View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

JOURNAL OF COMPUTER AND SYSTEM SCIENCES 47, 350–373 (1993)

Effect of Connectivity in an Associative Memory Model

JÁNOS KOMLÓS*

Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903

AND

RAMAMOHAN PATURI

University of California, San Diego, La Jolla, California 92093

Received June 1, 1991; revised April 27, 1992

We investigate how geometric properties translate into functional properties in sparse networks of computing elements. Specifically, we determine how the eigenvalues of the interconnection graph (which in turn reflect connectivity properties) relate to the quantities, number of items stored, amount of error-correction, radius of attraction, and rate of convergence, in an associative memory model consisting of a sparse network of threshold elements or neurons. © 1993 Academic Press, Inc.

1. INTRODUCTION

In this paper, we consider the ability to recall an item from a partial description of its properties. Such an ability is referred to as *associative memory* or *content-addressable memory*. Of particular interest to us are the models of associative memory based on *distributed representation* of information. A good introduction to such distributed models can be found in Hinton and Anderson [7].

Models of memory based on distributed systems have been studied by various researchers (see Hinton and Anderson [7], Kohonen [10] and Kohonen [11] for history and references), and, more recently, by Hopfield [8]. A closely related model is studied by Little [14], and Little and Shaw [15]. Both use Hebb's rule [6] of learning through gradual change of correlations.

We are going to use Hopfield's model in this paper. We want to emphasize that we only intend to make a technical analysis of some properties of the dynamical systems involved and do not believe that this has much relevance to actual brain functioning.

We first make some preliminary remarks about systems exhibiting associative memory properties. Assume that a dynamical system has a large number of stable

* Work partially supported by the Hungarian National Foundation for Scientific Research #1905.

0022-0000/93 $5.00

Copyright © 1993 by Academic Press, Inc.
All rights of reproduction in any form reserved.

states with a substantial domain of attraction around them. That is, the system started at *any* state in the domain of attraction would converge to the stable state. We can then regard such a system as an associative memory. In this framework, stored items are represented by stable states, nearby states represent partial information given a suitable metric. The process of retrieving full information from partial information corresponds to a state in the domain of attraction converging to the stable state. One can think of associative memory as correcting errors in a noisy input.

Many times, full information is not obtainable. Often, we can relax the requirement that a stored item corresponds to a stable state. We merely require selected states to have large domains of attraction around them such that if we start anywhere in the domain, we will eventually get within a small distance from the stored item (*residual error* in recall). What is important is that we have a significant amount of error-correction.

Another desirable feature of such a system is a *learning* mechanism, by which the system adapts itself to remember new items. With this general picture in mind, we now look at the specific details of the Hopfield model.

The model consists of a system of *fully* interconnected neurons or linear threshold elements where each interconnection is symmetric and has a certain weight. Each neuron in the system can be in one of two states ± 1. The state of the entire system can be represented by an n-dimensional vector, where n is the number of neurons in the system. The components of the vector denote the states of the corresponding neurons. The weight of each interconnection is given by real numbers w_{ij} with $w_{ij} = w_{ji}$.

Each neuron updates its state based on whether a linear form of the current states of the other neurons, computed with the weights of the interconnections, is above or below its threshold value. We will assume in this paper that all thresholds are zero. Hence, with the system in state x, neuron i resets its state to $\operatorname{sgn}(\sum_{j \neq i} w_{ij} x_j)$, where the function sgn is defined as

$$\operatorname{sgn}(x) = \begin{cases} +1 & \text{if } x \geqslant 0 \\ -1 & \text{otherwise.} \end{cases}$$

We consider two modes of dynamic operation of the system. In the *synchronous* mode, at every time step, every neuron updates its state simultaneously. In the *asynchronous* mode, at any instance, at most one neuron can update its state with each neuron eventually getting its turn.

A state v is called *stable* if no transition out of it is possible. More precisely, for each i, $v_i = \operatorname{sgn}(\sum_{j \neq i} w_{ij} v_j)$. Note that the notion of stability does not depend on the mode of operation.

Hopfield described the (asynchronous) dynamics of the system using an energy surface. The energy of the system is given by the negative of the quadratic form associated with the weight matrix. More precisely, the *energy* $\mathscr{E}(x)$ of the system in state x is given by $-\frac{1}{2} \sum_{i,j} w_{ij} x_i x_j$.

Using this energy function, we have, for any two vectors x, v,

$$\mathscr{E}(x)-\mathscr{E}(v)=2\sum_{\substack{i\in D\\ j\notin D}} w_{ij}v_iv_j,$$

where D is the set of indices where x and v differ.

If x differs from v only in the ith coordinate then

$$\mathscr{E}(x)-\mathscr{E}(v)=2v_i\sum_{j\neq i} w_{ij}v_j.$$

This shows, on the one hand, that stable states are local minima of the energy landscape ($\mathscr{E}(x)-\mathscr{E}(v)\geqslant 0$), and on the other hand, that an asynchronous step does not increase the energy of the system. This guarantees that the system, when operated asynchronously, will eventually reach a stable state. (Note that the definition $\operatorname{sgn}(0)=+1$ guarantees that we cannot get into a cycle.) Such a convergence is not guaranteed in the case of synchronous operation.

Hopfield used the following Hebb type rule to select the weights of the interconnections. To store a vector in the system, we require that each interconnection remember the correlation of the states of the two neurons it interconnects. More precisely, we set the weights $w_{ij}=v_iv_j$ to remember a single vector $v=(v_1, v_2, ..., v_n)$. With this choice of weights, the system has a stable point at state v.

Moreover, for this choice of weights, when the system is started at a state x within a distance $n/2$ from the stored vector v, it gets into state v in one synchronous step. In the asynchronous mode of operation, the state of the system converges monotonically to the stable state v. (As is customary, we measure the distance between two ± 1 vectors or states by their Hamming distance: the number of components in which they differ.) It is this *attracting* nature of the system that gives it an *error-correcting* capability.

If we wish to store several vectors in the system, we add the corresponding weights. More precisely, if we want to store the vectors $v^1, v^2, ..., v^m$, the weight w_{ij} is defined as $w_{ij}=\sum_{k=1}^{m} v_i^k v_j^k$. The hope is that if the stored vectors are sufficiently different, such a linear addition of weights would not cause much interference in the error-correcting behavior of the system. We call each such stored vector a *fundamental* memory.

When we store a number of fundamental memories in the system, we expect each of them to be stable and to attract all the vectors within a ρn distance for some constant $\rho>0$. Or more generally, we consider the system to be error-correcting, if every vector within a distance ρn from a fundamental memory eventually ends up within a distance of εn for some $\varepsilon<\rho$. We call this εn *residual error*. We also consider the existence of a domain of attraction around each fundamental memory. We say that a state x has a domain of attraction of radius ρn if the system started in any state y within distance ρn from x, would eventualy converge to x (residual error is 0).

Another important characteristic of the system is the *rate of convergence*.

It should be emphasized that all the parameters considered so far may depend on whether the system is operated synchronously or asynchronously.

Given *any* set of m fundamental memories, we would like to store them in the system. But this requirement is somewhat at odds with the requirement of error-correction. For we cannot expect to store vectors which are too close to each other. (Closeness is not the only potential problem. If we require stability of the stored vectors, then, even for $m=4$, there exist m pairwise distant vectors that cannot be stored as fundamental memories, at least not by using the above storage method of Hopfield.) A reasonable minimal requirement is that we would like to store almost all sets of m vectors. Therefore, we will take a set of m *random vectors* as our set of fundamental memories and expect the system to remember them with *probability near* 1. This randomness is often achieved by coding the input.

When $m=1$, we have already seen that the fundamental memory is a stable state of the system which attracts all vectors within a distance $n/2$ in one synchronous step. When we have a number of fundamental memories, the retrieval of a memory wil be disturbed by the *noise* created by the other fundamental memories. Yet, we hope that this noise is not overwhelming when the number of fundamental memories is not too large. Hence, the main question is to determine the amount of error-correction and the rate of convergence as a function of the number m of fundamental memories.

