Dynamics of Analog Neural Networks

A thesis presented

by

Charles Masamed Marcus

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

May, 1990

ABSTRACT

This thesis explores a variety of topics concerning the dynamics, stability and performance of analog neural networks. Techniques used to study these systems include global and local stability analysis, statistical methods originally developed for Ising-model spin glasses and neural networks, numerical simulation, and experiments on a small (8-neuron) electronic neural network. Attention is focused mostly on networks with symmetric connections. The analog neurons are taken to have a smooth and monotonic transfer function, characterized by a gain (i.e. maximum slope) β.

The electronic network includes time delay circuitry at each neuron. Additional circuitry allows measurement of the basins of attraction for fixed points and oscillatory attractors. Stability criteria for analog networks with time-delayed neuron response are derived based on local analysis. These results agree well with numerics and experiments on the electronic network.

A global stability analysis is presented for analog networks with parallel updating of neuron states. It is shown that symmetric networks satisfying the criterion: $1/\beta > -\lambda_{min}$ for all neurons, where λ_{min} is the minimum eigenvalue of the connection matrix, can be updated in parallel with guaranteed convergence to a fixed point. Based on this criterion, and a new analysis of storage capacity, phase diagrams for the Hebb and pseudo-inverse rule associative memories are derived. Analysis of parallel dynamics is then extended to a multistep updating rule that averages over M previous time steps. Multistep updating allows oscillation-free parallel dynamics for networks that have period-2 limit cycles under standard parallel updating.

It is shown analytically and numerically that lowering the neuron gain greatly reduces the number of local minima in the energy landscapes of analog neural networks and spin

glasses. Eliminating fixed-point attractors by using analog neurons has beneficial effects similar to stochastic annealing and can be easily implemented in a deterministic dynamical system such as an electronic circuit.

Finally, a numerical study of the distribution of basin sizes in the Sherrington-Kirkpatrick spin glass is presented. It is found that basin sizes are distributed roughly as a power law and that using analog state variables selectively eliminates small basins.

ACKNOWLEDGMENTS

It is with pleasure and sincere thanks that I acknowledge the help, support, and encouragement I have received while working on this thesis.

First of all, Bob Westervelt has been a terrific advisor. His intuition and sense of good, clean science has had a strong and positive influence on me. His guidance was always present but never overbearing, and his insistence on clarity helped me greatly to refine the ideas in this thesis into an intelligible form. Bob has also managed to create a working environment that is filled with exciting physics and nice people. As a result, his lab has always been a fun place to spend time.

Steve Strogatz has been a great friend, teacher, and (to use his word) partner. His contributions to this thesis and to my education are too many to count.

Much of the work in this thesis was done in collaboration with Fred Waugh. Working with Fred has been a real pleasure: easy, fun and remarkably productive.

I am also especially grateful to the following people for their many contributions to my education: Larry Abbott, John Anderson, Ken Babcock, Roger Brockett, Jim Clark, Jeff Dunham, Francis Everitt, Tom Kepler, Chris Lobb, Isaac Silvera, Paul Sokol, and Alan Yuille.

I am grateful to AT&T Bell Laboratories for generously providing four years of financial support.

Finally, I would like to thank my parents, my sister, and my friends for their support and love.

Table of Contents

Chapter 1

INTRODUCTION

Our world is filled with complex phenomena which emerge, as if by magic, out of interactions among many simple elements. Sometimes (rarely) an intellectually satisfying picture can be painted, allowing us to claim that we *understand* how the magic comes about. For certain static phenomena, statistical mechanics provides such a picture, and makes clear how the interaction of microscopic elements can give a large system a "life of its own," with well-defined properties that are not obviously present when the system is observed element by element. Statistical mechanics also provides a justification for the empirical fact that large systems can be characterized by a few well-chosen quantities; one does not need to keep track of all 10^{23} variables. This feature is essential for rendering large systems understandable.

Extending the principles and techniques of statistical mechanics to include complex, dynamic phenomena on a macroscopic scale remains an outstanding challenge and a problem of great current interest in many areas of physics. Ultimately, one would like to understand the complexity of the real world within a framework linking statistical mechanics and dynamical systems theory. The hope for such a synthesis, however, rests on the hypothesis that microscopic processes can be described by simple models. If the complexity of nature must be accounted for at all size scales, we certainly have no hope of understanding big systems.

Neural networks research certainly represents the most extreme test of the hypothesis that complexity can emerge directly out of the interaction of a large number of simple elements. It asks the question, "Can the operation of the most complicated object known

be described as an emergent property of maximally simple elements interacting according to simple rules?" This line of inquiry does not presuppose that the answer is yes, but rather seeks to discover just how far such a principle can go. Judging from our current level of knowledge about even the simplest biological systems, we may not learn whether the approach is justified for some time, let alone reap (and market) the fruits of the endeavor.

In addition to the role of neural networks as a paradigm for understanding biology, there is a purely technological motivation for developing highly parallel dynamical systems that can solve difficult problems. Simply put, the standard computer architecture is coming to the end of its rope. Many problems of great technological interest cannot be solved with acceptable speed using the fastest conventional computers, and even allowing factors of 100 or 1000 in speed, the present technology remains ill-suited to certain applications. It is interesting to note that many tasks which are routinely performed by humans with almost trivial ease seem to be the most challenging for computers. Our ability to quickly recognize a face, or to infer the shapes and distances of objects from visual information illustrates the astronomical superiority of biological computation over current computer technology. What makes this superiority more remarkable still is the fact that the fundamental time scale in biology is around a millisecond, some four orders of magnitude slower than standard computer cycle times.

The efficiency of biological computation suggests that perhaps by simply emulating biology's basic design - without necessarily duplicating it - we may realize revolutionary technological advances. Which qualities constitute biology's "basic design" is currently anybody's guess, though massive parallelism and fault tolerance seem to be two such basic principles, at least in the cortex. (Neither feature is part of the basic design of current computers.) If nothing else, biological neural networks serve as working demonstrations that a vastly superior technology is possible in principle .

We will not review the long and interesting history of neural networks here. Instead, we refer the reader to several good reviews which have appeared recently [Lippmann, 1987; Grossberg, 1988; Amit, 1989; Hirsch, 1989; Abbott, 1990; reprints of many of the classic articles can be found in Shaw and Palm, 1988]. The various accounts of neural networks are remarkably disparate, especially in their historical perspective, so it is necessary to read several versions in order to appreciate the breadth of the subject.

A common feature of nearly all neural network models, dating back to their modern origin in the work of McCulloch and Pitts [1943], is the sum-and-threshold device known as a formal neuron or simply a neuron for short[1]. The basic neuron we consider is shown in Fig. 1.1 (many variations will appear later). The output of the formal neuron can be binary ({0,1} or {-1,1}) or continuous, but the sigmoidal (s-shaped) nonlinearity of the input-output transfer function is standard. Much more will be said later regarding the shape of the neuron transfer function. A neural network is typically a collection of these formal neurons arranged in some architecture, with neuron inputs connected to external signals and to the outputs of other neurons. Connections between neurons are characterized by a set of connection weights, which may be negative, positive, or zero. In addition, one must specify a dynamic rule defining how the states of the neurons change in time.

The hard part of the problem, of course, is figuring out how to connect the neurons to each other so that the resulting dynamical system will do something interesting or useful. The idea, however, is not to set the connections "by hand," but rather to develop learning algorithms so that the network can respond to external stimuli by modifying its own connections in an effective way. Loosely speaking, we would like the network to learn from its own experience. Considerable progress in this area has been made, particularly for the task of associative memory, beginning with the work of Hebb [1949], and

[1]Henceforth, the word "neuron" will be taken to mean a formal neuron; any references to real (biological) neurons will be explicitly stated as such.

continuing through to the recent work of the PDP Group [Rumelhart *et al.*, 1986], E. Gardner [1988], and others. Hebb's primary contribution was to postulate a remarkably simple, yet effective mechanism for modifying connections which will store a particular neural state as a memorized pattern. The rule is the following: Impose a pattern of stimulus onto the network, and incrementally increase the connection weight between neurons with coincident activity. This modification will eventually cause the imposed pattern of neuronal activity to become a stable configuration of the system after the stimulus is removed. There is evidence that Hebb's mechanism is realized in biological systems, though at present this is an issue of considerable debate [Lynch, 1986].

An important milestone in the understanding of neural network models was the recent work of Hopfield [1982; 1984; Hopfield and Tank, 1985; 1986]. Hopfield (and later, Hopfield and Tank) emphasized four ideas, all of which have proven extremely fertile. Those ideas were: (1) there is a close analogy between neural network models with extensive feedback and random magnetic systems known as spin glasses; (2) the dynamical aspects of networks can be analyzed in terms of an energy function; (3) simple neural networks can be mapped onto traditionally difficult computational problems, yielding good, fast results; and (4) neural network models can be naturally realized in analog electronics. It could be argued that, in fact, none of these ideas was new. The connection between neural networks and magnetism, for example, dates back to the 1950s [Cragg and Temperley, 1954], and the energy function idea had also been used by Cohen and Grossberg [1983; Carpenter *et al.*, 1987]. The combination of ideas, however, along with tangible results and a clear exposition in a style familiar to physicists, managed to generate an excitement within the physics community which has proven to be both contagious and self-sustaining.

Many of the topics addressed in this thesis spring directly from the four ideas of Hopfield mentioned above. Most of the thesis will focus on various dynamical

properties of analog neural networks of the type described by Hopfield (see Fig. 1.1), with an emphasis on the practical, rather than the biological. The goal throughout will be to discover useful - and whenever possible, simple - results which can serve as guidelines for the design of fast, parallel computing devices. Occasionally, the relevance of a result to biology will be mentioned, but we stress at the outset that such insights are of secondary importance in this work.

The main conclusion of the thesis is that the input/output transfer function of the individual neurons greatly influences the collective dynamics of the whole network. Furthermore, for the restricted class of models considered, the nature of this influence can be analyzed and described quantitatively. From a practical point of view, this thesis demonstrates that analog neural networks have important computational advantages over corresponding models constructed from binary neurons. Thus, in emulating biology to make fast parallel computing machines, one is likely to find that the analog character of the neuron is an important aspect of the computational power of the system as a whole.

Chapter 2 gives a more detailed overview of the topics presented in this thesis, chapter by chapter.

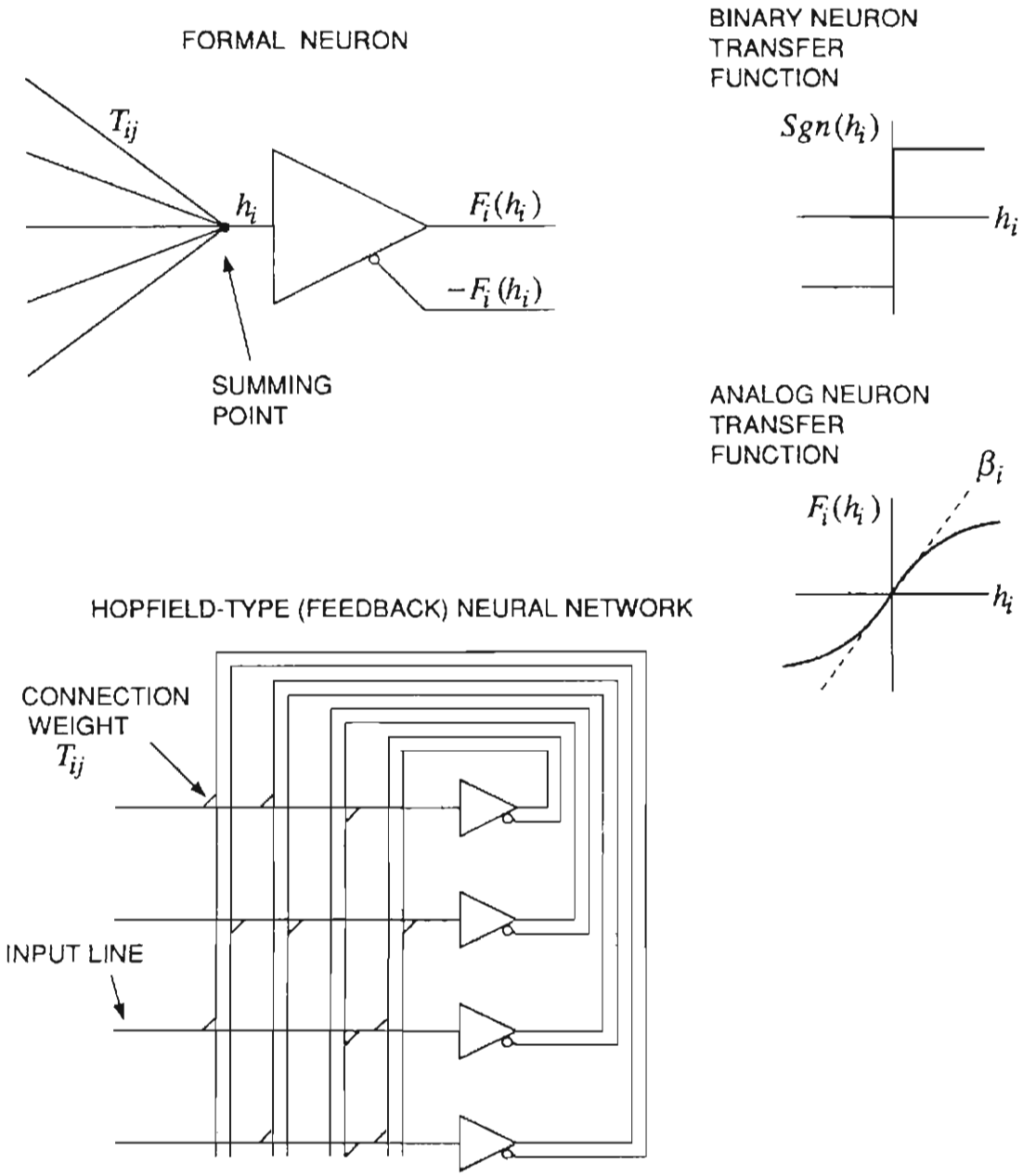


Fig. 1.1. The basic elements of the neural network model discussed in this thesis. Many variations will appear later. The formal neuron (or just "neuron") has an input h_i which is a weighted sum of the outputs from other neurons. The connection weight from neuron j to neuron i is given by the matrix element T_{ij}. The neuron output is a nonlinear function $F_i(h_i)$ of the input. This function is typically either binary or a continuous sigmoid (s-shaped) function. The network architecture we consider has extensive feedback, and we will frequently impose the symmetry condition ($T_{ij} = T_{ji}$).

Chapter 2

OVERVIEW OF THESIS[1]

This thesis addresses a variety of topics, mostly involving the dynamics, stability, and performance of analog neural networks. There are, however, many sections and even whole chapters (Ch. 8 and Ch. 10, for example) which depart from this subject. A consistent theme throughout the work concerns the dynamical behavior of nonlinear systems with many degrees of freedom. A variety of techniques will be used to explore these systems, including experiments, numerical investigations, and mathematical analysis.

In **chapter 3**, we describe an electronic analog neural network consisting of eight neurons built using operational amplifiers with nonlinear feedback, and accompanying circuitry to allow fast measurements of the basins of attraction for fixed points and oscillatory modes. After providing construction details, we present several measurements of the shapes of the basins of attraction in an analog associative memory. A notable feature of the network is the inclusion of charge-coupled device delay lines in each neuron. Delays are adjustable over nearly two orders of magnitude, allowing delay-induced instabilities to be studied experimentally, and critical values of delay to be measured in a variety of network configurations, and as other network parameters are varied.

In **chapter 4**, we consider the effect of time delay on the stability of symmetrically-connected analog neural networks from a more mathematical point of view. We present

[1] References have been stripped from this chapter to keep it short and easy to read. See subsequent chapters for references.

two stability criteria, based on local stability analysis, that give critical values of delay above which sustained collective oscillation appears. The surprising result is that the critical delay depends on only a few network parameters: the characteristic time of the network, the neuron gain, and the extremal eigenvalues of the connection matrix. Results are applied to several network configurations, including symmetrically connected rings, two-dimensional lattices of neurons, randomly connected networks, and associative memory networks. Results are found to be in good agreement with numerics and experiments performed on the electronic network. Finally we discuss chaotic dynamics in time-delay networks, and give an example of a three-neuron circuit with delay-induced chaos.

In **chapter 5** we study the stability and associative-memory capabilities of a discrete-time, analog neural network with *parallel* updating of neuron states. Parallel operation is crucial to the design of fast neural networks. The usual practice for discrete-time systems with binary neurons, however, is to update sequentially in order to prevent unwanted oscillation. We show that all oscillatory modes can be eliminated from a parallel-update analog neural network with symmetric connections by lowering the neuron gain below a certain critical value. The result is stated as a simple, global stability criterion relating the maximum neuron gain and the minimum eigenvalue of the connection matrix. This criterion allows "safe" parallel dynamics, with guaranteed convergence to a fixed point. Following this, we apply the analog network to the problem of associative memory, and present novel phase diagrams (in terms of neuron gain and the ratio of stored patterns to neurons) for the Hebb and pseudo-inverse learning rules. To our knowledge, these are the first reported analytical results of storage capacity for analog neural networks. Within the "recall" regions of the phase diagrams, where memory patterns are stable and have large basins, we find numerically that the performance of the associative memory improves as the neuron gain is lowered. This important observation, also noted by Hopfield and Tank and others, suggests the possibility of deterministic analog annealing.

In **chapter 6** we generalize the stability analysis of chapter 5 to include analog networks with an update rule based on an average over M previous time steps, for arbitrary M. Standard parallel updating corresponds to $M = 1$. The important result is that the critical value of neuron gain is increased for the multiple-time-step update rule by a factor of M, compared to standard parallel updating. Some applications to associative memories are then given. We also present a simple analysis of the convergence rate of the multiple-time-step network as a function of M.

In **chapter 7** we study the number of local minima in the dynamical (energy) landscape of the analog spin glass and the analog associative memory. We show that the expected number of local minima $\langle N_{fp} \rangle$ for both systems increases exponentially with the size of the system N, as $\langle N_{fp} \rangle \approx exp(aN)$. The scaling exponent a depends on the neuron gain for the case of the analog spin glass, and depends on both the neuron gain and the ratio of patterns to neurons for the analog associative memory. Analytical values for a are given for both systems. As neuron gain decreases, the value of a (for both systems) also decreases, which has the effect of dramatically reducing the number of local minima. These results provide an analytical framework for understanding how lowering the neuron gain can lead to improved performance in analog associative memories. Numerical observations of this effect are also presented in Ch. 5. Theoretical values for the scaling exponent a agree reasonably well with numerical values found by directly counting the fixed points in a large sample of computer-generated realizations.

In **chapter 8** we explore the basin structure of the deterministic (zero-temperature) SK spin glass. This model has been studied extensively and is known to possess an extremely rich energy landscape. The main result of this chapter is that the numerically measured distribution of basin sizes, averaged over realizations, obeys a power law with exponent near -3/2 over a wide range of basin sizes. The exponent of the power law appears to be independent of N. Some consequences of this power law are then

considered. The distribution of basin sizes in the deterministic SK model is qualitatively *different* from other closely related distributions which, among themselves, show certain universal features. Apparently, and perhaps surprisingly, the universality seen in these other distributions is *not* shared by the distribution observed here. We end this chapter by showing (again, numerically) that the distribution of basin sizes is strongly affected by the use of analog state variables. We find that reducing the gain in an analog spin glass selectively eliminates fixed points with small basins of attraction.

In **chapter 9**, we give some brief conclusions and remarks concerning unsolved problems and interesting future directions.

Chapter 10 is an appendix containing two papers on the dynamics of charge-density waves (CDWs). This work is essentially unrelated to neural networks, though it shares with the previous chapters a general theme of collective dynamics in nonlinear, many-body systems. The main idea in these papers is that a simple modification to allow phase slip in a previously-studied mean-field model of CDW dynamics causes the smooth depinning transition to become discontinuous and hysteretic. The behavior of the phase-slip model is very suggestive of *switching*, which is observed experimentally in certain CDW systems. The way that phase slip is introduced in this model has the added virtue of making the system analytically tractable.

Chapter 3

THE ELECTRONIC ANALOG NEURAL NETWORK

3.1. INTRODUCTION: WHY BUILD ELECTRONIC HARDWARE?

Soldering gives a person lots of time to think. One particularly deep question to think about while soldering together an electronic neural network is what distinguishes an experiment from a simulation, or, in other words, why build this circuit? Among neural networks researchers, there is a large camp of non-apologists who view the mathematical system as *the* neural network, rather than considering the equations to be a simplified description of some physical reality [see the discussion of Maddox, 1987]. From this perspective, an electronic neural network serves as a fast analog computer for simulating the "real" (mathematical) system. A more engineering-minded line of thought emphasizes the potential for building powerful computational devices. Because microelectronics, and particularly VLSI, is the likely medium for implementing these devices [Mead, 1989], it is important (the argument goes) to learn as much as possible about real circuits and the behavior of large, interconnected electronic networks. By this reasoning, building an electronic network from discrete components is progress toward the ultimate goal of building a "real" neural network (i.e. a large, fast and truly useful piece of electronic hardware).

Apart from these bigger questions of motivation is the simple fact that many important problems in neural networks (especially analog neural networks) are difficult to treat analytically or by conventional numerical simulation. Occasionally, such problems can

be studied easily and directly in a small electronic network. After presenting the details of our circuit in §3.2, we will consider two problems of this sort. They are:

(1) *What are the shapes - not just the volumes - of the basins of attraction for the recall states of an associative memory?* In a well-designed associative memory, the basins of attraction for recall states should be large, but that is not sufficient: the basins must also be roughly spherical (by some appropriate measure) and centered about the recall states. If the basins of attraction are diffuse or disconnected in state space, the memory will not be useful. In § 3.4 we show that the shapes of the basins for recall states are in fact somewhat irregular when the network is overloaded with memories.

(2) *How does time delay affect the transients, attractors and basins of attraction in a neural network?* This problem is of particular interest to the engineering-minded camp, as the operating speed of VLSI circuitry will likely be limited by switching-delay-induced instabilities (for a discussion of delays in VLSI, see [Mukherjee, 1985, Ch. 6]). Much of the mathematical analysis of networks with time delay that appears in chapter 4 was suggested by or confirmed using the electronic network.

Electronic circuits have also been used to find and characterize chaotic behavior in analog neural networks [Marcus and Westervelt, 1989b; Kepler *et al.*, 1989]. This application will be discussed in § 4.6.

3.2. CIRCUITRY

In this section we provide a detailed description of the electronic neural network circuit. First, though, we give a quick overview of the circuit's main features:

The electronic network consists of eight analog neurons (nonlinear amplifiers) connected via 128 manual switches and resistors. Connections between pairs of neurons can be noninverting, inverting, or open, depending on the positions of these 128

switches. Each neuron has an independently adjustable gain and saturation level, and has a time delay section based on a charge coupled device (CCD) analog delay line. (The reader may wish to glance ahead at Fig. 3.4 at this point.)

The dynamical equations for the voltages $u_i(t)$ on the input capacitors of the neurons (nonlinear amplifiers) are

$$C_i \dot{u}_i(t') = -\frac{1}{R_i} u_i(t') + \sum_{j=1}^{N} T'_{ij} f_j\left(u_j\left(t' - \tau'_j\right)\right) \qquad i = 1,\ldots, N, \qquad (3.1)$$

where C_i is the neuron input capacitance and $R_i = (\Sigma_j |T'_{ij}|)^{-1}$ is the resistance to the rest of the circuit at the input of neuron i, and f_i is a smooth sigmoid function describing the transfer function of the i^{th} neuron. Equation (3.1) is identical to the analog system described by Hopfield [1984], with the inclusion of time delay. It is not equivalent, however, to some other hardware implementations which have the input capacitor across, rather than in front of, the nonlinear amplifier [Denker, 1986c; Amit, 1989; Kepler *et al.*, 1989].

Digital timing circuitry and voltage-controlled analog switches are used to periodically open the feedback path from the resistor matrix to the neuron input and load initial conditions onto the neurons' input capacitors. The initial conditions are determined by eight independent voltages, any two of which can be raster-scanned using independent function generators. At the same time, the two function generators are used to position the beam of a storage oscilloscope (Conographic 611). When the state of the network matches some reference state (which has been set with manual switches), the beam of the storage oscilloscope is turned on, and the resulting pattern on the storage oscilloscope shows an image of the basin of attraction for the reference state in a two-dimensional slice of initial condition space. Alternately, the oscilloscope beam can be set to go on only when the circuit enters an oscillatory state, thus illuminating a slice of a basin of attraction

for oscillation. Neuron outputs can also be displayed directly (as X vs. Y, for any pair of neurons) using a second storage oscilloscope (Tektronix 611). The time scale for a complete load/run cycle is adjustable, and is typically ~10-40 *ms*.

The following three subsections provide the details of the various parts of the circuit.

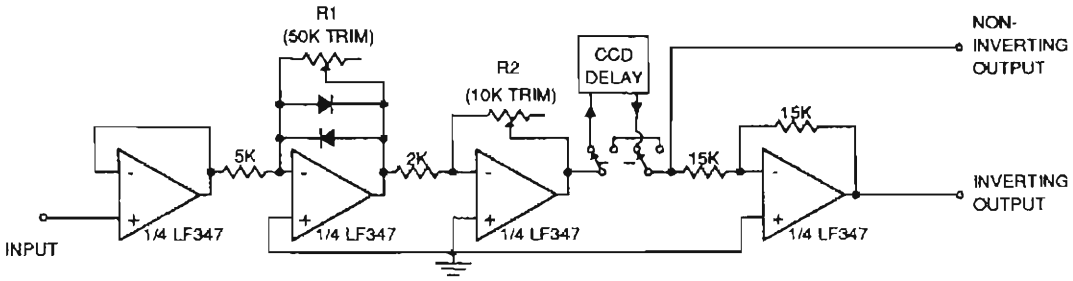
3.2.1. Neurons

The schematic for an individual analog neuron is shown in Fig. 3.1. Each neuron uses four JFET operational amplifiers (op-amps), all on a single 14-pin integrated circuit (National Semiconductor LF374N). Starting at the input side of the neuron, the first op-amp serves as a unity gain buffer, giving the neuron a high input impedance. The second op-amp, with diodes in the feedback, is the nonlinear part of the circuit, giving the neuron its sigmoidal or saturating transfer function, as discussed below. The third op-amp serves as a variable gain amplifier and sets the overall amplitude of the output. Next in the signal path is the CCD delay (see: § 3.2.2 below), which can be switched in or out independently for each neuron. Finally, the fourth op-amp inverts the output to allow inhibitory as well as excitatory connections.

