Journal of Statistical Physics, Vol. 69, Nos. 3/4, 1992

Rigorous Bounds on the Storage Capacity of the Dilute Hopfield Model

Anton Bovier[1] **and Véronique Gayrard**[2]

Received December 16, 1992; final May 26, 1992

We study a neural network model consisting of N neurons where a dendritic connection between each pair of neurons exists with probability p and is absent with probability $1-p$. For the Hopfield Hamiltonian on such a network, we prove that if $p \geqslant c[(\ln N)/N]^{1/2}$, the model can store at least $m = \alpha_c pN$ patterns, where $\alpha_c \approx 0.027$ if $c \geqslant \sim 3$ and decreases proportional to $1/(-\ln c)$ for c small. This generalizes the results of Newman for the standard Hopfield model.

KEY WORDS: Hopfield model; bond dilution; memory capacity.

1. INTRODUCTION

Over the last decade the study of neural network models has become a major, rapidly developing area of research in both physics and computer science (see, e.g., ref. 11 or 7 for a recent review). The typical model of a neural network functioning as an autoassociative memory consists of a set of, say, N neurons each of which may be in a certain number (typically two) states. This state space S_N is described by a set of N spins σ_i. Any element of the state space S_N can be chosen as a pattern one wants to memorize. Given a number m of such patterns, denoted $\xi^1,\ldots,\xi^m$, one wants to define a Hamiltonian function $H_N(\sigma)$ (which of course depends on the patterns ξ^i) on the space S_N in such a way that a Markovian time evolution governed by this Hamiltonian has m stationary states each associated

[1] SFB 237 Essen, Bochum, and Düsseldorf Institut für Mathematik, Ruhr-Universität Bochum, W-4630 Bochum, Germany.

[2] Centre de Physique Théorique, CNRS, Luminy Case 907, F-13288 Marseille Cedex 9, France, and Center for Mathematical Sciences Research, Rutgers University, New Brunswick, New Jersey 08903.

0022-4715/92/1100-0597$06.50/0

to one of these patterns. The classical choice of this Hamiltonian is that of the Hopfield model,[12] where one sets

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j=1}^{N} \sum_{\mu=1}^{m} \sigma_i \sigma_j \xi_i^\mu \xi_j^\mu \tag{1.1}$$

but many variants of this model are presently being used. An important question is to know the storage capacity of such a model, i.e., for which number m of patterns a network of given size N will function properly. It has been found numerically[12] as well as by nonrigorous analytical computations[1, 19] that there exists a sharp α_c such that if $m \leqslant \alpha_c N$, the original patterns can be retrieved provided a small fraction of errors is allowed, while above this number of patterns the memory fails. The number of rigorous results on this question is fairly limited: McEliece *et al.*[17] have shown that if *no errors* are allowed in the retrieved patterns, then m has to be smaller than $N/\ln N$. On the other hand, Newman[18] (see also ref. 13) proved that there exist local minima in the Hamiltonian function (1.1) near each original pattern which are surrounded by energy barriers of height εN, provided $m \leqslant \alpha_c N$, with α_c at least 0.055. We will give a precise statement of this result later. These results have been generalized recently to the Potts–Hopfield model. Results concerning the actual existence of invariant measures for some dynamics are available only for much smaller numbers $m \leqslant \alpha \ln N$ of patterns.[15, 10]

An important feature in the Hamiltonian (1.1) is that it assumes a connection between any pair of two neurons, an architecture that is clearly not practical in very large networks and certainly not the one used, e.g., in the brain, where the number of dendrites connecting to a given neuron is only of the order of 10^4 while the total number of neurons is of the order of 10^9. To study a more realistic architecture, one may consider a model in which each pair of neurons is connected at random with (small!) probability p, where p will be allowed to depend on N. The Hamiltonian of this so-called dilute Hopfield model is given as

$$H_N(\sigma) = -\frac{1}{pN} \sum_{i,j=1}^{N} \sum_{\mu=1}^{m} \varepsilon_{ij} \sigma_i \sigma_j \xi_i^\mu \xi_j^\mu \tag{1.2}$$

where $\varepsilon_{ij} \equiv \varepsilon_{ji}$ are, for $i > j$, independent identically distributed random variables (i.i.d.r.v.'s) which take the value one with probability p and zero with probability $1-p$, with the natural interpretation that the neurons i and j are connected by a dendrite if $\varepsilon_{ij} = 1$ and disconnected otherwise. Note that here we assume *symmetric* connections of the neurons. As we shall indicate in the last section, this assumption may be removed, however, without altering our results.

Dilute neural networks are considered frequently in the literature (see, e.g., refs. 14, 2, 6, and 8). Most references concern the highly dilute model of ref. 6, whose dynamic permits exact solutions, but which does not work as a memory in the sense of the normal Hopfield model. Concerning the weakly dilute model, there exist some sparse numerical results and some remarks based on the replica method,[2, 3] but we are not aware of any rigorous results nor of a careful and systematic numerical investigation of this model.

One would now like to answer the question of how the storage capacity depends on the two parameters p and N, and in particular how small p is allowed to be taken to obtain a functioning memory. It has been suggested[16, 19] that the memory capacity should in general be proportional to the number of synapses, which in our case would suggest $m \sim \alpha p N$, but it is clear that this linear regime cannot possibly remain valid for arbitrarily small values of p, as we shall discuss in a moment. In the present paper we extend the analysis of Newman to the Hamiltonian (1.2). We will show that for $p > c[(\ln N)/N]^{1/2}$, with $c^2 \approx 7$, one may indeed store up to $\alpha_c pN$ patterns, where $\alpha_c \approx 0.027$. If c decreases beyond this value, α_c begins to decrease roughly proportional to $1/(-\ln c)$, and for p that decreases with N even more rapidly we have no results.

The appearance of a critical value for p in our bounds is quite intriguing. In fact, the very nature of the random network on which our model is based makes it plain that for some p small enough, the memory function must fail. Let us explain this: It is known that for random graphs (see, e.g., the text by Bollobas[4]) with connectivity rate p there exist a number of threshold values at which the nature of a "typical" graph changes: First, for $p < 1/N$, the graph is made of a large number of *finite* connected components. At $p = 1/N$, a so-called "giant component" appears, whose size is at first proportional to $\ln N$, and which grows, as p grows above $1/N$, to a fraction of N, until, at $p = \ln N/N$, it engulfs the entire graph. Thus $p = \ln N/N$ appears to be the lowest value for which one may reasonably expect the network to function in a normal way. We will see that this value of p also appears as the critical value in a number of our estimates. Of course, mere connectedness is not sufficient for a functioning of a neural network, and it may not be too surprising that the critical threshold we get is considerably higher. While our estimates are certainly not optimal and we cannot prove upper bounds for the storage capacity, we conjecture that there is indeed a critical dilution rate of the order of $1/\sqrt{N}$ and we will give some argument supporting this conjecture.

To be able to make precise statements, let us introduce some notation. We write $\boldsymbol{\sigma}$ for the configuration of spins $(\sigma_i,..., \sigma_N)$, ξ^μ for the pattern $(\xi_1^\mu,..., \xi_N^\mu)$ and $\xi|_{N,m}$ for the family of patterns $(\xi^1,..., \xi^m)$. The patterns will be considered as "random", i.e., each family of patterns $\xi|_{N,m}$ will be considered as the restriction of a random variable ξ on a probability space $(\Omega, \mathscr{F}, \mathbb{P}_\xi)$, where Ω is the space of sequences $\{\xi_i^\mu\}_{i\in\mathbb{N},\mu\in\mathbb{N}}$, with $\xi_i^\mu\in\{-1,+1\}$, $\mathscr{F}$ is the corresponding σ-algebra, and $\mathbb{P}_\xi$ is taken as the product measure $\mathbb{P}_\xi=\prod_{i\in\mathbb{N}}\prod_{\mu\in\mathbb{N}}\mathbb{P}_{\xi_i^\mu}$ with $\mathbb{P}_{\xi_i^\mu}$ the symmetric Bernoulli distribution concentrated on ± 1. Note that we prefer to introduce a probability space of infinite sequences and consider for N and m finite, $\xi|_{N,m}$ as cylinder variables rather than to introduce different probability spaces for finite systems. We denote by M the $N\times N$ symmetric random matrix whose elements are $M_{ij}=M_{ji}=\varepsilon_{ij}$, where for $i\geqslant j$ the ε_{ij} are i.i.d.r.v.'s whose common distribution ρ_p assigns the value one with probability p and the value zero with probability $1-p$. We denote the corresponding probability space by $(\Gamma, \mathscr{F}', \mathbb{P}_M)$.

Let us introduce the notation $\mathbb{J}$ for the $N\times N$ matrix all of whose elements are equal to one. Finally, we will write $\boldsymbol{\xi}^\mu\boldsymbol{\sigma}$ for the vector $\boldsymbol{\xi}^\mu\boldsymbol{\sigma}=(\xi_1^\mu\sigma_1,\ldots,\xi_N^\mu\sigma_N)$. The Hamiltonian for the dilute Hopfield model can then be written in the simple form

$$H_N(\boldsymbol{\sigma})=-\frac{1}{pN}\sum_{\mu=1}^{m}(\boldsymbol{\xi}^\mu\boldsymbol{\sigma}, M\boldsymbol{\xi}^\mu\boldsymbol{\sigma}) \tag{1.3}$$

Note that the values this Hamiltonian function takes are random variables on the probability space $(\Omega, \mathscr{F}, \mathbb{P}_\xi)\times(\Gamma, \mathscr{F}', \mathbb{P}_M)$; however, we do not keep track of this fact in our notation. The standard Hopfield model results as the special case where $p=1$ and thus $M=\mathbb{J}$.

We define on the space of spin configurations the usual Hamming distance,

$$d(\boldsymbol{\sigma}, \boldsymbol{\sigma}')=\tfrac{1}{2}[N-(\boldsymbol{\sigma}, \boldsymbol{\sigma}')] \tag{1.4}$$

that is, the number of components of the spins $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ that disagree. For any $\boldsymbol{\sigma}$ and any number $\delta\in[0,1]$ we denote by $\mathscr{S}(\boldsymbol{\sigma},\delta)$ the sphere of radius δN centered at $\boldsymbol{\sigma}$, i.e.,

$$\mathscr{S}(\boldsymbol{\sigma},\delta)\equiv\{\boldsymbol{\sigma}'\,|\,d(\boldsymbol{\sigma},\boldsymbol{\sigma}')=[\delta N]\} \tag{1.5}$$

where $[\delta N]$ denotes the largest integer smaller than or equal to δN (in the sequel we will often write simply δN instead of $[\delta N]$ whenever it is clear that the corresponding quantity must be an integer). Let us set

$$h_N(\boldsymbol{\sigma},\delta)\equiv\min_{\boldsymbol{\sigma}'\in\mathscr{S}(\boldsymbol{\sigma},\delta)}H_N(\boldsymbol{\sigma}') \tag{1.6}$$

We will say that there exists an energy barrier of height εN centered at ξ^μ if for some $\delta \in (0, 1/2)$,

$$h_N(\xi^\mu, \delta) \geqslant H_N(\xi^\mu) + \varepsilon N \tag{1.7}$$

We are now in the position to announce our main result:

Theorem 1. Suppose $p \geqslant c[(\ln N)/N]^{1/2}$. Then there exists $\alpha_c \geqslant 0$ such that if $m \leqslant \alpha_c pN$, then there exists $\varepsilon > 0$ and $0 < \delta < 1/2$ such that there exists $\gamma > 0$ such that

$$\mathbb{P}_M\left[\mathbb{P}_\xi\left[\bigcap_{\mu=1}^{m} \{h_N(\xi^\mu, \delta) > H_N(\xi^\mu) + \varepsilon N\}\right] \geqslant 1 - e^{-\gamma N}\right] \to 1 \qquad \text{as} \quad N \uparrow \infty \tag{1.8}$$

where the convergence in (1.8) is exponentially fast in N. Moreover, we have the following bounds on α_c:

$$\alpha_c \approx (16 \ln\{2[8(1+a)]^{1/2}\} \ln\{2[8(1+a)]^{1/2}\})^{-1}$$

where:

(i) $a \approx 0$ if $(p^2 N/\ln N) \uparrow \infty$.