Worst-Case and Random Errors. So far, we have considered the following error-correcting behavior: all the vectors within a certain distance of the fundamental memory will eventually come closer to the fundamental memory. We can relax this requirement and ask that a *randomly chosen* vector within ρn distance from the fundamental memory come closer to it. In most applications, correcting such random errors may be satisfactory. Yet, it is interesting to find out if stable fundamental memories can attract *all* the vectors within a distance of ρn for some positive constant ρ. In other words, we are interested in establishing a domain of attraction of radius ρn around each fundamental memory. More generally, it is interesting to find out if *every* state within distance ρn from a fundamental memory ends up within a distance of εn. Such a requirement guarantees that even in the *worst case*, we make a significant error-correction. For dealing with worst-case errors, one cannot rely on simulations since simulations (due to the prohibitively large number of error patterns) can only reveal the behavior of the system in the presence of random errors.

Moreover, quantitative behavior of the system in the case of worst-case errors is different from that of the system in the case of random errors. There cannot be a one synchronous-step convergence in the presence of arbitrary ρn errors, not even for arbitrary $\sqrt{m}$ errors. The idea behind this observation is the following: One-step convergence would mean, e.g., for each i, $\sum_{j\neq i} w_{ij}x_i y_j \geqslant 0$ for all y close to x. By changing the jth bit, we change the quantity $x_i \sum_{j\neq i} w_{ij} y_j$ by $2w_{ij}x_i y_j$, which is of the order $\sqrt{m}$. Since $x_i \sum_{i\neq i} w_{ij}x_j = O(n)$, by changing appropriate $cn/\sqrt{m}$ bits of x, we can get a y such that $x_i \sum_{j\neq i} w_{ij} y_j < 0$ for some i.

One cannot also have radius of attraction near $\frac{1}{2}$; $\rho > \frac{1}{8}$ is already impossible,

even when m is very small as shown by Montgomery and Vijaya Kumar [20]. Thus, one can only hope for a gradual convergence and a domain of attraction of smaller radius in the case of worst-case errors. In the following section, we present some of the known answers regarding the random error and the worst-case error convergence analysis.

2. Previous Results

Several researchers have described and predicted the features of the model using simulations and approximate calculations based on some independence assumptions. Also, this model is related to models of spin glasses. We refer the reader to Hopfield [8], Amit, Gutfreund, and Sompolinsky [2], and Mezard, Parisi, and Virasoro [19] for a wealth of information.

Basic questions about the absolute stability of the global pattern formation in dynamical systems have been studied by Grossberg [5] and Cohen and Grossberg [3], using Liapunov functions.

In the following, we survey some of the *rigorously* proved results in the case when the system of neurons is **fully** interconnected. Let m denote the number of fundamental memories.

McEliece, Posner, Rodemich, and Venkatesh [18] determined the maximum number of stable fundamental memories and the convergence properties in the presence of random errors:

- If $m < n/(4 \log n)$, then (with probability near 1) *all* fundamental memories will be stable. Also, for any fundamental memory, the system can correct *most* patterns of less than $n/2$ errors in one synchronous step.
- If $n/(4 \log n) < m < n/(2 \log n)$, then still *most* fundamental memories will be stable with the above described capability of correcting most patterns of errors.

When m is larger than $cn/\log n$, in particular, when $m = \alpha n$, the fundamental memories are not retrievable exactly, but one still may find stable states in their vicinity. This is suggested by the "energy landscape" results of Newman [21]. In particular, Newman proves that

- for all fundamental memories, all the vectors which are exactly at a distance of ρn from the fundamental memory have energy in excess of at least μn^2 above the energy level of the fundamental vector.

Thus, when starting from a fundamental memory, the system cannot wander away too far.

Komlós and Paturi [12] addressed the question of worst-case errors and proved the following results.

There are absolute constants α_s, α_a, ρ_s, $\rho_a < \rho_b$ such that the following properties hold for almost all choices of the fundamental memories:

- In the synchronous case, if $m \leqslant \alpha_s n$ and if the system is started *anywhere* within a distance of $\rho_s n$ from a fundamental memory v, then, in about $\log(n/m)$

synchronous steps, it will end up within a distance $ne^{-n/(4m)}$ from v. In particular, when $m < n/(4 \log n)$, the system will *converge* to v in $O(\log \log n)$ synchronous steps.

- In the asynchronous case, if $m \leqslant \alpha_a n$ and if the system is started *anywhere* within a distance of $\rho_a n$ from a fundamental memory v, then it will *converge* to a stable state within a distance of $ne^{-n/(4m)}$ from v. In particular, when $m < n/(4 \log n)$, the system will converge to v.

- For any fundamental memory v, the maximum energy of any state *within* a distance of $\rho_a n$ from v is less than the minimum energy of any state at a distance of $\rho_b n$ from v, and there are no stable states in the annuli defined by the radii $\rho_b n$ and $ne^{-n/(4m)}$ centered at the fundamental memories.

3. General Interconnections—Summary

The previous models seem to rely on their dense interconnections for associative memory properties. Neither the physiological data nor the VLSI technology support such dense interconnections. In this paper, we consider models in which the neurons are less densely interconnected. We try to determine the properties of the underlying interconnection graph that are responsible for the emergence of associative memory.

Below, we try to indicate how the geometry of the interconnections influences the degree of error-correction and the rate of convergence. For contrast, first we recall some features of **fully interconnected** systems:

- When storing only one fundamental memory, one can retrieve it even in the presence of $n/2$ errors. Furthermore, this error-correction takes only one synchronous step.

- When storing as few as two vectors, there is already enough noise to slow down the system to a $\log \log n$ convergence. But, as we showed in [12], this $\log \log n$ convergence time is retained even when we have cn fundamental memories.

When the system is **not fully interconnected**, storing even one vector introduces new problems:

- If the graph does not have sufficiently good connectivity properties, then, for a d-regular interconnection graph, as few as $d/2$ errors can make full retrieval impossible (see Examples 1 and 2).

- Even if the graph does have good connectivity properties, the one-step error correction found in fully interconnected graphs gives way to a more gradual error correction. In fact, convergence time is roughly equal to the diameter of the graph.

When storing several fundamental memories, the main difficulty in the fully interconnected case was in showing that the noise introduced by the combined Hebb weights of the other fundamental memories would not completely destroy the error-correction capabilities observed in the simple case of storing one vector, even

though the presence of noise slowed convergence down. In the case of general connections, in addition to the connectivity properties that ensure good error-correction in the one-vector case, we need a graph that is not too sparse. It can be seen (Example 3) that

- If we store at least two vectors, then mere stability of these vectors requires that every neuron be connected to more than $\log n$ other neurons.

It turns out that the requirements mentioned above are sufficient in the general case:

- When we have several fundamental memories, good connectivity and large degrees give us convergence; in fact, the convergence time is bounded by the sum of the diameter and $\log \log n$.

So far, we have tacitly assumed that the system has not exceeded its *memory capacity*. We now address the question of capacity. We distinguish between *full capacity*, the maximum number of fundamental memories that can be stored such that they are fully retrievable, and *partial capacity*, the maximum number of fundamental memories that can be stored such that a large fraction of the bits can be retrieved with few residual errors.

In the case of *fully* interconnected systems, the residual error is given by $ne^{-n/(4m)} = ne^{-1/(4\alpha)}$, where n is the number of neurons, m is the number of fundamental memories stored, and $\alpha = m/n$. Hence, the full capacity is $O(n/\log n)$, and the partial capacity is linear in n.

In the case of *general* connections (with good connectivity properties and sufficiently large degrees), the residual error is governed by a similar formula with $\alpha = m/d$. Hence,

- the full capacity is $O(d/\log n)$, and the partial capacity is linear in d.

This phenomenon of diminished capacity has been observed before by other researchers (McClelland, [17], Kinzel [9]). But, fortunately, the degree of error correction does not decrease.

- If the graph is highly connected, one can still recover the fundamental memories in the presence of *arbitrary* error patterns, and the number of errors allowable is still *proportional to n*.

Also, the decrease of errors is so rapid that terminating the synchronous algorithm after a small number of steps leaves negligibly few errors in the retrieval.

It is, of course, necessary to explain what we mean by good connectivity properties in all the above statements. The key parameter we will work with is the ratio of the second eigenvalue to the first (see the next section for technical definitions). This ratio was shown by Alon and Milman [1] to reflect connectivity properties.