The neuron transfer function f (dropping the subscript i) is defined by the relation $f(input) = output$, where *output* refers to the neuron's noninverting output. The function f is made interesting by the diodes in the feedback path of the second op-amp. To derive an expression for f, we start with a simple form for the current-voltage (I-V) characteristic of a diode [see: Sze, 1981, § 2.4]

$$I = I_s\left[exp\left(\frac{V}{V_T}\right) - 1\right] \quad . \tag{3.2}$$

The parameters $I_s = 2.9 \times 10^{-5} mA$ and $V_T = 5.9 \times 10^{-2} V$ were determined by a least-



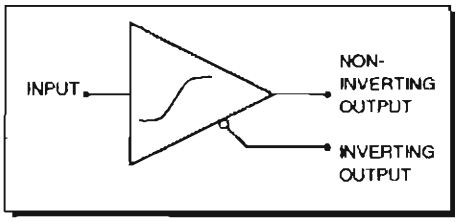


Fig. 3.1. Schematic diagram of analog neuron.

square fit to data in the manufacturer's data sheet for the diode used, which was the 1N914. Equation (3.2) and standard op-amp circuit analysis (i.e., the principle of virtual null) give the following implicit expression for f,

$$input = [5k\Omega]\left(\frac{V}{R1} + 2I_s \sinh\left(\frac{V}{V_T}\right)\right); \tag{3.3a}$$

$$V = \left(\frac{[2k\Omega]}{R2}\right) output, \tag{3.3b}$$

where the resistance values from Fig. 3.1 are shown in square brackets. Figure 3.2 shows the neuron *output* as a function of its *input* as given by Eq. (3.3) for different values of $R1$, with $R2$ held fixed at 2 $k\Omega$ and numerical values for I_s and V_T inserted. For large and small signals, Eq. (3.3) can be expanded to leading order to give

$$output = \left(\frac{R1\ R2}{10(k\Omega)^2}\right) input \quad \text{(for small signals, } \textit{linear} \text{ regime)}, \tag{3.4a}$$

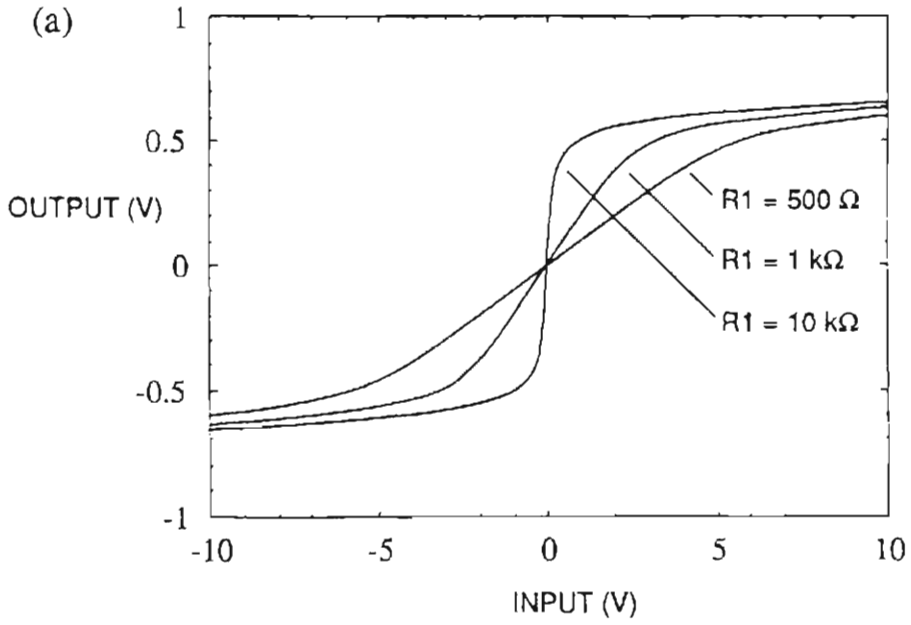
$$output = \frac{R2}{[2k\Omega]} ln\left(\frac{input}{[5k\Omega]I_s}\right) \text{ (for large signals, } \textit{saturated} \text{ regime)}. \tag{3.4b}$$

The crossover from the linear to the saturated regime occurs when

$$input \sim \left(\frac{[5k\Omega]}{R1}\right) V_T . \tag{3.5}$$

The maximum slope of the neuron transfer function, defined as the *neuron gain* β, will be very important for all sorts of analysis in later chapters. From (3.4a), we can

(a)



(b)

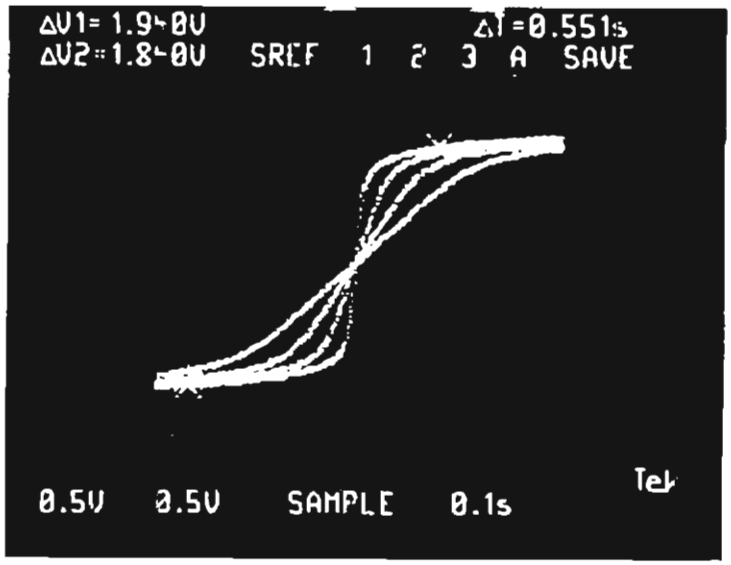


Fig. 3.2. The nonlinear neuron transfer function. (a) Theoretical transfer functions based on Eq. (3.4) for different values of $R1$, with $R2 = 2k\Omega$ (see Fig. 3.1). (b) Transfer function measured in the electronic neuron for different values of $R1$.

immediately identify the neuron gain,

$$\beta = \left(\frac{R1\,R2}{10(k\Omega)^2} \right). \tag{3.6}$$

Notice from Eq. (3.4b) that for large signals, the neuron output saturates to a logarithmic function of the input. This behavior is different than the *tanh* function - the canonical sigmoid - which saturates to a hard limit at large argument. In practice this difference does not appear to be significant.

3.2.2. Analog delay

The schematic for the analog delay circuit is shown in Fig. 3.3. Each neuron has its own independent delay circuit, which can be switched in or out manually. The heart of the circuit is a charge-coupled device (CCD) analog delay line, RD5106A, manufactured by EG&G-Reticon. The RD5106A chip is a so-called "bucket-brigade" device. The device operates by charging an input capacitor to the instantaneous (analog) input voltage, and then passing that charge along a brigade of 256 subsequent capacitors, with each transfer of charge triggered by a pair of pulses from an external clock. At the end of the brigade of capacitors, the charge is converted back to an analog voltage which constitutes the output signal. The time τ' taken to traverse the entire brigade (i.e. the delay time) is related to the clock frequency f_{clock} by:

$$\tau'\,[ms] = \frac{0.512}{f_{clock}[MHz]}. \tag{3.7}$$

The shortest delay available from this chip is nominally 300μs (f_{clock} = 1.7 MHz),

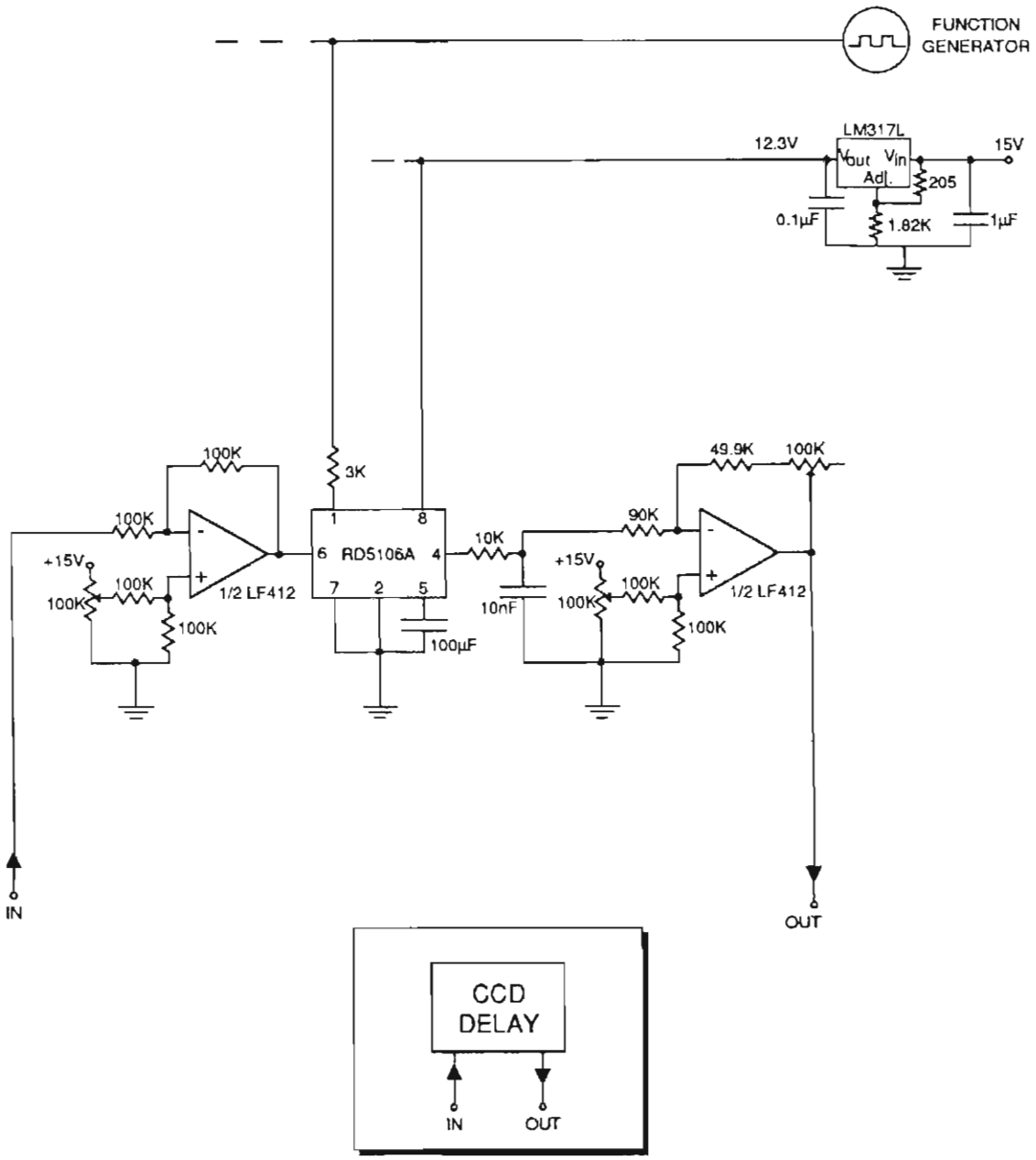


Fig. 3.3. Schematic for analog delay circuit. The entire circuit (excluding the voltage regulator section) is duplicated for each neuron. The heart of the circuit is the EG&G Reticon RD5106A charge coupled device analog delay line.

although the appearance of dc offsets limits the usable range to $\tau' > \sim 450\ \mu s$ ($f_{clock} < 1.2\ MHz$). On the long-delay end, the chip itself is good up to delays exceeding one second, but in practice the delay is limited by a "bucket discretization noise" at a frequency $f_{noise}[kHz] = 256/(\tau'[ms])$. The network itself will filter out this noise as long as f_{noise} is well above the network's bandwidth, which is in the range 1 - 8kHz depending on the connection matrix (see next subsection). This gives a range of delay covering nearly two orders of magnitude:

$$450\mu s \ < \ \tau' \ < \ \sim 30ms\,. \tag{3.8}$$

Because all delay circuits are clocked using the same function generator, all delays (when switched in) are identical. It would be very simple to construct individual on-board trigger circuits using LM555's to allow independent delays.

The rest of the delay circuitry is mostly used to get around one unfortunate aspect of the CCD delay line, which is that input voltages must be positive. The first op-amp is used to add an adjustable dc offset to the input signal, and the second op-amp is used to remove that offset. The second op-amp also has adjustable gain to compensate for the RD5106A not being exactly unity gain. All offsets and gains are independently adjustable via three trim-pots per delay circuit. There is also a 10 kHz low pass filter section between the delay line and the second op-amp, to remove bucket discretization noise. Finally, a single voltage regulator (National Semiconductor LM317L) is used to supply the required 12.3 V to all eight RD5106A's.

3.2.3. Network, measurement and timing circuitry

A. The network

Figure 3.4 shows the layout of the entire network. A single circuit element in a box represent multiple identical components: 8 delay-neurons, 8 initial condition loaders, 8 output buffers (LM741's), and 64 resistor-switch interconnect circuits. The characteristic relaxation time of the network (without delay) is determined by the interconnect resistances and neuron input capacitance. For interconnect resistances $(T'_{ij})^{-1} = 100k\Omega$ and input capacitances $C_i = 10nF$, the network relaxation time is $1/n$ $[ms]$, where n is the number of neurons connected to the input of any given neuron. The characteristic relaxation time for the electronic network can easily be varied over a few orders of magnitude by replacing the input capacitors, which are installed using plug-in connectors. Indeed, the entire circuit could be sped up considerably, with characteristic times in the tens of microseconds, without pushing the bandwidth of any of the integrated circuits; the limiting factor is the delay, which cannot be less than 450 μs. The characteristic time to load initial conditions is $(10\ k\Omega)(10\ nF) = 100\ \mu s$.

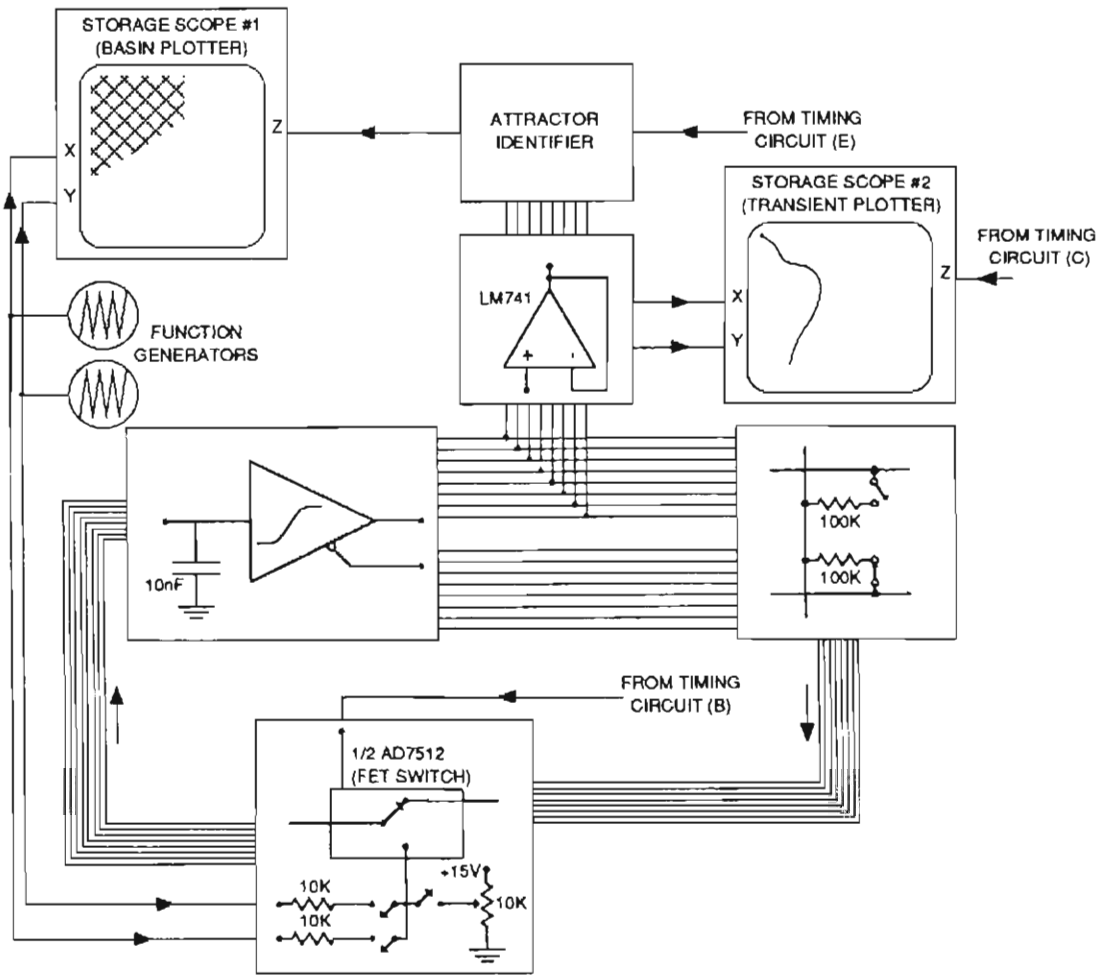


Fig. 3.4. Schematic diagram of the entire network. Single devices in boxes represent multiple devices: 8 each of the neurons, run/load sections, and LM741 buffer; 64 of the resistor-switch pairs. Storage oscilloscope #1 is used to plot slices of basins of attraction; storage oscilloscope #2 is used to plot the outputs of any pair of neurons as X vs. Y. The source of the various timing signals is shown in Fig. 3.6.

B. Attractor Identifier

The box marked "attractor identifier" in Fig. 3.4 is shown in detail in Fig. 3.5. This circuit is used to test if the network has settled onto a specified fixed point attractor or, alternately, to determine if the network is in an oscillatory mode.

The part of the attractor identifier circuit marked "fixed-point attractor identifier" (the larger dashed box in Fig. 3.5) works as follows: First, eight comparators (LM311's) are used to convert the analog state of the network into eight thresholded digital (TTL) signals. This eight-bit digital state is then compared bit by bit to a reference state, which has been selected by positioning eight manual switches. The reference state might be, for example, a programmed memory pattern. If *all* eight bits of the network state match the reference state, then the line leaving the fixed-point section is set high, otherwise, it is set low.

The part of the attractor identifier marked "oscillation detector" (the smaller dashed box in Fig. 3.5) uses a retriggerable 1-shot (96LS02) with a high-time that is set to be longer than the period of oscillation under investigation. If the comparator undergoes a state transition within the high-time of the 1-shot, the output of the 1-shot remains in a high state. If the output of the comparator remains fixed - because the neuron being observed has stopped oscillating - the 1-shot goes low at the end of its current high-time. The high-time of the 1-shot can be continuously varied from 4.3 *ms* to 8.6 *ms* via an external potentiometer. Note that an oscillating neuron must cross zero output in its excursion to trigger the oscillation detector.

Where the fixed-point and oscillation detectors come together there is a bit more digital logic, and another manual switch. Depending on the position of this switch, the TTL output which goes to a sample/hold (AD583KD) indicates one of the two following conditions:

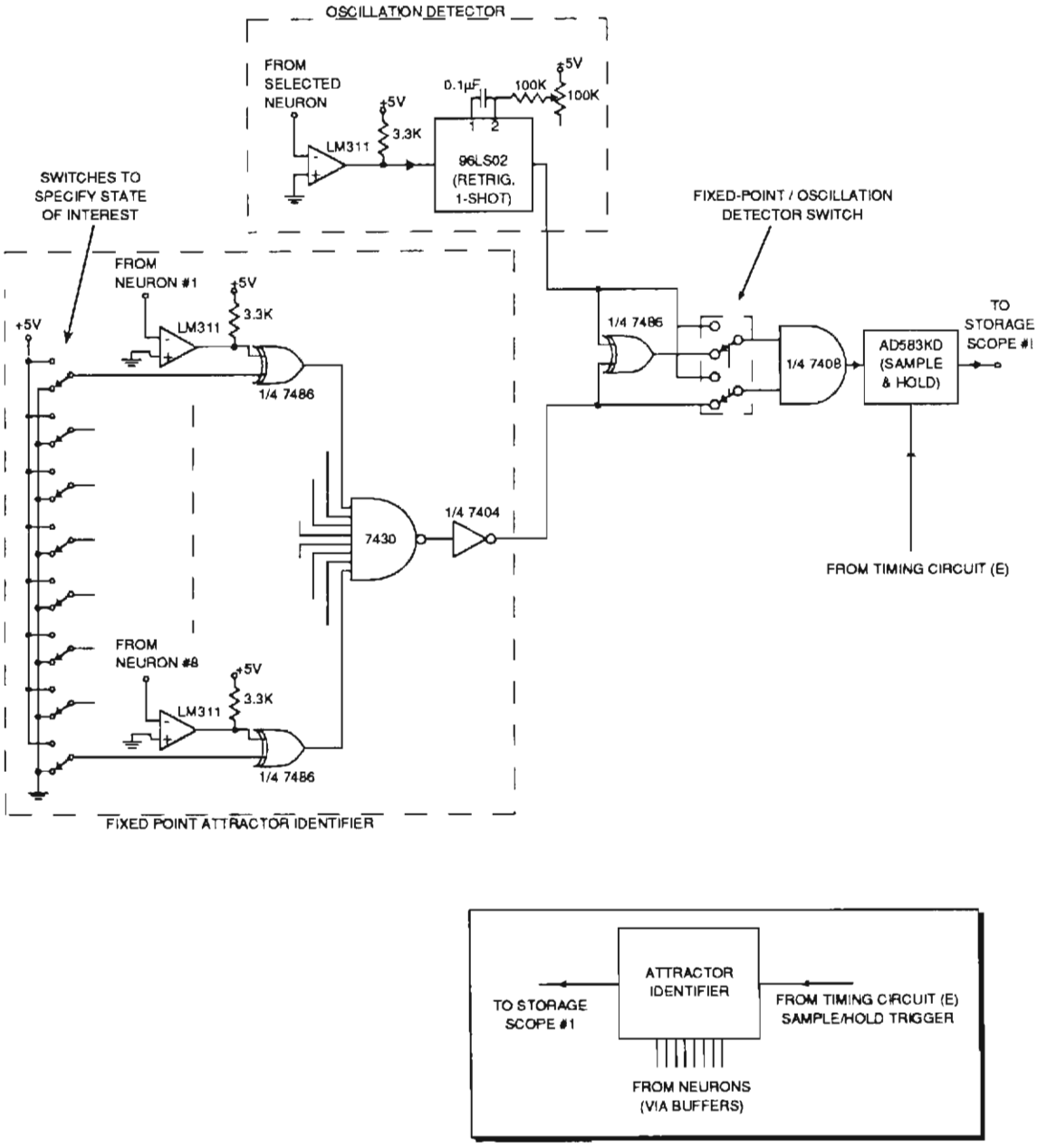


Fig. 3.5. Schematic for attractor identifier circuit.

Switch up: [selected neuron is oscillating]

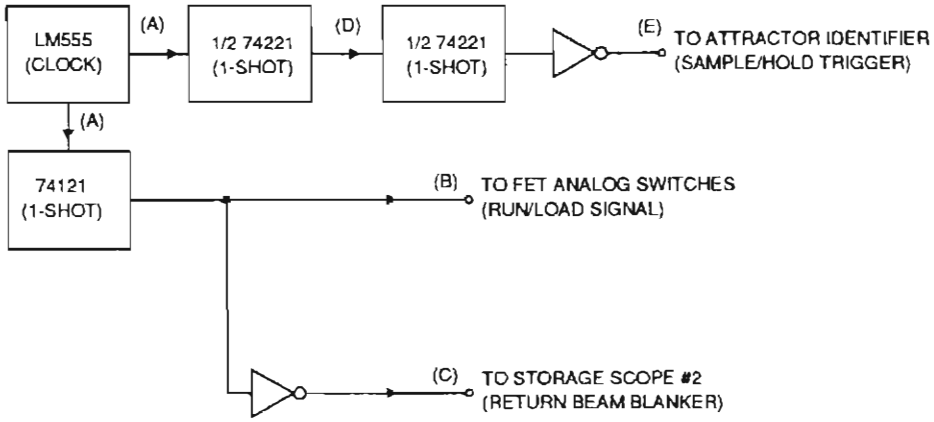
Switch down: {[network state matches reference state]
AND
[selected neuron is not oscillating]}.

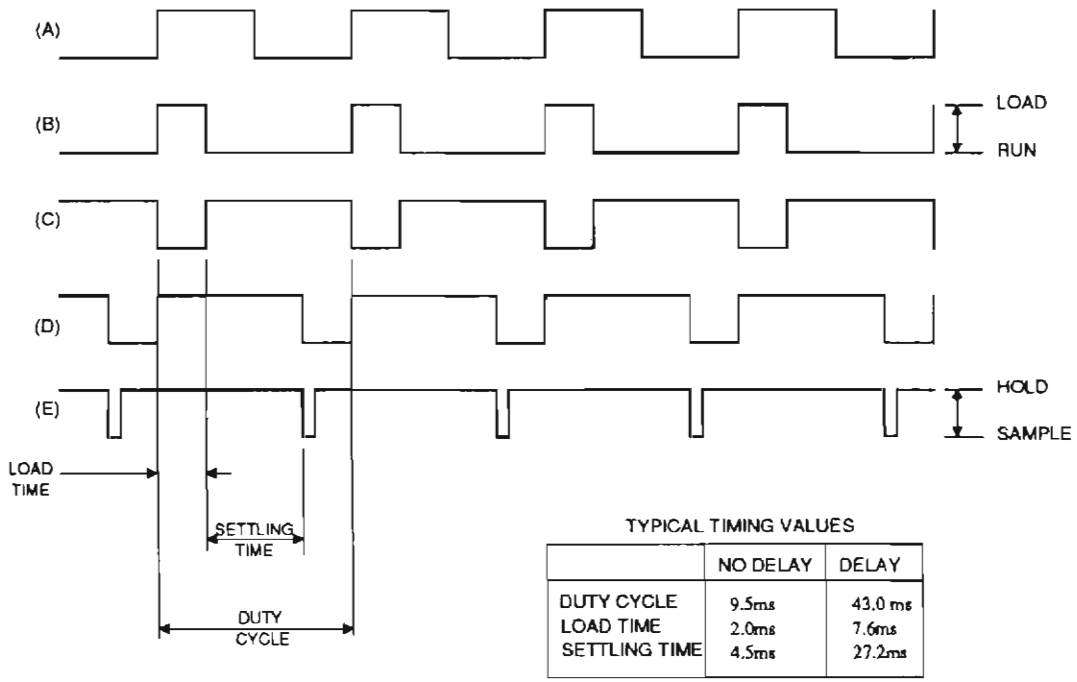
The combined logic for the *Switch down* position insures that all matches to the reference state are actual fixed points, not oscillatory modes or transients.

C. Timing

The timing signals appearing in Figs. 3.4 and 3.5 are supplied by the digital (TTL) circuit shown in Fig. 3.6. Also shown in Fig. 3.6 are TTL logic states (low = 0*V*, high = 5 *V*) as a function of time at several points in the timing circuit, labeled (A) - (E). Trace (B) is the run/load signal sent to the voltage-controlled analog switches; trace (E) is the "sample now" signal sent to the sample/hold in the attractor identifier circuit. The time between when (B) goes low (network feedback path reconnected) and when (E) goes low (state of the system sampled by sample/hold) defines the allowed settling time of the network. This value is adjusted to be ~5-10 times the network relaxation time, so that nearly all transients have died out by the time a new "sample now" signal is sent. When delays are used, the allowed settling time must be quite long, as indicated in the table of Fig. 3.6.

Examples of network dynamics are shown in Figs. 3.7 and 3.8. Figure 3.8 shows that in addition to creating sustained oscillatory modes, delay can induce extremely long and complicated transients when the circuit converges to a fixed point. Long, delay-induced transients were also investigated by Babcock and Westervelt [1986b].

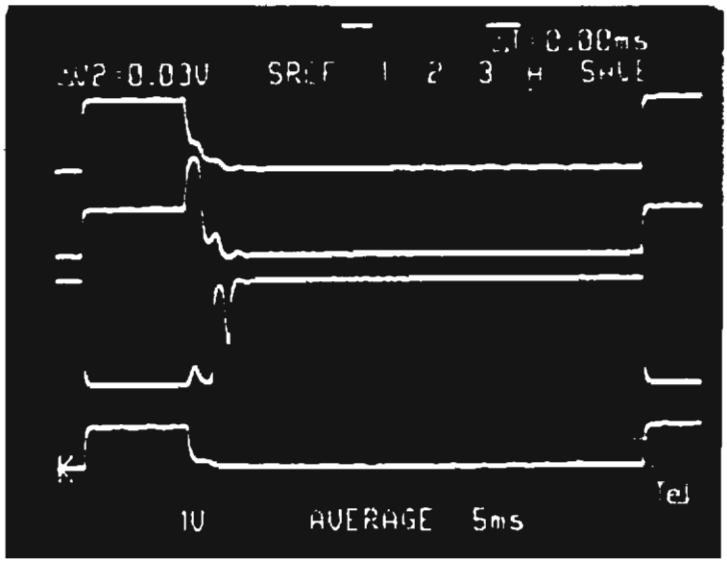




TYPICAL TIMING VALUES

	NO DELAY	DELAY
DUTY CYCLE	9.5ms	43.0 ms
LOAD TIME	2.0ms	7.6ms
SETTLING TIME	4.5ms	27.2ms

Fig. 3.6. Schematic for timing circuit. (a) The TTL signals labeled (A) - (E) control various parts of the circuit and appear in the other schematics. (b) TTL timing diagrams and typical timing values (insert).