(ii) $a < \frac{1}{2}$ if $c^2 > \sim 7$.

(iii) $a = 1 + 2/c$ otherwise.

Remark. In the Hopfield case (i.e., $p = 1$), such a result was first obtained by Newman.[18]

Remark. Our bounds on the probabilities in Theorem 1 also imply, by the Borel–Cantelli lemma,[5] that

$$\lim_{N \uparrow \infty} \mathbb{P}_\xi\left[\bigcap_{\mu=1}^{m} \{h_N(\xi^\mu, \delta) > H_N(\xi^\mu) + \varepsilon N\}\right] = 1 \qquad \mathbb{P}_M\text{-a.s.} \tag{1.9}$$

and that

$$\liminf_{N \uparrow \infty} \inf_{0 \leqslant \mu \leqslant m} \{h_N(\xi^\mu, \delta) - H_N(\xi^\mu) - \varepsilon N\} \geqslant 0 \qquad \mathbb{P}_M \times \mathbb{P}_\xi\text{-a.s.} \tag{1.10}$$

The original patterns are not the only local minima for our Hamiltonian, but there exist others corresponding to certain linear combinations of finitely many original patterns. These were first found in the context of the replica method by Amit *et al.*,[1] and Newman[18] has proven a theorem analogous to Theorem 1 establishing their existence. We will show that under the same conditions on p as in Theorem 1, this result, too, carries over to the dilute model. A precise statement of this theorem will be given in Section 4.

The remainder of this article is organized as follows. In Section 2 we present the proof of Theorem 1, assuming a result on the largest eigenvalues of certain submatrices of the random matrix M (Proposition 4), whose proof will then be given in Section 3. In Section 4 we investigate the "mixed memories" corresponding to superpositions of finitely many of the original patterns and prove a theorem analogous to Theorem 1 for them. In Section 5 we discuss a number of generalizations of our results, in particular to nonsymmetric networks and to networks with several types of dendrites.

2. PROOF OF THEOREM 1

In this section we present the backbone of the proof of Theorem 1. Let us denote by I a subset of $\{1, 2,..., N\}$ and for any vector $\boldsymbol{\sigma}$ let $\boldsymbol{\sigma}_I$ denote the vector in $\mathscr{S}(\boldsymbol{\sigma}, |I|)$ that differs from $\boldsymbol{\sigma}$ exactly on the coordinates $i \in I$. Then (we suppress the subscript N in the sequel whenever no confusion may arise)

$$\begin{aligned} 1 - \mathbb{P}_\xi \left[\bigcap_{\mu=1}^{m} \{h(\xi^\mu, \delta) > H(\xi^\mu) + \varepsilon N\} \right] &= \mathbb{P}_\xi \left[\bigcup_{\mu=1}^{m} \bigcup_{\substack{I \subset \{1,...,N\} \\ |I| = \delta N}} \{H(\xi_I^\mu) \leqslant H(\xi^\mu) + \varepsilon N\} \right] \\ &\leqslant \sum_{\mu=1}^{m} \sum_{\substack{I \subset \{1,...,N\} \\ |I| = \delta N}} \mathbb{P}_\xi[H(\xi_I^\mu) \leqslant H(\xi^\mu) + \varepsilon N] \end{aligned} \tag{2.1}$$

Notice that $\mathbb{P}_\xi$ here denotes the probability with respect to the random variable ξ only and that these probabilities are in itself still random variables on the probability space Γ of the matrices M. We have to show that the right-hand side of (2.1) goes to zero as N goes to infinity, for suitably chosen δ and ε, provided that M is in a subset Γ_{good} of the probability space Γ whose measure we will later show to go to 1 as $N \uparrow \infty$. Notice that the number of terms in the sum in (2.1) is $m\binom{N}{\delta N}$. We will therefore need to get uniform estimates on the probabilities in the sum that are small even when multiplied by this very large number. Note that bounding the probability of the union of events in (2.1) by the sum of the probabilities introduces a considerable overestimation, since the events corresponding to sets I and I' which are not disjoint are not independent. It appears, however, difficult to take this fact into account.

Let us now consider the probabilities $\mathbb{P}_\xi[H(\xi_I^\mu) \leqslant H(\xi^\mu) + \varepsilon N]$. Using that for symmetric matrices M,

$$(a, Ma) - (b, Mb) = ((a-b), M(a+b))$$

we have

$$\begin{aligned} H(\xi^\mu) - H(\xi_I^\mu) = & -\frac{1}{pN}((\xi^\mu\xi^\mu, M\xi^\mu\xi^\mu) - (\xi^\mu\xi_I^\mu, M\xi^\mu\xi_I^\mu)) \\ & -\frac{1}{pN}\sum_{\nu\neq\mu}((\xi^\mu\xi^\nu, M\xi^\mu\xi^\nu) - (\xi^\mu\xi_I^\nu, M\xi^\mu\xi_I^\nu)) \\ = & -\frac{1}{pN}((\xi^\mu\xi^\mu - \xi^\mu\xi_I^\mu), M(\xi^\mu\xi^\mu + \xi^\mu\xi_I^\mu)) \\ & -\frac{1}{pN}\sum_{\nu\neq\mu}((\xi^\mu\xi^\nu - \xi^\mu\xi_I^\nu), M(\xi^\mu\xi^\nu + \xi^\mu\xi_I^\nu)) \end{aligned} \tag{2.2}$$

It is easy to check that

$$(\xi^\mu\xi^\nu)_i - (\xi^\mu\xi_I^\nu)_i = \begin{cases} 2(\xi_i^\mu\xi_i^\nu) & \text{if} \quad i \in I \\ 0 & \text{if} \quad i \in I^c \end{cases} \tag{2.3}$$

and

$$(\xi^\mu\xi^\nu)_i + (\xi^\mu\xi_I^\nu)_i = \begin{cases} 0 & \text{if} \quad i \in I \\ 2(\xi_i^\mu\xi_i^\nu) & \text{if} \quad i \in I^c \end{cases} \tag{2.4}$$

where I^c denotes the complement of I in $\{1,..., N\}$.

We write for $\nu \neq \mu$, $\mathbf{y}^\nu = \xi^\mu\xi^\nu$. Notice that for μ fixed, the components of these vectors form a family

$$\{\mathbf{y}_i^\nu\}_{\substack{i \in \{1,...,N\} \\ \nu \in \{1,...,m\}/\{\mu\}}}$$

of i.i.d. random variables with $\mathbb{P}_\xi[y_i^\nu = \pm 1] = \frac{1}{2}$. Let us define the $N \times N$ matrix E_I whose matrix elements e_{ij} are given by

$$e_{ij} = \begin{cases} \varepsilon_{ij} & \text{if} \quad i \in I \text{ and } j \in I^c \\ \varepsilon_{ij} & \text{if} \quad i \in I^c \text{ and } j \in I \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

With this notation we may write

$$H(\xi^{\mu}) - H(\xi^{\mu}_{I}) = -\frac{2}{pN}\left\{(\mathbf{1}, E_I\mathbf{1}) + \sum_{\nu\neq\mu}(\mathbf{y}, E_I\mathbf{y}^{\nu})\right\} \tag{2.6}$$

where $\mathbf{1}$ denotes the N-vector all of whose components are equal to 1.

This allows us to write

$$\begin{aligned}
&\mathbb{P}_{\xi}[H(\xi^{\mu}) - H(\xi^{\mu}_{I}) \geqslant -\varepsilon N]\\
&= \mathbb{P}_{\xi}\left[-\left(\frac{(\mathbf{1}, E_I\mathbf{1})}{N\sqrt{\delta(1-\delta)}} + \sum_{\nu\neq\mu}\frac{(\mathbf{y}^{\nu}, E_I\mathbf{y}^{\nu})}{N\sqrt{\delta(1-\delta)}}\right) \geqslant \frac{-p\varepsilon N^2}{2N\sqrt{\delta(1-\delta)}}\right]\\
&\leqslant \inf_{t\geqslant 0}\exp\left\{\frac{p\varepsilon N^2 t}{2N\sqrt{\delta(1-\delta)}}\right\}\mathbb{E}_y\exp\left\{-t\left(\frac{(\mathbf{1}, E_I\mathbf{1})}{N\sqrt{\delta(1-\delta)}} + \sum_{\nu\neq s}\frac{(\mathbf{y}^{\nu}, E_I\mathbf{y}^{\nu})}{N\sqrt{\delta(1-\delta)}}\right)\right\}\\
&= \inf_{t\geqslant 0}\exp\left\{\frac{p\varepsilon N^2 t}{2N\sqrt{\delta(1-\delta)}} - t\frac{(\mathbf{1}, E_I\mathbf{1})}{N\sqrt{\delta(1-\delta)}}\right\}\mathbb{E}_y\exp\left\{-t\sum_{\nu\neq\mu}\frac{(\mathbf{y}^{\nu}, E_I\mathbf{y}^{\nu})}{N\sqrt{\delta(1-\delta)}}\right\}\\
&= \inf_{t\geqslant 0}\exp\left\{\frac{p\varepsilon N^2 t}{2N\sqrt{\delta(1-\delta)}} - t\frac{(\mathbf{1}, E_I\mathbf{1})}{N\sqrt{\delta(1-\delta)}}\right\}\\
&\quad\times\left[\mathbb{E}_{\mathbf{y}}\exp\left\{-t\frac{(\mathbf{y}^{\nu}, E_I\mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\}\right]^{m-1}
\end{aligned} \tag{2.7}$$

Here $\mathbb{E}_y$ denotes the expectation with respect to the family of independent random variables $\{\mathbf{y}^{\nu}\}$ introduced above, and $\mathbb{E}_{\mathbf{y}}$ denotes the expectation with respect to a random N-vector with the same distribution. We have used the exponential Markov inequality[5] and in the last line the independence of the variables $\mathbf{y}^{\nu}$. Also, we have divided all quantities by the factor $N[\delta(1-\delta)]^{1/2}$ for later convenience.

We have to estimate the expectation appearing in the last line of (2.7). In the case of the standard Hopfield model, the corresponding quantities were shown by Newman[18] to be essentially Gaussian expectations by virtue of the central limit theorem. The following lemma shows that in our more general situation, the expectations in (2.7) can still be estimated by Gaussian expectations.