EXAMPLE 1. Let G be the d-dimensional hypercube, where d is an odd integer. We assume, without loss of generality, that we want to store the vector of all ones. Let I be a set of neurons which form a $\lceil d/2 \rceil$-dimensional subcube. It is easy to see

that if the system is initiated with negative ones for the neurons in I and positive ones for the remaining neurons, the state remains unchanged. This shows that we cannot even correct $2^{d/2} = \sqrt{n}$ worst-case errors in a hypercube. As we will see later, this inability to correct errors is due to the fact that the largest two eigenvalues of the hypercube graph (equal to d and $d-2$) are too close to each other, which reflects a low degree of connectivity.

EXAMPLE 2. Assume that the interconnection graph has a clique of size $d/2+1$, and we store the vector of all ones. When the system is initiated with negative ones for the neurons corresponding to this clique, the neurons in the clique will never change their state.

EXAMPLE 3. It is easy to see that two stable states cannot be identical in the neighbourhood of a vertex i and yet different at i. Now, let the interconnection graph be d-regular. The probability that two randomly selected fundamental memories have unequal values at i, but equal values at i's neighbours, is $2^{-(d+1)}$. Thus, if $n \gg 2^{d+1}$, then the probability that both fundamental memories are stable is near zero.

The structure of the remainder of the paper is the following. First, we introduce the required graph-theoretical notions. Then, we state the precise results; we first consider the simple case of storing one vector to gain an understanding of the connectivity properties required for error-correction followed by the general case of storing many fundamental memories. Finally, we present the proofs of our theorems.

4. SPECTRA OF GRAPHS

In this paper, we use the graph spectrum to capture the connectivity structure of G, thus bringing the techniques of linear algebra into play. In the following, we introduce some notation to deal with graphs and linear spaces. We then present the basic facts concerning the spectrum of a graph. For more information, we refer the reader to Lancaster and Tismenetsky [13], or Lovász [16].

We use the undirected graph $G = \langle V, E \rangle$ to represent the symmetric interconnections among the neurons. The set V of vertices represents the collection of neurons, and the set E of *unordered* pairs represents the symmetric interconnections among them. $N(j)$ stands for the neighbourhood of vertex j, and $d(j) = |N(j)|$ denotes the degree of j. The average degree in G is $\delta(G) = \sum_{j \in V} d(j)/|V| = 2|E|/|V|$.

For $X, Y \subset V$, we use $G\{X, Y\}$ to denote the subgraph of G determined by the edges with one endpoint in X and the other in Y. We use $E\{X, Y\}$ to denote the set of edges in $G\{X, Y\}$ and $e\{X, Y\}$ to denote the cardinality of $E\{X, Y\}$.

In contrast with the set $E\{X, Y\}$ of unordered pairs, we denote by $E(X, Y)$ the set of ordered pairs $E(X, Y) = \{(x, y);\ x \in X,\ y \in Y\}$.

If I is a set of vertices, we let $\bar{I}$ denote the set $V - I$ of vertices not in I.

We regard A, the adjacency matrix of the graph G, as a linear transformation on

the real space $\mathbf{R}^n$. For vectors $x, y \in \mathbf{R}^n$, we use $(x, y) = (y, x)$ to denote the inner product of x and y. The l_2 norm of a vector $x \in \mathbf{R}^n$ is $|x| = (x, x)^{1/2}$.

The *spectrum* of a graph G is defined to be the spectrum of its adjacency matrix A, that is, the set of eigenvalues (or characteristic values) or A. Here, we present some elementary facts about the eigenvalues and eigenvectors of a matrix. In particular, we concentrate on real symmetric non-negative matrices, as adjacency matrices of undirected graphs enjoy these properties.

An $n \times n$ symmetric real matrix has n (not necessarily distinct) real eigenvalues. In fact, a symmetric real matrix has an orthonormal basis of real eigenvectors. Let $x_1, x_2, ..., x_n$ be an orthonormal basis of eigenvectors corresponding to the eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. The matrix U consisting of these orthonormal eigenvectors as columns, diagonalizes the matrix A; $UAU^{-1} = D$, where D is the diagonal matrix with $D_{ii} = \lambda_i$.

We use the following principle of Rayleigh to determine the eigenvalues of a matrix A. Let $\mathcal{V}_i$ be the subspace of $\mathbf{R}^n$ spanned by the vectors $x_1, x_2, ..., x_i$.

RAYLEIGH PRINCIPLE. *Let A be a real-symmetric matrix. Then, the largest eigenvalue λ_1 is given by*

$$\lambda_1 = \max_{0 \neq x \in \mathbf{R}^n} (x, Ax)/|x|^2,$$

and, for $i \geqslant 2$, the ith largest eigenvalue is given by

$$\lambda_i = \max_{0 \neq x \perp \mathcal{V}_{i-1}} (x, Ax)/|x|^2.$$

In the case of non-negative matrices, the Perron–Frobenius theory [13] provides more information about the eigenvalues. The largest eigenvalue of a non-negative matrix is non-negative, and there is a non-negative eigenvector belonging to it. It also has the maximum absolute value among all eigenvalues.

If a graph G is connected, then its largest eigenvalue $\lambda_1(G)$ has multiplicity 1, and there is a strictly positive eigenvector belonging to it. Also λ_1 satisfies

$$\text{average degree} \leqslant \lambda_1 \leqslant \text{maximum degree}.$$

In particular, if G is d-regular, we have $\lambda_1 = d$, and the vector $\mathbf{1}$ is a corresponding eigenvector.

The largest eigenvalues of subgraphs of G behave monotonically. More specifically, if H is a subgraph of G, we have that $\lambda_1(H) \leqslant \lambda_1(G)$.

In addition to the largest eigenvalue λ_1, the second largest eigenvalue reveals important connectivity properties of the graph. For example, if a d-regular graph is disconnected, then λ_1 occurs with multiplicity more than one (that is, $\lambda_2 = \lambda_1$). In fact, the multiplicity of λ_1 gives us the number of connected components of the graph. A large difference between the two largest eigenvalues corresponds to a high degree of connectivity of the graph. (For a complete graph on n vertices, $\lambda_1 = n - 1$, and $\lambda_i = -1$, $2 \leqslant i \leqslant n$.) We will also use the quantity $\mu = \max_{i \geqslant 2} |\lambda_i|$.

Remark. We will always assume that the interconnection graph G is d-regular, so that $\lambda_1 = d$. This assumption is technical. For non-regular G, the standard technique is to consider the eigenvalues of the matrix Q rather than those of the adjacency matrix A. Here $Q = D - A$, where D is the diagonal matrix containing the degrees of G in the diagonal. Q is positive (semi)definite, and has smallest eigenvalue zero with eigenvector $(1, 1, ..., 1)$. The second smallest eigenvalue λ (equal to $\lambda_1 - \lambda_2$ for regular graphs) is critical for expanding properties. For more information, see Alon and Milman [1].

We also use the following additional notations in this paper:

$c_1, c_2, ...$ denote absolute constants.

For $0 < \rho < 1$, we use the *entropy* function $h(\rho)$ defined as $h(\rho) = \rho \log(1/\rho) + (1-\rho) \log 1/(1-\rho)$.

log refers to the natural logarithm, but we truncate it from below, so that its value is always at least one.

We define the function χ on truth values as $\chi(\text{TRUE}) = +1$ and $\chi(\text{FALSE}) = -1$.

5. Results

Let $G = \langle V, E \rangle$ be the undirected graph that represents the interconnections among the neurons, and let A be the adjacency matrix of G. We will assume that G is a d-regular graph, so that $\lambda_1 = d$.

