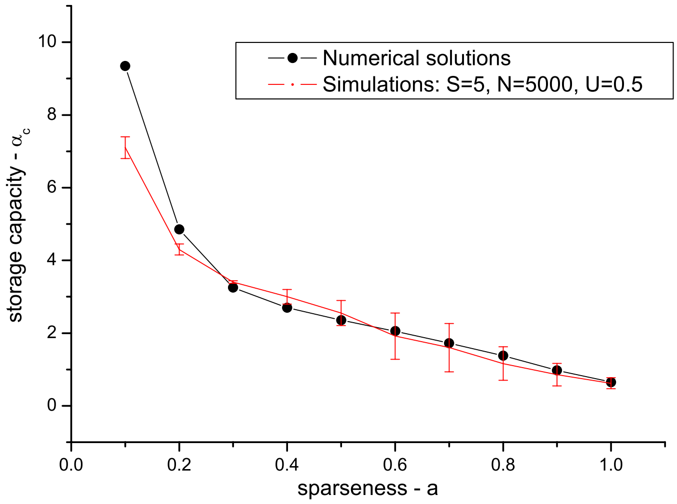

Figure 2. Dependence of the storage capacity of a sparse Potts network of $N = 5000$ units on the sparseness a. The black dots show numerical solutions of equations (16) and (17), while the red line shows the result of simulations. For very sparse simulations (low values of a) finite size effects are observed, which make the storage capacity lower than predicted by the equations.

in figure 2 for the parameters $U = 0.5$, $S = 5$ and varying sparseness, contrasting it with simulations of a network of $N = 5000$ units. This figure shows quite a good agreement between simulations and numerical solutions for a region of the sparseness parameter a, whereas for $a < 0.3$ finite size effects appear, resulting in a lower storage capacity than predicted theoretically.

4.2. Limit case

Given that the equations presented in the previous subsection are quite complex, we now analyse the simpler and interesting limit case $\tilde{a} \ll 1$. Though it is not evident from the equations, the normalized storage capacity $\tilde{\alpha}_c$ goes to zero logarithmically as $\tilde{a}$ goes to zero, which means that the storage capacity is not as high as the simple signal-to-noise analysis of section 3 might suggest. Our analysis of the replica equations for the symmetric Potts model (equation (12)) showing logarithmic corrections is an example of this. We now analyse as another example the sparse Potts model in the case $U = 0.5$.

For the limit of $\tilde{a} \ll 1$ one can approximate equations (17) by

$$m \approx \phi(y - x) \tag{18a}$$

$$q \approx \frac{(1-a)}{\tilde{a}}\phi(-x) + \phi(y - x) \tag{18b}$$

$$C\sqrt{r} \approx \frac{1}{\sqrt{2\pi\tilde{\alpha}}}\left\{\frac{(1-a)}{\tilde{a}}\exp\left(-\frac{x^2}{2}\right) + \exp\left(-\frac{(y-x)^2}{2}\right)\right\} \tag{18c}$$

which is still quite a complex system. We can now make some self-consistent assumptions. First we note that, considering x and y as variables that diverge logarithmically as $\tilde{a}$ goes to zero, equations (18b) and (18c) indicate that $\sqrt{q} \gg C\sqrt{r}$. Second, for $U = 1/2$ it is

J. Stat. Mech. (2005) P08010

possible to consider $x \approx y$, and thus, from equation (18a), $y \approx 1/\sqrt{2\tilde{\alpha}}$ and $x \approx \varepsilon/\sqrt{2\tilde{\alpha}}$, where ε is a correcting factor for x which is close to 1. With this in mind, and taking into account that $\tilde{\alpha}$ goes to zero with $\tilde{a}$, we can approximate equations (18b) and (18c) by keeping only the second term in the first case and only the first term in the second. The equations for y and x can be derived from equations (18b) and (16)

$$\begin{aligned} y &= \sqrt{\frac{\phi(y-x)}{\tilde{\alpha}}} \\ x &= \left[2U - \frac{1-a}{\tilde{a}}\sqrt{\frac{\tilde{\alpha}}{\pi}}\exp\left(-\frac{x^2}{2}\right)\right]\frac{1}{\sqrt{2\tilde{\alpha}}}. \end{aligned} \tag{19}$$

Replacing x by $\varepsilon/\sqrt{2\tilde{\alpha}}$ (and ε by 1 where irrelevant) we can approximate $\tilde{\alpha}$ as

$$\tilde{\alpha} = \frac{1}{4\ln\left(1/((2U-\varepsilon)\tilde{a})\right)}. \tag{20}$$