Lemma 1. Let $\{z_i\}_{i=1,\ldots,N}$ be a family of i.i.d.r.v.'s whose common distribution is the standard normal distribution (i.e., Gaussian with mean zero and variance one). Let $\mathbf{z} = (z_1,..., z_N)$ and let $\mathbb{E}_{\mathbf{z}}$ denote the expectation with respect to these variables. Then

$$\mathbb{E}_{\mathbf{y}} \exp\left\{-t \frac{(\mathbf{y}, E_I \mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\} \leqslant \mathbb{E}_{\mathbf{z}} \exp\left\{t \frac{(\mathbf{z}, E_I \mathbf{z})}{N\sqrt{\delta(1-\delta)}}\right\} \tag{2.8}$$

Proof. Denote by $\mathbb{E}_A$ the expectation with respect to the random variables $\{y_i\}_{i \in A}$. Then the left-hand side of (2.8) can be written more explicitly as

$$\begin{aligned} &\mathbb{E}_{\mathbf{y}} \exp\left\{-t \frac{(\mathbf{y}, E_I \mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\} \\ &\quad = \mathbb{E}_I \mathbb{E}_{I^c} \exp\left\{-2t \sum_{i \in I} \sum_{j \in I^c} \frac{y_i}{\sqrt{\delta N}} \frac{y_j}{\sqrt{N(1-\delta)}} e_{ij}\right\} \end{aligned} \tag{2.9}$$

Denoting by Z_j the sum

$$Z_j \equiv \sum_{i \in I} \frac{y_i \varepsilon_{ij}}{\sqrt{\delta N}}$$

we get

$$\mathbb{E}_{\mathbf{y}} \exp\left\{-t \frac{(\mathbf{y}, E_I \mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\} = \mathbb{E}_I \mathbb{E}_{I^c} \exp\left\{-2t \sum_{j \in I^c} \frac{y_j Z_j}{\sqrt{N(1-\delta)}}\right\} \tag{2.10}$$

Now

$$\begin{aligned} \mathbb{E}_{I^c} \exp\left\{-2t \sum_{j \in I^c} \frac{y_j Z_j}{\sqrt{N(1-\delta)}}\right\} &= \prod_{j \in I^c} \cosh\left(\frac{2tZ_j}{\sqrt{N(1-\delta)}}\right) \\ &\leqslant \prod_{j \in I^c} \exp\left\{\frac{1}{2}\left(\frac{2tZ_j}{\sqrt{N(1-\delta)}}\right)^2\right\} \\ &= \prod_{j \in I^c} \mathbb{E}_{z_j} \exp\left\{\frac{2tz_j Z_j}{\sqrt{N(1-\delta)}}\right\} \end{aligned} \tag{2.11}$$

where we have used the identity

$$\exp\left(\frac{1}{2}x^2\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}z^2 + xz\right) dz$$

and the inequality $\cosh x \leqslant \exp(\frac{1}{2}x^2)$. Thus we get that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{y}} \exp\left\{-t\frac{(\mathbf{y}, E_I \mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\} \\
&\quad\leqslant \mathbb{E}_{\{z_j\}_{j\in I^c}} \mathbb{E}_I \exp\left\{\sum_{j\in I^c} \frac{2tz_j Z_j}{\sqrt{N(1-\delta)}}\right\} \\
&\quad= \mathbb{E}_{\{z_j\}_{j\in I^c}} \mathbb{E}_I \exp\left\{\sum_{i\in I} \frac{2ty_i}{\sqrt{\delta N}}\left(\sum_{j\in I^c} \frac{z_j \varepsilon_{ij}}{\sqrt{N(1-\delta)}}\right)\right\} \\
&\quad= \mathbb{E}_{\{z_j\}_{j\in I^c}} \mathbb{E}_I \exp\left\{\sum_{i\in I} \frac{2ty_i \tilde{Z}_i}{\sqrt{\delta N}}\right\} \qquad (2.12)
\end{aligned}$$

where we have set

$$\tilde{Z}_i \equiv \sum_{j\in I^c} \frac{z_j \varepsilon_{ij}}{\sqrt{N(1-\delta)}}$$

Now, just as before,

$$\mathbb{E}_I \exp\left\{\sum_{i\in I} \frac{2ty_i \tilde{Z}_i}{\sqrt{\delta N}}\right\} \leqslant \mathbb{E}_{\{z_j\}_{j\in I}} \exp\left\{\sum_{i\in I} \frac{2tz_i \tilde{Z}_i}{\sqrt{\delta N}}\right\} \qquad (2.13)$$

and we finally we arrive at

$$\begin{aligned}
&\mathbb{E}_{\mathbf{y}} \exp\left\{-t\frac{(\mathbf{y}, E_I \mathbf{y})}{N\sqrt{\delta(1-\delta)}}\right\} \\
&\quad\leqslant \mathbb{E}_{\{z_i\}_{i\in I}} \mathbb{E}_{\{z_j\}_{j\in I^c}} \exp\left\{2t \sum_{i\in I} \sum_{j\in I^c} \frac{z_i}{\sqrt{\delta N}} \frac{z_j}{\sqrt{N(1-\delta)}} \varepsilon_{ij}\right\} \qquad (2.14)
\end{aligned}$$

which proves the lemma. ∎

Notice that the right-hand side of (2.8) reads, in explicit form,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{z}} \exp\left\{t\frac{(\mathbf{z}, E_I \mathbf{z})}{N\sqrt{\delta(1-\delta)}}\right\} \\
&\quad= \int \frac{d^N \mathbf{z}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{z}, \left[id - \frac{2t}{N\sqrt{\delta(1-\delta)}} E_I\right] \mathbf{z}\right)\right\} \qquad (2.15)
\end{aligned}$$

This integral exists, provided the matrix

$$id - \frac{2t}{N\sqrt{\delta(1-\delta)}} E_I$$

is positive definite, and in this case is nothing but the inverse of the square root of the determinant of this matrix. Now, for t positive, the positive definiteness of this matrix depends on the maximal eigenvalue of the random matrix E_I. These will be investigated in the next section. Assuming these eigenvalues to be given, we get the following result.

Lemma 2. Let $\lambda_1,\dots,\lambda_N$ be the eigenvalues of E_I. Choose $\gamma>0$. Then, for all $t\geqslant 0$ such that for all $i=1,\dots,N$,

$$1-\frac{4t^2\lambda_i^2}{N^2\delta(1-\delta)}\geqslant\gamma \tag{2.16}$$

the inequality

$$\mathbb{E}_{\mathbf{z}}\exp\left\{t\frac{(\mathbf{z},E_I\mathbf{z})}{N\sqrt{\delta(1-\delta)}}\right\}\leqslant\exp\left\{\frac{t^2}{\gamma}\frac{\operatorname{tr}E_I^2}{N^2\delta(1-\delta)}\right\} \tag{2.17}$$

holds.

Proof. Under the assumptions of the lemma, the matrix

$$id-\frac{2t}{N\sqrt{\delta(1-\delta)}}E_I$$

is positive definite, so that by the remarks above

$$\begin{aligned}\mathbb{E}_{\mathbf{z}}\exp\left\{t\frac{(\mathbf{z},E_I\mathbf{z})}{N\sqrt{\delta(1-\delta)}}\right\}&=\det\left(id-\frac{2t}{N\sqrt{\delta(1-\delta)}}E_I\right)^{-1/2}\\&=\prod_{i=1}^{N}\left(1-\frac{2t}{N\sqrt{\delta(1-\delta)}}\lambda_i\right)^{-1/2}\end{aligned} \tag{2.18}$$

Notice further that the spectrum of E_I is always symmetric, i.e., if λ is a nonzero eigenvalue of E_I with eigenvector $\mathbf{v}$, then the vector $\mathbf{w}$, where $w_i=-v_i$ for $i\in I$ and $w_j=v_j$ for $j\in I^c$, is an eigenvector for the eigenvalue $-\lambda$. We can thus arrange the λ_i in such a way that $\lambda_i\geqslant 0$ for $i\leqslant[N/2]$. If N is odd, since $\operatorname{tr}E_I=0$, E_I has at least one eigenvalue zero. Thus,

$$\begin{aligned}&\det\left(id-\frac{2t}{N\sqrt{\delta(1-\delta)}}E_I\right)^{-1/2}\\&=\prod_{i=1}^{[N/2]}\left(1-\frac{2t}{N\sqrt{\delta(1-\delta)}}\lambda_i\right)^{-1/2}\prod_{i=1}^{[N/2]}\left(1+\frac{2t}{N\sqrt{\delta(1-\delta)}}\lambda_i\right)^{-1/2}\\&=\prod_{i=1}^{[N/2]}\left(1-\frac{4t^2}{N^2\delta(1-\delta)}\lambda_i^2\right)^{-1/2}\end{aligned} \tag{2.19}$$

This latter product can be further bounded, using our assumption (2.16), by

$$\begin{aligned}
&\prod_{i=1}^{[N/2]}\left(1-\frac{4t^2}{N^2\delta(1-\delta)}\lambda_i^2\right)^{-1/2}\\
&\quad=\prod_{i=1}^{[N/2]}\left(1+\frac{\{4t^2/N^2\delta(1-\delta)\}\lambda_i^2}{1-\{4t^2/N^2\delta(1-\delta)\}\lambda_i^2}\right)^{1/2}\\
&\quad\leqslant\prod_{i=1}^{[N/2]}\left(1+\frac{4t^2\lambda_i^2}{N^2\delta(1-\delta)\gamma}\right)^{1/2}\\
&\quad\leqslant\exp\left\{\prod_{i=1}^{[N/2]}\frac{2t^2\lambda_i^2}{N^2\delta(1-\delta)\gamma}\right\}=\exp\left\{\frac{t^2}{\gamma N^2\delta(1-\delta)}\operatorname{tr}E_I^2\right\}
\end{aligned}\tag{2.20}$$

which proves the lemma. ▮

Remark. The estimates used in particular in (2.20) may look somewhat grotesque. However, for the range of t values we will finally use they give away rather little and the convenient form of the bound obtained is well worth the concessions we make. Of course, our final numerical bounds on the storage capacity could in principle be improved a little by refining these bounds.

In the case of the standard Hopfield model (i.e., $p=1$), the matrices E_I have two eigenvalues equal to $\pm N[\delta(1-\delta)]^{1/2}$ and all other eigenvalues equal to zero. Moreover, $\operatorname{tr} E_I^2=2N^2\delta(1-\delta)$. One may then choose, e.g., $\gamma=\frac{1}{2}$ to obtain the bound $\exp(4t^2)$ for all $t\leqslant 1/\sqrt{8}$, from which the result of Newman[18] follows easily. The essential new feature in the dilute model is that we need to obtain probabilistic bounds on the largest eigenvalue of the random matrices E_I, as well as on the traces of E_I^2. This will be done in the next section. An essential difficulty that arises there is that we need such bounds uniformly for all subsets I of given volume. As we shall see, this is the main reason why we need to restrict ourselves to dilution rates $p>c[(\ln N)/N]^{1/2}$.

Let us now assume that our dilution matrix M is such that *for all* $I\subset\{1,..., N\}$ such that $|I|=\delta N$, the following conditions hold:

(C1) $\max_i \lambda_i^2\leqslant \delta N^2p^2(1+a)$.

(C2) $2N^2p\delta(1-\delta)(1-x)\leqslant \operatorname{tr} E_I^2\leqslant 2N^2p\delta(1-\delta)(1+x)$.