One Vector Case. Without loss of generality, we will assume that we want to remember the vector of all ones. In this case, the *synchronous algorithm amounts to each neuron resetting its state to the majority state among its adjacent neurons.* Let x be a vector which has ρn -1's and $(1-\rho)n$ 1's in it. Let $I \subset V$ be the set of the ρn neurons with the state -1. For the system in state x to improve itself and converge to the vector of all 1's, we need that many neurons be adjacent to more neurons outside of I than inside of I. This seems to imply that G should have good expansion properties, which, in turn, are related to the eigenvalues. In fact, it is clear from Examples 1 and 2 that graphs which have a small subset of vertices with many of their incident edges concentrated within the subset may have trouble correcting errors. We measure this concentration using the *average degree* of subgraphs in relation to their size. Since the average degree of any graph is bounded from above by its largest eigenvalue, measuring the concentration is reduced to finding upper bounds on the largest eigenvalues of subgraphs of G. We find several such bounds in Lemma 2.

The importance of concentration was noted earlier by Pippenger [22], and Dwork, Peleg, Pippenger, and Upfal [4] in the construction of fault-tolerant networks.

We will relate now eigenvalues of the graph to convergence properties of the system.

LEMMA 1. *If G has $\rho n < n/2$ errors, then at most $\rho' n$ errors remain after one synchronous step, where*

$$\rho' = c_1 \rho \left(\frac{\mu}{d}\right)^2.$$

Thus, in the one vector case, we see that even a bounded degree graph with good connectivity properties (small μ/d) guarantees the recovery of the memory in the presence of a constant fraction of errors. In fact, it takes $O(\log n/\log(d/\mu))$ steps for synchronous convergence.

Let us see now how the error-correction behavior and the rate of convergence will be modified when we store a number of fundamental memories.

General Case. Let us choose m fundamental memories, $v^1, v^2, ..., v^m$, independently from a uniform distribution. Write $\alpha = m/d$, where d is the degree of the graph. Let W^i, $i = 1, 2, ..., m$, be the weight matrices corresponding to v^i,

$$W^i_{jk} = \begin{cases} v^i_j v^i_k & \{j, k\} \in E \\ 0 & \text{otherwise} \end{cases}$$

for $j, k = 1, 2, ..., n$. The weight matrix of the system is $W = \sum_{i=1}^m W^i$.

We define the energy of the system at state x as $\mathscr{E}(x) = \sum_{i=1}^m \mathscr{E}^i(x) = -\frac{1}{2}\sum_{i=1}^m (x, W^i x)$. The term $\mathscr{E}^i$ represents the energy component due to the ith fundamental memory. It can be seen easily that an asynchronous step never increases the energy. This fact guarantees asynchronous convergence in any system with symmetric weights.

Now we present our results in a precise form for systems with arbitrary interconnections. Note their similarity to the results in the fully interconnected case: β_s, β_a, ρ_s, ρ_a, ρ_b are absolute constants. Let

$$\beta = (m + \mu)/d = \alpha + (\mu/d), \qquad \varepsilon = e^{-c_2/\beta}.$$

The parameter β plays an analogous role to that of the parameter $\alpha = m/n$ in the case of fully connected systems. ε measures the residual error in the recall procedure.

THEOREM 1. *The following statement holds with probability $1 - o(1)$ (as $n \to \infty$). In the synchronous case, if $m + \mu \leqslant \beta_s d$, and if the system is started within a distance of $\rho_s n$ from a fundamental memory, then, in*

$$O\left(\log(1/\beta) + \frac{1/\beta}{\log(d\beta/\mu)}\right) = O\left(\log\log(1/\varepsilon) + \frac{\log(1/\varepsilon)}{\log[d/(\mu \log(1/\varepsilon))]}\right)$$

synchronous steps, it will end up within a distance εn from the fundamental memory; that is, it will get within distance εn of the fundamental memory, and then it will remain (forever) within that distance. When $m+\mu=c_3 d/\log n$, the system will converge *to the fundamental memory in time*

$$O\left(\log\log n+\frac{\log n}{\log[d/(\mu\log n)]}\right).$$

In addition, if $\mu=O(d^{1-\xi})$ (which is the case with $\xi=\frac{1}{2}$ for most d-regular graphs), and $d>(\log n)^{2+\eta}$, then the system will converge to the fundamental memory in time

$$O\left(\log\log n+\frac{\log n}{\log d}\right). \tag{1}$$

Note that in the case $d=n-1$, the last formula gives back our earlier result [12] of a $\log\log n$ synchronous convergence.

THEOREM 2. *The following statement holds with probability $1-o(1)$. In the asynchronous case, if $m+\mu\leqslant\beta_a d$, and if the system is started within a distance of ρ_a from a fundamental memory, then it will converge to a stable state within a distance of εn from the fundamental memory. In particular, when $m+\mu=c_4 d/\log n$, the system will converge to the fundamental memory.*

THEOREM 3. *The following statement holds with probability $1-o(1)$. For any fundamental memory v, the maximum energy of any state* within *a distance of $\rho_a n$ from v is less than the minimum energy of any state at a distance of $\rho_b n$ from v, and there are no stable states in the annuli defined by the radii $\rho_b n$ and εn centered at the fundamental memories.*

6. PROOFS

An important step in proving our theorems is to characterize the *error-correction dynamics* of the system. Let us first consider the synchronous case. Let x be a vector at a distance of $\rho n\leqslant\rho_s n$ from some fundamental memory. In one synchronous step, the system, started at state x, will move to a state x', which is at a distance of $\rho' n$ from that fundamental memory. Our goal is to find the relationship between ρ and ρ'; we describe it in the main lemma. This relationship will completely determine the behaviour of the synchronous algorithm. It turns out that the *convergence in the synchronous case is monotone.*

In the asynchronous case, we establish energy barriers for vectors within a distance of $\rho_a n$ from a fundamental memory. We use these barriers to show that the system cannot go too far from the fundamental memory, and consequently, it converges near or to the fundamental memory, since the main lemma guarantees

the existence of an annulus centered at the fundamental memory, which is free of stable states. Here, we use the simple fact that in any symmetrically interconnected system asynchronous convergence is guaranteed.

In proving these lemmas, we face the task of estimating the sum of $m-1$ independent quantities of the form $S=\sum_{(i,j)\in E_H}(y_i, y_j)$, where y_i are independent random vectors with uniform distribution, and H is a subgraph of G with E_H as its set of edges. For this purpose, we derive an upper bound on the moment generating function of S. This will show that the tail of the distribution of S is governed by the largest eigenvalue of the subgraph H. Hence, we also need estimates on the eigenvalues of the subgraphs of G. In the following section, we derive these combinatorial tools. In the subsequent sections, we present the *main lemma* and the *energy-barrier lemma*, and finally prove the theorems stated in the earlier section.

6.1. *Moment Generating Function*

The theorem below is crucial in extending our results from the one vector case to the general case. We think it is interesting on its own right.

Let G be a simple graph with n vertices and N edges. We assign independent random variables X_i to the vertices of G, each taking the values ± 1 with equal probabilities. We define the sum

$$S=\sum_{\substack{\{i,j\} \text{ is an} \\ \text{edge of } G}} X_iX_j$$

(thus $\mathbf{E}S=0$ and $\sigma^2=\mathrm{Var}(S)=N$).

Let $K_2=\frac{1}{2}N$, and for $r\geqslant 3$, let K_r denote the number of simple cycles of length r in G.

The following theorem gives an estimate on the Laplace transform of the distribution of S.

THEOREM 4. *The moment generating function* $\mathbf{E}e^{tS}$ *of* S *can be bounded as*

$$\mathbf{E}e^{-tS}\leqslant \mathbf{E}e^{tS}\leqslant e^{\sum_{r\geqslant 2}K_rt^r}\leqslant e^{(1/2)Nt^2/(1-\lambda_1 t)} \qquad \text{for} \quad 0\leqslant t<1/\lambda_1.$$

The theorem says that $\mathbf{E}e^{tS}\leqslant e^{cNt^2}$ as long as $|t|<1/\lambda_1$. Thus, λ_1 determines the behaviour of the tail of the distribution of S. Namely, $S/\sigma(S)$ behaves approximately as standard normal in the range $|x|\ll\sigma/\lambda_1$.