Next, we posit that $\tilde{a}^{-1}$ is the larger factor in the logarithm, while $(2U-\varepsilon)^{-1}$ gives a correction. A rough approximation for α_{c} is then

$$\alpha_{\mathrm{c}} = \frac{S^2}{4\,a\ln(1/\tilde{a})} \tag{21}$$

which, inserted in (19), gives

$$(2U-\varepsilon) = (1-a)\left[4\pi\ln\left(\frac{1}{\tilde{a}}\right)\right]^{-1/2}.$$

This expression can be re-inserted into (20) in order to get a more refined approximation

$$\alpha = \frac{S^2}{4\,a\ln\left(2/\tilde{a}\sqrt{\ln(1/\tilde{a})}\right)}. \tag{22}$$

We show in figure 3 that the approximation given by equation (22) fits the numerical solution of the storage capacity for the sparse Potts model quite well, particularly for very low values of $\tilde{a}$.

5. Diluted networks

In this section we present two modifications to our model which make the network biologically more plausible in terms of connectivity.

First, after considering, to a zeroth-order approximation, the long-range cortical network as a *fully connected* network, we now wish to describe it, to a better approximation, as a network in which the probability that two units are connected is c_M/N. Traditionally, analytic studies have focused on two soluble cases: the fully connected, which we have studied in the previous sections ($c_M = N$), and the highly diluted ($c_M \lesssim \log(N)$). A recent work has shown, however, that the intermediate case is also analytically treatable and that the storage capacity of an intermediate random network, regardless of the symmetry in the weights, stands between the storage capacity of the limit cases [15]. Supported by this result, we will focus on the (easier) solution for

J. Stat. Mech. (2005) P08010

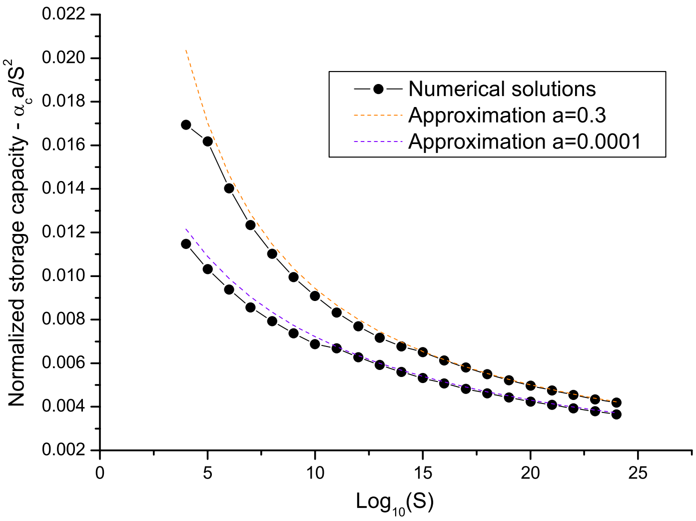

Figure 3. Corrections to the S^2/a behaviour of the storage capacity of a sparse Potts network for very low values of $\tilde{a}$ in the $U = 0.5$ case. The normalized storage capacity $\alpha_c a/S^2$ is represented with black dots from numerical solutions of equations (16) and (17) for two values of the sparseness, $a = 0.3$ and 0.0001, and with colour lines from the corresponding approximation given by equation (19).

the highly diluted case, and consider any intermediate situation to be between the two limits.

The second modification reflects the notion that, although the function of long-range connections is to transmit information about the state of a local network to another one, this transmission might not be perfectly efficient. We thus introduce an efficacy e, the probability that, in the reduced Potts model, a given state of the pre-synaptic unit is connected with a given state of the post-synaptic one.

Introducing these two modifications, the weights of the sparse Potts model become

$$J_{ij}^{kl} = \frac{C_{ij}^{kl}}{c_M ea(1-\tilde{a})} \sum_\mu v_{\xi_i^\mu k} v_{\xi_j^\mu l}$$

where C_{ij}^{kl} is 1 with probability $c_M e/N$ and 0 otherwise.