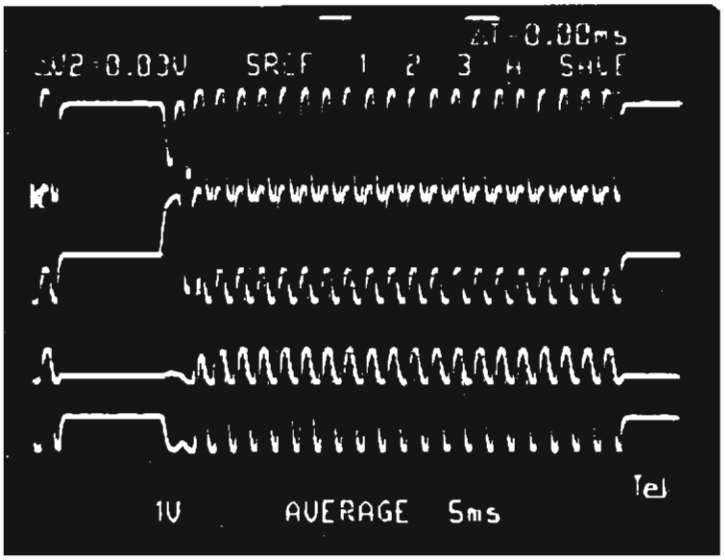


Fig. 3.7. The output voltages of 4 of the 8 neurons as a function of time for the network with randomly selected symmetric connections and delay. The two pictures are for the same network, the only difference is a slight change in initial conditions. (a) Initial conditions lead to a fixed point after ~ 5 *ms*. The end of the trace shows a return to the initial condition as the run/load cycle is repeated. (b) A slight change in initial conditions for the same circuit leads to a sustained in-phase oscillatory mode.

Fig. 3.8. A delayed neuron response function can induce long, complicated transients, as illustrated here, in addition to inducing sustained oscillation.

3.3. BASINS OF ATTRACTION IN 2-D SLICES

A neural network consists of more than a just set of attractors embedded into the state space of a dynamical system. To design a well-behaved network, one must also consider the structure of the basins of attraction. Indeed, one standard figure of merit for an associative memory network is the average size of the basins of attraction for the embedded memories [Forrest, 1988]. Size, in this context, means the volume of state space which flows to a particular attractor. One also speaks of a basin's radius, which is the distance from an attractor (in some appropriate metric, for example the number of differing bits in a binary network) at which the probability of flowing to that attractor drops off quickly. It has been demonstrated that the radius of a basin of attraction is intimately related to the strength with which a pattern is embedded by a learning rule [Kepler and Abbott, 1988].

Figure 3.9 shows a highly schematic view of different basins of attraction to illustrate how radius alone does not fully characterize the quality of a basin of attraction for producing good associative recall. Aside from having a large radius, a good basin should also be compact, spherical (roughly equal radii in all directions), centered on the attractor, and smooth.

The shapes of basins of attraction for Hebb-rule associative memories have been investigated by Keeler [1986] for large (N = 200) networks of binary neurons with sequential dynamics. Keeler used a clever scheme to reduce the high-dimensional state space to only two dimensions by projecting distances from a reference state (e.g. the locus of an attractor) onto a random direction and its complement. This projection scheme preserves topology such that neighboring points in the full N-dimensional space are also neighbors in this projection. Keeler found that as the number of stored patterns approaches the storage capacity of the network, the basins of attraction for the recall

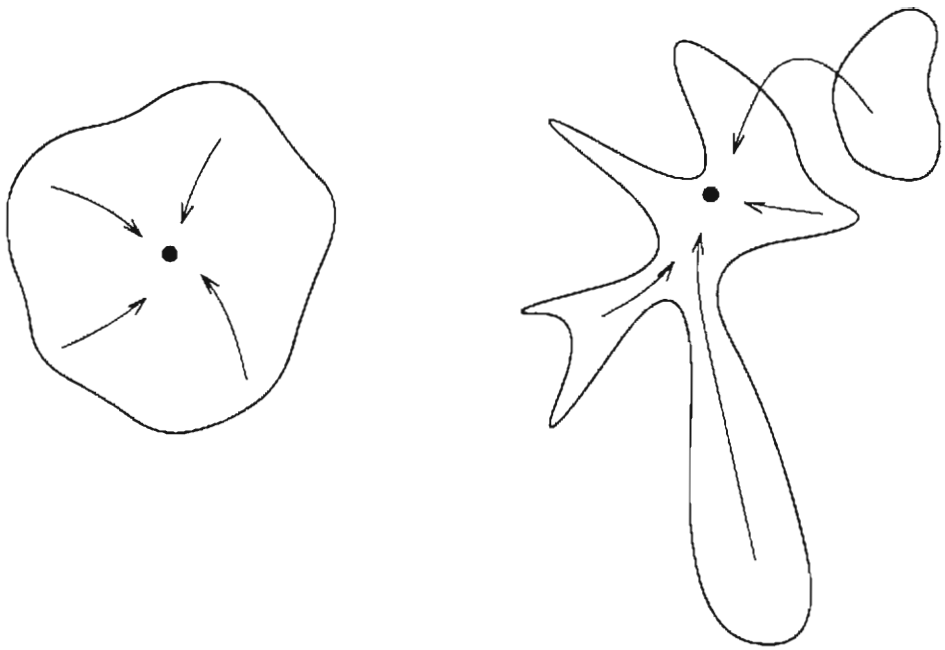


Fig. 3.9. A highly schematic representation of two basins of attraction having roughly the same volume, but different shapes. Designing a network to have large basins is not sufficient: basins must also be smooth, regular in shape, and centered on the attractor in order to yield reliable performance. The basin on the right fails to meet these criteria.

states - as seen in this representation - become highly irregular, disconnected and filled with "large crevices and holes," in his words. It is unclear whether Keeler's findings are an artifact of his algorithm for compressing a 200-dimensional space into a 2-dimensional slice, or if they indicate an important and previously unsuspected shortcoming of Hebb-rule associative memories.

In analog networks, where the state space is continuous, one faces the additional complication of having to consider dynamics on the interior of the hypercubic state space. What is the basin structure for analog networks *within* the hypercube? Is the space cleanly cleaved and parcelled evenly among the memories, or is the inside of the hypercube a tangled knot of intersecting hypersurfaces?

A final question concerns the effect of delay on the basin structure. Certain nonlinear delay-differential systems are known to possess fractal basins of attraction [Aguirregabiria and Etxebarria, 1987]; one might suspect that neural networks with delay could show similar behavior. A fractal basin structure would be highly undesirable in an associative memory.

To address these questions, we have measured the basins of attraction in two dimensional slices of state space for both fixed-point and oscillatory attractors using the electronic network and basin identification circuitry. The present method of slicing up state space is simpler than the one used by Keeler - slices are viewed directly on the storage oscilloscope - and is designed to probe the basin structure on the interior of the hypercube. Each slice is generated by holding fixed all but two of the initial voltages sent to the neurons, while initial voltages sent to the remaining two neurons are raster-scanned using a pair of function generators with triangle-wave output. The raster periods (~1 *s*) are chosen to be much longer than the run/load cycle time (see Fig 3.6), so that roughly 100 data points (beam on or off) are generated each time the beam crosses the screen. Changing one of the non-rastered initial voltages moves the location of the slice in the

direction in state space associated with the neuron receiving that initial condition. There are (N - 2) directions perpendicular to the plane of the slice - one for each of the neurons receiving non-rastered initial voltages. From a series of slices, one can infer the basin structure in higher dimensions. This is illustrated in Fig. 3.10 for the simple case of three neurons with symmetric positive (ferromagnetic) coupling. Notice that the basins of attraction for the two ferromagnetic states - all neurons saturated positive or all saturated negative - divide state space in a smooth, symmetric way.

3.4. MEASUREMENTS WITHOUT DELAY

We have investigated the basin structure for an eight-neuron associative memory using a clipped form of the Hebb rule [Denker, 1986],

$$T'_{ij} = \frac{1}{100k\Omega}\ Sgn\left[\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu\right], \quad i,j = 1,\ldots,8. \tag{3.9}$$

Figure 3.11 shows a series of slices through the 8-D state space for an associative memory storing three patterns (thus six programmed attractors, including the inverses of the memories). The slices shown are in the plane defined by rastering on neurons 1 and 2. In each of the four pictures, the initial condition on neuron 5 was set to a different dc value, while the initial conditions of the other neurons (3,4,6,7,8) were fixed at 0 V. Different basins in a single picture were distinguished by a using a different raster pattern, as determined by the relative frequencies of the two function generators. Several basins were imaged in the same picture by manually disconnecting the attractor identifier (Fig. 3.5) from the oscilloscope after generating the first basin image, then resetting the switches on the attractor identifier to the next memory state, changing one of the function

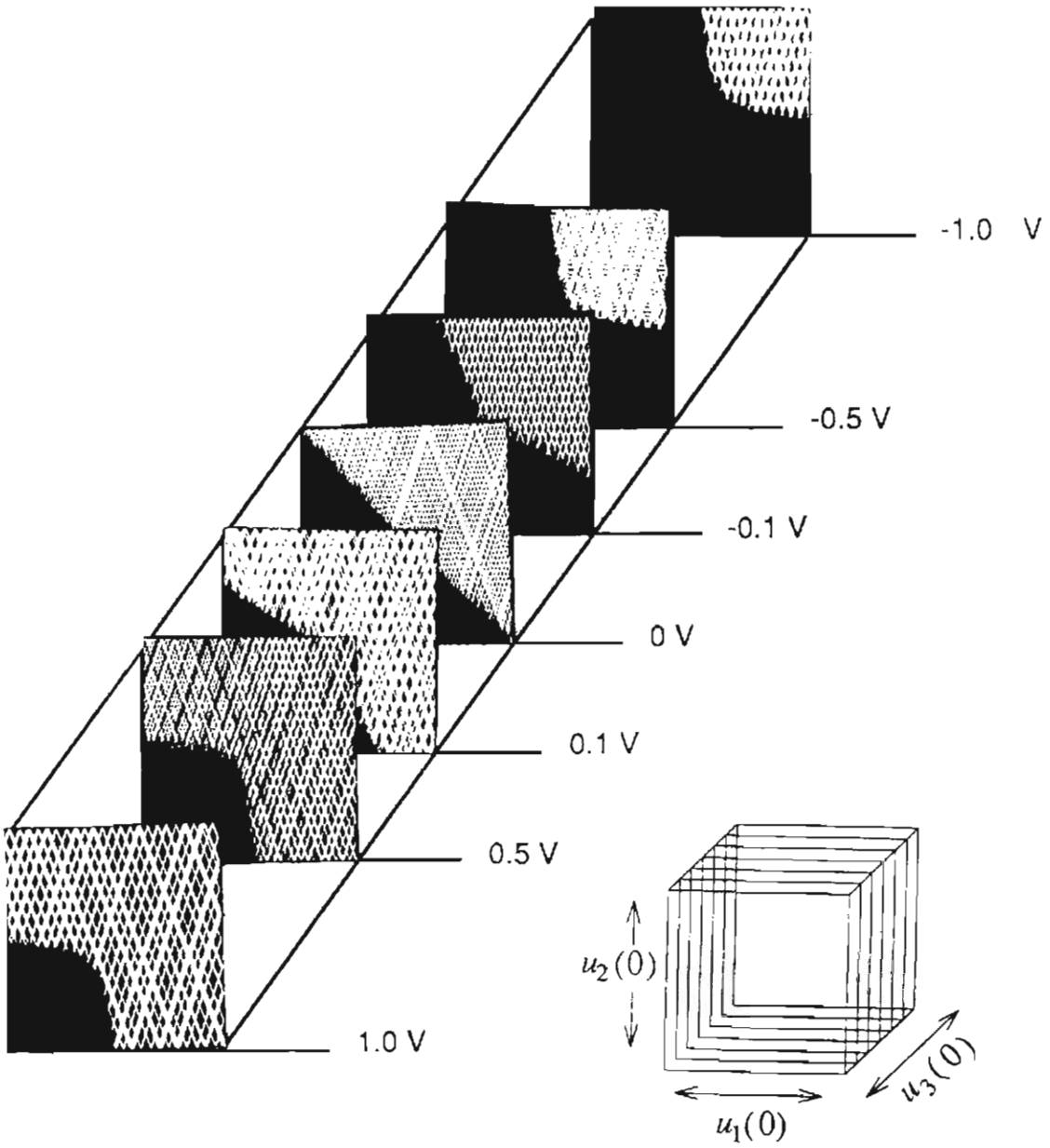


Fig. 3.10. Basin structure for three neurons with symmetric positive (ferromagnetic) coupling. Slices were produced by rastering the initial voltages for neurons 1 and 2 between ± 1*V*, while the initial voltage on neuron 3 was held fixed in each slice. The value of the initial voltage on neuron 3 differs in each slice, as indicated. The hatched region marks initial conditions leading to the attractor with all neurons saturated positive $(\uparrow\ \uparrow\ \uparrow)$, the black region marks initial conditions leading to the attractor with all neurons saturated negative $(\downarrow\ \downarrow\ \downarrow)$. Neuron gains were all $\beta \sim 10$.

generator frequencies to make a new raster pattern, and then reconnecting the attractor identifier to the oscilloscope to generate the next basin image. By repeating this process, many basins could be shown in a single picture, although usually no more than four basins were present in any one slice (six was the most observed for any network configuration).

Figure 3.11 suggests that the electronic network works extremely well as an associative memory, despite the fact that with three memories and eight neurons, it is loaded well above the nominal storage capacity for the clipped Hebb rule, $p/N \cong 0.1$ [Sompolinsky, 1986]. When initial conditions lie outside the hypercube (defined by the saturation voltages of the neurons), the basin shapes become more distorted. This is illustrated in Fig. 3.12 for the same connection matrix as in Fig. 3.11, only now the slice is in the place defined by rastering on neurons 2 and 4. To the extent that this distortion is a problem, it can easily be avoided by limiting initial conditions to lie within the range of the neuron outputs.

The take-home message of this subsection is that the electronic associative memory works extremely well, despite clipping and overloading. So well, in fact, that the results are somewhat uninteresting: the network did just what one might guess (or hope) that it would do. Grossly distorted basins or attraction were not observed, even when the connection matrix was deliberately corrupted by randomly altering several matrix elements. In all cases, basin boundaries appear smooth, and, within the hypercube, they are also quite straight. Far outside the bounds of the hypercube, basin shapes become somewhat irregular, but not very much so; certainly they do not appear to be disconnected or fractal.

We emphasize the difference between our measurements and those of Keeler [1986]. In Keeler's 2-D slices, each *point* in the slices represents a corner of the hypercube, and the interior of the hypercube is not part of the state space. Therefore, our results do not

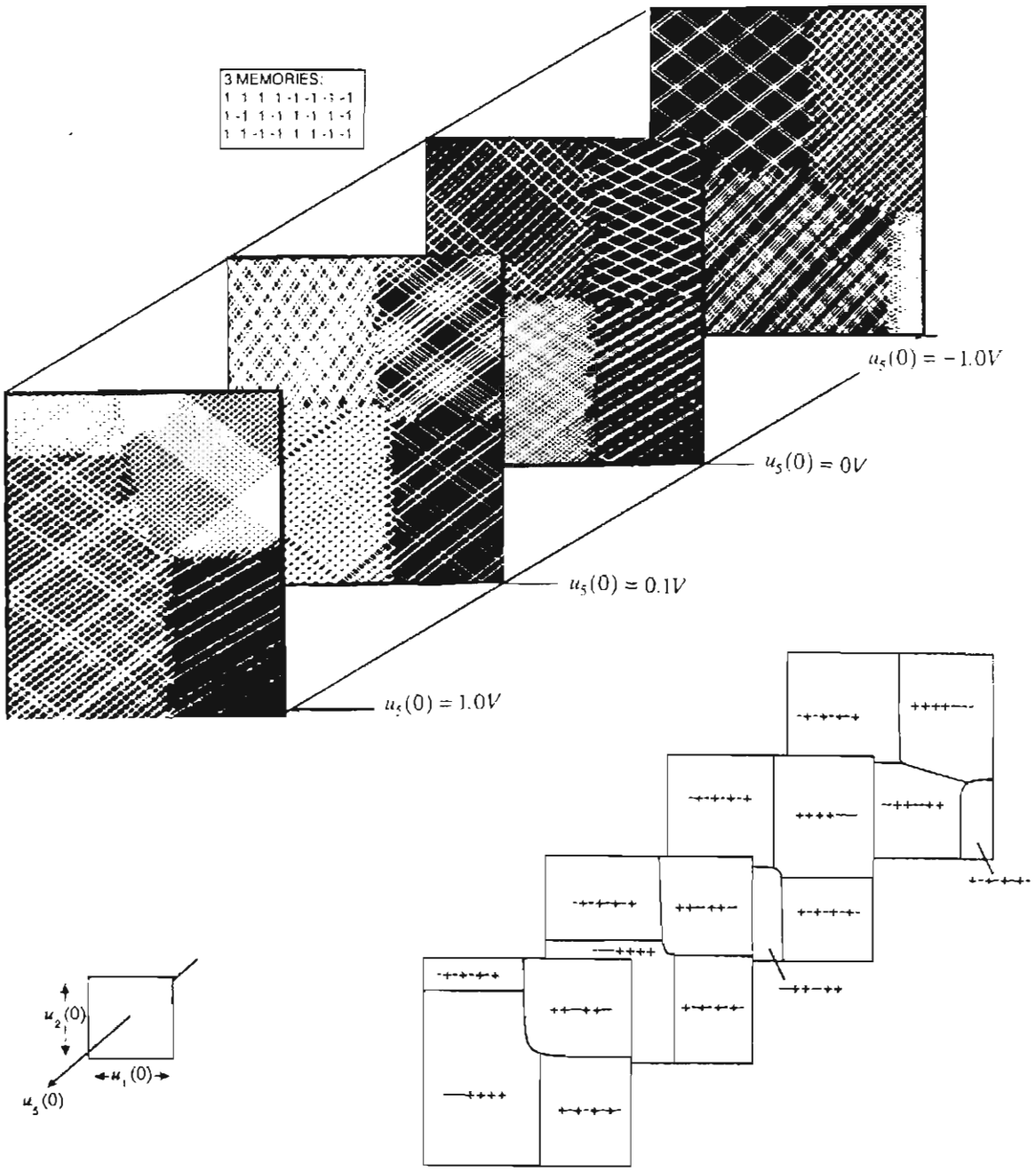


Fig. 3.11. Basin structure for eight-neuron circuit storing three memory patterns with a clipped Hebb rule, Eq. (3.9). Different rastering patterns mark initial conditions leading to the various memory patterns and their inverses. The attractors associated with each region are indicated: + means saturated positive, - means saturated negative. Despite overloading, no spurious attractors are observed and basin shapes appear regular.

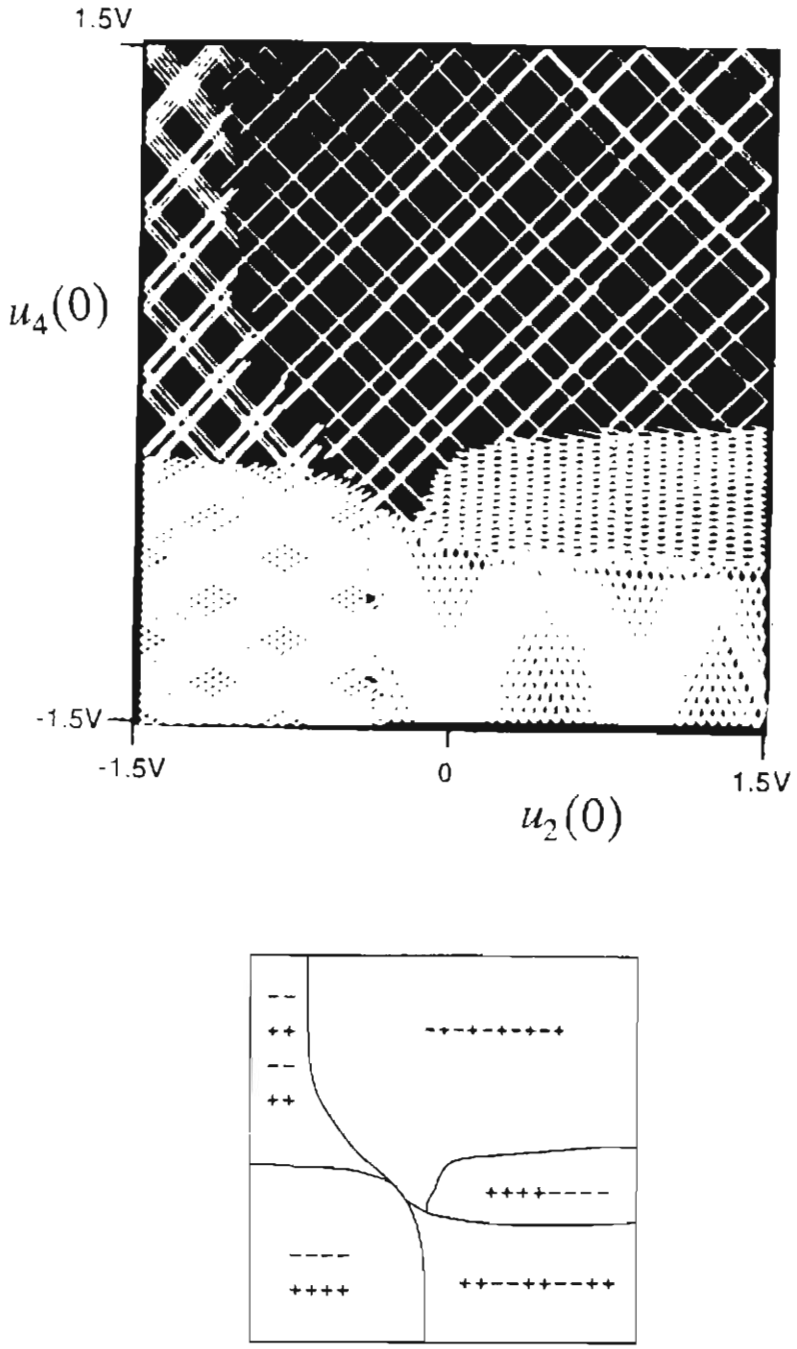


Fig. 3.12. Basin structure becomes more convoluted when some of the initial conditions lie outside the range of the neuron outputs. The network configuration here is identical to that of Fig. 3.11. Non-rastered initial conditions are: $u_3(0)$ = -0.0V (?); $u_5(0)$ = –4.6V; $u_6(0)$ = -1.2V; $u_7(0)$ = -2.26V; $u_8(0)$ = -0.06V.

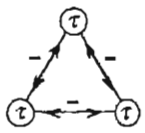
contradict those of Keeler, as the spaces represented in the two studies are entirely different. Furthermore, it may be that the complicated basin structure is only seen for systems considerably larger than $N = 8$.

3.5. MEASUREMENTS WITH DELAY

The basin structure becomes more interesting when delay is introduced into the response of the neurons. We concentrate here on symmetrically connected networks, which possess only fixed points and simple periodic attractors. Chaotic behavior is observed when connections are nonsymmetric, as discussed in §4.6. We have not studied the basins of attraction in chaotic networks; this would certainly be an interesting area to investigate.

The simplest symmetric network that shows delay-induced sustained oscillation (in the absence of self-coupling) is the all-inhibitory triangle: three delayed neurons all connected to each other via inverting, or inhibitory, connections:

$$T'_{ij} = \frac{1}{100k\Omega}\begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}. \qquad (3.10)$$



The network defined by (3.1) and (3.10) is analyzed in detail in Ch. 4, and a phase diagram is given in Fig. 4.8. The analysis shows that for sufficient delay, $\tau \equiv \tau'/R_iC_i > ln(2) = 0.693..$, the all-inhibitory triangle has an oscillatory attractor along the (1,1,1) direction - that is, with all neurons oscillating *in phase*. For sufficient gain (see Fig. 4.8) the oscillatory mode is not the only attractor; there are also several fixed-point attractors, each with its own basin of attraction.

Figure 3.13 shows two slices of the basin of attraction for the *oscillatory mode* of

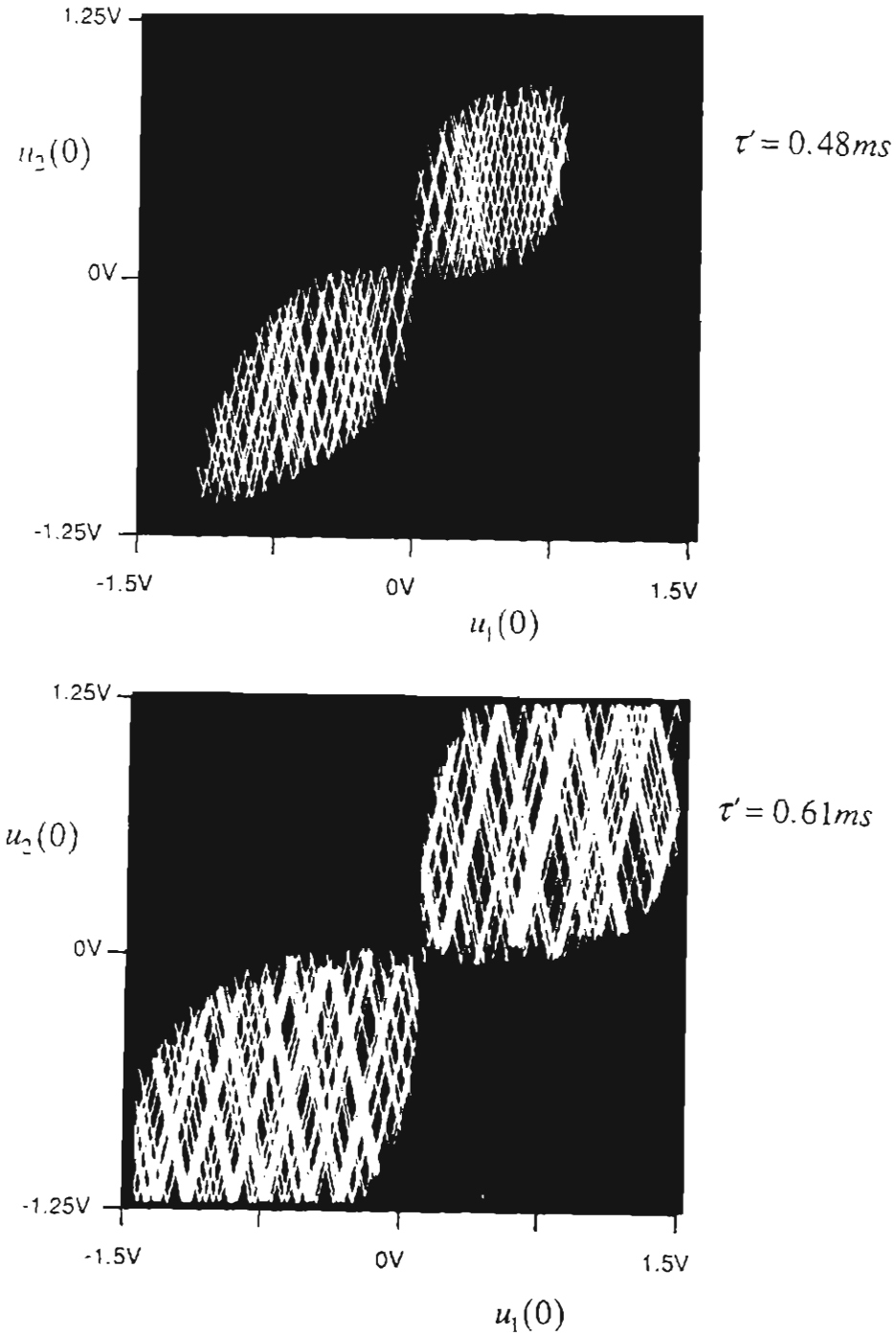


Fig. 3.13. Basin of attraction for coherent oscillatory mode (hatched region) for three delayed-output neurons with symmetric inhibitory (i.e. negative or antiferromagnetic) coupling. Black region indicates initial conditions leading to a fixed point. As the delay is increased, the basin for the oscillatory mode expands to fill more of the state space. The delay τ' should be compared to the network characteristic time $R_iC_i = 0.5\ ms$.

the all-inhibitory triangle in the regime where both fixed-point attractors and the in-phase oscillatory attractor exist. The two slices shown are for different values of normalized delay $\tau \equiv \tau'/R_iC_i = \tau'/[0.5ms]$. The slices are in the plane defined by the initial condition $u_3(0) = 0V$, with initial conditions on neurons 1 and 2 raster scanned as described above.[1] The first thing to notice in Fig. 3.13 is that a larger delay yields a larger basin of attraction for the oscillatory mode. As the normalized delay is reduced towards 0.693, the basin shrinks and finally disappears. At that point, the oscillatory mode itself goes unstable, in accordance with the analysis of Ch. 4. The second thing to notice in Fig. 3.13 is the two-lobed shape of the basin as seen in these slices. The basin structure leading to this interesting shape is revealed by shifting the position of the slice, which is done by changing the dc initial condition on u_3 , as shown in Fig. 3.14.
From the three images in Fig. 3.14, and the symmetry of the state space, we can deduce that the basin of attraction for the oscillatory mode forms a cylinder centered about the (1,1,1) direction that pinches together at the origin ($u_i = 0$ for all i). This structure is shown schematically in Fig. 3.15. This figure explains the two-lobed pattern seen in Figs. 3.13 and 3.14: the pattern marks the intersection of the pinched-cylindrical basin with the planes of the slices u_3 = constant. From these pictures, we can deduce the curvature of the basins near the origin basin from the shape of the lobes. We infer that near the origin, the basin looks like two paraboloids back to back, aligned along the (1,1,1) direction, as illustrated in Fig. 3.15.