Remark. The proof below will actually show that the above inequality

$$\mathbf{E}e^{-tS}\leqslant \mathbf{E}e^{tS}\leqslant e^{(1/2)\sigma^2t^2/(1-\lambda_1 t)} \qquad \text{for} \quad 0\leqslant t<1/\lambda_1 \tag{2}$$

holds for $S=\sum_{1\leqslant i<j\leqslant n}a_{ij}X_iX_j$, where $A=\{a_{ij}\}$ is an *arbitrary* non-negative symmetric matrix, λ_1 is the largest eigenvalue of A, and N was replaced in general by $\sigma^2=\sum_{i<j}a_{ij}^2$, the variance of S. As a matter of fact, A could even have negative entries, but then λ_1 should be defined as the largest eigenvalue of the matrix, whose entries are the absolute values of the corresponding entries of A.

COROLLARY 1. *For any* $y > 0$,

$$\mathbf{P}(S \geqslant y) \leqslant e^{-(1/2)y^2/(\sigma^2 + \lambda_1 y)}.$$

Proof of Theorem 4. We start with the identity

$$\mathbf{E}e^{tS} = \sum_{k \geqslant 0} \mathbf{E}(tS)^k/k! = 1 + \sum_{k \geqslant 2} \mathbf{E}(tS)^k/k!.$$

Now, a product term in S^k has a zero expectation, unless the edges involved cover every vertex an even number of times.

Such a collection of edges is a union of cycles (a two cycle is defined as an edge taken twice). Thus, if $C(k)$ denotes the number of ways to select an ordered sequence of k edges such that they form a union of cycles, then (writing $C(0) = 1$)

$$\mathbf{E}e^{tS} = \sum_k C(k)t^k/k!,$$

the exponential generating function of the sequence $C(k)$.

Using standard counting techniques, we first show that

$$\sum_k C(k)t^k/k! \leqslant \exp\left\{\sum_{r \geqslant 2} K_r t^r\right\}. \tag{3}$$

Indeed, let us say that a partition of a set into non-empty parts is of *type* $(T_1, T_2, T_3, \ldots)$ if the number of parts of size r is T_r. Then, the number of partitions of the integers $\{1, 2, \ldots, k\}$ of type $(T_1, T_2, T_3, \ldots)$ is equal to

$$P_k(T_1, T_2, \ldots) = k! \Big/ \prod_{r \geqslant 1} [(r!)^{T_r} T_r!],$$

where, of course, $\sum rT_r = k$. Given a partition of k of type $(0, T_2, T_3, \ldots)$, we can generate an ordered collection of edges which is a union of cycles by assigning to each part of size r an ordered collection of edges which form a simple cycle of length r. This can be done in $\prod_{r \geqslant 2} (K_r r!)^{T_r}$ ways. Since each element of $C(k)$ is generated this way at least once, we therefore have,

$$C(k) \leqslant \sum P_k(0, T_2, T_3, \ldots) \prod_{r \geqslant 2} (K_r r!)^{T_r},$$

where the sum is extended for all $r \geqslant 2$, $T_r \geqslant 0$ such that $\sum_{r \geqslant 2} rT_r = k$. Thus,

$$\sum_k C(k)t^k/k! \leqslant \prod_{r \geqslant 2} \sum_{T \geqslant 0} (K_r t^r)^T/T! = \exp\left\{\sum_{r \geqslant 2} K_r t^r\right\},$$

proving the first part of (3).

For the second part, we use the inequality

$$K_r \leqslant \frac{1}{2r} \operatorname{trace}(A^r). \tag{4}$$

Indeed, $\operatorname{trace}(A^r)$ counts the number of closed walks of length r in G, and simple cycles are counted $2r$ times (for $r=2$, we actually have equality in (4)). Thus,

$$\begin{aligned}\sum_{r\geqslant 2} K_r t^r &\leqslant \sum_{r\geqslant 2} \frac{1}{2r} \operatorname{trace}((tA)^r) \leqslant \frac{1}{4} \sum_{r\geqslant 2} \sum_{j=1}^{n} (\lambda_j t)^r \\ &= \frac{1}{4} \sum_{j=1}^{n} \lambda_j^2 t^2/(1-\lambda_j t) \leqslant \frac{1}{4} \sum_{j=1}^{n} \lambda_j^2 t^2/(1-\lambda_1 t) \\ &= \frac{1}{4} \sum_{i,j} a_{i,j}^2 t^2/(1-\lambda_1 t) = \frac{1}{2} Nt^2/(1-\lambda_1 t).\end{aligned}$$

To prove the corollary, select $t = y/(\sigma^2 + \lambda_1 y)$ in the theorem and use the inequality

$$\mathbf{P}(S \geqslant y) \leqslant e^{-ty} \mathbf{E} e^{tS}. \quad \blacksquare$$

6.2. *Estimates of Eigenvalues*

To apply Theorem 4 to subgraphs of G, we need to bound the largest eigenvalues of subgraphs of G. Also, the largest eigenvalue is useful in bounding the number of edges in a graph, since the average degree in a graph is bounded from above by the largest eigenvalue of that graph. In the following lemma, we give upper bounds on the largest eigenvalue of, and the number of edges in, a subgraph. It is a standard application of linear algebra, and some of these simple inequalities have been used before.

For a subgraph H of G, we define the neighbourhood graph $N\{H\}$ of H as the graph determined by those edges of G (and the implied vertices) which are incident to at least one vertex in H. The largest eigenvalue of a subgraph H of G is denoted by $\lambda_1(H)$.

LEMMA 2. *If H is a subgraph of G with ρn vertices, then the maximal eigenvalue of H satisfies*

$$\lambda_1(H) \leqslant \rho\lambda_1(G) + (1-\rho)\lambda_2(G). \tag{5}$$

Consequently, the number of edges in H is at most $\frac{1}{2}[\rho^2\lambda_1(G) + \rho\lambda_2(G)]n$. Also,

$$\lambda_1(N\{H\}) \leqslant 2[\sqrt{\rho}\,\lambda_1(G) + (1-\sqrt{\rho})\mu(G)]. \tag{6}$$

In general, if the matrix B is obtained from A (the adjacency matrix of G) by keeping only ρn rows and γn columns, and replacing all other elements of A with 0's, then, for any vector u of unit length,

$$|Bu| \leqslant \sqrt{\rho\gamma}\,\lambda_1(A) + (1-\sqrt{\rho\gamma})\mu(A) \tag{7}$$

or, which is the same, for any two vectors x and y,

$$|(x, By)| \leqslant [\sqrt{\rho\gamma}\,\lambda_1 + (1-\sqrt{\rho\gamma})\mu]\,|x|\,|y|.$$

Consequently, if I and J are two subsets of the vertex set of G with $|I| = \rho n$, $|J| = \gamma n$, then the number of (directed) edges going from J to I is at most

$$e(J, I) = \sum_{j \in J} e(j, I) \leqslant [\rho\gamma\lambda_1 + \sqrt{\rho\gamma}\,\mu]n$$

and the largest eigenvalue of the graph H determined by the (undirected) edges of G going from J to I is bounded as

$$\lambda_1(H) \leqslant 2[\sqrt{\rho\gamma}\,\lambda_1(G) + (1-\sqrt{\rho\gamma})\,\mu(G)]. \tag{8}$$

Proof. We start with some inequalities. Write v for the eigenvector corresponding to λ_1, and

$$x = \bar{x} + \frac{(x, v)}{|v|^2}v, \qquad y = \bar{y} + \frac{(y, v)}{|v|^2}v,$$

where $\bar{x}$ and $\bar{y}$ are orthogonal to v. Then,

$$|x|^2 = |\bar{x}|^2 + \frac{|(x, v)|^2}{|v|^2}, \qquad |y|^2 = |\bar{y}|^2 + \frac{|(y, v)|^2}{|v|^2}$$

and we have the inequalities

$$(x, Ax) \leqslant \lambda_2 |x|^2 + (\lambda_1 - \lambda_2)\frac{|(x, v)|^2}{|v|^2} \tag{9}$$

and

$$|(x, Ay)| \leqslant \mu\,|x|\,|y| + (\lambda_1 - \mu)\frac{|(x, v)(y, v)|}{|v|^2}. \tag{10}$$

Indeed,

$$(x, Ax) = (\bar{x}, A\bar{x}) + \lambda_1\frac{|(x, v)|^2}{|v|^2} \leqslant \lambda_2|\bar{x}|^2 + \lambda_1\frac{|(x, v)|^2}{|v|^2}$$

$$= \lambda_2|x|^2 + (\lambda_1 - \lambda_2)\frac{|(x, v)|^2}{|v|^2}$$

$$|(x, Ay)| = \left|(\bar{x}, A\bar{y}) + \lambda_1\frac{(x, v)(y, v)}{|v|^2}\right| \leqslant \mu\,|\bar{x}|\,|\bar{y}| + \lambda_1\frac{|(x, v)(y, v)|}{|v|^2}$$

$$\leqslant \mu\,|x|\,|y| + (\lambda_1 - \mu)\frac{|(x, v)(y, v)|}{|v|^2}.$$

We can clearly assume that H is a spanned subgraph of G, since adding edges can only increase the largest eigenvalue (for it has a non-negative corresponding eigenvector.)