The local field for the unit i and the state k can be analysed into a signal, a noise and a threshold part, just as in equation (3)

$$h_i^k = \sum_{jl} J_{ij}^{kl} \delta_{\sigma_j l} - (1-\delta_{k0})U = (1-\delta_{k0}) \left\{ (\delta_{\xi_i^1 k} - \tilde{a}) m_i^k + N_k - U \right\} \qquad (23)$$

where

$$m_i^k \equiv \frac{1}{c_M ea(1-\tilde{a})} \sum_j C_{ij}^{k\sigma_j} (\delta_{\xi_j^1 \sigma_j} - \tilde{a})(1-\delta_{\sigma_j 0}).$$

Generally, when studying highly diluted networks, the noise term N_k can be treated directly as a uniform distributed random variable, because the states of different neurons are uncorrelated. In this case, N_k cannot be considered as a random variable but rather

J. Stat. Mech. (2005) P08010

as a weighted sum of normally distributed random variables η_l,

$$N_k \equiv \sum_{l=0}^{S}(\delta_{lk} - \tilde{a}) \left\{ \sum_{\mu>1} \frac{\delta_{\xi_i^\mu l}}{c_M e a(1-\tilde{a})} \sum_j C_{ij}^{k\sigma_j} (\delta_{\xi_j^\mu \sigma_j} - \tilde{a})(1-\delta_{\sigma_j 0}) \right\} \equiv \sum_l (\delta_{lk} - \tilde{a})\eta_l.$$

The mean of η_l is zero for all l and its standard deviation is

$$\langle \eta_l^2 \rangle = \frac{N\alpha P_l q_i^k}{(1-\tilde{a})c_M e}$$

with

$$q_i^k \equiv \frac{1}{c_M e a} \sum_j C_{ij}^{k\sigma_j}(1-\delta_{\sigma_j 0}).$$

Note that m_i^k and q_i^k are analogous to m_ρ^ν and $q_{\rho\lambda}$ used in section 4. If $c_M e$ is large enough these quantities tend to be independent of i and k.

$$m_i^k \to m \equiv \frac{1}{Na(1-\tilde{a})} \sum_j (\delta_{\xi_j^1 \sigma_j} - \tilde{a})(1-\delta_{\sigma_j 0})$$
$$q_i^k \to q \equiv \frac{1}{N\,a} \sum_j (1-\delta_{\sigma_j 0}).$$

Following the analysis of highly diluted networks in [16], the retrievable stable states of the network are given by the equations

$$m = \frac{1}{a(1-\tilde{a})} \left\langle\!\!\left\langle \int \mathrm{D}\mathbf{z} \sum_\sigma v_{\xi\sigma} \left[\frac{1}{1+\sum_{\rho\neq\sigma} \exp\{\beta(h_\rho^\xi - h_\sigma^\xi)\}} \right] \right\rangle\!\!\right\rangle$$
$$q = \frac{1}{a} \sum_{\sigma\neq 0} \left\langle\!\!\left\langle \int \mathrm{D}\mathbf{z} \left[\frac{1}{1+\sum_{\rho\neq\sigma} \exp\{\beta(h_\rho^\xi - h_\sigma^\xi)\}} \right] \right\rangle\!\!\right\rangle$$

where the local field, as in equation (23) is

$$h_\rho^\xi = m\, v_{\xi\rho} - U(1-\delta_{\rho 0}) + \sum_k \sqrt{\frac{\alpha\, N}{c_M e} \frac{q\, P_l}{(1-\tilde{a})}}\, z_k v_{k\rho}.$$

These equations are equivalent to those obtained with the replica method (which in the zero temperature limit are equations (15) and (14) respectively) if one considers $C=0$ (and, thus, $r=q$) and an effective value of α given by $\alpha_{\mathrm{eff}} = p/(c_M e)$.

Comparing this result with that for the fully connected model one notes that, as $\tilde{a} \to 0$, the influence of C in the overall equations becomes negligible (this can be guessed already from equation (16)). Therefore if the coding is very sparse, the fully connected and the highly diluted networks become equivalent, and consequently also the intermediate networks. We show this in figure 4. As the parameter $\tilde{a}$ goes to zero, the storage capacity of the fully connected and the highly diluted limit models converge.