An analysis of the all-inhibitory network, which will be presented in § 4.3, explains the basin structure described above. We briefly mention some relevant features here.

---

[1] In delay systems, the initial state of each neuron must be specified over the entire interval of time $[-\tau',0]$. We took care that the initial condition load time (see Fig. 3.6) was much longer than the neuron delay, so that initial functions were nearly constant over this time interval. Of course, this particular choice is arbitrary, and is itself only a "slice" of an infinite-dimensional space of possible initial conditions. One might wonder if other choices - say, for example, wildly oscillating initial functions over the interval $[-\tau',0]$ - would not lead to undiscovered dynamics. It appears, based on tests of just this sort, that nothing interesting happens when non-constant initial functions are used.

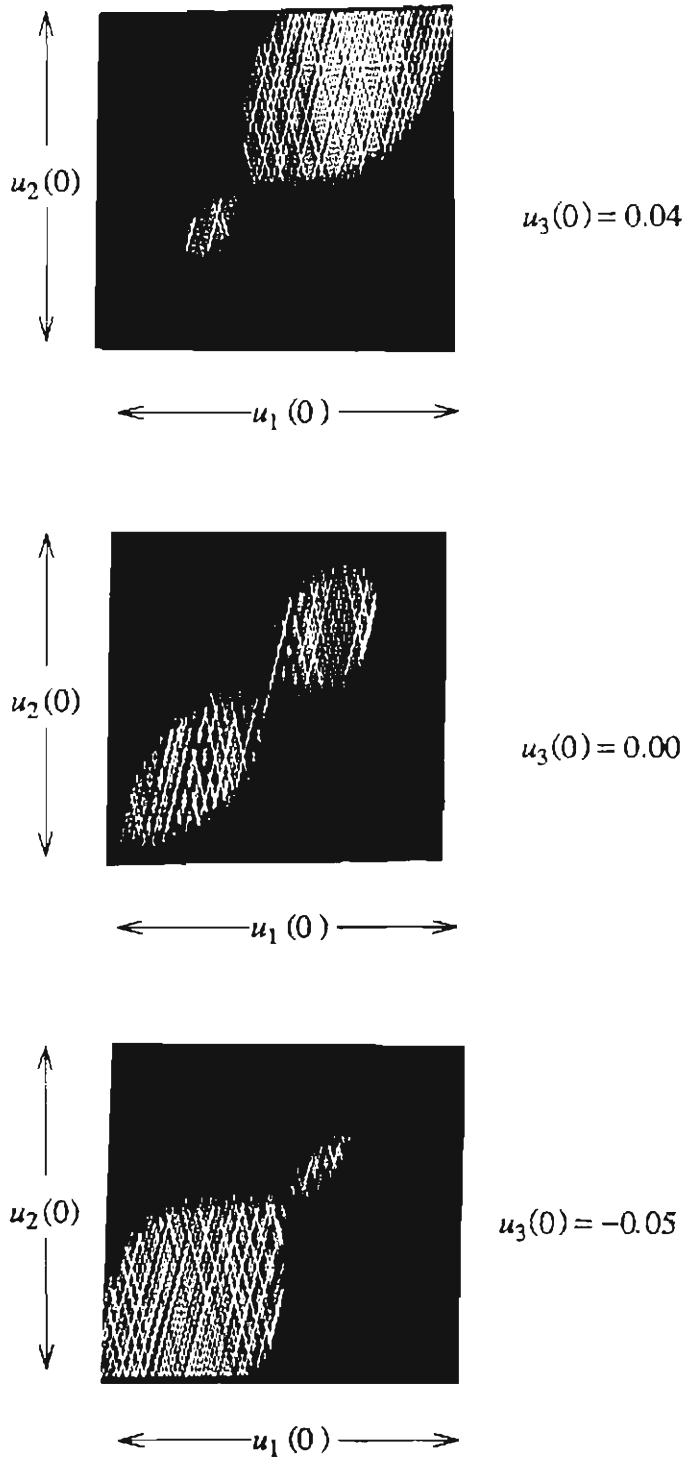


Fig. 3.14. The overall shape of the basin for sustained oscillation (hatched region) in the all-inhibitory triangle (same circuit as in Fig. 3.13) is revealed by shifting the slice in the direction of neuron 3.

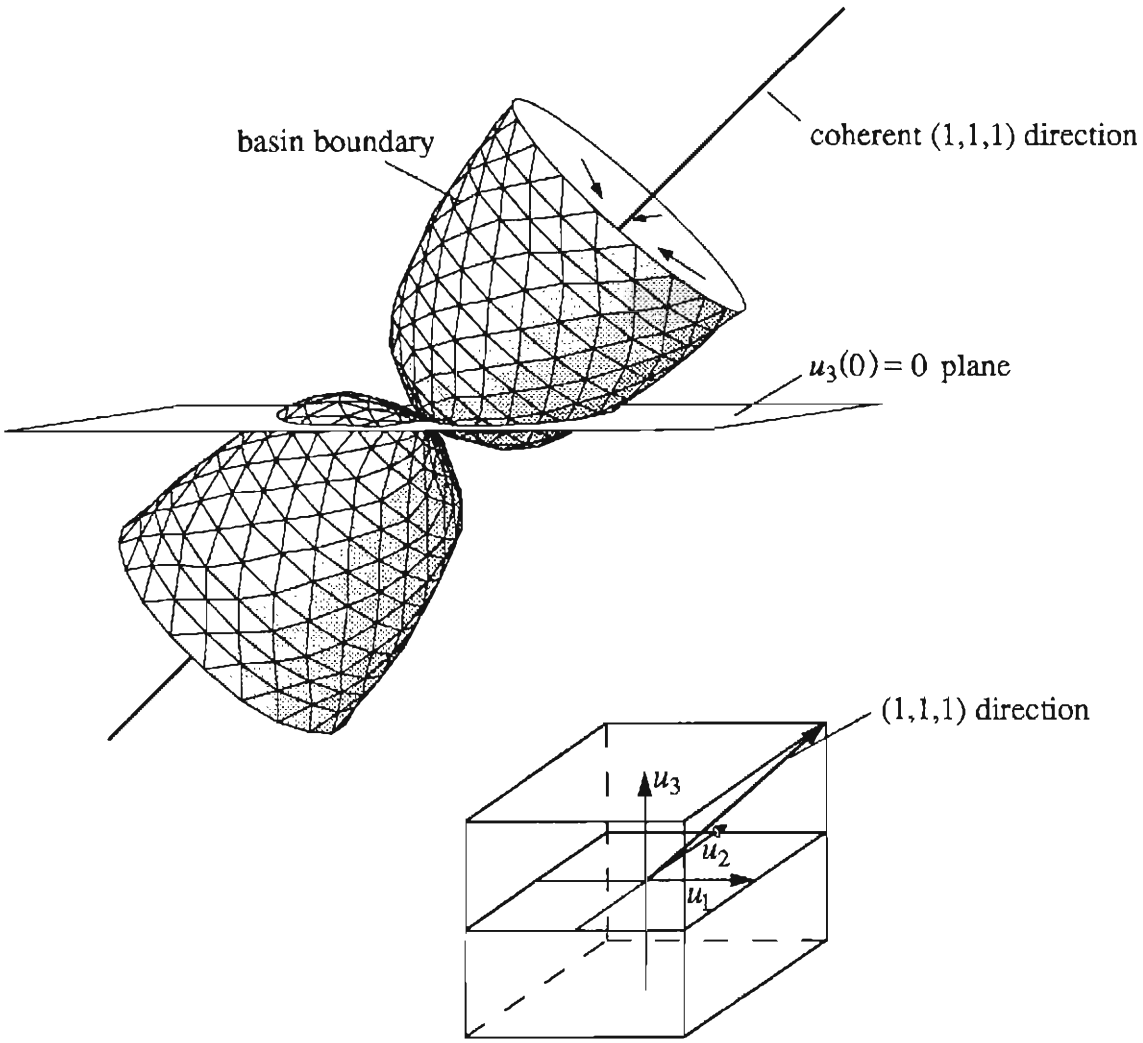


Fig. 3.15. The characteristic two-lobed basin shape seen in Figs. 3.13 and 3.14 is explained by a cylindrical basin oriented along the (1,1,1) direction, and pinched at the intersection with the plane $\Sigma_i\, u_i = 0$ (see text).

(These features apply for any N, not only $N = 3$.) In the regime where multiple fixed points and a coherent oscillatory mode coexist, dynamics in the vicinity of the origin is characterized by N-1 eigenvectors spanning the hyperplane $\sum_i u_i = 0$. The eigenvalues associated with these eigenvectors are degenerate and greater than one, so the entire hyperplane $\sum_i u_i = 0$ is a degenerate outset of the origin. The remaining eigenvector is in the (1,1,...,1) direction and has a large *negative* eigenvalue. This negative eigenvalue makes the (1,1,...,1) direction an outset of the origin as well, but in this direction the instability is oscillatory. Initial conditions near the (1,1,...,1) direction will be pulled onto the oscillatory attractor, giving rise to a cylindrical basin of attraction about the (1,1,...,1) direction. Initial conditions near the hyperplane $\sum_i u_i = 0$ are pulled away from the origin and onto this plane towards fixed points. As a result of the centrifugal dynamics within the hyperplane, the cylindrical basin for oscillation is pinched as the (1,1,...,1) vector crosses the hyperplane at the origin. As delay is reduced, the relative strength of the centrifugal dynamics in the hyperplane become sufficient to even rip apart the oscillatory attractor. Analyzing this event yields a value for the critical delay for sustained oscillation (see § 4.4).

To further check the inferred basin structure for the general all-inhibitory network, we have constructed a special circuit which allows the basin of attraction for the oscillatory mode to be sliced along the (1,1,...,1) direction. This circuit, shown in Fig. 3.16, supplies the initial conditions to the analog switches. It replaces the independent (Cartesian) initial condition rastering scheme shown in Fig. 3.4. In the present scheme the Y coordinate gives the component of the initial condition vector along the (1,1,...,1) direction; the X coordinate gives the component of the initial condition vector perpendicular to this direction - into the $\sum_i u_i = 0$ hyperplane. This excursion into the hyperplane is chosen to be in a direction in which only two of the possible N components deviate from (1,1,...,1). The slice generated can be thought of as an axial

cut down the cylindrical basin, as shown in Fig. 3.17(a). Images generated using this initial condition circuit are shown in Fig. 3.17(b). The network configuration used to produce these images was the $N = 5$ all-inhibitory network; the two images are for different values of delay. These images confirm the inferred shape of the basin of attraction for oscillation, and reveal the pinched cylinder in its natural coordinate system.

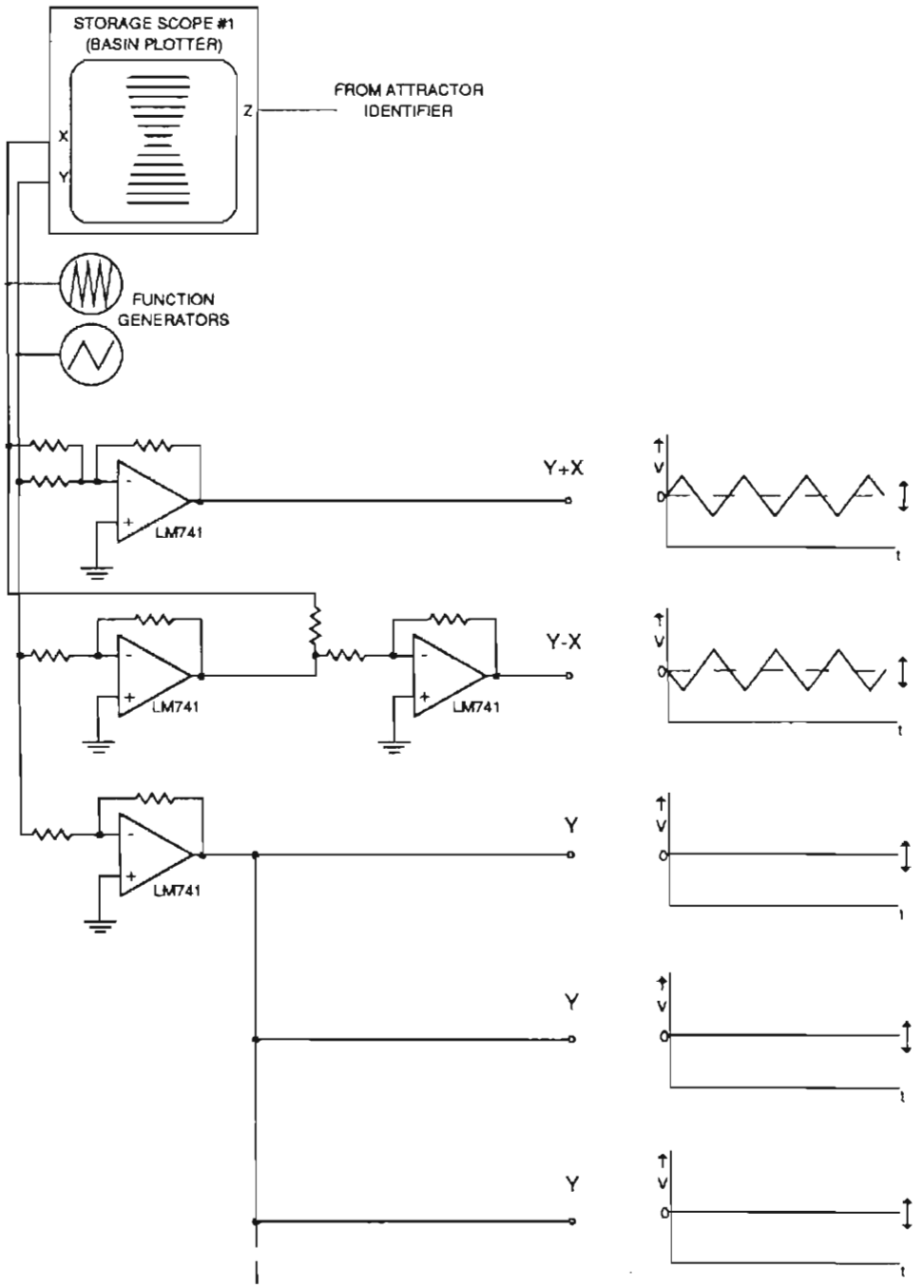


Fig. 3.16. Schematic of circuit to provide rastering in the coherent direction (measured by Y) and perpendicular to the coherent direction (measured by X). All initial conditions contain an equal amount of Y, and two others have added voltages X and -X, respectively. Note that the direction $u_i(0)$ = X and $u_j(0)$ = -X for any i and j constitutes a particularly simple excursion into the plane $\Sigma_i u_i = 0$, which is perpendicular to the coherent direction (1,1, ..., 1).

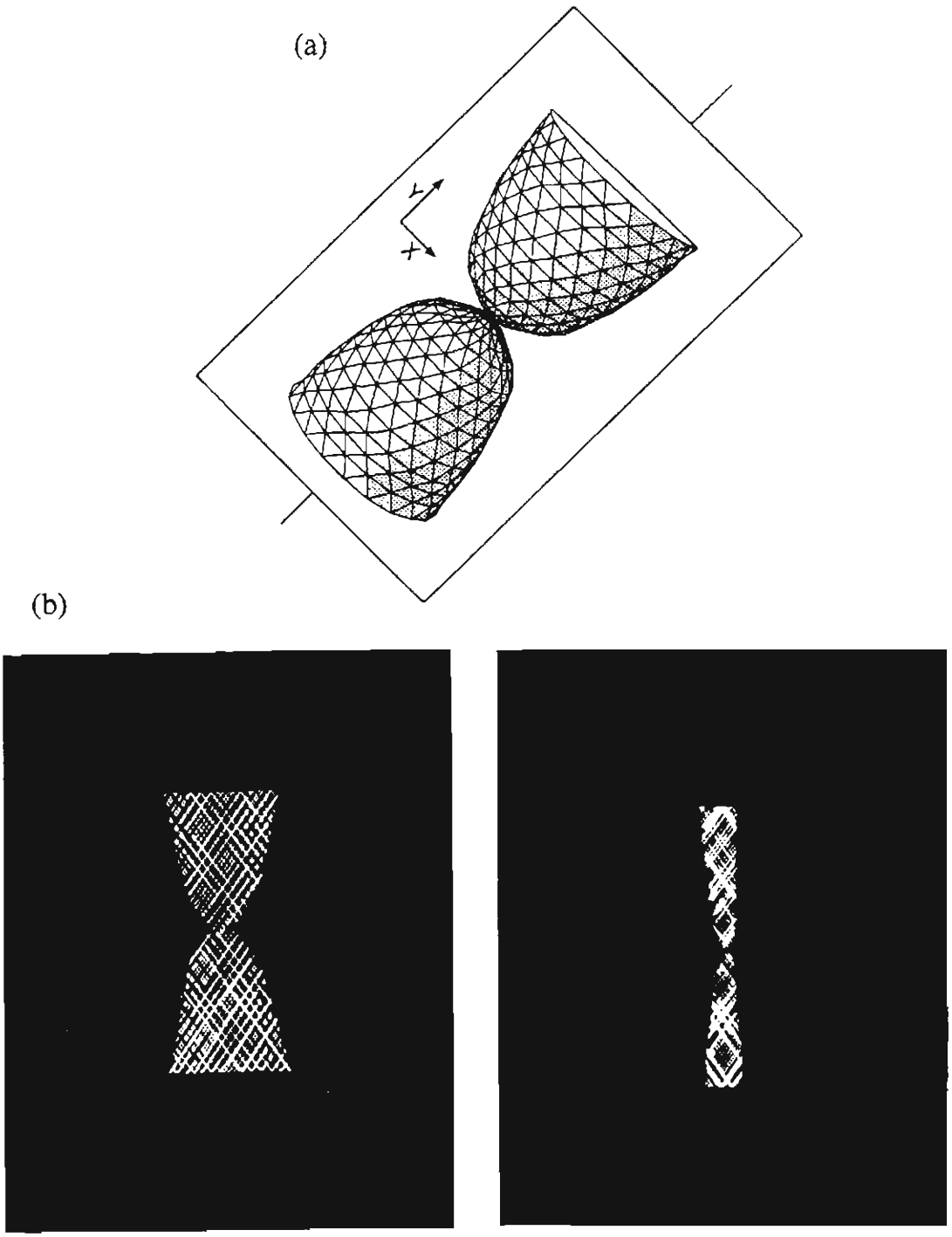


Fig. 3.17. (a) The plane swept out by the rastering circuit of Fig. 3.16 is shown in relation to the proposed basin structure for the coherent oscillatory mode. (b) For the five-neuron all-inhibitory (antiferromagnetic) network, the observed basin of attraction for sustained oscillation (hatched region) confirms the general shape inferred from the standard rastering scheme. Left: $\tau' = 0.73\ ms$, Right: $\tau' = 0.51\ ms$. Characteristic time: $R_i C_i = 0.5\ ms$.

Chapter 4

ANALOG NEURAL NETWORKS WITH TIME DELAY

4.1 INTRODUCTION

It is well known that symmetrically connected networks of analog neurons operating in continuous time will always settle onto a fixed-point attractor [Cohen and Grossberg, 1983; Hopfield 1984]. This important result assumes, however, that neurons communicate and respond *instantaneously*. As demonstrated in the previous chapter, all bets are off regarding network stability once time delay is introduced into the response of the neurons. Designing an electronic neural network to operate as quickly as possible will increase the relative size of the intrinsic delay and can eventually lead to oscillation or chaos. In the world of microelectronics, delays due to the finite switching speed of amplifiers are well characterized, and constitute an important aspect of analog and digital VLSI circuit design [Mukherjee, 1985]. In biological neural networks, it is known that time delay can cause an otherwise stable system to oscillate [Coleman and Renninger, 1975; Coleman and Renninger, 1976; Hadeler and Tomiuk, 1977; an der Heiden, 1979; an der Heiden,1980; Glass and Mackey, 1988]. Instabilities introduced by delays have also been analyzed in the context of control theory and electrical engineering [Kolmanovskii and Nosov, 1986].

The goal of this chapter is to develop an understanding of how a delay in the response of the neurons in a network can induce sustained oscillation and chaos. For the case of symmetrically connected networks, we find that for some connection topologies,

delays much less than the network relaxation time can lead to sustained oscillation, while for other topologies even very long delays will not induce oscillation. Furthermore, for those network configurations which can oscillate at small delay, there is a critical value of delay below which the network will not support sustained oscillation.

The results reported in this chapter show that the existence of oscillatory modes in symmetric networks with delay has a surprisingly simple dependence on the neuron gain and delay, and on the size and connection topology of the network. These results are stated as stability criteria which extend the famous result: "symmetric connections implies no oscillation" to the case of time delay networks. Results derived in this chapter are based on local rather than global stability analysis and therefore do not provide a rigorous guarantee that all initial states will converge to fixed points. Rather, we support our results with extensive numerical and experimental evidence suggesting that the stability criteria presented here are valid under the conditions investigated. In addition to using standard numerical integration to test the theoretical results, we have measured critical delays for sustained oscillation in the electronic network described in Ch. 3.

In the chapter following this one, Ch. 5, we consider a network with discrete-time parallel dynamics. This network is equivalent to the long-delay limit of the continuous-time network considered here. In the discrete-time limit, we are able to analyze the dynamics globally and thus provide a rigorous stability criterion guaranteeing that all attractors are fixed points. It is reassuring that the local results presented here limit properly at long delay to the global results derived in Ch. 5.

The rest of the chapter is organized as follows: In § 4.2, we write down a general system of delay-differential equations starting from the circuit equations for an electronic network and describe the simplifying assumptions of our model. In § 4.3 we present a linear stability analysis about the point where all neurons have zero input and steepest transfer function. This point is defined as the origin of an N dimensional space where

each direction represents the input voltage of a neuron. For sufficiently large neuron gain, the origin loses stability in either a pitchfork bifurcation, which creates fixed points away from the origin, or in a Hopf bifurcation [Chaffee, 1971], which creates an attractor for sustained oscillation. Which sort of bifurcation occurs first depends on the largest and smallest eigenvalues of the connection matrix and on the normalized delay. Experimentally, we find that the Hopf bifurcation marks the appearance of sustained oscillation in symmetric networks. The analysis in § 4.3 is formulated as a design criterion that will yield fixed-point dynamics in a delay network as long as the ratio of delay to relaxation time is kept below a critical value.

In § 4.4, we consider networks operating in a large-gain regime where fixed point attractors away from the origin and oscillatory attractors coexist, each with large basins of attraction. We restrict our attention in this regime to networks which oscillate coherently (defined below), and present a novel nonlinear stability analysis of the coherent oscillatory attractor which yields a critical delay for sustained oscillation in these networks. The results of the linear and nonlinear stability analyses presented in § 4.3 and § 4.4 are compared with numerical integration of the delay-differential equations and experiments in the electronic delay network; good agreement is found between theory, experiment and numerics.

In § 4.5, we discuss stability for several specific network topologies: symmetric rings of neurons, two-dimensional lateral inhibition networks, random symmetric networks, and associative memory networks based on the Hebb rule [Hebb, 1949; Hopfield, 1982]. A particularly important result is that Hebb rule networks are stable for long delays, but that clipping algorithms which limit the connection strengths to a few values can yield an connection matrix with large negative eigenvalues which can lead to sustained oscillation.

In § 4.6, we discuss chaotic dynamics in asymmetric neural networks, and give an example of a small (three neuron) network which shows delay-induced chaos. Finally, a summary of useful results is given in § 4.7.

4.2. DYNAMICAL EQUATIONS FOR ANALOG NETWORKS WITH DELAY

In this section we derive a general system of delay-differential equations, Eq. (4.3), starting from the circuit equations for the electronic network discussed in Ch. 3. The network consists of N saturating voltage amplifiers with delayed output coupled via a resistive interconnection matrix, and is identical with the analog network described by Hopfield [1984], with the addition of a delay τ'_j:

$$C_i \dot{u}_i(t') = -\frac{1}{R_i} u_i(t') + \sum_{j=1}^{N} T'_{ij} f_j\left(u_j\left(t' - \tau'_j\right)\right) . \tag{4.1}$$

The variable $u_i(t')$ in (4.1) represents the voltage on the input of the i^{th} neuron. Each neuron is characterized by an input capacitance C_i, a delay τ'_i, and a nonlinear transfer function f_i. The transfer function f_i is taken to be sigmoidal, saturating at ± 1 with maximum slope at $u = 0$. The connection matrix element T'_{ij} has a value $+1/R_{ij}$ when the noninverting output of j is connected to the input of i through a resistance R_{ij}, and a value $-1/R_{ij}$ when the inverting output of j is connected to the input of i through a resistance R_{ij}. The parallel resistance at the input of each neuron is defined as $R_i = (\Sigma_j |T'_{ij}|)^{-1}$. We consider the case of identical neurons, $C_i = C$, $f_i = f$, $\tau'_i = \tau'$, and also assume each neuron is connected to the same total input resistance, defining $R \equiv R_i$ for all i. With these assumptions, the equations of motion become

$$RC\,\dot{u}_i(t') \;=\; -u_i(t') \;+\; R\sum_{j=1}^{N} T'_{ij}\, f\big(u_j(t'-\tau')\big)\;. \tag{4.2}$$

Rescaling time, delay and T'_{ij} gives the following new variables: $t = t'/RC$; $\tau = \tau'/RC$; $T_{ij} = RT'_{ij}$. This definition of T_{ij} has a normalization $\Sigma_j |T_{ij}| = 1$. In terms of these scaled variables the delay system takes on the simple form

$$\dot{u}_i(t) \;=\; -u_i(t) \;+\; \sum_{j=1}^{N} T_{ij}\, f\big(u_j(t-\tau)\big)\,. \tag{4.3}$$

All times in Eq. (4.3) are in units of the characteristic network relaxation time RC.

As mentioned in Ch. 3, the initial conditions for a delay-differential system must be specified as a function on the time interval $[-\tau,0]$. All experimental and numerical results presented take all initial functions to be constant on this interval, though not necessarily the same for different i. A cursory numerical investigation suggests that the stability results presented below do not depend on the particulars of the initial function.