Let u be an arbitrary real unit vector and I the set of rows (and columns) to which the matrix A is restricted to obtain B, the adjacency matrix of H.

Write u_I for the vector obtained from u by replacing with 0 all coordinates u_i, $i \notin I$. Then, by (9),

$$(u, Bu) = (u_I, Au_I) \leqslant \lambda_2 + (\lambda_1 - \lambda_2) \frac{|(u_I, v)|^2}{|v|^2}$$

$$= \lambda_2 + (\lambda_1 - \lambda_2) \frac{|(u, v_I)|^2}{|v|^2} \leqslant \lambda_2 + (\lambda_2 + (\lambda_1 - \lambda_2) \frac{|v_I|^2}{|v|^2}.$$

Using the last inequality for $v = \mathbf{1}$, we get

$$\lambda_1(B) = \max_{|u|=1} (u, Bu) \leqslant \lambda_2 + (\lambda_1 - \lambda_2)\rho$$

which proves (5).

To get an estimate on the number of edges in H, it remains to use the fact that the average degree in H not more than the largest eigenvalue of H; or, a direct proof, $|E_H| = \frac{1}{2}(\mathbf{1}_I, B\mathbf{1}_I) \leqslant \frac{1}{2}\lambda_1(B)\,|\mathbf{1}_I|^2 = \frac{1}{2}\lambda_1(B)\rho n$.

If we write B for the matrix obtained from A by keeping only the intersection of rows with indices in J and columns with indices in I, and replace all other entries by 0, then, by (10),

$$|(x, By)| = |(x_J, Ay_I)| \leqslant \mu\,|x|\,|y| + (\lambda_1 - \mu) \frac{|(x_J, v)(y_I, v)|}{|v|^2}$$

$$= \mu\,|x|\,|y| + (\lambda_1 - \mu) \frac{|(x, v_J)(y, v_I)|}{|v|^2}$$

$$\leqslant |x|\,|y| \left[\mu + (\lambda_1 - \mu) \frac{|v_J|\,|v_I|}{|v|^2} \right].$$

Using the last inequality for $v = \mathbf{1}$, we get

$$|(x, By)| \leqslant |x|\,|y|\,[\mu + (\lambda_1 - \mu)\sqrt{\rho\gamma}]$$

which proves (7). In particular,

$$e(J, I) = \sum_{j \in J} e(j, I) = (1_J, A1_I) = (1_J, B1_I) \leqslant |1_J|\,|1_I|\,[\mu + \lambda_1 \sqrt{\rho\gamma}]$$

$$= [\rho\gamma\lambda_1 + \sqrt{\rho\gamma}\,\mu]n.$$

Equation (6) easily follows from (7). ∎

6.3. *Error Correction Lemma*

In this section, we present the lemma that determines the dynamics of the system in the synchronous case.

MAIN LEMMA. *Let* $\rho_s > 0$ *and*

$$\varepsilon = e^{-c_2 d/(\mu + m)}$$

The following holds with probability $1 - o(1)$: *Let* x *be a vector at a distance of* ρn *from a fundamental memory, where* $\varepsilon \leqslant \rho \leqslant \rho_s$. *Let* x' *be the resulting state after one step of the synchronous algorithm, given that the system is started in state* x. *Then, the distance of* x' *from the fundamental memory is at most* $f(\rho)n$, *where* $f(\rho)$ *is given by*

$$f(\rho) = c_5 \rho \left(h(\rho) + \left[\frac{\mu}{d} \log \frac{1}{\rho} \right]^{2/3} \right). \tag{11}$$

COROLLARY 2. *The number of residual errors in the synchronous case is at most* εn.

This lemma brings out the two key parameters: the ratio μ/d that governs the convergence, and the number of fundamental memories $m = \alpha d$, which together with μ/d determine the number of remaining errors. It is not hard to see that the parameter μ could actually be replaced by λ_2 here and throughout the whole paper.

Proof of the Main Lemma. For notational convenience, let us assume that the fundamental memory in question is v^1. (This assumption will later be removed by multiplying the probabilities with m.) We specify x by the set of co-ordinates I, $|I| = \rho n$, in which x and v^1 differ. Let x' be the vector resulting after one step of the algorithm, and let J, $|J| = \rho' n = f(\rho)n$, be a set of co-ordinates in which x' and v^1 differ. In other words, J is a set of components j such that $v_j^1 (Wx)_j \leqslant 0$, where W is the weight matrix of the system. This implies $T = \sum_{j \in J} v_j^1 (Wx)_j \leqslant 0$. Since W is the sum of W^i, the contribution of the ith fundamental memory, we write $T = \sum_{i=1}^m T^i = \sum_{i=1}^m \sum_{j \in J} v_j^1 (W^i x)_j$. We consider each of the terms T^i.

If ρ_s is chosen sufficiently small, x tends to be closer to v^1 than to other fundamental memories, and consequently T^1, the "tendency" towards v^1, will be the dominating term. Indeed,

$$T^1 = \sum_{j \in J} v_j^1 (W^1 x)_j = \sum_{j \in J} \sum_{k \in N(j)} v_k^1 x_k = \sum_{j \in J} \{e(j, \bar{I}) - e(j, I)\}$$

$$= \sum_{j \in J} \{e(j, V) - 2e(j, I)\}.$$

Here $e(J, V) = d|J|$, since the graph G of interconnections is d-regular. We estimate $e(J, I)$ using Lemma 2. We then get

$$T^1 \geqslant d|J| - 2\left(\frac{|I|\,|J|}{n} d + \sqrt{|I|\,|J|}\,\mu\right) = d\rho' n\left(1 - 2\rho - 2\sqrt{\frac{\rho}{\rho'}}\frac{\mu}{d}\right) \tag{12}$$

$$\geqslant c_6 d\rho' n \tag{13}$$

if ρ' satisfies

$$\rho' \geqslant c_7 \rho \left(\frac{\mu}{d}\right)^2 \tag{14}$$

For the other terms T^i, $i \geqslant 2$, we have

$$T^i = \sum_{j \in J} v_j^1 (W^i x)_j = \sum_{(j,k) \in E(J, \bar{I})} (v_j^1 v_j^i)(v_k^1 v_k^i) - \sum_{(j,k) \in E(J, I)} (v_j^1 v_j^i)(v_k^1 v_k^i).$$

Let $u_j^i = v_j^1 v_j^i$ for $j = 1, 2, \ldots, n$. From now on, we will consider the vector v^1 fixed, although we will not indicate this conditioning in our notations. The estimates we get on moment generating functions and probabilities will not depend on the choice of v^1, so we get the same estimates for the unconditional probabilities (or moment generating functions).