Then we may choose $\gamma>0$ and use (2.17) for all

$$0\leqslant t\leqslant T\equiv\frac{1}{2p}\left\{\frac{(1-\gamma)(1-\delta)}{1+a}\right\}^{1/2}$$

and therefore get

$$\mathbb{P}_{\xi}[H(\xi^{\mu})-H(\xi^{\mu}_{I})\geqslant -\varepsilon N]$$

$$\leqslant \inf_{0\leqslant t\leqslant T} \exp\left\{\frac{\varepsilon p N^2 t}{2N\sqrt{\delta(1-\delta)}}-\frac{t \operatorname{tr} E_I^2}{N\sqrt{\delta(1-\delta)}}+m\frac{t^2}{\gamma N^2\delta(1-\delta)}\operatorname{tr} E_I^2\right\} \tag{2.21}$$

Let us now set $m\equiv \alpha p N$ and define

$$\tilde{t}\equiv\frac{p}{\sqrt{\delta(1-\delta)}},\qquad \tilde{T}\equiv\frac{1}{2\sqrt{\delta}}\left(\frac{1-\gamma}{1+a}\right)^{1/2}$$

Then the right-hand side of (2.21) becomes

$$\inf_{0\leqslant \tilde{t}\leqslant \tilde{T}} \exp\left\{N\left(\frac{\varepsilon\tilde{t}}{2}-\tilde{t}\frac{\operatorname{tr} E_I^2}{pN^2}+\frac{\alpha}{\gamma}\frac{\operatorname{tr} E_I^2}{pN^2}\tilde{t}^2\right)\right\}$$

$$\leqslant \inf_{0\leqslant \tilde{t}\leqslant \tilde{T}} \exp\left\{N\left(\tilde{t}\left[\frac{\varepsilon}{2}-2\delta(1-\delta)(1-x)\right]+\frac{\alpha}{\gamma}2\delta(1-\delta)(1+x)\tilde{t}^2\right)\right\} \tag{2.22}$$

where we have inserted the bounds (C2) for the trace of E_I^2. The exponent on the right of (2.22) is minimized for

$$\tilde{t}=t^*\equiv\frac{\gamma}{2\alpha}\left[\frac{1-x}{1+x}-\frac{\varepsilon}{4\delta(1-\delta)(1+x)}\right] \tag{2.23}$$

and thus, provided $t^*\leqslant\tilde{T}$, we get

$$\mathbb{P}_{\xi}[H(\xi^{\mu})-H(\xi^{\mu}_{I})\geqslant -\varepsilon N]$$

$$\leqslant \exp\left\{-N\frac{\gamma}{2\alpha}\delta(1-\delta)(1+x)\left[\frac{1-x}{1+x}-\frac{\varepsilon}{4\delta(1-\delta)(1+x)}\right]^2\right\} \tag{2.24}$$

Notice that this bound is uniform in I and so

$$\sum_{|I|=\delta N}\mathbb{P}_{\xi}[H(\xi^{\mu})-H(\xi^{\mu}_{I})\geqslant -\varepsilon N]$$

$$\leqslant \binom{N}{\delta N}\exp\left\{-N\frac{\gamma}{2\alpha}\delta(1-\delta)(1+x)\left[\frac{1-x}{1+x}-\frac{\varepsilon}{4\delta(1-\delta)(1+x)}\right]^2\right\}$$

$$\leqslant \exp\left\{-N[\delta\ln\delta+(1-\delta)\ln(1-\delta)]\right.$$

$$\left.-N\frac{\gamma}{2\alpha}\delta(1-\delta)(1+x)\left[\frac{1-x}{1+x}-\frac{\varepsilon}{4\delta(1-\delta)(1+x)}\right]^2\right\} \tag{2.25}$$

In order for this probability to converge to zero as $N \uparrow \infty$, the coefficient of N in the exponent must be negative. This, together with the constraint $\tilde{t}^* \leqslant \tilde{T}$, implies the following two conditions on our parameters:

$$-\ln \delta - (1-\delta)\frac{\ln(1-\delta)}{\delta} \leqslant \frac{\gamma}{2\alpha}(1-\delta)(1+x)\, Y \tag{2.26}$$

and

$$0 \leqslant \frac{\gamma}{2\alpha}\, Y \leqslant \frac{1}{2\sqrt{\delta}}\left(\frac{1-\gamma}{1+a}\right)^{1/2} \tag{2.27}$$

Here we have put

$$Y \equiv \left[\frac{1-x}{1+x} - \frac{\varepsilon}{4\delta(1-\delta)(1+x)}\right]$$

By choosing ε, we may vary Y between 0 and $(1-x)/(1+x)$, and for x small (which will be the only case we consider), we may obtain, e.g., $Y = \frac{1}{2}$. We can choose γ between 0 and 1, but since we are interested in obtaining α as large as possible, we should not choose γ very small. In fact, little can be gained over a choice $\gamma = \frac{1}{2}$. With these choices, (2.26) and (2.27) are compatible if

$$-\ln \delta - (1-\delta)\frac{\ln(1-\delta)}{\delta} \leqslant \frac{1}{4\sqrt{\delta}}\left[\frac{1}{2(1+a)}\right]^{1/2}(1-\delta)(1+x) \tag{2.28}$$

It is easy to see that for any a fixed, there exists $\delta_0 > 0$ such that for $\delta \leqslant \delta_0$, (2.28) is satisfied. In fact, such δ_0 exists even if a is allowed to grow as $\delta \downarrow 0$, provided only

$$|\ln \delta|\ [\delta(1+a_\delta)]^{1/2} \downarrow 0 \qquad \text{as} \quad \delta \downarrow 0 \tag{2.29}$$

It is not difficult to get reasonable estimates for δ_0. Since we anticipate $\delta_0 \ll 1$, and putting $4[2(1+a)]^{1/2} \equiv b$, δ_0 is essentially the solution of

$$\sqrt{\delta_0}\ |\ln \delta_0| = \frac{1}{b} \tag{2.30}$$

whose solution can be obtained by a standard iteration procedure to arbitrary accuracy. The first nontrivial approximant yields

$$\sqrt{\delta_0} \approx \frac{1}{2b \ln 2b} \tag{2.31}$$

The maximal allowed α such that (2.26) is satisfied is then obtained if δ is chosen as small as possible, i.e., equal to δ_0. Then

$$\alpha_c \approx \frac{1}{16\ |\ln \delta_0|} \approx \frac{1}{16 \ln(2b \ln 2b)} \tag{2.32}$$

We collect the results of the foregoing discussion in the following statement.

Proposition 3. Suppose the matrix M satisfies conditions (C1) and (C2) with $0 \leqslant x \ll 1$ and a arbitrary but independent of δ. Then there exists

$$\alpha_c \approx (16 \ln\{2[8(1+a)]^{1/2}\} \ln\{2[8(1+a)]^{1/2}\})^{-1}$$

such that if $m \leqslant \alpha_c pN$, there exists $\varepsilon > 0$ and $\delta > 0$ such that

$$\mathbb{P}_\xi \left[\bigcap_{\mu=1}^{m} \{h_N(\xi_I^\mu, \delta) > H_N(\xi_I^\mu) + \varepsilon N\} \right] \to 1 \qquad \text{as} \quad N \uparrow \infty \tag{2.33}$$

Moreover, the rate of convergence to 1 is exponential in N.

Theorem 1 now follows immediately from the following result:

Proposition 4. Assume $p \geqslant c[(\ln N)/N]^{1/2}$ for any constant $c > 0$. Then

$$\mathbb{P}_M[\forall_{I:|I| = \delta N} \{(\mathrm{C1}) \wedge (\mathrm{C2})\}] \uparrow 1 \qquad \text{as} \quad N \uparrow 1 \tag{2.34}$$

where in (C2) $x > 0$ may be chosen arbitrarily small, while the choice of a in (C1) depends on p, namely:

(i) If $p^2 N/\ln N \uparrow \infty$, any a such that $(1-\delta)(1+a) > 1$ suffices.

(ii) If $Np^2 = c^2 \ln N$, with c^2 sufficiently large (~ 7), a may be chosen less than $\sim \frac{1}{2}$.

(iii) Otherwise, we must choose $a > 1 + 2/c$.

The rate of convergence in (2.34) is faster than any power of $1/N$ in case (i). In the other cases the convergence is like a power of $1/N$, which depends on the choice of a and which can be made as large as desired.

The proof of Proposition 4 will be given in the next section. Assuming this proposition, the proof of Theorem 1 is now finished. ∎

3. BOUNDS ON THE EIGENVALUES OF E_I

In this section we provide the necessary probabilistic estimates on the traces and eigenvalues of the matrices E_I which will yield Proposition 4 and thus conclude the proof of our main Theorem 1. We begin by stating the following technical lemma, which will be convenient for later use:

Lemma 5. Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a family of i.i.d.r.v.'s such that $\mathbb{P}(\varepsilon_i=1)=q$ and $\mathbb{P}(\varepsilon_i=0)=1-q$. Then:

(i) If $c=Mq(1+y)$, $0\leqslant y\leqslant 1$, then

$$\mathbb{P}\left[\sum_{i=1}^{M}\varepsilon_i\geqslant c\right]\leqslant\exp\left\{-Mq\frac{y^2}{2}\left(1-\frac{y}{3}e^y\right)\right\}\leqslant\exp\left\{-Mq\frac{y^2}{2}\left(1-\frac{e}{3}\right)\right\} \tag{3.1}$$

(ii) If $c=Mq(1+y)$, $y>1$, then

$$\mathbb{P}\left[\sum_{i=1}^{M}\varepsilon_i\geqslant c\right]\leqslant\exp\left\{-Mq\left(y-\frac{3+e}{6}\right)\right\} \tag{3.2}$$

(iii) If $c=Mq(1-y)$, $y>0$, then

$$\mathbb{P}\left[\sum_{i=1}^{M}\varepsilon_i\leqslant c\right]\leqslant\exp\left(-Mq\frac{y^2}{2}\right) \tag{3.3}$$

Proof. We first prove (i) and (ii). Using the exponential Markov inequality[5] and the independence of the r.v.'s ε_i, we have that

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{M}\varepsilon_i\geqslant c\right]&\leqslant\inf_{t\geqslant 0}\exp(-ct)\,[\mathbb{E}\exp(t\varepsilon_1)]^M\\&=\inf_{t\geqslant 0}\exp(-ct)\exp\{M\ln[q(e^t-1)+1]\}\end{aligned} \tag{3.4}$$

Now for all $t\geqslant 0$,

$$\ln[q(e^t-1)+1]\leqslant qt+\frac{qt^2}{2}+\frac{qt^3}{6}e^t \tag{3.5}$$

Therefore

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{M}\varepsilon_i\geqslant c\right]&\leqslant\inf_{t\geqslant 0}\exp\left\{-t(c-Mq)+Mq\left[\frac{t^2}{2}+\frac{t^3}{6}e^t\right]\right\}\\&\leqslant\inf_{t\geqslant 0}\exp\left\{-Mq\left[ty-\frac{t^2}{2}+\frac{t^3}{6}e^t\right]\right\}\\&\leqslant\inf_{0\leqslant y_1\leqslant y}\exp\{-Mqy(y-y_1)\}\\&\quad\times\exp\left\{Mq\left[\frac{(y-y_1)^2}{2}+\frac{(y-y_1)^3}{6}e^{y-y_1}\right]\right\}\end{aligned} \tag{3.6}$$

where the last inequality was obtained by setting $t = y - y_1$. If $y \leqslant 1$, we may bound (3.6) by putting $y_1 = 0$, which gives (3.1). On the other hand, if $y > 1$, we set $y_1 = y - 1$, which gives (3.2). This latter bound is good for large y and could of course be somewhat improved for y only slightly bigger than 1, but we will not need this here. Now to prove (iii), just note that

$$\mathbb{P}\left[\sum_{i=1}^{M} \varepsilon_i \leqslant c\right] \leqslant \inf_{t \geqslant 0} \exp(ct) \, [\mathbb{E} \exp(-t\varepsilon_1)]^M$$

$$= \inf_{t \geqslant 0} \exp(ct) \exp\{M \ln[q(e^{-t} - 1) + 1]\} \tag{3.7}$$

and since for all $t \geqslant 0$,

$$\ln[q(e^{-t} - 1) + 1] \leqslant -qt + \frac{qt^2}{2} \tag{3.8}$$

this time we get simply

$$\mathbb{P}\left[\sum_{i=1}^{M} \varepsilon_i \leqslant c\right] \leqslant \inf_{t \geqslant 0} \exp\left\{-t(Mq - c) + Mq\frac{t^2}{2}\right\}$$

$$\leqslant \exp\left(-Mq\frac{y^2}{2}\right) \tag{3.9}$$

where the last line is obtained by setting $t = y$. This concludes the proof of Lemma 5. ∎

The first simple application of Lemma 5 gives the following bounds on the traces of E_I^2.