4.3. LINEAR STABILITY ANALYSIS

We consider the stability of Eq.(4.3) near the origin ($u_i = 0$ for all i). Linearizing $f_i(u)$ about the origin gives

$$\dot{u}_i(t) \;=\; -u_i(t) \;+\; \sum_{j=1}^{N}\beta\; T_{ij}\; u_j(t-\tau)\,, \tag{4.4}$$

where the gain β is defined as slope of $f_i(u)$ at $u = 0$. It is convenient to represent the linearized system of N delay equations as amplitudes φ_i (i = 1, ..., N) along the N

eigenvectors of the connection matrix T_{ij},

$$\dot{\varphi}_i(t) = -\varphi_i(t) + \beta\, \lambda_i\, \varphi_i(t-\tau)\,, \tag{4.5}$$

where λ_i ($i = 1, .., N$) are the eigenvalues of the connection matrix T_{ij}. The λ_i will be referred to as the *connection eigenvalues* to avoid confusion with the roots of the characteristic equation that will be derived from Eq. (4.5). In general, these connection eigenvalues are complex; when T_{ij} is a symmetric matrix, the λ_i are real. Assuming exponential time evolution of the φ_i, we introduce the complex characteristic exponents s_i and define $\varphi_i(t) = \varphi_i(0)e^{s_i t}$. Substituting this form of $\varphi_i(t)$ into Eq. (4.5) gives the characteristic equation

$$\left(s_i + 1\right) e^{s_i \tau} = \beta\, \lambda_i\ . \tag{4.6}$$

The origin is asymptotically stable when $\mathrm{Re}(s_i) < 0$ for all i [Bellman and Cooke, 1963]. When $\mathrm{Re}(s_k) > 0$ for some k, the origin is unstable to perturbations in the direction of the eigenvector associated with s_k.

4.3.1 Linear stability analysis with $\tau = 0$

When the neurons have zero delay ($\tau = 0$), Eq. (4.6) reduces to $(s_i + 1) = \beta\lambda_i$. In this case, the origin is the unique attractor as long as all connection eigenvalues λ_i have real part less than $1/\beta$ as shown in Fig. 4.1. For a symmetric connection matrix, the λ_i are real and the bifurcation is of the pitchfork type: For $\beta > 1/\lambda_k$ the origin becomes a saddle and a pair of stable fixed points appears on opposite sides of the origin in the direction of the k^{th} eigenvector of T_{ij}. In neural networks language, this new pair of

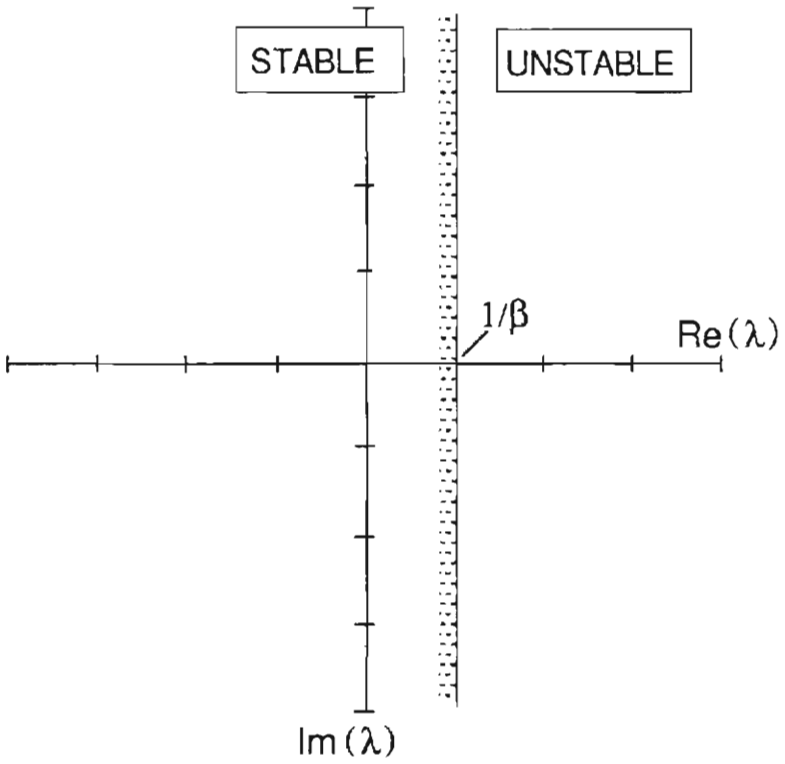


Fig. 4.1. The stability of the origin for zero delay is determined by the condition $\mathrm{Re}(\lambda_i) < 1/\beta$ for all i, where λ_i are the eigenvalues of the connection matrix T_{ij} which appears in Eq. (4.3). The border of the stability region is shown as a vertical line in the complex λ plane.

fixed points away from the origin is a memory.

As an example of linear stability analysis with $\tau = 0$, consider the $N \times N$ *all-excitatory* - or ferromagnetic - interaction matrix ($T'_{ij} = +1/R$; $T'_{ii} = 0$)

$$T_{ij} = \frac{1}{N-1}\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}. \tag{4.7}$$

The connection eigenvalues for this matrix are

$$\lambda_i = \begin{cases} 1 & \text{[once]} \\ -\dfrac{1}{N-1} & \text{[}(N-1)\text{- fold degenerate]} . \end{cases} \tag{4.8}$$

Notice that because T_{ij} is symmetric the λ_i are real. When $\beta < 1/\lambda_{max}$, where λ_{max} is the maximum connection eigenvalue, the origin is the only attractor. When $\beta > 1/\lambda_{max}$ the origin is unstable, and two fixed points appear on either side of the origin along the eigenvector associated with λ_{max}. In the present example, $\lambda_{max} = 1$ from Eq. (4.8) and the eigenvector associated with λ_{max} is the *ferromagnetic* direction ($u_i = 1$ for all i).

A second example is the $N \times N$ *all-inhibitory* or antiferromagnetic connection matrix

$$T_{ij} = \frac{1}{N-1}\begin{pmatrix} 0 & -1 & \cdots & -1 \\ -1 & 0 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & 0 \end{pmatrix}. \tag{4.9}$$

This network configuration is important in neural networks as a model of lateral inhibition (see § 4.5.2) and as a so-called winner-take-all circuit. The eigenvalues for the all-inhibitory network are

$$\lambda_i = \begin{cases} \dfrac{1}{N-1} & [(N-1)\text{-fold degenerate}] \\ -1 & [\text{once}]\,. \end{cases} \tag{4.10}$$

For this network configuration, the origin does not become unstable and fixed points away from the origin do not appear until $\beta > 1/\lambda_{max} = N\text{-}1$. Thus the origin for a large all-inhibitory network is very stable for zero delay. The eigenvector associated the minimum eigenvector λ_{min} is in the in-phase, or ferromagnetic, direction ($u_i = 1$ for all i). The N-1 eigenvectors associated with the degenerate λ_{max} all satisfy the condition $\sum_i u_i = 0$ which defines a hyperplane perpendicular to the ferromagnetic direction.

4.3.2. Frustration and equivalent networks

A symmetric matrix with connection strengths limited to three values - positive, negative and zero - can be represented as an undirected signed graph with a neuron at each vertex. An important property of the all-inhibitory network discussed above is that every loop formed from three neurons in the connection graph has an odd number of negative (inhibitory) edges. A connection graph containing loops with an odd number of negative edges is said to be frustrated. Frustration is important in systems with competing interactions [Toulouse, 1977], and is considered essential in the formation of a spin-glass state in magnetic systems [Binder and Young, 1986; Mezard *et al.*, 1987]. We suspect, though have not proven, that frustration is also essential for delay-induced oscillation when there is no self connection, i.e. $T_{ii} = 0$. Because every triangular loop

in the all-inhibitory network has an odd number of negative edges, this configuration is said to be fully frustrated. There are 2^{N-1} other networks that are also fully frustrated; these other configurations are related by the Mattis transformation [Mattis, 1976]: For any i let $u_i \rightarrow -u_i$ and $T_{ij} \rightarrow -T_{ij}$ for all j. All 2^{N-1} fully frustrated networks have identical dynamics, up to changes of sign. Similarly, there are 2^{N-1} networks equivalent to the ferromagnetic network, Eq. (4.7), all of which are nonfrustrated.

4.3.3. Linear stability analysis with delay

In this section, we show that for $\tau > 0$ the stability region, defined by the condition $\mathrm{Re}(s_i) < 0$, is no longer a simple vertical line at $1/\beta$ in the complex λ-plane as in Fig. 4.1, but forms a closed teardrop-shaped region that becomes smaller and more circular as the delay is increased as shown in Fig. 4.2. This idea is also discussed by May [1974]. As $\tau \rightarrow 0$, the region of stability expands to fill the half plane $\mathrm{Re}(\lambda) < 1/\beta$, recovering Fig. 4.1; as $\tau \rightarrow \infty$ the stability region becomes a circle centered at $\lambda = 0$ with radius $1/\beta$. A circular stability region is characteristic of iterated-map dynamics just as a half-plane stability region is characteristic of differential equation dynamics; thus as delay is increased from $\tau << 1$ to $\tau >> 1$ the local stability condition of the delay-differential system goes from that of continuous-time, differential equation dynamics to iterated-map or parallel-update dynamics [May, 1974]. The dynamics of the iterated-map analog network: $u_i(t+1) = \Sigma_j T_{ij} f(u_j(t))$, where t is the index of *discrete* time, corresponds to the long delay limit of Eq. (3.1). A global stability criterion for the iterated-map network will be given in Ch. 5. The iterated-map stability criterion agrees with the local analysis presented here in the long-delay limit $\tau \rightarrow \infty$.

The exact shape of the stability region at any value of delay can be found by substituting $s_j = \sigma_j + i\omega_j$ $(i = \sqrt{-1})$ into Eq. (4.6) and finding the condition $\sigma_j = 0$. The loci of points on the border of the stability region can be written in polar coordinates

as

$$\lambda_{\text{border}} = \Lambda(\theta)\, e^{i\theta} \ , \tag{4.11}$$

where $\Lambda(\theta) > 0$ is the radial distance from the point $\lambda = 0$ to the border of the stability region at an angle θ from the positive Re(λ) axis. Putting Eq. (4.11) and the condition $\sigma_j = 0$ into Eq. (3.3) gives

$$\left(i\omega_j + 1\right) e^{i\omega_j \tau} \ = \ \beta\, \Lambda(\theta)\, e^{i\theta} \ . \tag{4.12}$$

Solving for $\Lambda(\theta)$ gives the border of the stability region as an implicit function of delay:

$$\Lambda(\theta) = \frac{1}{\beta}\sqrt{\omega_j^2 + 1} \ , \tag{4.13a}$$

$$-\omega_j \ = \ \tan\left(\omega_j \tau - \theta\right) , \tag{4.13b}$$

where ω_j is in the range $(\theta - \pi/2) \le \omega_j \tau \le \theta$ modulo 2π. We are interested in the *smallest* root ω_j of Eq. (4.13b) for a given value of θ and τ. Large roots of Eq. (4.13b) produce large values of $\Lambda(\theta)$ by Eq. (4.13a), which lie outside of the stability region defined by the smaller roots. Only the part of the λ-plane inside the *smallest* stability region is actually stable. The stability region for the origin is plotted for several values of delay in Fig. 4.2.

Because the stability region closes in the negative half-plane for $\tau > 0$, it is possible for the origin to lose stability due to large *negative* connection eigenvalues - even purely real ones. The intersection of the stability region border and the Re(λ) axis in the negative half-plane is given by the solution to Eq. (4.13a) at $\theta = \pi$. We define this

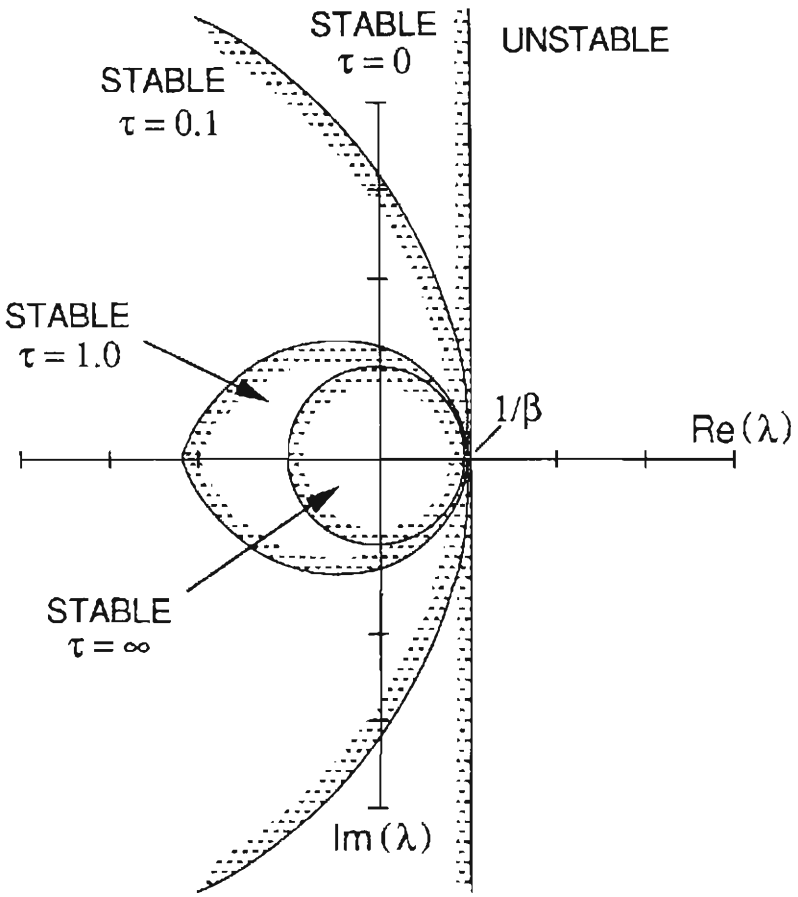


Fig. 4.2. The stability of the origin in the delay network lies within a closed region in the complex plane of eigenvalues of the connection matrix T_{ij}. Regions of stability are plotted for different values of delay: For $\tau = 0$, the border is a vertical line at $\mathrm{Re}(\lambda) = 1/\beta$ as in Fig. 4.1 ; For $\tau = \infty$, the stability region is a circle of radius $1/\beta$ centered at the origin of the λ plane. At finite delay, the stability region is teardrop shaped, crossing the real axis in the positive half-plane at $1/\beta$ and crossing the real axis in the negative half-plane at a delay-dependent value Λ. The tick marks along both axes are in units of $1/\beta$.

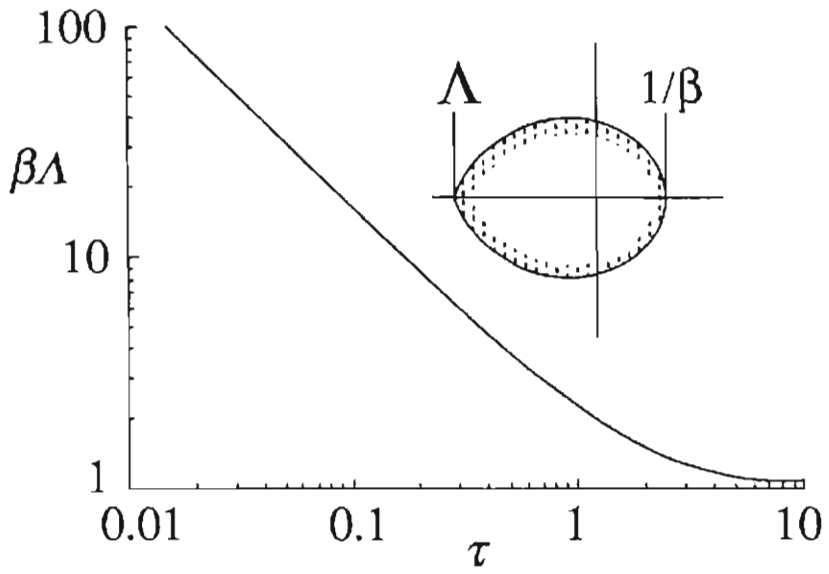


Fig. 4.3. The border of the stability region crosses the Re(λ) axis in the negative half plane at Λ for $\tau > 0$. The product $\Lambda\beta$, where β is the neuron gain, is plotted as a function of normalized delay τ. The value of Λ is particularly important for symmetric networks where the eigenvalues are confined to the Re(λ) axis.

solution as Λ, dropping the argument for the special case $\theta = \pi$. The value of Λ is inversely proportional to the gain of the neurons and is a transcendental function of delay defined implicitly by Eq. (4.13). A plot of the product $\Lambda\beta$, which depends only on delay, is shown in Fig. 4.3. For large and small delay, Λ can be approximated as an explicit function of delay and gain:

$$\Lambda \cong \begin{cases} \frac{1}{\beta}\left(\frac{\pi}{2\tau}\right) & \tau << 1, \qquad (4.14a) \\ \frac{1}{\beta}\sqrt{1+\left(\frac{\pi}{\tau+1}\right)^2} & \tau >> 1. \qquad (4.14b) \end{cases}$$

For a symmetric connection matrix (λ_i real) the origin will be unstable when $\lambda_{max} > 1/\beta$ or $\lambda_{min} < -\Lambda$. The bifurcation at $\lambda_{max} = 1/\beta$ is a pitchfork (as it is for $\tau = 0$) corresponding to a single real root s_i of Eq. (4.6) passing into the half plane $\mathrm{Re}(s_i) > 0$. The bifurcation at $\lambda_{min} = -\Lambda$ corresponds to a Hopf bifurcation [Chaffee, 1971] of the origin, with a complex pair of roots s_i passing into the half-plane $\mathrm{Re}(s_i) > 0$ at $\pm\omega_i$. The imaginary component $\omega_i = (\beta\Lambda - 1)^{1/2}$ at the bifurcation gives the approximate frequency of the oscillatory mode that results from this bifurcation.

4.3.4. Symmetric networks with delay

Figure 4.4 shows the evolution of the stability region of the origin for a delay network at three different values of gain. Each frame also shows schematically a distribution of eigenvalues for one of two types of symmetric networks: The eigenvalues on the left side of Fig. 4.4 are skewed negative, that is $|\lambda_{max}/\lambda_{min}| < 1$, while the eigenvalues on the right side are skewed positive, with $|\lambda_{max}/\lambda_{min}| > 1$. At low gain

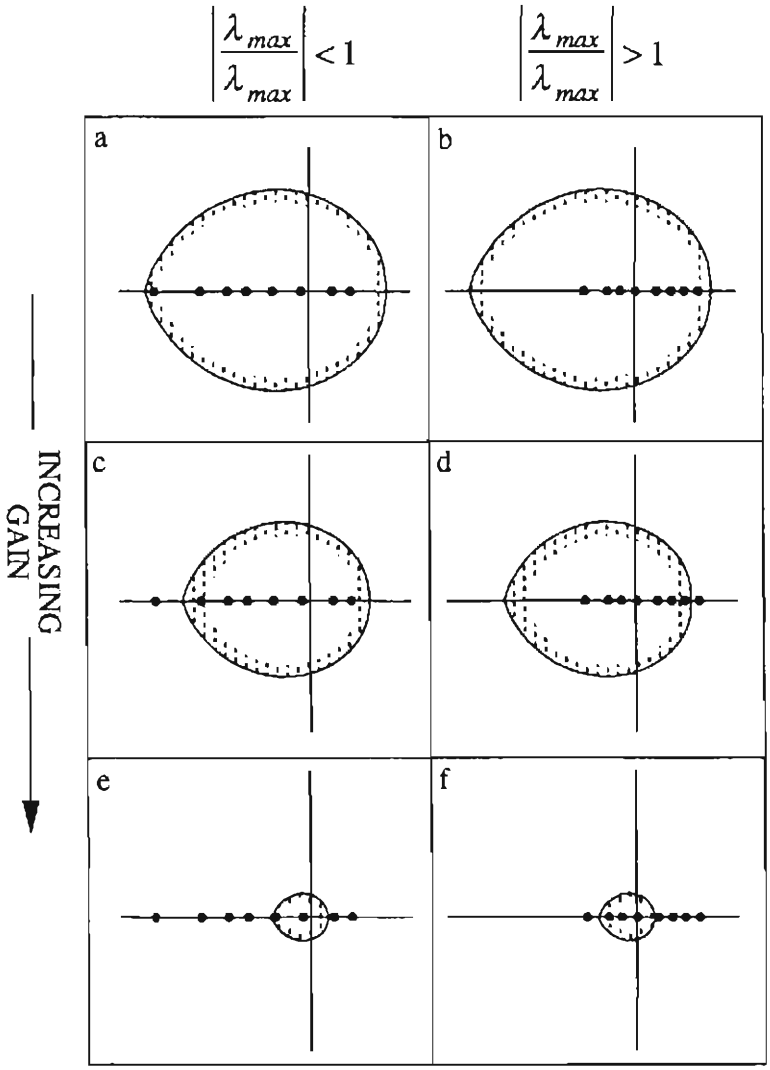


Fig. 4.4. The stability region of the origin and two different types of eigenvalue distributions (filled circles) are shown schematically. On the left (a,c,e), the eigenvalues satisfy $|\lambda_{max}/\lambda_{min}| < 1$; on the right (b,d,f), the eigenvalues satisfy $|\lambda_{max}/\lambda_{min}| > 1$. As the gain is increased, the stability region decreases in size and the origin loses stability. The bifurcations for each type of distribution are explained in the text.

(Figs. 4.4(a) and (b)) all eigenvalues lie within the large stability region and the origin is the unique fixed point and is stable. As the gain is increased, the size of the stability region decreases as $1/\beta$. The first eigenvalue to leave the stability region will either be the most negative, λ_{min}, as in Fig. 4.4(c), or the most positive, λ_{max}, as in Fig. 4.4(d). For the case in Fig. 4.4(d), a pair of attracting fixed points appear on either side of the origin along the eigenvector associated with λ_{max} and the origin becomes a saddle. For the case in Fig. 4.4(c), an oscillatory attractor exists along the eigenvector associated with the eigenvalue λ_{min}. The value of gain at which λ_{min} leaves the stability region in Fig. 4.4(c) is given by

$$\beta = -\frac{\sqrt{\omega^2+1}}{\lambda_{min}}, \tag{4.15a}$$

where

$$\omega = -tan(\omega\tau) \quad , \quad \frac{\pi}{2} < \omega\tau < \pi . \tag{4.15b}$$

In the limit of small delay, this value of gain is

$$\beta \equiv -\frac{\pi}{2\tau\lambda_{min}} \qquad (\tau << 1), \tag{4.16}$$

and the period of oscillation is approximately $2\pi/\omega$ ($\equiv 4\tau$ for $\tau << 1$).

For an eigenvalue distribution which satisfies $|\lambda_{max}/\lambda_{min}| < 1$, the first bifurcation to occur as the gain is increased can be either a pitchfork bifurcation, as λ_{max} leaves the stability region, or a Hopf bifurcation as λ_{min} leaves the stability region, depending on the value of delay. For an eigenvalue distribution which satisfies $|\lambda_{max}/\lambda_{min}| > 1$,

λ_{max} will always leave the stability region before λ_{min} regardless of delay.

A stability criterion for symmetric networks based on linear stability analysis can be formulated by requiring that λ_{min}, the minimum eigenvalue of T_{ij}, remain inside of the negative border of the stability region of the origin. In terms of the notation we have defined, this criterion requires $-\Lambda < \lambda_{min}$. The condition can be simplified by noting that Λ is always larger than its small-delay limit of $\pi/(2\tau\beta)$. The stability criterion for symmetric networks with delay can thus be stated:

$$\tau < -\frac{\pi}{2\,\beta\,\lambda_{min}} \quad \Rightarrow \quad \text{no sustained oscillation.} \tag{4.17}$$

This criterion lacks the rigor of a global stability condition, which exists for $\tau = 0$ [Cohen and Grossberg, 1983; Hopfield, 1984] and $\tau \to \infty$ [Marcus and Westervelt, 1989c] but is supported by considerable numerical and experimental evidence.

Figs. 4.4(e) and 4.4(f) show the situation when the gain is sufficiently large that eigenvalues have left the stability region through both negative and positive borders, indicating that Eq. (4.17) is violated and that fixed points exist away from the origin. In this regime the system possesses multiple basins of attraction for coexisting fixed-point and oscillatory attractors.

We find experimentally and numerically that delay networks in the large-gain regime may or may not show sustained oscillation, depending on the value of delay and the eigenvalue distribution. The observed behavior at large gain can be classified according to the ratio $|\lambda_{max}/\lambda_{min}|$: Networks with $|\lambda_{max}/\lambda_{min}| > 1$, as in Fig. 4.4(f), either do not oscillate at all or will oscillate only when the delay is much larger than the relaxation time. We have never observed sustained oscillation at $\tau < 1$ in any network satisfying $|\lambda_{max}/\lambda_{min}| > 1$ experimentally or numerically. This result remains empirical, but is consistent with the analysis in § 4.4 for delay networks that oscillate *coherently*.

In contrast, all networks investigated that satisfy $|\lambda_{max}/\lambda_{min}| < 1$ will oscillate for sufficient delay. At large gain, as in Fig. 4.4(e), these networks show coexisting fixed-point and oscillatory attractors. The basins of attraction for the oscillatory attractors are large for large delay but shrink as the delay is decreased, as seen in Ch. 3. For delay less than a critical value τ_{crit}, the oscillatory attractors disappear and only fixed-point dynamics are observed. A value for τ_{crit} cannot be found by the linear stability analysis described in this section because of the importance of the nonlinearity in the large-gain regime. An expression for τ_{crit} for networks that oscillate coherently is derived in § 4.4. The critical delay τ_{crit} found in this case diverges as $|\lambda_{max}/\lambda_{min}| \to 1$, in agreement with the empirical results mentioned above.

4.3.5. Self connection in delay networks

Including a delayed self connection affects the dynamics by shifting the distribution of connection eigenvalues and by decreasing the relaxation time of the network. As an example, consider the effect of adding a *delayed* self-connection term γ to the all-inhibitory network.[1] With the self-connection, the properly normalized connection matrix and eigenvalues are

$$T_{ij} = \frac{1}{N-1+|\gamma|}\begin{pmatrix} \gamma & -1 & \cdots & -1 \\ -1 & \gamma & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & \gamma \end{pmatrix}, \tag{4.18}$$

[1] A different normalization for the self connection is introduced in Ch. 5. Notice that in the present usage γ is the *relative* strength of the self connection compared to the strength of interneuron connections.

$$\lambda_i = \begin{cases} \dfrac{1+\gamma}{N-1+|\gamma|} & [(N-1)\text{- fold degenerate}] \\ \\ \dfrac{1-N+\gamma}{N-1+|\gamma|} & [\text{once}]\,. \end{cases} \tag{4.19}$$

The connection eigenvalues λ_{max} and λ_{min} for the all-inhibitory network are shown as functions of the self-connection γ in Fig. 4.5. Notice that adding a negative self connection ($\gamma < 0$) does not change λ_{min}, thus the value of delay where the Hopf bifurcation occurs in the all-inhibitory network is not changed by a negative self-connection. Adding a positive self-connection ($\gamma > 0$) will bring λ_{min} closer to zero and will increase the delay necessary for the Hopf bifurcation to occur. The condition $|\lambda_{max}/\lambda_{min}| > 1$ is satisfied in (4.19) when γ exceeds ($N/2$ - 1).