Given the vector v^1, the numbers u_j^i, $2 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, are (conditionally) independent and uniformly distributed ± 1 random variables, since the fundamental memories are chosen independently and uniformly. We then have

$$T^i = \sum_{j \in J}\left(\sum_{(j,k) \in E(j, \bar{I})} u_j^i u_k^i - \sum_{(j,k) \in E(j, I)} u_j^i u_k^i\right)$$

$$= \sum_{(j,k) \in E(J, V)} u_j^i u_k^i - 2 \sum_{(j,k) \in E(J, I)} u_j^i u_k^i.$$

Each one of the sums above is a sum over ordered pairs of vertices. We rewrite these sums as sums over unordered pairs, resulting in

$$T^i = \sum_{\{j,k\} \in E\{J, V\}} u_j^i u_k^i - 2 \sum_{\{j,k\} \in E\{J, I\}} u_j^i u_k^i$$

$$+ \sum_{\{j,k\} \in E\{J, J\}} \chi(j \notin I \cap J \text{ or } k \notin I \cap J)\, u_j^i u_k^i \equiv S_1^i + 2S_2^i + S_3^i.$$

The term T^i, for $i \geqslant 2$, represents the tendency of x to go towards the ith fundamental memory. We want to show that, with a large probability, the combined tendency, $\sum_{i \geqslant 2} T^i$, towards the other fundamental memories, is no more than the tendency T^1 towards the fundamental memory v^1. For this purpose, we use the moment generating function derived in Thorem 4. Given the vector v^1, the terms T^i, $i \geqslant 2$, are conditionally independent, thus, the (conditional) moment

generating function of the sum $\sum_{i\geq 2} S_k^i$ is obtained by taking the product of the (conditional) moment generating functions of each of the S_k^i. We will then show that each $\sum_{i\geq 2} S_k^i$ is bounded from below by a constant fraction $-\gamma_k T^1$ of the main term T^1.

Let $\gamma_1, \gamma_1, \gamma_3 \geq 0$ be such that $\gamma_1 + 2\gamma_2 + \gamma_3 = 1$. We observe that the selection of γ_k affects only the constants involved. In the following, we will estimate each of the quantities $\sum_{i\geq 2} S_k^i$ separately.

From the corollary to Theorem 4, it follows that the probability of $\sum_{i\geq 2} S_1^i < -\gamma_1 T^1$ is at most

$$\exp\left\{-\frac{1}{2}\frac{(\gamma_1 T^1)^2}{\gamma_1 T^1 \lambda_1(G_1) + e\{J, V\}m}\right\}$$

with $G_1 = G\{J, V\}$, the subgraph determined by the edges incident upon J.

There are m possible ways to choose our fundamental memory (which was assumed to be v^1 only to simplify notation), and there are $\binom{n}{|J|}$ ways to choose the set J. Furthermore, $\binom{n}{|J|} \leq \exp\{h(|J|/n)n\} = \exp\{h(\rho')n\}$. Therefore, the probability that there exists a fundamental memory and a set J, $|J| = \rho' n$, such that $\sum_{i\geq 2} S_1^i \leq -\gamma_1 T^1$, is $o(1)$ if

$$\frac{1}{2}\frac{(\gamma_1 T^1)^2}{\gamma_1 T^1 \lambda_1(G_1) + e\{J, V\}m} \geq h(\rho')n + \log m + \Delta(n), \tag{15}$$

where $\Delta(n)$ is any unboundedly increasing function of n. Note that $\rho' \geq 1/n$ implies $h(\rho')n + \log m + \Delta(n) \leq 3h(\rho')n$.

To simplify the above condition, we use the estimates (Lemma 2):

$$\lambda_1(G_1) \leq 2(\sqrt{\rho'}\, d + \mu), \qquad e\{J, V\} \leq e(J, V) = d\,|J| = d\rho' n.$$

It is now easy to see that the inequalities $\rho' \geq \varepsilon = e^{-c_2 d/(m+\mu)}$ and $T^1 \geq c_6\, d\rho' n$ imply the above condition (15). Similarly, $\sum_{i\geq 2} S_2^i \leq -\gamma_2 T^1$ holds only with a probability not exceeding

$$\exp\left\{-\frac{1}{2}\frac{(\gamma_2 T^1)^2}{\gamma_2 T^1 \lambda_1(G_2) + e\{J, I\}m}\right\},$$

where $G_2 = G\{J, I\}$, the subgraph determined by the edges from J to I. Since we have at most $m\binom{n}{|I|+|J|}$ ways to select the fundamental memory and the sets I and J, the probability that there exists a fundamental memory and sets I, $|I| = \rho n$ and J, $|J| = \rho' n$ such that $\sum_{i\geq 2} S_2^i \leq -\gamma_2 T^1$ is $o(1)$ if

$$\frac{1}{2}\frac{(\gamma_2 T^1)^2}{\gamma_2 T^1 \lambda_1(G_2) + e\{J, I\}m} \geq h(\rho + \rho')n + \log m + \Delta(n). \tag{16}$$

Using the estimates $\lambda_1(G_2) \leqslant 2(\sqrt{\rho\rho'}d + \mu)$ and $e\{J, I\} \leqslant (\rho\rho' d + \sqrt{\rho\rho'}\mu)n$, from Lemma 2, we get that condition (16) will be satisfied if

$$\rho' \geqslant c_8 \rho (h(\rho))^2 \tag{17}$$

$$\rho' \geqslant c_9 \frac{\mu}{d} h(\rho) \tag{18}$$

$$\rho' \geqslant c_{10} \alpha \rho h(\rho) \tag{19}$$

$$\rho' \geqslant c_{11} \rho \left[\frac{\mu}{d} \alpha \log \frac{1}{\rho} \right]^{2/3}. \tag{20}$$

To estimate S_3^i, we observe that we have a bound on the moment generating function of S_3^i, which in turn is bounded from above by the moment generating function of the sum $\sum_{\{j,k\} \in E\{J,J\}} u_j^i u_k^i$. Hence, the probability that $\sum_{i \geqslant 2} S_3^i \leqslant -\gamma_3 T^1$ is at most

$$\exp\left\{ -\frac{1}{2} \frac{(\gamma_3 T^1)^2}{\gamma_3 T^1 \lambda_1(G_3) + e\{J, J\} m} \right\},$$

where $G_3 = G\{J, J\}$, the subgraph induced by J. Thus, the probability that there exists a fundamental memory and a set J, $|J| = \rho' n$, such that $\sum_{i \geqslant 2} S_3^i \leqslant -\gamma_3 T^1$ is $o(1)$ if

$$\frac{1}{2} \frac{(\gamma_3 T^1)^2}{\gamma_3 T^1 \lambda_1(G_3) + e\{J, J\} m} \geqslant h(\rho') n + \log m + \Delta(n). \tag{21}$$

From Lemma 2, we get $\lambda_1(G_3) \leqslant \rho' d + \mu$, and $e\{J, J\} \leqslant \frac{1}{2}(\rho' d + \mu)\rho' n$. Then, the assumptions $\rho' \geqslant \varepsilon = e^{-c_2 d/(\mu + m)}$ and $T' \geqslant c_6 d\rho' n$ are sufficient to satisfy (21).

Finally, we observe that the conditions (14) and (17)–(20) are satisfied with the value of $f(\rho)$ as given in the lemma. Hence, the lemma is proved. ∎

6.4. *Energy-Barrier Lemma*

In the following, we give bounds for the energy of the system in the vicinity of a fundamental memory.

ENERGY-BARRIER LEMMA. *With probability* $1 - o(1)$, *for all* ρ, $\varepsilon \leqslant \rho \leqslant \rho_b$, *and all vectors* x *at a distance* ρn *from a fundamental memory,*

$$|\mathscr{E}(x) - \mathscr{E}(v) - 2e(I, \bar{I})| \leqslant d\rho n, \tag{22}$$

where I, $|I| = \rho n$, *is the set of co-ordinates in which* x *and* v *differ.*

COROLLARY 3. *The following holds with probability* $1-o(1)$: *For all* ρ_1, ρ_2, $\varepsilon \leqslant \rho_1 < \rho_2 \leqslant \rho_b$, $\rho_1 < \rho_2/4$, *and for every fundamental memory* v, *the energy of any state* within *a distance of* $\rho_1 n$ *from* v *is less than the energy of any state at a distance of* $\rho_2 n$ *from* v.

Proof of the Energy-Barrier Lemma. As before, let us assume that the fundamental memory in question is v^1. (We will later remove this condition by multiplying the probability with m.)