Lemma 6. Suppose $p \geqslant \ln N/N$. Then for any $x > 0$,

$$\mathbb{P}[\exists_{I:|I| = \delta N} \operatorname{tr} E_I^2 \geqslant 2(1 + x) \, pN^2\delta(1 - \delta)] \downarrow 0 \qquad \text{as} \quad N \uparrow \infty \tag{3.10}$$

and

$$\mathbb{P}[\exists_{I:|I| = \delta N} \operatorname{tr} E_I^2 \leqslant 2(1 - x) \, pN^2\delta(1 - \delta)] \downarrow 0 \qquad \text{as} \quad N \uparrow \infty \tag{3.11}$$

where in both cases the rate of convergence is faster than exponential in N.

Proof. We will just bound the probability of the union of events by the sum of the respective probabilities. Doing this and using the symmetry of our probability space, we get

$$
\begin{aligned}
&\mathbb{P}[\exists_{I:|I|=\delta N} \operatorname{tr} E_I^2 \geqslant 2(1+x)\, pN^2\delta(1-\delta)] \\
&\quad\leqslant \sum_{\substack{I\subset\{1,\ldots,N\}\\ |I|=\delta N}} \mathbb{P}\left[2\sum_{\substack{i\in I\\ j\in I^c}} \varepsilon_{ij} \geqslant 2(1+x)\, pN^2\delta(1-\delta)\right] \\
&\quad\leqslant \binom{N}{\delta N} \mathbb{P}\left[\sum_{i=1}^{\delta N}\sum_{j=1}^{(1-\delta)N} \varepsilon_{ij} \geqslant (1+x)\, pN^2\delta(1-\delta)\right] \\
&\quad\leqslant \binom{N}{\delta N} \exp\left\{-pN^2\delta(1-\delta)\left[\frac{x^2}{2}-\frac{x^3}{6}e^x\right]\right\}
\end{aligned}
\tag{3.12}
$$

where the last inequality is an immediate application of Lemma 5. (We have assumed $x<1$, which of course implies *a fortiori* the result for larger x). Now for any x and δ fixed, the exponential in (3.12) decays at least as fast as $\exp(-cN\ln N)$, while the binomial factor is bounded by

$$\binom{N}{\delta N} \leqslant \exp\{N(\delta\,|\ln\delta| + (1-\delta)\,|\ln(1-\delta)|)\}$$

from which (3.10) follows. Condition (3.11) is proven in just the same way, using part (iii) of Lemma 5. ∎

We turn now to the bounds on the maximal eigenvalue of the matrices E_I. The following lemma provides the basis for our later probabilistic bounds:

Lemma 7. If E_I has a maximal eigenvalue λ_0, then there exists $i\in I$ such that

$$\sum_{j\in I^c} \varepsilon_{ij} + \sum_{l\in I/\{i\}} \sum_{j\in I^c} \varepsilon_{ij}\varepsilon_{jl} \geqslant \lambda_0^2 \tag{3.13}$$

Proof. Note that all matrix elements of E_I are nonnegative. Therefore, by the Perron–Frobenius theorem, there exists an eigenvector with only nonnegative components corresponding to the maximal eigenvalue λ_0. If we denote the components of this vector by v_i and take into account the fact that the matrix E_I^2 is block diagonal with respect to the index sets I and I^c, the eigenvalue equation $E_I^2\mathbf{v} = \lambda_0^2\mathbf{v}$ implies that for all $i\in I$,

$$
\begin{aligned}
\lambda_0^2 v_i &= \sum_{l\in I}\sum_{j\in I^c} \varepsilon_{ij}\varepsilon_{jl} v_l \\
&= \sum_{j\in I^c} \varepsilon_{ij} v_i + \sum_{l\in I/\{i\}}\sum_{j\in I^c} \varepsilon_{ij}\varepsilon_{jl} v_l \\
&\leqslant \sum_{j\in I^c} \varepsilon_{ij} v_i + \max_{l\in I}(v_l) \sum_{l\in I/\{i\}}\sum_{j\in I^c} \varepsilon_{ij}\varepsilon_{jl}
\end{aligned}
\tag{3.14}
$$

In particular, for the $i \in I$ for which $v_i = \max_{l \in I}(v_l)$, this gives

$$\lambda_0^2 \leqslant \sum_{j \in I^c} \varepsilon_{ij} + \sum_{l \in I/\{i\}} \sum_{j \in I^c} \varepsilon_{ij} \varepsilon_{jl} \tag{3.15}$$

which proves Lemma 7. ∎

Lemma 8. Assume $p \geqslant c[(\ln N)/N]^{1/2}$ for any constant $c > 0$. There exists $a > 0$, depending on c, such that

$$\mathbb{P}[\exists_{I:|I| = \delta N}\, \lambda_0^2(E_I) \geqslant p^2 N^2 \delta(1-\delta)(1+a)] \downarrow 0 \qquad \text{as} \quad N \uparrow \infty \tag{3.16}$$

Moreover, the constant a may be chosen in the following ways:

(i) If $p^2 N/\ln N \uparrow \infty$, any a such that $(1-\delta)(1+a) > 1$ suffices.

(ii) If $Np^2 = C \ln N$, with C sufficiently large (~ 7), a may be chosen less than $\sim \frac{1}{2}$.

(iii) Otherwise, we must choose $a > 1 + 2/\sqrt{c}$.

The rate of convergence in (3.16) is faster than any power of $1/N$ in case (i) and can be made as fast as any desired power of $1/N$ in cases (ii) and (iii) by appropriately choosing a large enough.

Proof. First observe that by Lemma 7

$$\begin{aligned} &\mathbb{P}[\exists_{I:|I| = \delta N}\, \lambda_0^2(E_I) \geqslant c] \\ &\quad \leqslant \mathbb{P}\left[\exists_{I:|I| = \delta N} \exists_{i \in I} \left(\sum_{j \in I^c} \varepsilon_{ij} + \sum_{l \in I/\{i\}} \sum_{j \in I^c} \varepsilon_{ij} \varepsilon_{jl} \right) \geqslant c \right] \\ &\quad \leqslant \mathbb{P}\left[\exists_{I:|I| = \delta N} \exists_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \geqslant \rho c \right] \\ &\qquad + \mathbb{P}\left[\exists_{I:|I| = \delta N} \exists_{i \in I} \sum_{l \in I/\{i\}} \sum_{j \in I^c} \varepsilon_{ij} \varepsilon_{jl} \geqslant (1-\rho) c \right] \end{aligned} \tag{3.17}$$

where $0 < \rho < 1$ can be chosen arbitrarily. To bound the first term in (3.17), notice simply that

$$\begin{aligned} \mathbb{P}\left[\exists_{I:|I| = \delta N} \exists_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \geqslant \rho c \right] &\leqslant \mathbb{P}\left[\exists_{i \in \{1,\ldots,N\}} \exists_{I: i \in I} \sum_{j \in I^c} \varepsilon_{ij} \geqslant \rho c \right] \\ &\leqslant N \mathbb{P}\left[\exists_{I: 1 \in I} \sum_{j \in I^c} \varepsilon_{1j} \geqslant \rho c \right] \\ &\leqslant N \mathbb{P}\left[\sum_{j=2}^{N} \varepsilon_{1j} \geqslant \rho c \right] \end{aligned} \tag{3.18}$$

We may write

$$\rho p^2 N^2 \delta(1-\delta)(1+a) = pN\{1+[\rho\delta(1-\delta)(1+a)\,pN-1]\}$$

and use case (ii) from Lemma 5 to get that

$$\begin{aligned}\mathbb{P}\left[\exists_{I:|I|=\delta N}\exists_{i\in I}\sum_{j\in I^c}\varepsilon_{ij}\geqslant \rho p^2N^2\delta(1-\delta)(1+a)\right]\\ \leqslant N\exp\{-pN[\rho(1+a)\,\delta(1-\delta)\,pN-2]\}\end{aligned}\tag{3.19}$$

which for any ρ, a, and δ positive converges to zero at least as fast as $e^{-cN\ln N}$ under our assumptions on p.

We now turn to the second term in (3.17). Here we use that

$$\begin{aligned}&\mathbb{P}\left[\exists_{I:|I|=\delta N}\exists_{i\in I}\sum_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl}\geqslant(1-\rho)c\right]\\ &\quad\leqslant\mathbb{P}\left[\exists_{I:|I|=\delta N}\exists_{i\in I}\exists_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl}\geqslant\frac{(1-\rho)c}{\delta N}\right]\\ &\quad\leqslant\mathbb{P}\left[\exists_{i\in\{1,\dots,N\}}\exists_{l\neq i}\exists_{I:i,l\in I}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl}\geqslant\frac{(1-\rho)c}{\delta N}\right]\\ &\quad\leqslant N(N-1)\,\mathbb{P}\left[\sum_{j=3}^{N}\varepsilon_{1j}\varepsilon_{j2}\geqslant\frac{(1-\rho)c}{\delta N}\right]\end{aligned}\tag{3.20}$$

The variables $\varepsilon_j\equiv\varepsilon_{1j}\varepsilon_{j2}$ are i.i.d.r.v.'s that take the values 1 with probability p^2 and 0 with probability $1-p^2$. Lemma 5 can thus again be used to estimate the last probability in (3.20). With $c=(1+a)\,\delta(1-\delta)\,p^2N^2$ and Y defined as $Y=(1+a)(1-\delta)(1-\rho)-1$

$$\mathbb{P}\left[\sum_{j=3}^{N}\varepsilon_{1j}\varepsilon_{j2}\geqslant(1-\rho)(1+a)(1-\delta)\,p^2N\right]\leqslant\exp\left\{-Np^2\left[\frac{Y^2}{2}-\frac{Y^3}{6}e^Y\right]\right\}\tag{3.21}$$

if $Y\leqslant 1$, and

$$\mathbb{P}\left[\sum_{j=3}^{N}\varepsilon_{1j}\varepsilon_{j2}\geqslant(1-\rho)(1+a)(1-\delta)\,p^2N\right]\leqslant\exp\{-Np^2[Y-1]\}\tag{3.22}$$

if $Y>1$. As we have seen, ρ can be chosen as close to zero as we wish, and we really need a result for very small δ, so that Y is essentially equal to a. In order to show that (3.20) goes to zero as $N\uparrow\infty$, we need to choose Y sufficiently big, such that either (3.21) or (3.22) goes to zero faster than N^2.