4.4. CRITICAL DELAY IN THE LARGE-GAIN LIMIT

In this section, we find a critical delay for sustained oscillation in the large-gain regime, where fixed point attractors away from the origin coexist with a single coherent oscillatory attractor. The main result, Eq. (4.23), applies to networks in which the oscillatory attractor is along a coherent direction. *Coherence* is defined by the condition that all $|u_i|$ are equal. Equivalently, a coherent oscillatory attractor lies along a vector extending from the origin to any corner of an N dimensional hypercube centered at the origin. When the eigenvector associated with λ_{min} is in a coherent direction, then the most robust oscillatory mode - that is, the one that will exist at the smallest delay - will be coherent. In this case, the network will not oscillate when the delay is smaller than the critical delay derived below. Connection topologies which have a coherent direction associated with λ_{min} include all fully frustrated networks: the all-to-all, one- and two-

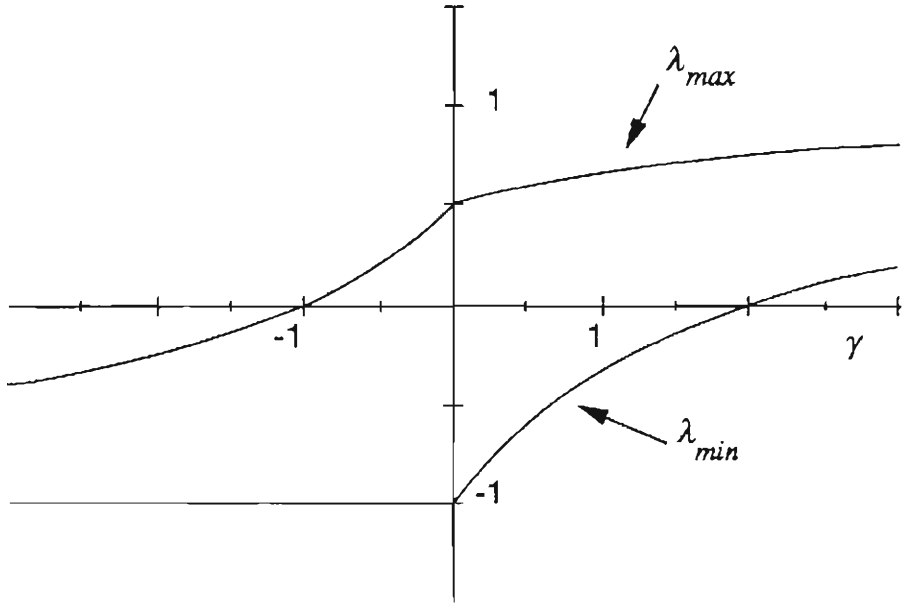


Fig. 4.5. The largest and smallest eigenvalues for the all-inhibitory network, Eq. (4.18), plotted as a function of the diagonal element γ. The values indicated at the axis crossings are for a general N, but the scale of the drawing is correct for the case $N = 3$. The asymptotic value for all eigenvalues as $\gamma \rightarrow \pm \infty$ is ± 1.

dimensional inhibitory networks treated in § 4.5, as well as all Mattis transformations [Mattis, 1976] of these networks. For other networks discussed in § 4.5, including the diluted inhibitory network and the negative-only clipped Hebb rule, the eigenvector associated with λ_{min} appears numerically to approach coherence at large N, though this has not been proven rigorously.

The stability criterion of § 4.3.4, stated as Eq. (4.17), applies at all values of gain but becomes useless in the large-gain limit. In particular, Eq. (4.17) requires that the delay go to zero as the gain diverges in order to prevent oscillation. Experimental and numerical investigation suggest that this requirement is too severe, and that there is a gain-independent critical delay τ_{crit} such that for $\tau < \tau_{crit}$ sustained oscillation disappears. Apparently, this critical delay results from an instability of the oscillatory attractor itself. Below, we derive a value for the critical delay τ_{crit} for coherent oscillation in the large-gain limit by considering the stability of the oscillatory attractor. This novel stability criterion agrees very well with experimental and numerical data.

4.4.1. Effective gain along the coherent oscillatory attractor

The basic idea of the derivation is that neurons with saturating output can be regarded as having an *effective gain* β_{eff} which is not constant as the state moves along the oscillatory attractor, and can be finite even when $f(u)$ is infinitely steep at $u = 0$. The effective gain is defined as $\beta_{eff} = f(u(t))/u(t)$. Note that β_{eff} is defined as the ratio of the neuron output $f(u(t))$ divided by the input $u(t)$ which gave rise to *that particular* output; this is a significant distinction for delay network, since the output $f(u(t))$ due to the input $u(t)$ does not appear at the output until a delay time τ has elapsed. This definition of β_{eff} reduces to the usual gain β when $f(u)$ is linear (with or without delay). We assume that the oscillatory attractor loses stability when the

effective gain is sufficiently large *at all points on along the attractor* that perpendicular perturbations will always lead the system off of the attractor. This instability occurs when the minimum value of β_{eff} along the attractor exceeds a critical value related to flow perpendicular to the oscillation direction.

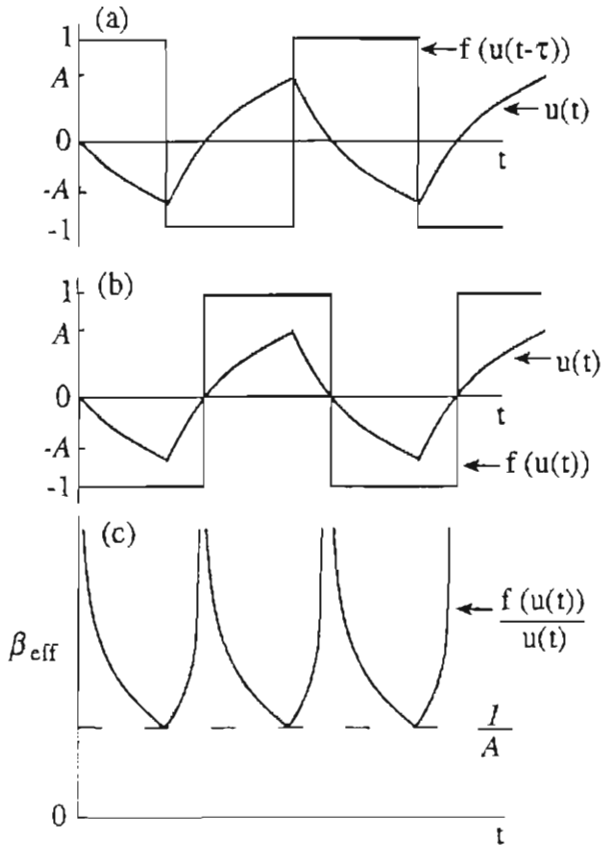
When the large-gain network is oscillating coherently, neuron outputs swing between ± 1 in the form of a square wave, while the inputs alternately charge and discharge exponentially with a time constant equal to the relaxation time of the network as shown in Fig. 4.6(a). The smallest value of β_{eff} occurs when the input amplitude is at an extremum of its charge-discharge oscillation and the corresponding output is saturated at ± 1. At this point, β_{eff} is the reciprocal of this input amplitude. The maximum amplitude A_i at the i^{th} input depends on the delay and is given by

$$A_i = \left| \sum_{j=1}^{N} T_{ij}\, sgn(u_j) \right| \left(1 - e^{-\tau}\right). \tag{4.20}$$

For coherent oscillation along the direction associated with λ_{min} all of the A_i in Eq. (4.20) will be the same (defined as A, with no subscript) and the term in the absolute value of (4.20) will equal $-\lambda_{min}$ (> 0). In this case β_{eff} will be bounded below by $1/A$, as shown in Fig. 4.6(c):

$$\beta_{eff} \geq \frac{1}{A} = -\frac{1}{\lambda_{min}\left(1 - e^{-\tau}\right)}. \tag{4.21}$$

Flow perpendicular to the oscillatory attractor is described by Eq. (4.5) with the λ_i ($i = 1, \ldots, (N\text{-}1)$) equal to the N-1 eigenvalues of T_{ij} excluding λ_{min} and with $\beta = \beta_{eff}$. The least stable of the N-1 directions perpendicular to the oscillatory attractor is along the eigenvector associated with λ_{max}. Thus the oscillatory attractor will



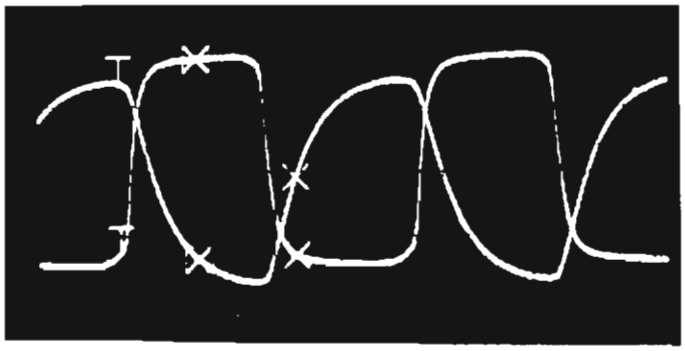


Fig. 4.6. (a) The input $u(t)$ (triangular wave) and output $f(u(t-\tau))$ (square wave) for a saturating infinite-gain neuron with delay in an oscillatory state. The value A, given by Eq. (4.20), is the maximum amplitude of the input. (b) The same input and output waveforms as above with the offset between input and output due to delay suppressed. (c) The effective gain β_{eff}, defined as the ratio of $f(u(t))/u(t)$, takes on finite values even when $f(u)$ is infinitely steep at $u = 0$. The minimum value of β_{eff} is where the input is an extremum; at this point $\beta_{eff} = 1/A$. (d) The input and output of a delayed neuron from the electronic circuit (Ch. 3) in a state of sustained coherent oscillation. Compare this to the idealized form in used in (a).

lose stability when $\lambda_{max}\beta_{eff} > 1$ at all points along the trajectory. From Eq. (4.21), this condition is satisfied when $\lambda_{max}/A > 1$. The critical delay τ_{crit}, defined by the condition $\lambda_{max}/A = 1$, is thus given by

$$\lambda_{max}\left(\frac{1}{-\lambda_{min}\left(1-e^{-\tau_{crit}}\right)}\right) = 1. \qquad (4.22)$$

Solving (4.22) for τ_{crit} gives the main result of § 4.4:

$$\tau_{crit} = -ln\left(1+\frac{\lambda_{max}}{\lambda_{min}}\right); \qquad \left(0 < \lambda_{max} < -\lambda_{min}\right). \qquad (4.23)$$

To illustrate this result we again consider the $N \times N$ all-inhibitory network (4.9) in the large-gain limit. This network has connection eigenvalues $\lambda_{max} = 1/(N\text{-}1)$, $\lambda_{min} = -1$, giving a large-gain critical delay

$$\tau_{crit} = ln\left(\frac{N-1}{N-2}\right) \quad [\sim 1/N \text{ for large } N]. \qquad (4.24)$$

Fig. 4.7 shows τ_{crit} for the all-inhibitory networks as a function of the size of the network N. The solid line is from Eq. (4.24), the circles are data from numerical integration with $\beta = 40$ indicating the smallest delay that would support sustained oscillation. The rapid decrease in τ_{crit} as the size of the network increases indicates that the all-inhibitory network is very prone to oscillation for large N.

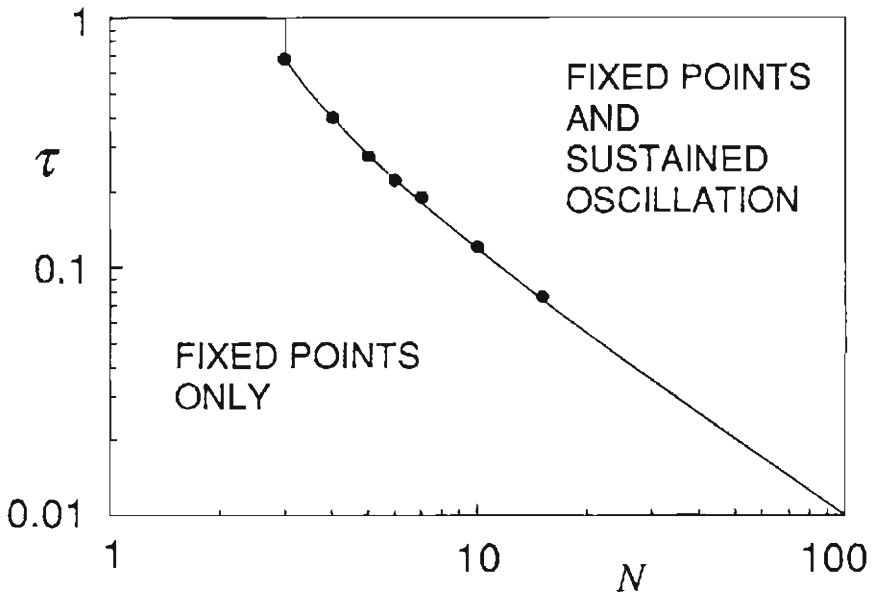


Fig. 4.7. Large-gain critical delay τ_{crit} for the all-inhibitory network plotted against N, the size of the network. Solid curve is the theory from Eq. (4.23), the filled circles are from numerical integration of the delay equations at $\beta = 40$. Numerical integration data were obtained by starting the system with initial functions ϕ_i:$[-\tau,0]$ along the eigenvector associated with λ_{min} and constant over the time interval $[-\tau,0]$. The delay-differential equations were integrated using a modified Euler method: A stack of 10 - 40 previous states was maintained for each neuron. Upon each Euler step, the elements in the stack were moved down one position and a new state was added to the top of the stack. The step size and size of the stack were chosen so that a state reached the bottom of the stack at precisely the specified delay, and could then be used as the neuron's delayed output. The system was checked for oscillation after many (up to 10^4) time constants. The critical delay was found by repeating the integration using a 10-split binary search in the value of delay.

4.4.2. Crossover from low gain to high gain regime

We have now found two critical values of delay: For small gain ($\beta < \lambda_{max}$) the network does not oscillate for $\tau < \tau_H$, where τ_H is the value of delay where the Hopf bifurcation occurs. For small delay,

$$\tau_H \equiv -\frac{\pi}{2\,\beta\,\lambda_{min}} \quad . \tag{4.25}$$

At large gain, the delay network does not oscillate for $\tau < \tau_{crit}$, where τ_{crit} is given by Eq. (4.23). We now consider the crossover from the small-gain regime to the large-gain regime for the specific example of an all-inhibitory triangle of neurons. For this network,

$$T_{ij} = \frac{1}{2}\begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}; \; \lambda_{max} = \tfrac{1}{2} \; ; \quad \lambda_{min} = -1 \, . \tag{4.26}$$

Fig. 4.8(a) shows the two theoretical curves for each of the two regimes. The data points are the values of delay where the oscillatory attractor disappears as measured in the analog circuit (open circles) and by numerically integrating the delay equations (filled circles). Fig. 4.8(b) shows four regions of the β - τ plane, each with distinct dynamical properties. For $\beta < 2$ and $\tau < \tau_H$, where τ_H is found by setting λ_{min} = -1 in Eq. (4.15), there is a single fixed point attractor at the origin. For $\beta < 2$, $\tau > \tau_H$, the fixed point at the origin is unstable and there is a single oscillatory attractor. At $\beta =$ 2 fixed points away from the origin appear. At this crossover point, $\tau_H \equiv 1.209$. For $\beta > 2$, the Hopf bifurcation line no longer marks the critical delay for sustained

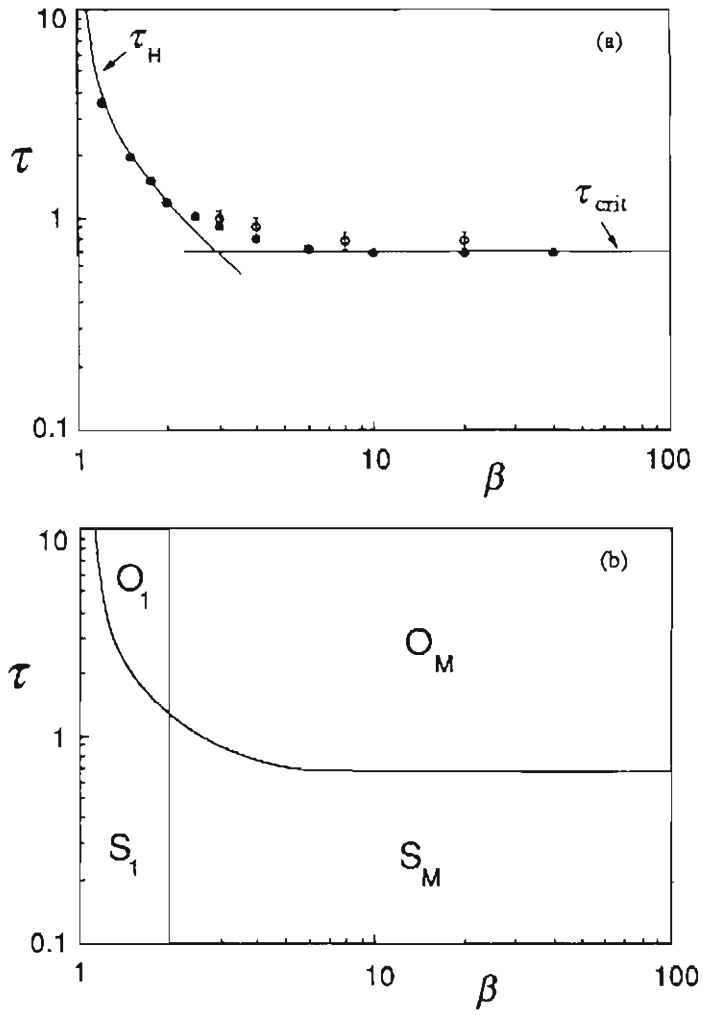


Fig. 4.8. Phase diagram for the all-inhibitory (or frustrated) triangle of delay neurons. (a) Two theoretical curves are shown. The curve labelled τ_H indicates the value of delay and gain where the origin undergoes a Hopf bifurcation, from Eq. (4.17); the line labelled τ_{crit} indicates the large-gain critical delay where the oscillatory mode loses stability. For $\tau < \tau_{crit}$ only fixed-point attractors are stable. The data points are critical delays measured in the electronic network (open circles) and by numerical integration (filled circles) with $\beta = 40$. Numerical integration data were obtained as described in the caption of Fig. 4.7. (b) The four regions in the $\beta - \tau$ plane with qualitatively different dynamics are: S_1: Single fixed point attractor at the origin; O_1: Single coherent oscillatory attractor; S_M: Multiple fixed point attractors away from the origin, all fixed points; O_M: Multiple attractors away from the origin, including fixed points and a coherent oscillatory attractor.

oscillation. As β becomes large, the critical delay for sustained oscillation approaches the gain-independent theoretical value of τ_{crit}. From Eq. (4.24), $\tau_{crit}(N=3) = ln(2) \cong 0.693$.

4.5. STABILITY OF PARTICULAR NETWORK CONFIGURATIONS

In this section we consider sustained oscillation in four symmetric delay networks: (1) symmetrically connected inhibitory rings; (2) large two-dimensional arrays of nearest-neighbor lateral inhibition networks on square and hexagonal lattices; (3) spin-glass-like random symmetric networks; and (4) Hebb rule and clipped Hebb rule associative memories.

4.5.1. Symmetrically connected rings

A ring of neurons with symmetric connections, all of equal strength but of either sign, inhibitory or excitatory, has a spectrum of connection eigenvalues given by

$$\lambda_k = cos\left(\frac{2\pi}{N}(k+\varphi)\right) \quad ; \; k=0,1,\ldots,(N\text{-}1). \tag{4.27}$$

where φ = 1/2 for a frustrated ring, i.e. $Sgn(\prod_{ring} T_{ij}) = -1$, and $\varphi = 0$ for a nonfrustrated ring [Reger and Binder, 1985]. (The normalized matrix has elements $T_{ij} = T_{ji} = \pm 1/2$.) The ratio of maximum to minimum eigenvalues can be found directly from Eq. (4.27):

$$\left|\frac{\lambda_{max}}{\lambda_{min}}\right| = \begin{cases} cos(\pi/N) & [<1] \quad N \text{ odd; frustrated,} \\ sec(\pi/N) & [>1] \quad N \text{ odd; nonfrustrated,} \\ 1 & \quad N \text{ even; frustrated or nonfrustrated.} \end{cases} \tag{4.28}$$

Notice that only frustrated rings with odd N satisfy the condition $|\lambda_{max}/\lambda_{min}| < 1$, suggesting that only these configurations will show sustained oscillation. This conclusion is confirmed experimentally and numerically. The large-gain critical delay for the frustrated ring with odd N is found from Eq. (4.23),

$$\tau_{crit} = -ln\left(1 - cos\left(\frac{\pi}{N}\right)\right) \; ; \quad (N \text{ odd, frustrated}). \tag{4.29}$$

Notice also that τ_{crit} *increases* with increasing N for the symmetric ring, while for the all-inhibitory network τ_{crit} decreases with increasing N. Inhibitory rings are thus much less prone to oscillation than fully-connected inhibitory networks. The critical delays from numerical integration are compared to Eq. (4.29) in Fig. 4.9.

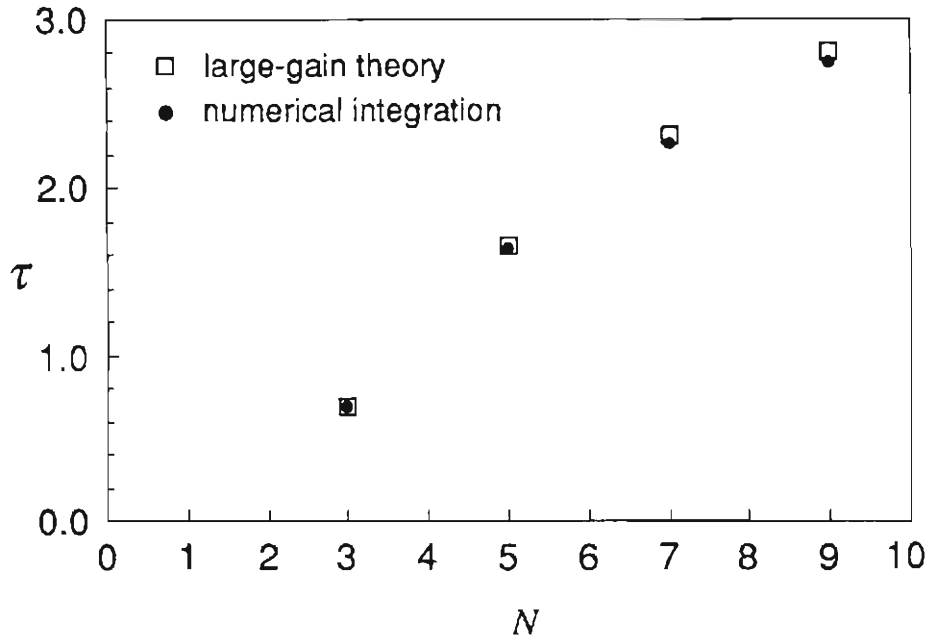


Fig. 4.9. Large-gain critical delay τ_{crit} for symmetrically connected frustrated rings with N = 3,5,7,9 from Eq. (4.23) (open squares) is plotted along with critical delay from numerical integration (filled circles) with β = 40. Numerical integration data were obtained as described in the caption of Fig. 4.7. Frustrated symmetric rings with even N do not satisfy $|\lambda_{max}/\lambda_{min}| < 1$ and therefore are not expected to oscillate for any delay within the large-gain theory. Numerically, frustrated rings with even N showed sustained oscillation only for very large delay ($\tau > 10$), though this is possibly a numerical artifact.

4.5.2 Two-dimensional lateral-inhibition networks

An important network configuration, especially to the study of real and artificial visual systems, is one in which each neuron inhibits the activity of its neighbors. This configuration, called *lateral inhibition*, is ubiquitous in vertebrate and invertebrate vision systems [Dowling, 1987], and is widely used in artificial vision systems for edge and feature detection. Lateral inhibition has also been incorporated into an electronic VLSI model of the retina [Mead, 1989]. The function of lateral inhibition is to enhance the contrast of edges in a visual scene [Ratliff, 1965; Dowling, 1987] and to broaden the dynamic range of a visual system by setting a local rather than global reference point for measuring relative intensity variations [Mead, 1989].

A case of lateral inhibition in which time delay is significant is in the compound eye of the horseshoe crab, *Limulus* [for a collection of papers see: Ratliff, 1974]. It is found experimentally that the individual eyelets (ommatidia) that form the compound eye of *Limulus* are mutually inhibitory, and that there is a significant time delay (~ 0.1 sec.) before lateral inhibition is activated between any pair of ommatidia. It is also found that under certain experimental conditions, a spatially uniform illumination over the entire eye will induce sustained coherent oscillation, with all ommatidia showing an in-phase periodic modulation in their output firing rate, with a period of ~ 0.3 sec. [Barlow and Fraioli, 1978].

Such experiments have stimulated several mathematical analyses addressing oscillation in delayed lateral inhibition systems [Coleman and Renninger, 1974, 1975, 1978; Hadeler and Tomiuk, 1977, an der Heiden, 1980]. These analyses have assumed uniform, all-to-all coupling between ommatidia, and further have assumed a coherent form for the oscillatory solution, which allows the problem to be reduced to a one-dimensional delay-differential equation for motion along the in-phase (1,1, ...,1)

direction. This second assumption does not allow an instability of the oscillatory mode to broken-symmetry states, and thus previous treatments have not predicted the instability of the coherent oscillatory solution which lead to our large-gain critical delay τ_{crit} in § 4.4.

Already in Ch. 4 we have considered two extremes of lateral inhibition networks: the all-to-all inhibitory network (Eq. (4.9) and (4.24)) and the one-dimensional laterally inhibiting ring, which is covered by the analysis in § 4.5.1. These two networks are seen to behave quite differently as the number of neurons becomes large. Specifically, as $N \to \infty$ the critical delay τ_{crit} from (4.23) tends to zero for the all-to-all inhibitory network and tends to infinity for the one-dimensional ring with nearest-neighbor inhibition (for now, we set the self-connection $\gamma = 0$):

$$\tau_{crit} \to 0 \quad \text{as} \quad N \to \infty \qquad \text{(all-to-all)}, \tag{4.30}$$

$$\tau_{crit} \to \infty \quad \text{as} \quad N \to \infty \qquad \text{(1-D ring)}. \tag{4.31}$$

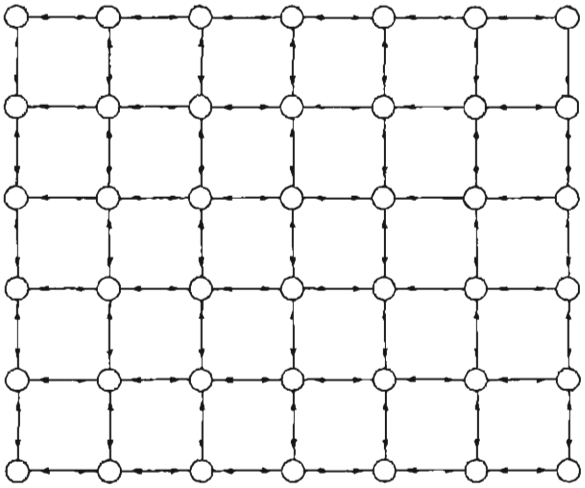
Of course, the case of most direct application to vision is neither of these extremes, but rather a large 2-D network. In this subsection, we show that the stability of large 2-D networks with *delayed nearest-neighbor lateral inhibition* depends crucially on the form of the lattice. With the neurons on a square lattice (Fig.4.10(a)), we find

$$\tau_{crit} \to \infty \quad \text{as} \quad N \to \infty \quad \text{(2-D square lattice)}. \tag{4.31}$$

That is, this configuration will not show sustained oscillation in the large-N limit. In contrast, when the neurons are placed on a triangular lattice (Fig. 4.10(b)) the critical delay for sustained oscillation is *finite* in the large-N limit, approaching the limit

$$\tau_{crit} \to ln(2) = 0.693... \quad \text{as} \quad N \to \infty \qquad \text{(2-D triangular lattice)}. \tag{4.32}$$

(a)



(b)

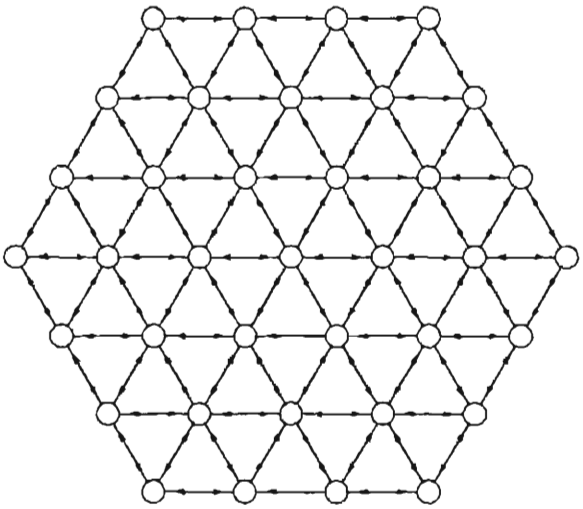


Fig. 4.10. Two-dimensional lattices with lateral inhibition. (a) Square lattice, (b) triangular lattice.