J. Stat. Mech. (2005) P08010

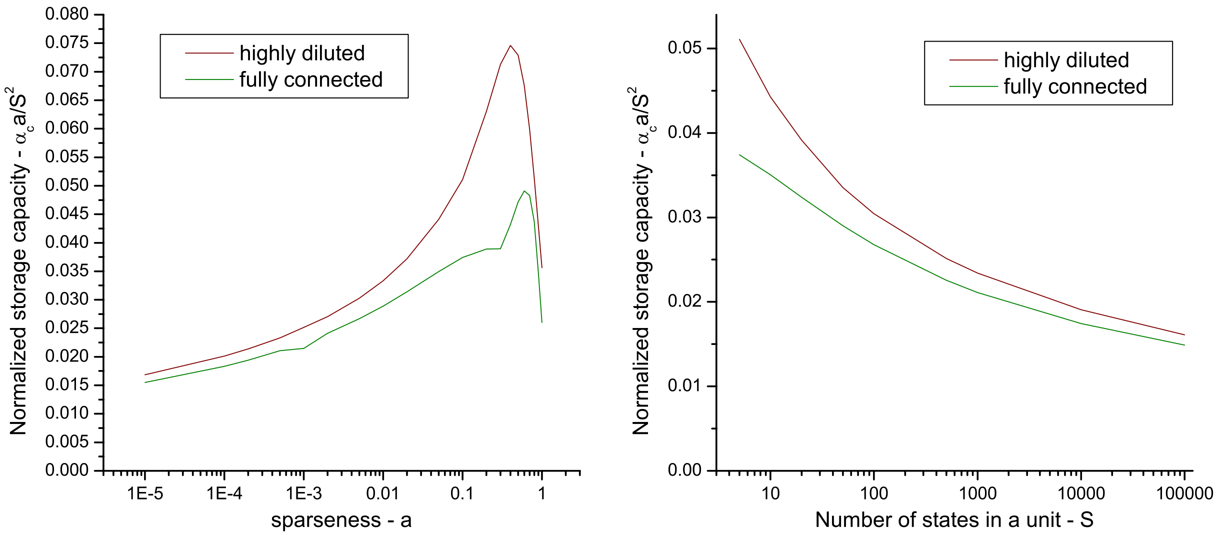

Figure 4. A comparison of the storage capacity of a fully connected and of a highly diluted sparse Potts networks. Numerical solutions to the corresponding equations with $U = 0.5$. Left, the dependence of the storage capacity, in the two cases, on the sparseness a, with $S = 5$. Right, the dependence on the number of states per unit S, with $a = 0.1$. In both cases we plot the normalized storage capacity, to focus only on the corrections to the S^2/a behaviour. Note that as $\tilde{a} \to 0$ the storage capacity of the two types of network converges to the same result.

6. Information capacity

We have shown that the storage capacity of well behaved models scales roughly like S^2/a, while in the two particular examples that we analysed in full with the replica method, equations (12) and (21), there is a correction that makes it

$$\alpha_c \propto \frac{S^2}{a\ \ln(1/\tilde{a})} \tag{24}$$

for high values of S and low values of a. We now discuss why this is reasonable in the general case from the information storage point of view.

It is widely believed, though not proved, that autoassociative memory networks can store a maximum of information equivalent to a fraction of a bit per synapse. In our model the total number of synaptic variables is given by the different combinations of indices of the weights J_{ij}^{kl}

$$\text{number of synaptic variables} = N\, c_M e S^2.$$

On the other hand, the information in a retrieved pattern is N times the contribution of a single unit, which, using the distribution in equation (1), can be bounded by Shannon's entropy

$$H = -\sum_{x \in \text{distribution}} P(x) \ln(P(x)) = -[(1-a)\ln(1-a) + a\ \ln(\tilde{a})].$$

J. Stat. Mech. (2005) P08010

The upper bound on the retrievable information over p patterns is then

$$\text{information} \leq -p\,N\,[(1-a)\ln(1-a) + a\ \ln(\tilde{a})].$$

The first term between brackets is negligible with respect to the second term provided a is small enough and S is large enough. In this way we can approximate

$$\frac{\text{information}}{\text{number of synaptic variables}} \leq -\frac{\alpha a \ln(\tilde{a})}{S^2} \leq -\frac{\alpha_c a \ln(\tilde{a})}{S^2}.$$

This result, combined with equation (24), shows that the storage capacity of our model is consistent with the idea that the information per synaptic variable is at most a fraction of a bit.