This will of course depend on the behavior of p. We distinguish three regimes:

(i) $p^2N/\ln N \uparrow \infty$. In this case we may choose Y as small as we wish and still have our probability go to zero faster than any power.

(ii) $p^2N \geqslant C \ln N$, where $C \approx 7$. In this case Y can still be chosen smaller than 1, in particular smaller than the value that maximizes $Y^2/2 - Y^3/6e^Y$.

(iii) $p^2N = C \ln N$, with $C < 7$. Here we need to use (3.22) and must choose a such that $Y > 1 + 2/C$.

The speed of convergence to zero is easily read off (3.22). This concludes the proof of Lemma 8. ∎

Lemma 8, and in particular the estimate (3.20), is the reason why we need to restrict the dilution rate p to be at least of the order $(\ln N/N)^{1/2}$. One may wonder whether this is intrinsic or an artefact of our estimates. Now, the criterion furnished by Lemma 7 yields rather reasonable bounds on the largest eigenvalue under the conditions of Lemma 8, i.e., they are of the order of the largest eigenvalue of the averaged matrix. On the other hand, the estimates of Lemma 8 cannot be substantially improved, as is shown by the following lemma.

Lemma 9. Assume that $p < \delta/\sqrt{N}$. Then, for all c such that $c \leqslant (1-x)\,p^2N^2$, with $x > 0$,

$$\mathbb{P}\left[\exists_{I:|I|=\delta N}\,\exists_{i\in I}\sum_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl} \geqslant c\right]\uparrow 1 \qquad \text{as} \quad N\uparrow\infty \tag{3.23}$$

Proof. Note that

$$\begin{aligned}&\mathbb{P}\left[\exists_{I:|I|=\delta N}\,\exists_{i\in I}\sum_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl} \geqslant c\right]\\ &\qquad = 1 - \mathbb{P}\left[\forall_{I:|I|=\delta N}\,\forall_{i\in I}\sum_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl} \leqslant c\right]\end{aligned} \tag{3.24}$$

Now

$$\begin{aligned}&\mathbb{P}\left[\forall_{I:|I|=\delta N}\,\forall_{i\in I}\sum_{l\in I/\{i\}}\sum_{j\in I^c}\varepsilon_{ij}\varepsilon_{jl} \leqslant c\right]\\ &\qquad \leqslant \mathbb{P}\left[\forall_{I:|I|=\delta N,\,1\in I}\sum_{l\in I/\{1\}}\sum_{j\in I^c}\varepsilon_{1j}\varepsilon_{jl} \leqslant c\right]\end{aligned} \tag{3.25}$$

Now the ε_{1j} are independent from all the other ε_{jl} that appear in the last expression, and therefore we may condition on the given realization of the ε_{1j} and sum over all realizations. This gives

$$\begin{aligned}&\mathbb{P}\left[\forall_{I:|I|=\delta N,\, 1\in I} \sum_{l\in I/\{1\}} \sum_{j\in I^c} \varepsilon_{1j}\varepsilon_{jl} \leqslant c\right]\\&\quad=\sum_{\{\varepsilon_{1,j}\}} \mathbb{P}\left[\forall_{I:|I|=\delta N,\, 1\in I} \sum_{l\in I/\{1\}} \sum_{j\in I^c} \varepsilon_{1j}\varepsilon_{jl} \leqslant c \,\middle|\, \{\varepsilon_{1j}\}\right] \mathbb{P}[\{\varepsilon_{1j}\}] \qquad (3.26)\end{aligned}$$

The sum over the $\{\varepsilon_{1j}\}$ can be seen as a sum over all sets $J \subset \{2,..., N\}$ on which $\varepsilon_{1j}=1$ [we will see that the conditioned probability in (3.26) depends only on the cardinality of that ensemble]. Identifying for simplicity the set J and the event $\{\varepsilon_{1j}=1$ if and only if $j\in J\}$, the last line in (3.26) can be written as

$$\sum_J \mathbb{P}\left[\forall_{I:|I|=\delta N,\, 1\in I} \sum_{l\in I/\{1\}} \sum_{j\in J\cap I^c} \varepsilon_{jl} \leqslant c \,\middle|\, J\right] \mathbb{P}[J] \qquad (3.27)$$

and, defining the random variables

$$A_l(J) \equiv \sum_{j\in J} \varepsilon_{jl} \qquad (3.28)$$

(3.27) is bounded by

$$\sum_J \mathbb{P}\left[\forall_{I:|I|=\delta N,\, 1\in I,\, I\subset J^c} \sum_{l\in I/\{1\}} A_l(J) \leqslant c \,\middle|\, J\right] \mathbb{P}[J] \qquad (3.29)$$

The conditioned probability in (3.29) can now be expressed as

$$\begin{aligned}&\mathbb{P}\left[\forall_{I:|I|=\delta N,\, 1\in I,\, I\subset J^c} \sum_{l\in I/\{1\}} A_l(J) \leqslant c \,\middle|\, J\right]\\&\quad=\sum_{r=0}^{|J^c|} \mathbb{P}\left[\forall_{I:|I|=\delta N, 1\in I, I\subset J^c} \sum_{l\in I/\leqslant 1\}} A_l(J) \leqslant c \,\middle|\, |\{l\in J^c : A_l(J)\geqslant 1\}|=r, J\right]\\&\quad\quad\times \mathbb{P}[|\{l\in J^c : A_l(J)\geqslant 1\}|=r \mid J] \qquad (3.30)\end{aligned}$$

We now choose $c\leqslant \delta N$. Then, for the terms where $r\geqslant c$, we always can find a set I including a subset of J^c on which the sum of the $A_l(J)$ exceeds c; thus, the corresponding probability is zero. If, on the other hand, $r\leqslant c$, then taking a set I containing all the index in J^c where the $A_l(J)$ are greater or equal to 1, we get the bound

$$\mathbb{P}\left[\forall_{I:|I|=\delta N,\,1\in I,\,I\subset J^c} \sum_{l\in I/\{1\}} A_l(J)\leqslant c \,\middle|\, J\right]$$
$$=\sum_{r\leqslant c} \mathbb{P}\left[\sum_{l\in J^c/\{1\}} A_l(J)\leqslant c \,\middle|\, |\{l\in J^c: A_l(J)\geqslant 1\}|=r, J\right]$$
$$\times \mathbb{P}[|\{l\in J^c: A_l(J)\geqslant 1\}|=r\,|\,J]$$
$$\leqslant \mathbb{P}\left[\sum_{l\in J^c/\{1\}} A_l(J)\leqslant c \,\middle|\, J\right] \tag{3.31}$$

Now

$$\mathbb{P}\left[\sum_{l\notin J} A_l(J)\leqslant c \,\middle|\, J\right]=\mathbb{P}\left[\sum_{l=2}^{\delta N}\sum_{j=\delta N+1}^{\delta N+m} \varepsilon_{jl}\leqslant c\right] \tag{3.32}$$

and inserting all this into (3.26) yields

$$\mathbb{P}\left[\forall_{I:|I|=\delta N,\,1\in I} \sum_{l\in I/\{1\}}\sum_{j\in I^c} \varepsilon_{1j}\varepsilon_{jl}\leqslant c\right]$$
$$\leqslant \sum_{m=0}^{N} \mathbb{P}\left[\sum_{l=m+2}^{N}\sum_{j=2}^{m+1} \varepsilon_{jl}\leqslant c\right]\mathbb{P}\left[\sum_{j=1}^{N} \varepsilon_{1j}=m\right] \tag{3.33}$$

The probabilities appearing in the last equation may all be estimated easily using the bounds from Lemma 5. Note that $\mathbb{P}[\sum_{j=1}^{N} \varepsilon_{1j}=m]$ is concentrated at $m=pN$ and that the random variable $\sum_{l=m+2}^{N}\sum_{j=2}^{m+1} \varepsilon_{jl}$ has mean $pm(N-m)$ with Gaussian fluctuations. Therefore, if c is chosen as $c=(1-x)p^2N^2$, with $x>0$, the sum in (3.33) goes to zero as $N\uparrow\infty$ and the lemma is proven. ∎

One may prove that the probability in (3.23) goes to zero if $c>p^2N^2$. However, such a bound does not suffice to obtain a sufficiently good estimate on $\mathbb{P}[H(\xi_I^\mu)\leqslant H(\xi^\mu)+\varepsilon N]$ to compensate for the exponential number of terms in (2.1). One might think that such a large eigenvalue is realized only for a small subset of the possible I, but checking through the proof of Lemma 9 will also convince one that the number of sets I that contribute in (3.23) is still exponentially large once p^2N^2 is small compared to δN. We take these results thus as an indication that there is a real transition in the functioning of the dilute Hopfield model occurring for p of the order of $1/\sqrt{N}$. It would certainly be interesting (although maybe difficult) to have a numerical check of this conjecture.

4. MINIMA ASSOCIATED TO MIXED PATTERNS

Besides the minima located near the original patterns, the Hamiltonian (2.2) possesses other local minima corresponding to mixtures of finitely many original patterns which are also surrounded by extensive energy barriers. Such minima have been discussed by Amit *et al.*[1] and their existence in the standard Hopfield model was proven by Newman.[18] In this section we extend his results to the dilute model.

For a given m-vector $\mathbf{v}$ let us denote by $\xi(\mathbf{v})$ the vector with components

$$\xi_i(\mathbf{v}) = \operatorname{sign}\{\mathbf{v}_1 \xi_i^1 + \cdots + \mathbf{v}_m \xi_i^m\} \tag{4.1}$$

We assume that[18]:

(1) The number D of nonzero components of $\mathbf{v}$ is finite (i.e., independent of N).

(2) $\pm \mathbf{v}_1 \pm \mathbf{v}_2 \cdots \pm \mathbf{v}_m \neq 0$, for any choice of the $\pm$.

(3) For all i and all s, $\mathbb{E}\xi_i^s \xi_i(\mathbf{v}) = \mathbf{v}_s$.

For any such $\mathbf{v}$ we have the following result:

Theorem 2. Suppose p satisfies the hypothesis of Theorem 1. Let $\mathbf{v}$ be a vector satisfying conditions (1)–(3). Then there exists $\alpha_c \geqslant 0$, depending on the number D of nonzero components of $\mathbf{v}$ and on the lower bound on the quantities appearing in (2) and on p, such that if $m \leqslant \alpha_c pN$, then there exists $\varepsilon > 0$ and $0 < \delta < \frac{1}{2}$ such that there exists $\gamma > 0$ such that

$$\mathbb{P}_M[\mathbb{P}_\xi[\{h(\xi(\mathbf{v}), \delta) > H(\xi(\mathbf{v})) + \varepsilon N\}] \geqslant 1 - e^{-\gamma N}] \to 1 \qquad \text{as} \quad N \uparrow \infty \tag{4.2}$$

where the convergence in (4.1) is exponentially fast in N. Moreover, the dependence of α_c on p is of the same nature as in Theorem 1.