Let v_I and $v_{\bar{I}}$ be such that $v_{\bar{I}} + v_I = v$ and $v_{\bar{I}} - v_I = x$ (thus, v_I is obtained from v by replacing all coordinates outside I with 0). Then, the difference of energy at x and v^1 is

$$\begin{aligned}\mathscr{E}(x) - \mathscr{E}(v^1) &= \tfrac{1}{2}[((v_I + v_I),\, W(v_{\bar{I}} + v_I)) - ((v_{\bar{I}} - v_I),\, W(v_{\bar{I}} - v_I))] \\ &= 2(v_I,\, Wv_I)\end{aligned}$$

since W is symmetric. Furthermore, we have

$$\begin{aligned}(v_I,\, Wv_I) &= (v_I,\, W^1 v_{\bar{I}}) + \sum_{i \geqslant 2} (v_I,\, W^i v_{\bar{I}}) \\ &= \sum_{(j,k) \in E(I,\bar{I})} v_j^1 (v_j^1 v_k^1) v_k^1 + \sum_{i \geqslant 2} \left[\sum_{(j,k) \in (I,\bar{I})} v_j^1 (v_j^i v_k^i) v_k^1 \right] \\ &= e(I,\bar{I}) + \sum_{i \geqslant 2} \left[\sum_{(j,k) \in E(I,\bar{I})} (v_j^1 v_j^i)(v_k^i v_k^1) \right].\end{aligned}$$

Given v^1, the numbers $(v_j^1 v_j^i)$, $2 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, are (conditionally) independent and uniformly distributed ± 1 random variables. As in the main lemma, we use the moment generating function derived in Theorem 4 to estimate the second term. Given a fundamental memory, there are at most $\binom{n}{\rho n} < e^{h(\rho)n}$ possible choices for x. Thus, the probability that there exists a fundamental memory and a set I, $|I| = \rho n$, such that $2\,|\sum_{i \geqslant 2} [\sum_{(j,k) \in E(I,\bar{I})} (v_j^1 v_j^i)(v_k^i v_k^1)]| \geqslant y$, is $o(1)$ as long as

$$\frac{1}{2} \frac{y^2}{e\{I, \bar{I}\} m + \lambda_1(G_4)\, y} \geqslant h(\rho) n + \log m + \Delta(n), \tag{23}$$

where $G_4 = G\{I, \bar{I}\}$, the subgraph determined by the edges from I to $\bar{I}$. We use the estimate $\lambda_1(G_4) \leqslant 2(\sqrt{\rho}\, d + \mu)$ from Lemma 2, and we bound $e\{I, \bar{I}\}$ by $d\rho n$. It can then be easily verified that the above condition (23) holds for the value of $y = d\rho n$ given in the lemma. Hence, the lemma is proved. ∎

To prove the corollary, use the idendity $e\{I, \bar{I}\} = d\rho n - 2e\{I, I\}$, together with the bound $2e\{I, I\} \leqslant (\rho d + \mu)\rho n$ from Lemma 2.

6.5. *Proofs of the Theorems*

To prove Theorem 1, we use the main lemma repeatedly until we are at a distance εn from the fundamental memory. To analyze the time complexity of this process, we need to analyze the recurrence relation

$$\rho \leftarrow c_4 \rho \left(h(\rho) + \left[\frac{\mu}{d} \log \frac{1}{\rho} \right]^{2/3} \right) \tag{24}$$

given in the main lemma. We consider each of the two terms in this relation separately. The first term gives us $O(\log(1/\beta)) = O(\log\log(1/\varepsilon))$ time steps. To analyze the contribution of the second term, we use the following lemma.

LEMMA 3. *Let a_t be defined by the recurrence*

$$a_0 = c_{12}$$

$$a_{t+1} = q a_t \log \frac{1}{a_t},$$

where $q < 1$. If y is such that $q < y < 1$, then we have $a_t \leqslant e^{-1/y}$ for $t \geqslant 1/(y \log(y/q))$.

Proof. We observe that $a_{t+k} \leqslant q^k a_t \prod_{i=0}^{k-1} [\log(1/a_t) + i \log(1/q)]$ and the lemma follows. ∎

From the lemma, we get that the contribution by the second term is $O(1/(\beta \log(\beta d/\mu))$. Hence, the time complexity is at most the sum of these two contributions, as given in the theorem.

To prove Theorem 3, we take $\rho_b = \rho_s$, the synchronous radius of convergence. We take ρ_a, the asynchronous radius of convergence, to be equal to $\rho_b/4$. Hence, from the corollary to the energy-barrier lemma, it follows that the maximum energy of any state within a distance of $\rho_a n$ from a fundamental memory is less than the minimum energy of any state at a distance of $\rho_b n$ from that fundamental memory.

Furthermore, the asynchronous process, when started from within a distance of $\rho_a n$ from a fundamental memory, will converge to a stable state, and in the process the energy keeps on decreasing monotonically. Thus, it can never leave the region with radius $\rho_b n = \rho_s n$. In addition, Theorem 1 guarantees that there are no stable states in the annuli defined by $\rho_s n$ and εn around the fundamental memories. Hence, the stable state that the system converges to must be within εn of the fundamental memory. Thus, Theorem 2 follows from Theorem 1 and Theorem 3.

ACKNOWLEDGMENT

We thank one of the referees for the detailed and helpful criticism of the manuscript.

REFERENCES

1. N. ALON AND V. D. MILMAN, Eigenvalues, expanders and superconcentrators, *in* "25th IEEE Symposium on the Foundations of Computer Science, 1984," pp. 320–322.
2. D. J. AMIT, H. GUTFREUND, AND H. SOMPOLINSKY, Statistical mechanics of neural networks near saturation, *Ann. Phys. (N.Y.)* **173** (1987), 30–67.
3. M. A. COHEN AND S. GROSSBERG, Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Trans. Systems Man Cybernet.* **13** (1983), 815–826.
4. C. DWORK, D. PELEG, N. PIPPENGER, AND E. UPFAL, Fault tolerance in networks of bounded degree, *in* "ACM Symposium on Theory of Computing, 1986," pp. 370–379.
5. S. GROSSBERG, "Studies of Mind and Brain," pp. 5–213, Reidel, Boston, 1982.
6. D. O. HEBB, "The Organization of Behavior," Wiley, New York, 1949.
7. G. E. HINTON AND J. A. ANDERSON (Eds.), "Parallel Models of Associative Memory," Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.
8. J. J. HOPFIELD, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. U.S.A.* **79** (1982), 2554–2558.
9. W. KINZEL, Learning and pattern recognition in spin glass models, *Z. Phys. B* (1985), 205–213.
10. T. KOHONEN, "Self-Organization and Associative Memory," Springer-Verlag, New York, 1984.
11. T. KOHONEN, State of the art in neural computing, *in* "IEEE First International Conference on Neural Networks I, 1987," pp. 77–90.
12. J. KOMLÓS AND R. PATURI, Convergence results in an associative memory model, *Neural Networks* **1** (1988), 239–250.
13. P. LANCASTER AND M. TISMENETSKY, "The Theory of Matrices: With Applications," Academic Press, New York, 1985.
14. W. A. LITTLE, The existence of persistent states in the brain, *Math. Biosci.* **19** (1974), 101–119.
15. W. A. LITTLE AND G. L. SHAW, Analytic study of the memory storage capacity of a neural network, *Math. Biosci.* **39** (1978), 281–290.
16. L. LOVÁSZ, "Combinatorial Problems and Exercises," North-Holland, Amsterdam, 1979.
17. J. L. MCCLELLAND, Resource requirements of standard and programmable nets, *in* "Parallel Distributed Processing," Vol. I, MIT Press, Cambridge, MA, 1986.
18. 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** (1987), 461–482.
19. M. MEZARD, G. PARISI, AND M. A. VIRASORO, "Spin Glass Theory and Beyond," World Scientific, Singapore, 1987.
20. B. L. MONTGOMERY AND B. V. K. VIJAYA KUMAR, Evaluation of the use of the Hopfield neural network model as a nearest-neighbor algorithm, *Appl. Opt.* **25** (1986), 3759–3766.
21. C. M. NEWMAN, Memory capacity in neural network models: Rigorous lower bounds, *Neural Networks* **1** (1988).
22. N. PIPPENGER, On networks of noisy gates, *in* "IEEE 26th Annual Symposium on the Foundations of Computer Science, 1985," pp. 30–38.

Printed in Belgium
Uitgever: Academic Press, Inc.
Verantwoordelijke uitgever voor België:
Hubert Van Maele
Altenastraat 20, B-8310 Sint-Kruis