For generality, we introduce a diagonal element γ in the connection matrix (before normalization), corresponding to a delayed self-inhibition ($\gamma < 0$) or self-excitation ($\gamma > 0$). The value of γ indicates relative strength of the delayed self-connection compared to the strength of the delayed lateral inhibition (as used in Eq. (4.18)). With the delayed self-connection γ, the critical delay for the 2-D triangular lattice becomes

$$\tau_{crit} \rightarrow ln\left(\frac{\gamma-6}{2\gamma-3}\right) \text{ as } N \rightarrow \infty. \tag{4.33}$$

We restrict the γ to the range $-3 < \gamma < 6$ to insure $0 < \lambda_{max} < -\lambda_{min}$, which was assumed in the analysis of § 4.4. Equation (4.33) indicates that the triangular lattice can oscillate even when there is an overall self-excitation, as long as $\gamma < 1.5$.

It may seem surprising at first that the type of lattice can so greatly affect the network dynamics. The key to understanding the difference is realizing that on the triangular lattice, lateral inhibition (or, equivalently, antiferromagnetism) is *frustrated*, but on the square lattice it is not. On the square lattice, in fact, lateral inhibition is exactly equivalent to lateral excitation via a Mattis transformation [Mattis, 1976]. This difference is also seen in 2-D magnetic models: While ferromagnets on square and triangular lattices behave nearly identically (both are nonfrustrated), the corresponding 2-D antiferromagnets are quite different, due to the presence of frustration in the triangular lattice, but not the square lattice [Wannier, 1950]. As discussed in § 4.3.2, the presence of frustration seems to be essential for a delay network to support sustained oscillation.

To derive the above results, Eqs. (4.31)-(4.33), we need the extremal eigenvalues of the connection matrix for nearest-neighbor inhibition on these 2-D lattices. The value of τ_{crit} can then be found immediately from Eq. (4.23). This sort of eigenvalue problem is frequently encountered in condensed matter physics, for example to describe the

vibrational modes of a 2-D lattice[2]. Therefore, the eigenvalue spectra will be presented without derivation (see, for example [Ashcroft and Mermin, 1976, Ch. 22]). We assume periodic boundary conditions and take N_1 and N_2 as the number of neurons along each of the two lattice vectors. (The total number of neurons is the product $N_1 N_2$.) As usual, the connection matrix obeys the normalization $\Sigma_j |T_{ij}| = 1$. The eigenvalues λ_{k_1,k_2} of T_{ij}, for the 2-D square lattice are given by

$$\lambda_{k_1,k_2}(T_{ij}) = \frac{\gamma - 2\left[\cos(2\pi k_1/N_1) + \cos(2\pi k_2/N_2)\right]}{|\gamma| + 4} , \tag{4.34}$$

where the indices range over the values $k_{1,2} = 0, 1, \ldots, (N_{1,2}-1)$. From (4.34), we find that for the square lattice, the ratio appearing in (4.23) limits to

$$\lim_{N_1,N_2 \to \infty} (\lambda_{max}/\lambda_{min}) = \frac{\gamma + 4}{\gamma - 4} , \tag{4.35}$$

To apply the analysis of § 4.4.1, which assumed $0 < \lambda_{max} < -\lambda_{min}$, we require $-4 < \gamma < 4$. From (4.35) and (4.23) we conclude that for $\gamma \geq 0$, $\tau_{crit} \to \infty$ for large 2-D square lattices, as $N_1, N_2 \to \infty$.

The eigenvalues for the triangular lattice are given by

$$\lambda_{k_1,k_2}(T_{ij}) = \frac{\gamma - 2\left[\cos(2\pi k_1/N_1) + \cos(2\pi k_2/N_2) + \cos\left(2\pi(k_1/N_1 - k_2/N_2)\right)\right]}{|\gamma| + 6} \tag{4.36}$$

where, again $k_{1,2} = 0, 1, \ldots, (N_{1,2}-1)$. In this case, the ratio in (4.23) limits to

$$\lim_{N_1,N_2 \to \infty} (\lambda_{max}/\lambda_{min}) = \frac{\gamma + 3}{\gamma - 6} . \tag{4.37}$$

[2]A very helpful discussion with R. D. Meade regarding this point is gratefully acknowledged.

and we require $-3 < \gamma < 6$ to insure $0 < \lambda_{max} < -\lambda_{min}$. Notice that for the triangular lattice,

$$\lim_{N_1, N_2 \to \infty} \left(\lambda_{max} / \lambda_{min} \right) > -1 \tag{4.38}$$

for $\gamma < 1.5$. Equations (4.38) and (4.23) indicate that a nearest-neighbor lateral inhibition network on a large triangular lattice has a finite value for τ_{crit} as long as the (delayed) self-excitation strength remains less than 1.5 times the lateral-inhibition strength. From (4.37) and (4.23), we can find the value of τ_{crit} given above in Eqs. (4.32) and (4.33).

The eigenmode associated with the most negative eigenvalue of (4.36) is $k_1 = 0$, $k_2 = 0$. This 0-wavevector mode is the in-phase (coherent) direction in state space, as shown in Fig. 4.11(a), which justifies the application of the large gain analysis and Eq. (4.23). Further consideration reveals that the mode associated with the most positive eigenvalue - the mode which first goes unstable to break the symmetry of the coherent oscillation - is the $\sqrt{3} \times \sqrt{3}$ mode shown in Fig. 4.11(b).

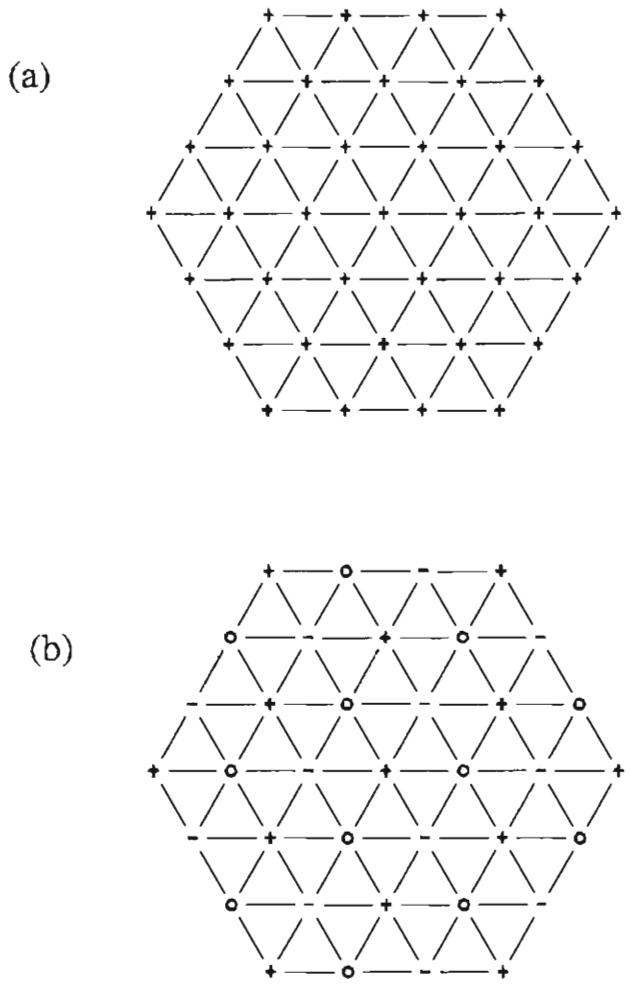


Fig. 4.11. Eigenmodes associated with the extremal eigenvalues of T_{ij} for the triangular lateral inhibition network, from Eq. (4.36). (a) The mode associated with λ_{min} is the coherent, in-phase mode, $k_1 = k_2 = 0$. This is the oscillatory mode which appears for the smallest value of delay. (b) The mode associated with λ_{max} is a $\sqrt{3} \times \sqrt{3}$ structure. The wave vectors for this mode are at the vertices of the hexagonal Brillouin zone. This is the first mode to break the symmetry of the coherent oscillatory mode, giving the value for τ_{crit}.

4.5.3. Random networks

Sustained oscillations in randomly connected neural networks have been considered previously for symmetric networks with parallel dynamics [Cabasino *et al.*,1988], which show at most period-2 oscillations [Peretto, 1984; Goles-Chacc *et al.*, 1985; Goles and Vichniac, 1986; Grinstein *et al.*, 1985; Frumkin and Moses, 1986; Marcus and Westervelt, 1989c] and for asymmetric networks with parallel dynamics [Amari, 1971; Shinomoto, 1986; Gutfreund *et al.*, 1988; Kürten, 1988], sequential dynamics [Shinomoto, 1986; Gutfreund *et al.*, 1988], and continuous-time dynamics [Amari, 1972; Kürten and Clark, 1986]. Periodic as well as chaotic dynamics in a mean field spin-glass model with delayed interaction have also been described [Choi and Huberman, 1983b].

We will only consider the effect of delay in *symmetric* random networks, and we find only simple (non-chaotic) oscillation above a critical delay. The absence of chaos in the symmetric continuous-time delay network (with monotonic nonlinearity) is not surprising, as the two limits of short and long delay are known to possess only fixed points and period-2 oscillations: A rigorous proof of this conjecture for the general delay-differential system has not been presented to our knowledge.

We consider a delay network with symmetric connection matrices whose elements T_{ij} (= T_{ji}) are independently fixed at one of three values (+,-,0). Any two neurons are connected by a positive connection with probability p_+ and by a negative connection with probability p_-. The *connectance* p is defined as $p = (p_+ + p_-)$; the *bias* q is defined as $q = (p_+ - p_-)$. The normalized matrix T_{ij}, has elements

$$T_{ij} = T_{ji} = \begin{cases} \pm\dfrac{1}{pN} & \textit{with probability } p_\pm \\ 0 & \textit{with probability } 1-p. \end{cases} \tag{4.39}$$

The eigenvalue spectrum of a random symmetric matrix is described by the famous Wigner semicircular law [Wigner, 1958; Edwards and Jones, 1976] (For a generalization of the semicircular law to random asymmetric matrices, see [Sommers *et al.*,1988]). The notation used here follows Edwards and Jones [1976]. For an $N \times N$ random symmetric matrix whose elements have a mean M_0/N and a variance σ^2/N, the spectrum of eigenvalues $\rho(\lambda)$ converges for large N to a continuous semicircular distribution. For $M_0 = 0$,

$$\rho_0(\lambda) = \begin{cases} \dfrac{\left(4\sigma^2 - \lambda^2\right)}{2\pi\sigma^2} & |\lambda| < 2\sigma \\ 0 & |\lambda| > 2\sigma \end{cases} \tag{4.40a}$$

and for $M_0 \neq 0$,

$$\rho(\lambda) = \begin{cases} \rho_0(\lambda) & |M_0| < \sigma \\ \rho_0(\lambda) + \dfrac{1}{N}\delta\left(\lambda - M_0 + \sigma^2/M_0\right) & |M_0| > \sigma \end{cases} \tag{4.40b}$$

For the (+,-,0) matrix, Eq. (4.39), we identify

$$M_0 \;\leftrightarrow\; \frac{q}{p}\,, \tag{4.41a}$$

$$\sigma^2 \;\leftrightarrow\; \frac{1}{p^2 N}\left(p - q^2\right). \tag{4.41b}$$

From Eq. (4.40) and Eq. (4.41), we can find the maximum and minimum eigenvalues of

T_{ij}. Setting $T_{ii} = 0$ adds a term of O(1/N) to all of the eigenvalues; we will neglect this and all terms O(1/N). These results are therefore valid only for large N, where $N^{1/2} << N$.

$$\lambda_{max} = \begin{cases} \frac{2}{p}\sqrt{\frac{p-q^2}{N}} + O\left(\frac{1}{N}\right) & for\ q < \sqrt{\frac{p}{N}} \\ \frac{q}{p} + O\left(\frac{1}{N}\right) & for\ q > \sqrt{\frac{p}{N}} \end{cases} \tag{4.42a}$$

$$\lambda_{min} = \begin{cases} -\frac{2}{p}\sqrt{\frac{p-q^2}{N}} + O\left(\frac{1}{N}\right) & for\ -q < \sqrt{\frac{p}{N}} \\ \frac{q}{p} + O\left(\frac{1}{N}\right) & for\ -q > \sqrt{\frac{p}{N}} \end{cases} \tag{4.42b}$$

The condition $|\lambda_{max}/\lambda_{min}| < 1$ is only satisfied when $-q > (p/N)^{1/2}$, suggesting that a symmetric random network must be biased sufficiently negative before it will oscillate for small delay ($\tau < \sim 1$).

4.5.4. Random symmetric dilution of the all-inhibitory network

An example of a random symmetric network that will oscillate for small delay is the randomly diluted inhibitory network. For this network $p_+ = 0$ and $p = -q = p_-$. To O(1/N), the maximum and minimum eigenvalues are

$$\lambda_{max} = \frac{2}{\sqrt{N}}\left(\frac{1}{p_-} - 1\right)^{1/2}, \tag{4.43a}$$

$$\lambda_{min} = -1. \tag{4.43b}$$

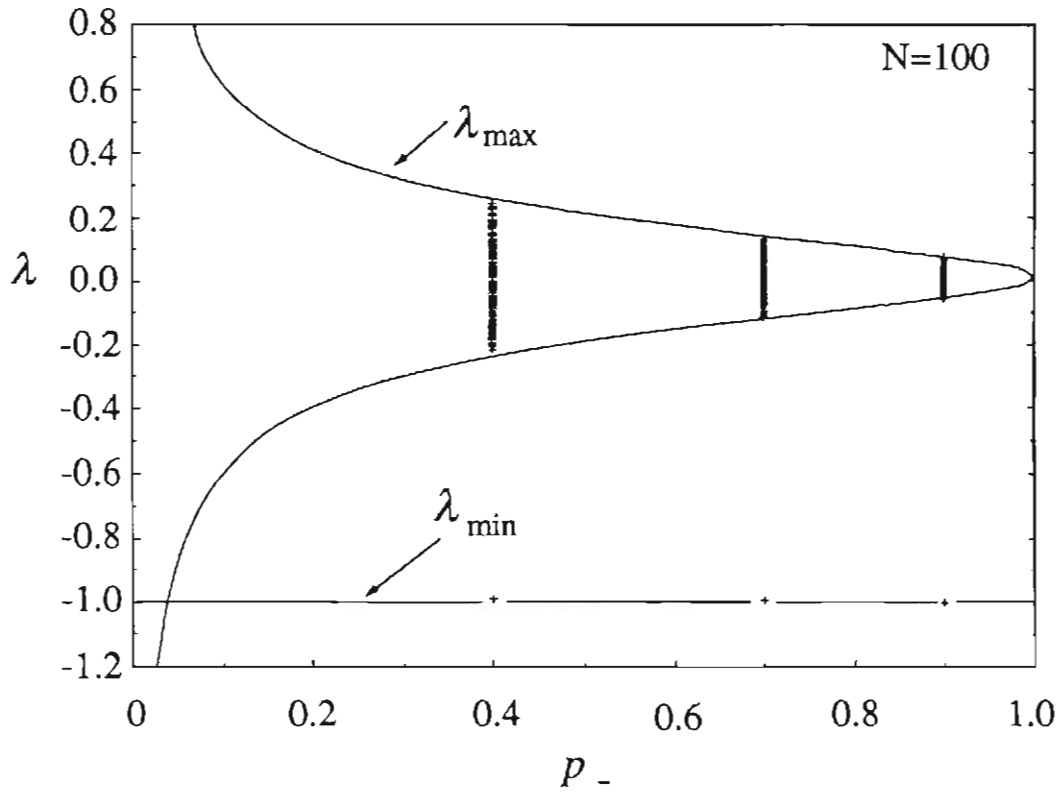


Fig. 4.12. The range of connection eigenvalues for a symmetrically diluted inhibitory network with N=100 from Eqs. (4.40) and (4.41) is plotted as a function of the connectance p_- (solid curves). The line at λ = -1 indicates a single eigenvalue λ_{min} lying outside of the quasi-continuous distribution. The small crosses are eigenvalues computed for randomly generated symmetric 100 × 100 matrices with p_- = 0.4, 0.7, and 0.9.

Fig. 4.12 shows the theoretical range of eigenvalues for a 100 × 100 randomly diluted inhibitory matrix as a function of connectance p_-. The small crosses are the eigenvalues of computer-generated random (-,0) matrices with $p_- = 0.4$, 0.7 and 0.9.

For the randomly diluted inhibitory network, with or without delay, the neuron gain at which the origin becomes unstable via a pitchfork bifurcation, creating fixed points away from the origin, is given by

$$\beta = \frac{\sqrt{N}}{2}\left(\frac{1}{p_-} - 1\right)^{-1/2} \qquad \text{(pitchfork).} \tag{4.44}$$

Because λ_{min} is independent of connectance, the delay at which the origin loses stability by a Hopf bifurcation is also independent of connectance. Inserting $\lambda_{min} = -1$ into Eq. (4.25) gives $\tau_H \equiv \pi/2\beta$, the small-delay limit being appropriate for large N and therefore large β.

The large-gain analysis of § 4.4 can be applied to the diluted inhibitory network when N is large. At large N the eigenvector associated with λ_{min} is nearly coherent, that is, the differences in $|u_i|$ along the eigenvector associated with λ_{min} are small compared to $|u_i|$ and appear numerically to vanish as $N \to \infty$. Applying Eq. (4.23) gives a gain-independent critical delay which depends on the connectance. From Eq. (4.23) and Eq. (4.43), the randomly diluted inhibitory network will not oscillate in the large-gain limit for $\tau < \tau_{crit}$, where

$$\tau_{crit} = -ln\left[1 - \frac{2}{\sqrt{N}}\left(\frac{1}{p_-} - 1\right)^{1/2}\right]. \tag{4.45}$$

Fig. 4.13 shows τ_{crit} as a function of connectance p_- for $N = 1000$. (At $p_- = 1$, the

result from Eq. (4.24) is used instead of Eq. (4.45) which neglects terms of O(1/*N*) and is therefore not valid precisely at $p_- = 1$.) Figure 4.13 shows that for a very mild dilution of connections, τ_{crit} is greatly increased, but additional dilution does little to increase τ_{crit} further. When the dilution $d \equiv (1 - p_-)$ is mild ($d << 1$), the right hand side of Eq. (4.45) can be expanded to yield

$$\tau_{crit} \cong \sqrt{\frac{4d}{N}} \qquad \left(N^{-1} << d << 1\right). \tag{4.46}$$

Eq. (4.46) can be compared to the critical delay for the undiluted all-inhibitory network, Eq. (4.24), to give a simple expression for the increase in critical delay due to random symmetric dilution:

$$\frac{\tau_{crit}^{(diluted)}}{\tau_{crit}^{(undiluted)}} \cong \sqrt{4dN} \qquad \left(N^{-1} << d << 1\right). \tag{4.47}$$

This result demonstrates how small random dilution of a large inhibitory network can be used to stabilize a network by increasing the critical delay for sustained oscillation.

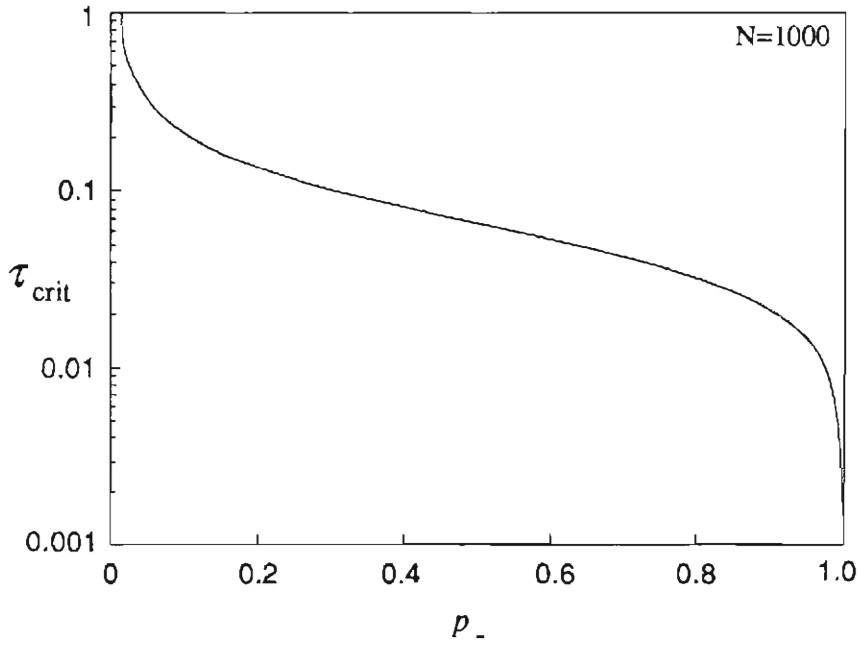


Fig. 4.13. Plot of the large-gain critical delay τ_{crit} as a function of connectance p_- for the diluted inhibitory network with N = 1000. Note that very mild dilution greatly increases τ_{crit}. At the point p_- = 1 the result of Eq. (4.24) is used instead of Eq. (4.45) which neglects terms of O(1/N), and is not correct at p_- = 1.

4.5.5. Associative memory networks

Associative memory networks are designed to converge to one of a set of specified fixed points away from the origin. Which memory pattern is retrieved depends on the initial state of the network. The existence of many attractors with large basins of attraction is essential to the dynamics of an associative memory.

A variety of algorithms for adjusting the interconnections to efficiently store memories have been developed [see, for example, Denker, 1986a; Amit, 1989] The simplest and most well studied scheme for storing a set of memory states ξ_i^μ ($i = 1, ..., N$; $\mu = 1, ..., p$) is the Hebb rule [Hebb, 1949; Hopfield, 1982],

$$T_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \, \xi_j^\mu \; ; \quad T_{ii} = 0, \tag{4.48}$$

where p is the number of stored memory patterns. The storage capacity and dynamic properties of an analog Hebb-rule network are discussed extensively in § 5.4; we only mention a few relevant facts here. For random uncorrelated patterns, the maximum number of patterns that the Hebb rule can store is $\sim 0.14\,N$ in the limit of $p, N >> 1$. This capacity is for large neuron gain; at lower gain the capacity is less [Marcus *et al.*, 1990]. For all $p/N < 1$, the Hebb rule matrix always satisfies $|\lambda_{max}/\lambda_{min}| > 1$, suggesting that the Hebb network with delay will not oscillate for any finite delay. (However, we will show in Ch. 5 that sustained oscillation is present in the infinite-delay limit - that is, in the analog iterated-map network with Hebb-rule connections.)

A variation of the Hebb rule that is important for hardware implementation is the clipped Hebb rule [Denker, 1986a; Sompolinsky, 1986; van Hemmen, 1987], which restricts the interconnection matrix to a few values. The distribution of eigenvalues for a

clipped Hebb matrix T_{ij}^{*} is greatly affected by the details of the clipping algorithm as seen in the numerical data of Fig. 4.14. Fig. 4.14(a) shows the distinct eigenvalues $\lambda(T_{ij}^{*})$ for the clipping algorithm $T_{ij}^{*} = (1/Z)Sgn(T_{ij})$, where Z is the normalization $Z = \Sigma_i|Sgn(T_{ij})|$. This clipping algorithm introduces large negative eigenvalues but still satisfies $|\lambda_{max}/\lambda_{min}| > 1$ for all values observed of p. We conclude that networks built with this clipping algorithm will not oscillate as long as the delay is not much longer than the relaxation time - to be safe, when $\tau < \sim 1$. Experimentally (in the electronic circuit) and numerically, we find that this clipping algorithm does not produce sustained oscillation until the delay is much longer than the relaxation time ($\tau >> 1$). Fig. 4.14(b) shows the distinct eigenvalues for the negative-only clipping algorithm: $-T_{ij}^{**} = (1/Z)\theta(-T_{ij})$, where θ is the Heaviside function and $Z = \Sigma_i\, \theta(-T_{ij})$. This clipping algorithm, which sets all positive elements of the unclipped matrix T_{ij} to 0 and all negative elements to -1/Z, has the hardware advantage of only requiring a single inverting output from each neuron, as pointed out by Denker [1986b]. As seen in Fig. 4.14(b), this algorithm unfortunately introduces a large negative eigenvalue which can lead to sustained oscillation for a neuron delay of the order of the relaxation time ($\tau < \sim 1$).

4.6. CHAOS IN TIME-DELAY NEURAL NETWORKS

4.6.1. Chaos in neural network models

Relaxing the constraint of symmetric connections greatly enriches the repertoire of neural network dynamics and provides powerful computational properties that are not available in symmetric networks. The most important novel feature of asymmetric networks (with or without delay) is that attractors need not be fixed points. Depending on the details of the connection matrix and the network dynamics, the attractors in

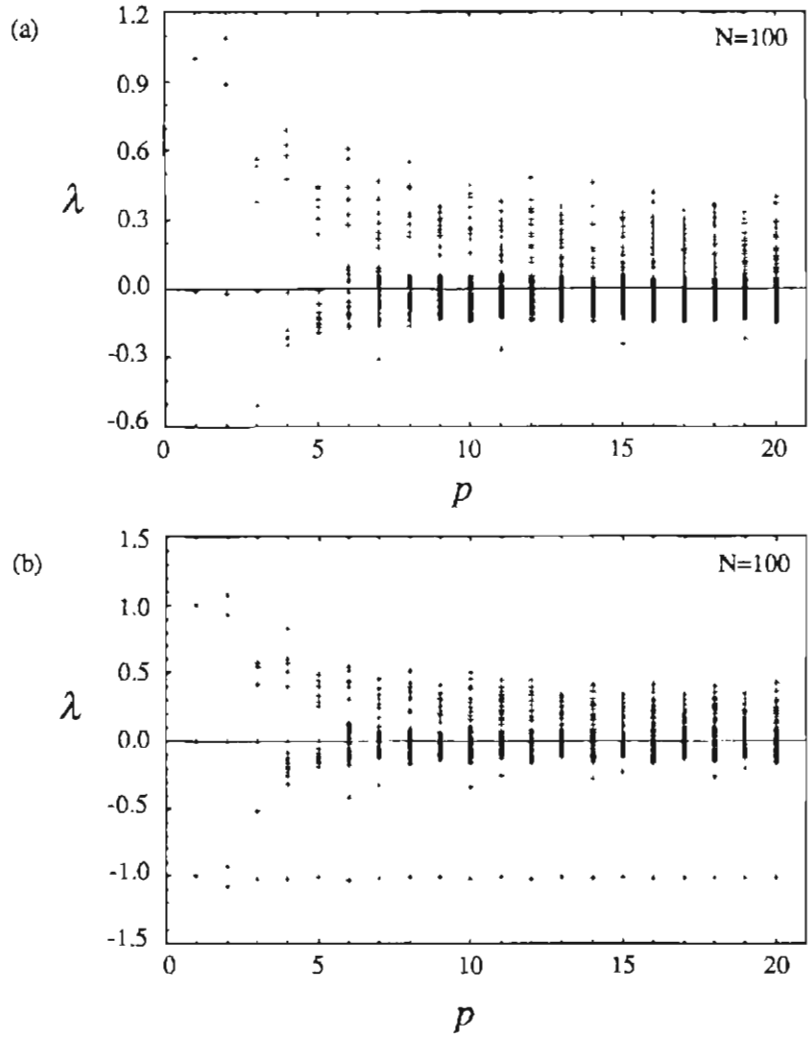


Fig. 4.14. Connection eigenvalues λ for clipped Hebb matrices plotted as a function of the number of stored random memories p, using two clipping algorithms discussed in the text. (a) Hebb matrix T_{ij} clipped according to $T_{ij}^{*} = (1/Z)sgn(T_{ij})$, with normalization $Z = \Sigma_i|Sgn(T_{ij})|$, gives an unbiased matrix and an eigenvalue distribution which satisfies $|\lambda_{max}/\lambda_{min}| > 1$ for all observed values of p. (b) Clipping algorithm which sets all positive T_{ij} to zero and all negative T_{ij} to $-1/Z$, with normalization $Z = \Sigma_i\, \theta(-T_{ij})$, has the advantage of only requiring a single output from each neuron, but produces a large negative eigenvalue that can lead to sustained oscillation. The data were obtained numerically for a 100×100 Hebb matrix T_{ij} with random memories as in Eq. (4.48).

asymmetric networks may be periodic or chaotic.