7. Discussion

The capacity to store information in any device, and in particular the capacity to store concepts in the human brain, is limited. We have shown in a minimal model of semantic memory, and in progressive steps, how one can expect the storage capacity to behave depending on the parameters of the system: a global parameter (the sparseness a) and a local parameter (the number of local retrieval states S, or, in other words, the storage capacity within a module). The S^2/a behaviour, with its corresponding logarithmic corrections, can be thought of as the combination of two separate results: the a^{-1} behaviour due to sparseness and the S^2 behaviour of the Potts model, which combine in a simple way. We have shown, however, that it is not trivial to define a model that combines these aspects correctly, and that the key is how the state operators are defined. From this study we have deduced the minimum requirements of any model of this kind in order to have a high capacity. Furthermore, through the argument of information capacity we present the well behaved family as representative of general Hebbian models with the same degree of complexity.

The featural representation approach has been so far successful in explaining several phenomena associated to semantic memory, like similarity priming, feature verification, categorization and conceptual combination [4, 17]. The present work demonstrates that the advantage of the use of features in allowing the representation of a large number of concepts can be realized in a simple associative memory network. More quantitatively, our calculation specifies that in the Potts model the number of concepts that can be stored is neither linear [2] nor an arbitrary power [18] of the number S of values a feature can take, but quadratic in S.

In the case of non-unitary sparseness, one can associate the necessity of introducing a threshold (U) term, whatever its exact form in the local field or the Hamiltonian, with a criterion of selectivity, which is actually observed in the representation of concepts in the brain, as pointed out in the introduction. The threshold behaviour, which is a typical characteristic of neurons, appears to be also necessary at the level of local networks in order to maintain activity low in the less representative modules. The origin of such a threshold has not been discussed in this paper. However, a comment on this issue can be made regarding the internal dynamics of local networks. One can show that, as extensively described in the literature [7], only when the state of a local autoassociative network is driven by external fields sufficiently close to an attractor (inside one of the S basins of attraction) may the local system end up retrieving a pattern on its own, a process

J. Stat. Mech. (2005) P08010

that from the global network point of view corresponds to the activation of a unit. The local basin boundary acts in the full system as an effective threshold, roughly equivalent to the simple U term we introduced in the local field of our reduced system. Whether this threshold mechanism is enough, or some addition must be made, can be assessed by studying, in the future, the complete multi-modular network without reducing it to Potts units.

Acknowledgments

This research was supported by Human Frontier Grant RGP0047/2004-C. We thank Yasser Roudi and David Hansel for very stimulating conversations.

J. Stat. Mech. (2005) P08010

References

[1] Braitenberg V and Schuz A, 1991 *Anatomy of the Cortex: Statistics and Geometry* (Berlin: Springer)
[2] O'Kane D and Treves A, 1992 *Network* **3** 379
[3] Pulvermuller F, 2002 *Prog. Neurobiol.* **67** 85
[4] McRae K, de Sa V and Seidemberg M, 1997 *J. Exp. Psychol. General* **126** 99
[5] O'Kane D and Treves A, 1992 *J. Phys. A: Math. Gen.* **25** 5055
[6] Fulvi Mari C and Treves A, 1998 *Biosystems* **48** 47
[7] Amit D J, 1989 *Modelling Brain Function: the World of Attractor Neural Networks* (Cambridge: Cambridge University Press)
[8] Mechelli A, Price C J, Noppeney U and Friston K J, 2003 *J. Cogn. Neurosci.* **15** 925 http://jocn.mitpress.org/cgi/reprint/15/7/925.pdf, URL http://jocn.mitpress.org/cgi/content/abstract/15/7/925
[9] Buhmann J, Divko R and Schulten K, 1989 *Phys. Rev.* A **39** 2689
[10] Tsodyks M V and Feigel'Man M V, 1988 *Europhys. Lett.* **6** 101
[11] Kanter I, 1988 *Phys. Rev.* A **37** 2739
[12] Bolle D, Cools R, Dupont P and Huyghebaert J, 1993 *J. Phys. A: Math. Gen.* **26** 549
[13] Treves A, 2005 *Cogn. Neuropsychol.* **6** 101
[14] Edwards S F and Anderson P W, 1975 *J. Phys. F: Metal Phys.* **5** 965
[15] Roudi Y and Treves A, 2004 *J. Stat. Mech.* P07010
[16] Derrida B, Gardner E and Zippelius A, 1987 *Europhys. Lett.* **4** 167
[17] McRae K, *Semantic memory: some insights from feature-based connectionist attractor networks*, 2005 *Psychology of Learning and Motivation* vol 45, ed B H Ross (Amsterdam: Elsevier) at press
[18] Schyns P and Rodet L, 1997 *J. Exp. Psychol.: Learn. Mem. Cogn.* **23** 681