Remark. If D is fixed as well as a finite lower bound on the moduli of sums appearing in condition (2), Theorem 2 may be slightly strengthened to state that the probability that the event in (4.2) holds *for all* such $\mathbf{v}$ tends to one.

Remark. One may of course also extract almost sure convergence statements as in the remark following Theorem 1.

Proof. The proof of Theorem 2 is quite similar to that of Theorem 1. We have to estimate from above the probabilities

$$\mathbb{P}_\xi[H(\xi_I(\mathbf{v})) - H(\xi(\mathbf{v})) \leqslant \varepsilon N] \tag{4.3}$$

For simplicity and without loss of generality we can assume that the first D components of $\mathbf{v}$ are those that are different from zero. Defining $\mathbf{y}^s \equiv \xi^s \xi(\mathbf{v})$, and repeating the calculations leading to (2.6), we may write

$$H(\xi(\mathbf{v})) - H(\xi_I(\mathbf{v})) = \frac{2}{pN}\left\{\sum_{s=1}^{D} (\mathbf{y}^s, E_I \mathbf{y}^s) + \sum_{s=D+1}^{m} (\mathbf{y}^s, E_I \mathbf{y}^s)\right\} \tag{4.4}$$

We have distinguished the terms with $s \leqslant D$ and $s > D$, since for the latter, the y_i^s form a family

$$\{y_i^s\}_{\substack{i=1,\dots,N\\ s=D+1,\dots,m}}$$

of i.i.d. random variables with $\mathbb{P}_\xi[y_i^s = \pm 1] = 1/2$. Note that, on the other hand, for $s \leqslant D$, we have $\mathbb{E} y_i^s = \mathbf{v}_s$.

Now clearly, for any constant c (to be chosen later), we have that

$$\begin{aligned}
&\mathbb{P}_\xi[H(\xi_I(\mathbf{v})) - H(\xi(\mathbf{v})) \leqslant \varepsilon N] \\
&\quad = \mathbb{P}_\xi\left[\sum_{s=1}^{D} (\mathbf{y}^s, E_I \mathbf{y}^s) + \sum_{s=D+1}^{m} (\mathbf{y}^s, E_I \mathbf{y}^s) \leqslant \varepsilon p N^2\right] \\
&\quad \leqslant \mathbb{P}_\xi\left[\sum_{s=D+1}^{m} (\mathbf{y}^s, E_I \mathbf{y}^s) \leqslant (\varepsilon - c\delta)\, pN^2\right] + \mathbb{P}_\xi\left[\sum_{s=1}^{D} (\mathbf{y}^s, E_I \mathbf{y}^s) \leqslant c\delta p N^2\right]
\end{aligned} \tag{4.5}$$

The estimation of the first probability in (4.5) proceeds exactly as in Section 2, We concentrate thus on the second term. Notice that

$$\begin{aligned}
\sum_{s=1}^{D} (\mathbf{y}^s, E_I \mathbf{y}^s) &= \sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} y_i^s y_j^s \\
&= \sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} y_i^s \mathbf{v}_s + \sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} y_i^s (y_j^s - \mathbf{v}_s)
\end{aligned} \tag{4.6}$$

Now note that for the first term we get

$$\begin{aligned}
\sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} y_i^s \mathbf{v}_s &= \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \left(\sum_{s=1}^{D} \xi_i^s \mathbf{v}_s\right) \xi_i(\mathbf{v}) \\
&= \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \left|\sum_{s=1}^{D} \xi_i^s \mathbf{v}_s\right| \\
&\geqslant K \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} = \frac{K}{2} \operatorname{tr} E_I^2
\end{aligned} \tag{4.7}$$

for some $K > 0$. Here we have used the definition of $\xi_i(\mathbf{v})$, and in the last line property (2) of the vector $\mathbf{v}$. Note that the constant K depends on $\mathbf{v}$. Since Lemma 6 provides sharp uniform bounds in probability for $\operatorname{tr} E_I^2$, we are done as far as this term is concerned.

For the second term in (4.6) we write again

$$\sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} y_i^s (y_j^s - \mathbf{v}_s)$$
$$= \sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \mathbf{v}_s (y_j^s - \mathbf{v}_s) + \sum_{s=1}^{D} \sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} (y_i^s - \mathbf{v}_s)(y_j^s - \mathbf{v}_s) \tag{4.8}$$

Then

$$\mathbb{P}_\xi \left[\sum_{s=1}^{D} (\mathbf{y}^s, E_I \mathbf{y}^s) \leqslant c\delta p N^2 \right]$$
$$\leqslant \sum_{s=1}^{D} \mathbb{P}_\xi \left[\sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} \mathbf{v}_s (y_j^s - \mathbf{v}_s) \leqslant \frac{c\delta p N^2 - K \operatorname{tr} E_I^2}{2D} \right]$$
$$+ \mathbb{P}_\xi \left[\sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} (y_i^s - \mathbf{v}_s)(y_j^s - \mathbf{v}_s) \leqslant \frac{c\delta p N^2 - K \operatorname{tr} E_I^2}{2D} \right] \tag{4.9}$$

Now, we may place ourselves in the subspace where conditions (C1) and (C2) from Section 2 hold and therefore use the uniform bound $\operatorname{tr} E_I^2 \geqslant (1-x)\, pN^2 \delta(1-\delta)$. Using the exponential Markov inequality, we get, with $K' \equiv (1-x)(1-\delta) K$,

$$\mathbb{P}_\xi \left[\sum_{i \in I} \sum_{j \in I^c} \varepsilon_{ij} (y_i^s - \mathbf{v}_s)(y_j^s - \mathbf{v}_s) \leqslant \frac{(c-K')\, \delta p N^2}{2D} \right]$$
$$\leqslant \inf_{t \geqslant 0} \exp\left\{ t \frac{(c-K')\, \delta p N^2}{2D} \right\}$$
$$\times \mathbb{E}_{\mathbf{y}^s} \exp\{ -t((\mathbf{y}^s - \mathbf{v}_s \mathbf{1}), E_I(\mathbf{y}^s - \mathbf{v}_s \mathbf{1})) \}$$
$$\leqslant \inf_{t \geqslant 0} \exp\left\{ t \frac{(c-K')\, \delta p N^2}{2D} \right\} \mathbb{E}_{\mathbf{z}} \exp\{ -t(1+\mathbf{v}_s)^2 (\mathbf{z}, E_I \mathbf{z}) \tag{4.10}$$

where $\mathbb{E}_{\mathbf{z}}$, as in Lemma 1, denotes the expectation with respect to a family of independent standard Gaussian random variables. In fact, (4.10) uses an immediate generalization of Lemma 1, where use is made of the following simple observation:

Lemma 10. Let y be a random variable taking the values ± 1 with mean v. Then

$$\mathbb{E} \exp\{(y-v)t\} \leqslant \exp\left\{ \frac{t^2}{2} (1+v)^2 \right\} \tag{4.11}$$

Proof. Just notive that

$$\mathbb{E}e^{(y-v)t} = \frac{v+1}{2}e^{t(1-v)} + \frac{1-v}{2}e^{-t(1+v)}$$

$$= \cosh(t(1+v)) + \frac{e^t}{2}(-2\sinh(vt) + ve^{-vt} - ve^{-(2+v)t)})$$

$$\leqslant \cosh(t(1+v)) \leqslant e^{(t^2/2)(1+v)^2} \quad ∎ \tag{4.12}$$

The estimation of the probability in (4.10) now proceeds as in Section 2. Using Lemma 2, and recalling that $T = (1/2p)[(1-\delta)/2(1+a)]^{1/2}$, we get that

$$\mathbb{P}_\xi\left[\sum_{i\in I}\sum_{j\in I^c}\varepsilon_{ij}(y_i^s-\mathbf{v}_s)(y_j^s-\mathbf{v}_s) \leqslant \frac{(c-K')\,\delta pN^2}{2D}\right]$$

$$\leqslant \inf_{0\leqslant t\leqslant T}\exp\left\{-t\frac{(K'-cx)\,\delta pN^2}{2DN\sqrt{\delta(1-\delta)}} + 2t^2(1+\mathbf{v}_s)^2\frac{\operatorname{tr}E_I^2}{N^2\delta(1-\delta)}\right\}$$

$$\leqslant \inf_{0\leqslant t\leqslant T}\exp\left\{-t\frac{K'-c}{2D}\left(\frac{\delta}{1-\delta}\right)^{1/2}pN + 2(1+\mathbf{v}_s)^2t^2p(1+x)\right\}$$

$$\leqslant \exp\left\{-\frac{K'-c}{2D}\left(\frac{\delta}{2(1+a)}\right)^{1/2}N + (1+\mathbf{v}_s)^2\frac{1}{2p}\frac{1+x}{(1-\delta)\,2(1+a)}\right\} \tag{4.13}$$

where we have put $t = T$ in the last line.

Finally we must deal with the first summand in (4.9). Using again the exponential Markov inequality together with Lemma 10, we get

$$\mathbb{P}_\xi\left[\sum_{i\in I}\sum_{j\in I^c}\varepsilon_{ij}\mathbf{v}_s(y_j^s-\mathbf{v}_s) \leqslant \frac{c\delta pN^2 - K\operatorname{tr}E_I^2}{2D}\right]$$

$$\leqslant \inf_{t\geqslant 0}\exp\left\{-t\frac{(K'-c)\,\delta pN^2}{2DN}\right\}\prod_{j\in I^c}\mathbb{E}_{y_j}\exp\left\{t\sum_{i\in I}\varepsilon_{ij}\mathbf{v}_s(y_j-\mathbf{v}_s)\right\}$$

$$\leqslant \inf_{t\geqslant 0}\exp\left\{-t\frac{(K'-c)\,\delta pN^2}{2DN} + \frac{t^2}{2}\mathbf{v}_s^2(1+\mathbf{v}_s)^2\sum_{h\in I^c}\left(\sum_{u\in I}\varepsilon_{ij}\right)^2\right\}$$

$$\leqslant \exp\left\{-\frac{(K'-c)^2\,p^2N^4\delta^2}{8D^2\mathbf{v}_s^2(1+\mathbf{v}_s)^2\sum_{j\in I^c}(\sum_{i\in I}\varepsilon_{ij})^2}\right\} \tag{4.14}$$

Now if

$$\sum_{j\in I^c}\left(\sum_{i\in I}\varepsilon_{ij}\right)^2 \leqslant b\delta N^3p^2$$

for some constant b, then the probability in (4.14) is bounded by

$$\mathbb{P}_\xi\left[\sum_{i\in I}\sum_{j\in I^c}\varepsilon_{ij}\mathbf{v}_s(y_j^s-\mathbf{v}_s)\leqslant\frac{c\delta pN^2-K\operatorname{tr}E_I^2}{2D}\right]$$
$$\leqslant\exp\left\{-\frac{(K'-c)^2N}{8D^2b\mathbf{v}_s^2(1+\mathbf{v}_s)^2}\right\} \tag{4.15}$$

which will suffice. For, combining (4.9), (4.13), and (4.15), we arrive at

$$\mathbb{P}_\xi\left[\sum_{i\in I}\sum_{j\in I^c}\varepsilon_{ij}(y_i^s-\mathbf{v}_s)(y_j^s-\mathbf{v}_s)\leqslant\frac{(c-K')\,\delta pN^2}{2D}\right]$$
$$\leqslant D\exp\left\{-\frac{K'-c}{2D}\left(\frac{\delta}{2(1+a)}\right)^{1/2}N+\frac{2}{p}\frac{1+x}{(1-\delta)\,2(1+a)}\right\}$$
$$D\exp\left\{-\frac{(K'-c)^2N}{32D^2b}\right\} \tag{4.16}$$

Here we have used the trivial bound $\mathbf{v}_s\leqslant 1$. It is clear that for c chosen small enough (in dependence on K') and for δ small enough (depending again on K' and on D), this bound can be summed over all I of size δN and the sum still converges to zero exponentially fast. For the first term in (4.5) one obtains, following exactly the calculations of Section 2, a bound

$$\mathbb{P}_\xi\left[\sum_{s=D+1}^{m}(\mathbf{y}^s,E_I\mathbf{y}^s)\leqslant(\varepsilon-c\delta)\,pN^2\right]\leqslant\exp\left\{-N\frac{\gamma}{2\alpha}\delta\left[c-\frac{\varepsilon}{\delta}\right]^2\right\} \tag{4.17}$$

For c and δ chosen according to the needs of (4.16) one may now choose ε and α small enough to render this probability summable over the I. Thus, all that is needed to complete the proof of Theorem 2 is the estimate leading to Eq. (4.15). It is provided by the next lemma.

Lemma 11. Assume that $p\geqslant c(\ln N/N)^{1/2}$. Then there exists a constant $b>1$ such that

$$\mathbb{P}_M\left[\exists_{I:|I|=\delta N}\sum_{jI^c}\left(\sum_{i\in I}\varepsilon_{ij}\right)^2\geqslant b\delta^2N^3p^2\right]\downarrow 0 \qquad \text{as} \quad N\uparrow\infty \tag{4.18}$$

where the convergence is exponentially fast in N.

Proof. The proof of this lemma is again an application of the exponential Markov inequality and of Lemma 5. Note that

$$\mathbb{P}_M\left[\sum_{j\in I^c}\left(\sum_{i\in I}\varepsilon_{ij}\right)^2\geqslant b\delta^2N^3p^2\right]$$
$$\leqslant\inf_{t\geqslant 0}\exp(-N^3p^2\delta^2bt)\prod_{j\in I^c}\mathbb{E}\exp\left\{t\left(\sum_{i\in I}\varepsilon_{ij}\right)^2\right\} \tag{4.19}$$

To bound the Laplace transform in (4.19), we write

$$\begin{aligned}
&\mathbb{E}\exp\left\{t\left(\sum_{i\in I}\varepsilon_{ij}\right)^2\right\}\\
&\quad=\int e^{tx^2}\mathbb{P}_M\left[\sum_{i\in I}\varepsilon_{ij}\in dx\right]\\
&\quad=\int_0^{2\delta pN} e^{tx^2}\mathbb{P}_M\left[\sum_{i\in I}\varepsilon_{ij}\in dx\right]+\int_{2\delta pN}^{\delta N} e^{tx^2}\mathbb{P}_M\left[\sum_{i\in I}\varepsilon_{ij}\in dx\right] \qquad (4.20)
\end{aligned}$$

where $\mathbb{P}_M[\sum_{i\in I}\varepsilon_{ij}\in dx]$ is understood to be the probability distribution of the variable $x=\sum\varepsilon_{ij}$. The two integrals correspond to the different bounds in Lemma 5 and will be treated separately. Our foremost concern is the second one, as it will impose restrictions on the permissible choices of t. Note that using (3.2) from Lemma 5, we get

$$\begin{aligned}
\int_{2\delta pN}^{\delta N} e^{tx^2}\,d\mathbb{P}\left[\sum_{i\in I}\varepsilon_{ij}=x\right]&\leqslant\int_{2\delta pN}^{\delta N} e^{tx^2}\,\mathbb{P}\left[\sum_{i\in I}\varepsilon_{ij}\geqslant x\right]\\
&\leqslant\int_{2\delta pN}^{\delta N} e^{tx^2-x+\delta pN(s+e)/6} \qquad (4.21)
\end{aligned}$$

We will bound this integral only for a restricted set of t values [which, however, will be seen to suffice to get a good bound in (4.20)]. For simplicity, we require that

$$tx^2\leqslant c^*x \qquad \text{and} \qquad (1-c^*)x-\delta pN\frac{8+e}{6}\geqslant 0 \qquad (4.22)$$

for all x in the range of integration in (4.21). This can be achieved by setting $c^*\equiv(4-e)/12$ and $t\leqslant c^*/\delta N$. Then

$$\begin{aligned}
&\int_{2\delta pN}^{\delta N}\exp\left(tx^2-x+\delta pN\frac{8+e}{6}\right)\\
&\qquad\leqslant\int_{2\delta pN}^{\infty}\exp\left\{-(1-c^*)x+\delta pN\frac{8+e}{6}\right\}\leqslant\frac{1}{1-c^*} \qquad (4.23)
\end{aligned}$$

Given the fact that we needed to restrict t to such a small range, there is no advantage to be gained over the following trivial estimate for the first integral in (4.20):

$$\int_0^{2\delta pN} e^{tx^2}\,d\mathbb{P}\left[\sum_{i\in I}\varepsilon_{ij}=x\right]\leqslant e^{t4\delta^2p^2N^2} \qquad (4.24)$$

Inserting these two bounds into (4.19) and putting $t=c^*/\delta N$, we get

$$\begin{aligned}\mathbb{P}_M&\left[\sum_{j\in I^c}\left(\sum_{i\in I}\varepsilon_{ij}\right)^2\geqslant b\delta^2N^3p^2\right]\\&\leqslant e^{-N^2p^2\delta b}\left[\frac{1}{1-c^*}+e^{4c^*\delta Np^2}\right]^{N(1-\delta)}\\&\leqslant e^{-N^2p^2c^*\delta(b-4)}\end{aligned}\tag{4.25}$$

Since under our assumptions $N^2p^2\sim N\ln N$, this probability is summable over the I and tends to zero exponentially fast, provided b is large enough. This concludes the proof of Lemma 11 and of Theorem 2. ∎

5. GENERALIZATIONS

So far we have considered a model where dendritic connections between two neurons are present or absent, but all existing dendrites have the same "strength" and connect the pair of neurons in a symmetric way, a fact expressed in the symmetry of the matrix M. In practical situations, it may be appropriate to drop these assumptions and to replace the matrix M by a more general matrix. Our results can easily be generalized to such models.

Let us first discuss the assumption of symmetry. Using the representation (1.3) of the Hamiltonian, we see that

$$\begin{aligned}H_N(\boldsymbol{\sigma})&=-\frac{1}{pN}\sum_{\mu=1}^m(\xi^\mu\boldsymbol{\sigma},M\xi^\mu\boldsymbol{\sigma})\\&=-\frac{1}{pN}\sum_{\mu=1}^m\left(\xi^\mu\boldsymbol{\sigma},\frac{1}{2}(M+M^t)\,\xi^\mu\boldsymbol{\sigma}\right)\end{aligned}\tag{5.1}$$

where $\tilde{M}\equiv\frac{1}{2}(M+M^t)$ is again symmetric. The study of networks with nonsymmetric matrices is thus reduced to the study of symmetric ones with different distributions of the matrix elements. For example, if we consider the completely asymmetric dilute model where M is the matrix all of whose elements are independent random variables ε_{ij} taking values 0 and 1 with probabilities p and $1-p$, then this is equivalent to the model with symmetric matrix $\tilde{M}$ where for $i>j$ the elements $\tilde{\varepsilon}_{ij}$ are i.i.d.r.v.'s that take the values 0, 1/2, and 1 with probabilities $(1/p)^2$, $2p(1-p)$, and p^2, respectively.

It is easy to verify that all our results can be proven if ε_{ij} are random variables with support on [0, 1] and mean p, For under these assumptions

$$\mathbb{E}e^{t\varepsilon_{ij}}\leqslant p(e^t-1)+1$$

which is all that is required in Section 3. The requirement that the ε_{ij} have bounded support on the positive half-line may even be slightly relaxed, but we see to immediate application for the cases tractable. On the other hand, if ε_{ij} is allowed to be substantially negative, the situation changes drastically as we approach a spin-glass-like model where our results clearly would not apply.

ACKNOWLEDGMENTS

A.B. thanks the Centre de Physique Théorique, Marseille, and V.G. thanks the SFB 237 and the Institut für Mathematik der Ruhr-Universität Bochum for their warm hospitality, which made this collaboration possible.

REFERENCES

1. D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, *Phys. Rev. A* **32**:1007–1018 (1985); Storing infinite numbers of patterns in a spin glass model of neural networks, *Phys. Rev. Lett.* **55**:1530–1533 (1985).
2. D. J. Amit and A. Treves, *J. Phys. A* **21**:3155 (1988).
3. D. J. Amit, G. Parisi, and S. Nicolis, Neural potentials as stimuli for attractor neural networks, *Network* **1**:75–88 (1990).
4. B. Bollobás, *Random Graphs* (Academic Press, 1985).
5. Y. S. Chow and H. Teicher, *Probability Theory* (Springer, New York, 1978).
6. B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**:167 (1987).
7. E. Domany, J. L. van Hemmen, and K. Schulte, *Models of Neutral Netyworks* (Springer, 1990).
8. H. Englisch, Y. Xiao, and K. Yao, NTZ-preprint 13/1991 (Leipzig).
9. P. Ferrari, S. Martinez, and P. Picco, A lower bound for the memory capacity in the Potts–Hopfield model, *J. Stat. Phys.* **66**:1643 (1992).
10. V. Gayrard, The thermodynamic limit of the q-state Potts–Hopfield model with infinitely many patterns, *J. Stat. Phys.* **68**:977 (1992).
11. E. Goles and S. Martinez, *Neural and Automata Networks* (Kluwer, 1990).
12. J. J. Hopfield, Neural networks and physical systems with emergent collective computationalabilities, *Proc. Natl. Acad. Sci. USA* **79**:2554–2558 (1982).
13. J. Komlos, Convergence results in a autoassociative memory model, *Neural Networks* **1**:239–250 (1988).
14. D. Kree and A. Zippelius, Asymmetric dilute neural networks, in *Modls of Neural Networks*, E. Domany, J. L. van Hemmen, and K. Schulte, eds. (Springer, 1990).
15. H. Koch and J. Piasko, Some rigorous results on the Hopfield neural network model, *J. Stat. Phys.* **55**:903 (1989).
16. W. A. Little and G. L. Shaw, Analytic study of the memory capacity of neutral networks, *Math. Biosci.* **59**:281–290 (1978).
17. R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, The capacity of the Hopfield associative memory, *IEEE Trans. Inform. Theory* **33**:461–482 (1987).
18. Ch. M. Newman, Memory capacity in neural network models: Rigorous lower bounds, *Neural Networks* **1**:223–238 (1988).
19. P. Peretto and J. J. Niez, Long term memory storage capacity of multiconnected neural networks, *Biol. Cybernet.* **54**:53–63 (1986).