Chaos is usually taken to mean quasi-random behavior in a deterministic dynamical system [Guckenheimer and Holmes, 1983; Bergé *et al.*, 1984]. Typically, though not always, the dynamical system of interest is low dimensional. This is, of course, not the case for delay systems [see, for example: Farmer, 1982] or for neural networks with large N. The term "chaotic" is used both to describe a dynamical system (perhaps with a particular set of parameter values) or to describe an attractor of a dynamical system. The distinction is a crucial one, however, since a chaotic attractor may occupy only a small volume of the system's state space. Chaotic and non-chaotic attractors often coexist in state space, each attractor having its own basin of attraction. This can make the presence of chaos in a high-dimensional system (such as a neural network) difficult to detect, since a particular set of initial conditions may, for example, lead to a fixed point, while a nearby chaotic attractor remains undetected.

A definitive signature of a chaotic attractor is sensitivity to initial conditions, which means that close-lying points on the attractor move away from each other as time evolves (for short times). In large random dynamical systems the corresponding signature of chaos is the vanishing of an average autocorrelation function [Sompolinsky *et al.*, 1988].

A. Chaos in large asymmetric networks

Chaos in large deterministic neural networks with random asymmetric connections has been studied extensively in several network models [Kürten and Clark, 1986; Shinomoto, 1986; Derrida, 1988a; Kürten, 1988; Sompolinsky, *et al.*, 1988; Gutfreund, *et al.* , 1988; Bauer and Martienssen, 1989; Spitzner and Kinzel, 1989; Renals and Rohwer, 1990]. Unfortunately, a consistent picture of when and how chaos arises from random connections has not yet emerged.

Sompolinsky *et al.* [1988] considered the large-N behavior of a network of analog neurons with continuous-time dynamics. They found that when the connection matrix elements are independent random variables with zero mean, and with zero correlations between T_{ij} and T_{ji}, the *only* attractors (in the large-N limit) are: (1) the fixed point at the origin for low neuron gain; and (2) a chaotic state above a critical value of gain. Numerical evidence in support of this claim is given by Bauer and Martienssen [1989], who also describe a transition to chaos via quasi-periodicity (see also [Renals and Rohwer, 1990]). However, Gutfreund *et al.*[1988], in an investigation of small random networks of binary neurons with discrete-time dynamics, found that long-period attractors exist only when the connection matrix is completely asymmetric, but that whenever there is a correlation between T_{ij} and T_{ji}, short-period attractors predominate. This result suggests that the presence of chaos in large random networks is not so common, being present only in fully asymmetric networks. Spitzner and Kinzel [1989] present numerical evidence to the contrary: They find that random networks of binary neurons with parallel updating show a sharp transition from a so-called frozen state to a chaotic state as a function of the correlation between T_{ij} and T_{ji}, and that the transition to the chaotic state occurs at a non-zero value of the correlation. At the opposite extreme, analytical and numerical work of Crisanti and Sompolinsky [1987] for the asymmetric spherical model (an approximation to analog neurons) suggests that as $N \rightarrow \infty$ all frozen states disappear, leaving only chaos, as soon as an infinitesimal amount of asymmetry is introduced into the connections .

Shinomoto [1986], extending early work by Amari [1971], considered the effect of random connections for distributions of connection strengths with non-zero mean. He presents a numerically derived phase diagram showing that randomly connected binary neurons with parallel updating are chaotic only when the mean of the distribution is within a narrow range. For a large negative mean, only period-2 attractors are observed;

for a positive mean, only fixed point attractors are observed.

An alternative approach to identifying chaos in large random asymmetric networks is based on the sensitivity to initial conditions described above [Derrida, 1988a]. The idea here is to follow the evolution of a statistically averaged distance between two initial states and observe whether this distance converges or diverges under the dynamics of the network. Derrida [1988a] treated a highly dilute asymmetric spin glass model in this manner and identified a transition to a chaotic phase, where pairs of initial conditions always diverge, as the connectivity of the network is decreased. Kürten [1988] showed that for dilute networks of binary neurons, Derrida's transition also marks a transition to a phase in which the mean length of limit cycles grows exponentially with the size of the system.

B. Chaos in small networks

Large system size is not necessary for the existence of chaos in neural networks. This fact has been demonstrated for continuous-time analog networks by Kepler *et al.*[1989] who used a 6-neuron electronic network with computer-controlled interconnections to rapidly test many random matrices for chaotic dynamics. They found that chaos was rare but present. They also identified some general characteristics of the connection matrices that result in chaotic networks. Matrices leading to chaos tend to have average loop correlations that obey the following trends: $\langle T_{ij}T_{ji}\rangle < 0$; $\langle T_{ij}T_{jk}T_{ki}\rangle \sim 0$; $\langle T_{ij}T_{jk}T_{km}T_{mi}\rangle > 0$. Babcock and Westervelt [1986a ;1986b] have shown that a simple analog network of two neurons with an inductive component in the coupling can become chaotic when driven by an oscillating external current.

C. Chaos in time delay networks

Neural networks with nonsymmetric connections and time delay can be configured as

associative memories for the storage and recall of sequences of patterns. These networks have been described and studied by a number of researchers [Grossberg, 1970; Kleinfeld, 1986; Somplinsky and Kanter 1986; Gutfreund and Mezard, 1988; Riedel, *et al.*, 1988; Herz, *et al.*, 1988; Kühn, *et al.*, 1989]. It has been shown that for certain parameter values, these large sequence-generating networks can also be chaos-generating [Riedel, 1988]. Chaos in scalar delay-differential systems will be discussed in § 4.6.3.

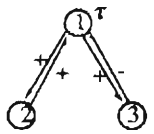
4.6.2. Chaos in a small network with a single time delay

The electronic analog network described in Ch. 3 shows endogenous chaos for particular connection configurations and network parameters. We now describe one such example using three neurons, one with delay. The dynamical equations for the chaotic network are

$$C_i \dot{u}_i(t') = -\frac{1}{R_i}u_i(t') + \sum_{j=1}^{N} T'_{ij} f_j\left(u_j\left(t' - \tau'_j\right)\right) , \quad i = 1,2,3. \tag{4.49a}$$

where

$$T_{ij}' = \frac{1}{10^5 \Omega}\begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} , \tag{4.49b}$$



$R_i = (\Sigma_j |T'_{ij}|)^{-1}$, and $C_1 = C_2 = C_3 = 10nF$. The characteristic relaxation times for the three neurons (in the absence of any delay) are $R_1C_1 = 0.5ms$, $R_2C_2 = 1.0ms$, $R_3C_3 = 1.0ms$. The neuron transfer functions are well-approximated by *tanh* functions with the following gains and amplitudes:

$$f_1(u) = 3.8\tanh(8.0u),$$
$$f_2(u) = 2.0\tanh(6.1u), \qquad (4.49c)$$
$$f_3(u) = 3.5\tanh(2.5u),$$

Only neuron 1 is delayed,

$$\tau_1' \equiv \tau;$$
$$\tau_2' = \tau_3' = 0. \qquad (4.49d)$$

The outputs of neurons 1 and 2 are shown in Fig. 4.15 for four values of the delay τ ranging from 0.64 ms to 0.97 ms. For $\tau < 0.64$ ms the system shows limit cycle behavior similar to that shown in Fig. 4.15(a). In the range $\tau =$ 0.64 ms - 0.97 ms the system undergoes a series of period doubling bifurcations leading to chaos. As the delay is increased beyond 0.97 ms, both chaotic and periodic regimes are found.

4.6.3. Chaos in delay systems with noninvertible feedback

The chaotic circuit described above is closely related to a well-studied class of chaotic delay-differential systems with noninvertible feedback [see, for example: Mackey and Glass, 1977; Farmer, 1982]. These systems are defined by a scalar delay-differential equation of the form

$$\dot{x}(t) = -ax(t) + h\big(x(t-\tau)\big), \qquad (4.50)$$

where $a, \tau > 0$, and the function h is noninvertible (also called "humped" or "mixed"). Equation (4.50) has been studied in the context of white blood-cell production

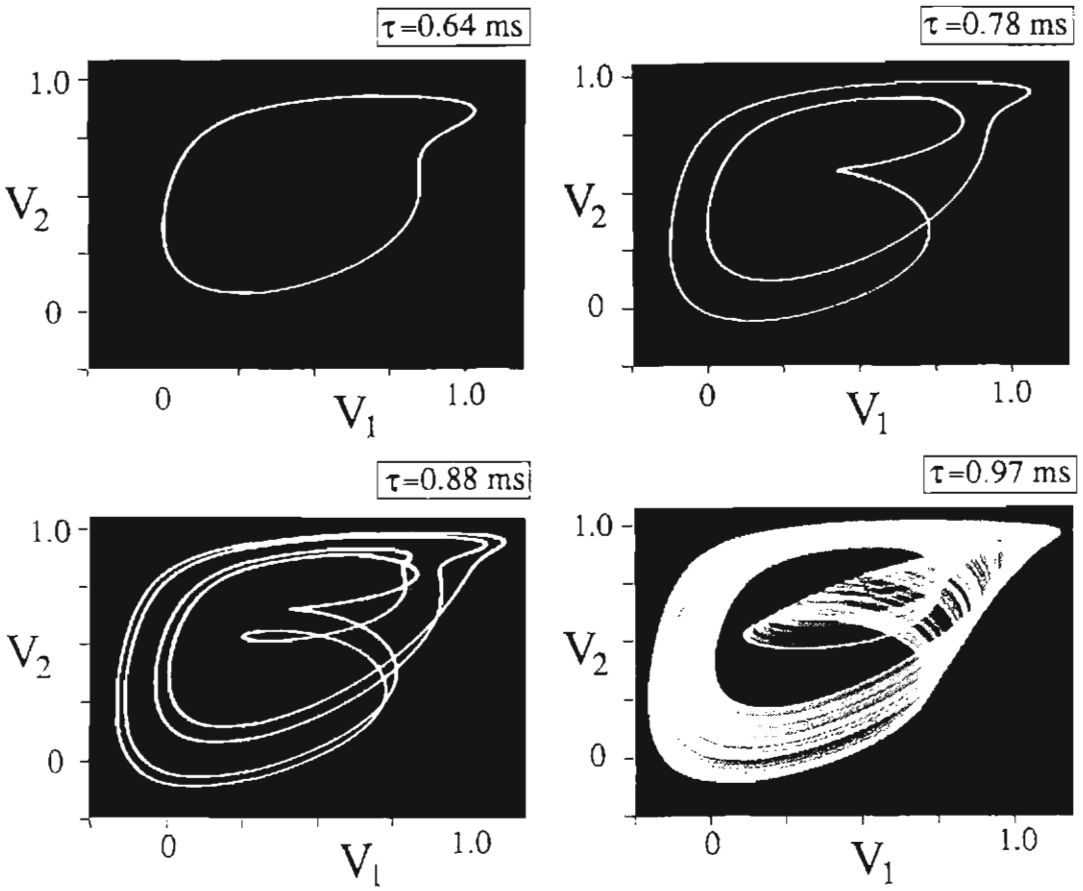


Fig. 4.15. Period doubling to chaos in the electronic analog network (described in Ch. 3) as the delay of neuron one is increased. The dynamical equations for this three-neuron circuit are given by Eq. (4.49). Pictures are photographs of the oscilloscope screen.

[Mackey and Glass, 1977; Glass and Mackey, 1988], recurrent inhibition in a three-cell circuit in the hippocampus [Mackey and an der Heiden, 1984], and other biological and ecological systems [Glass and Mackey, 1988 (and references therein)]. Analysis of (4.50) indicates that the noninvertiblity of h is crucial for chaos [an der Heiden and Walther, 1983; Hale and Sternberg, 1988].

The relation between (4.50) and the electronic network, with its three *monotonic* neurons, can be seen by plotting the total feedback f_{tot} to neuron 1,

$$f_{tot}(u_1(t-\tau)) = f_2\big(f_1(u_1(t-\tau))\big) - f_3\big(f_1(u_1(t-\tau))\big), \tag{4.51}$$

which is noninvertible, as shown in Fig. 4.16(b). Note that using neurons with different gains is necessary (in this example) to obtain noninverting feedback. Though the correspondence between the (4.49) and (4.50) is not perfect, we feel that the noninvertibility of f_{tot} lies at the heart of the chaotic behavior in the electronic network. In the limit $\tau \to \infty$, Eq. (4.50) and the equation for u_1 from (4.49) can both be written as a 1–dimensional iterated map:

$$x(t+1) = H(x(t)) \tag{4.52}$$

where the function H (= h/a or f_{tot}) is, again, a noninvertible or humped function. The logistic map [May, 1976] is a famous case of a noninvertible iterated map (4.52) whose behavior has been studied extensively [see Bergé *et al.*, 1984].

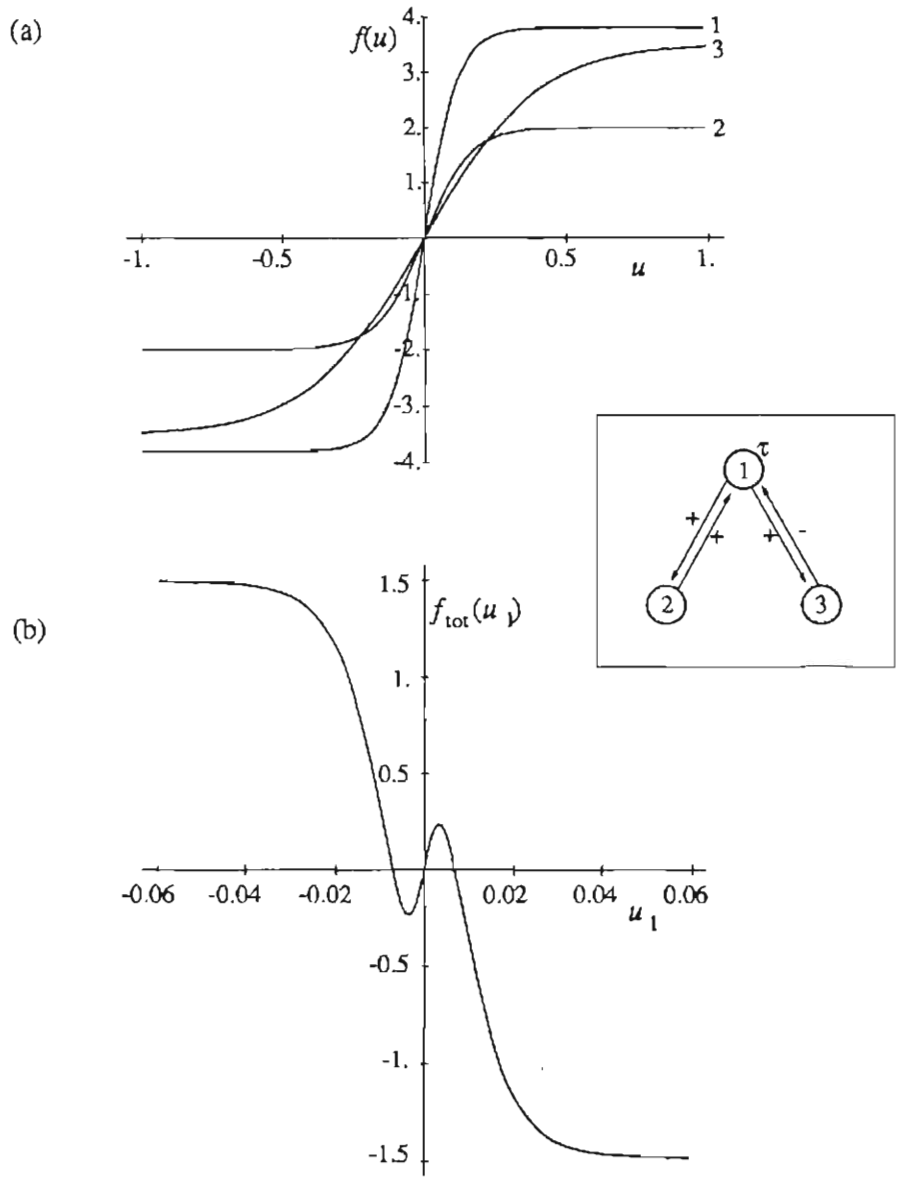


Fig. 4.16. Illustration of how monotonic nonlinearities can combine to give noninvertible feedback. (a) The three neuron transfer functions in Eq. (4.49c). (b) The total feedback to neuron $f_{tot}(u_1(t-\tau)) = f_2(f_1(u_1(t-\tau))) - f_3(f_1(u_1(t-\tau)))$. Note that f_{tot} is noninvertible.

4.7. DISCUSSION AND REVIEW OF RESULTS

This chapter is quite long and contains a number of new results. This final section is included to provide a summary of its contents.

We have considered the stability of analog neural networks with delayed response. The aim has been to extend the stability condition: "symmetric connection implies no oscillation" - which is valid when the neurons have instantaneous response - to a more realistic model of neural networks where time delay is included. We find that symmetrically connected networks can show sustained oscillation when the neurons have delayed output, but only when the ratio of delay to relaxation time exceeds a critical value.

At low neuron gain, linear stability analysis about the origin suggests that for $\tau < -\pi/(2\beta\lambda_{min})$ a symmetric network will not oscillate. In this inequality, τ is the neuron delay in units of the network relaxation time, β is the gain (maximum slope) of the neuron transfer function at the origin and λ_{min} is the minimum eigenvalue of the connection matrix T_{ij} as defined in Eq. (4.3).

The stability criterion based on linear stability analysis is valid at all values of gain but becomes overly conservative in the large-gain limit. We find experimentally and numerically that symmetric networks with extremal eigenvalues satisfying $|\lambda_{max}/\lambda_{min}| > 1$ do not oscillate as long as the delay is comparable to or less than the network relaxation time. In contrast, symmetric networks satisfying $|\lambda_{max}/\lambda_{min}| < 1$ do show coexisting fixed point and oscillatory attractors at large gain. There exists a critical delay τ_{crit} in the large-gain limit below which oscillatory attractors vanish and only fixed points attractors are observed. For symmetric networks in which the oscillatory mode present for the smallest delay is *coherent* (as defined in § 4.4.1), sustained oscillation vanishes for $\tau < \tau_{crit} = -ln(1+ \lambda_{max}/\lambda_{min})$. This result is independent of gain and

is useful as $\beta \rightarrow \infty$, unlike the linear stability result (Eq. (4.17)).

The stability criteria have been tested numerically and in the electronic neural network described in Ch. 3. Agreement between theory, experiment and numerics is very good.

Some results for particular network topologies:

(a) The all-inhibitory network is the most oscillation-prone configuration of the delay network. For this configuration, the critical delay in the large-gain limit is given by $\tau_{crit} = ln((N\text{-}2)/(N\text{-}1)) \sim 1/N$, where N is the size of the network. Diluting the all-inhibitory network by randomly - but symmetrically - setting a small fraction $d << 1$ of the interconnections (T_{ij} and T_{ji}) to zero will increase the critical delay by a factor of $(4dN)^{1/2}$.

(b) Rings of symmetrically connected delay neurons will oscillate only when the ring is frustrated ($Sgn(\prod_{ring} T_{ij}) = -1$) *and* when there is an odd number of neurons in the ring.

(c) The critical delay for large two-dimensional networks with nearest-neighbor lateral inhibition can be either finite or infinite, depending on the type of lattice. For zero self-connection, $\tau_{crit} \rightarrow \infty$ as $N \rightarrow \infty$ for a square lattice, and $\tau_{crit} \rightarrow ln(2) = 0.693...$ as $N \rightarrow \infty$ for a triangular lattice.

(d) The Hebb rule, Eq. (5.13), satisfies $|\lambda_{max}/\lambda_{min}| > 1$ and, as expected, does not show sustained oscillation numerically or in the electronic network for any observed (finite) delay. Clipping algorithms, which limit the interconnections to a few strengths, can introduce large negative connection eigenvalues and produce sustained oscillation in networks with delay smaller than the network relaxation time.

Finally, we have discussed chaotic dynamics in asymmetric neural networks. An example of a chaotic three-neuron network with a single time delay was presented. A connection was made between asymmetric networks of *monotonic* analog neurons and a well-studied chaotic system that has noninvertible or "mixed" delayed feedback.

Chapter 5

THE ANALOG ITERATED-MAP NEURAL NETWORK

5.1. INTRODUCTION

In this chapter, we analyze in detail the dynamics of an analog neural network with discrete-time parallel dynamics. Because the network's dynamical equations form a set of coupled iterated maps, we will refer to the system as an *iterated-map neural network*. The main purpose of this chapter is to show that the notorious problem of sustained oscillations associated with parallel dynamics can be eliminated by using analog neurons. Specifically, we present a global stability criterion that places an upper limit on the gain (maximum slope) of the neuron transfer function. When satisfied, this criterion guarantees that a symmetrically connected iterated-map network will always converge to a fixed point [Marcus and Westervelt, 1989c]. As an application, we treat the problem of associative memory, and present novel phase diagrams for analog associative memories based on the Hebb rule and the pseudo-inverse rule [Marcus *et al.*, 1990]. These results show that analog associative memories can be updated in parallel over a broad range of neuron gains and storage ratios while maintaining good recall and *guaranteed convergence to a fixed point*. This feature distinguishes analog networks from the standard Ising-spin networks (with or without temperature) which, in general, must be updated sequentially to prevent oscillation.

We will also discuss a second important advantage of analog associative memories, which is that lowering the neuron gain can greatly increase the chances that an initial state

far from all memories will correctly flow to a recall state without getting trapped in a spurious attractor.

The subsections of Ch. 5 are organized as follows. In § 5.2 we define the iterated-map neural network and prove that for a broad class of transfer functions and symmetric connections, the only attractors are period-2 limit cycles and fixed points. In § 5.3, we then show that all limit cycles can be eliminated by lowering the neuron gain below a critical value. In § 5.4, we investigate analog associative memories based on the Hebb rule [Hebb, 1949; Hopfield, 1982] and the pseudo-inverse rule [Personnaz *et al.*,1985; Kanter and Sompolinsky, 1987], and present phase diagrams in the parameter space of neuron gain β and memory storage ratio α. In § 5.5, numerical results for the associative memory networks are presented. These results agree well with the analytical results of § 5.4. The numerical results in § 5.5 also show that the probability of retrieval is increased at low analog gain, suggesting the use of analog annealing to enhance recall. Applications of these results and conclusions are presented in § 5.6. Some lengthy - but important! - details are presented in two appendices: In appendix 5A, the storage capacity of the Hebb rule for the analog iterated-map network is derived. This analysis generalizes the cavity method approach of Domany *et al.* [1989]. In appendix 5B, storage and recall properties of the pseudo-inverse rule are derived.

5.2. ITERATED-MAP NETWORK DYNAMICS

The dynamical system investigated in this chapter is an iterated-map neural network in which all neurons have continuous input-output transfer functions and updating is done in parallel [Marcus and Westervelt, 1989c]. The network is defined by the set of coupled nonlinear maps,

$$x_i(t+1) = F_i\left(\sum_j T_{ij}x_j(t) + I_i\right) \quad , \qquad i = 1, ..., N \tag{5.1}$$

where the real variables $x_i(t)$ describe the state of the system at time t. Time, in its present usage, is a discrete index: $t = 0,1,2,...$, and can equivalently be thought of as a layer index in a feed-forward network with identical coupling between each layer (cf. [Meir and Domany, 1987; 1988]). The interconnection matrix T_{ij} is assumed *real* and *symmetric*. We also assume that the neuron transfer functions F_i are all *single-valued* and *monotonic* (without loss of generality, we take all F_i to be monotonically *increasing*) and may be different for each i. Notice that the functions F_i can be concave-up or concave-down at any finite argument and do not need to saturate to a finite value. However, to insure that the Liapunov functions presented below and in § 5.3 are bounded below, we require that all F_i increase in magnitude slower than linear for large negative and positive argument. An example of a neuron transfer function that satisfies these conditions is illustrated in Fig. 5.1(a). The maximum slope of each F_i is defined as the *gain* β_i for that neuron, as shown in Fig. 5.1(a).

There is a completely equivalent form of Eq. (5.1) that has the neuron transfer functions inside of the sum:

$$u_i(t+1)/R_i \;=\; \sum_j T_{ij} f_j\left(u_j(t)\right) + I_i \;. \tag{5.2}$$

Equation (5.2) describes the evolution of the neuron inputs $u_i(t)$ rather than the outputs $x_i(t)$, and is related to (5.1) by the change of variables:

$$u_i(t)/R_i \equiv \sum_j T_{ij}x_j(t) + I_i \;;\; f_i(z) \equiv F_i(z/R_i). \tag{5.3}$$

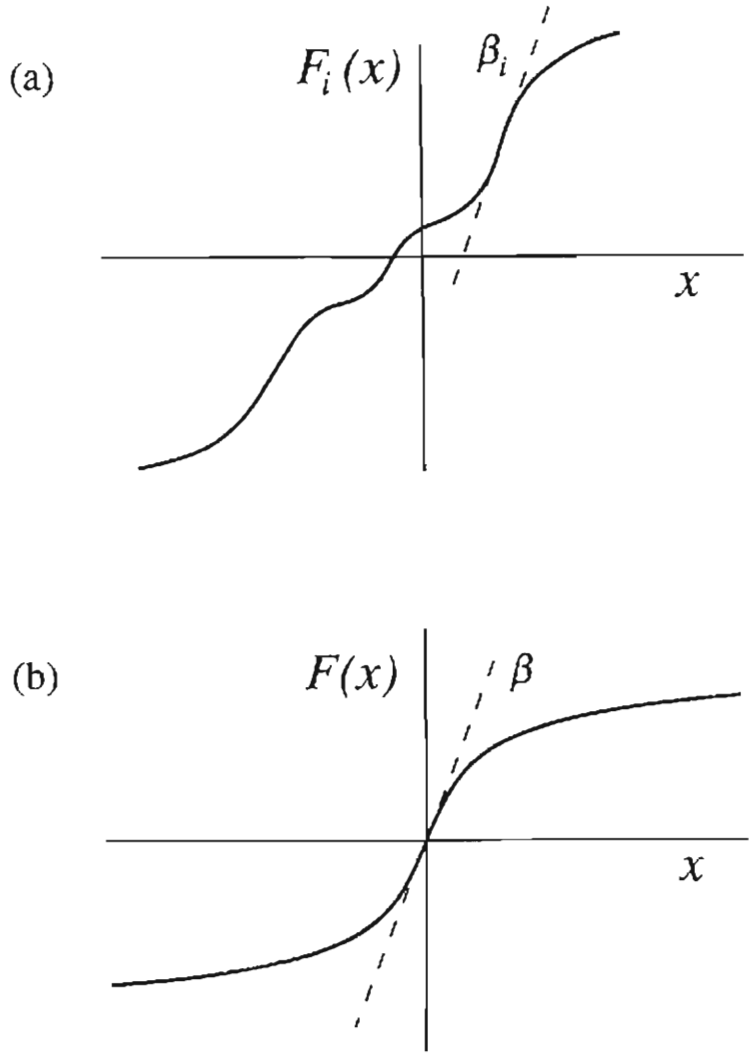


Fig. 5.1. (a) An example of a nonlinear neuron transfer function which meets the conditions for the dynamic properties given in § 5.2 and § 5.3. Those conditions are: Each function must be single valued and monotonic, and must grow in magnitude slower than linear in the limit of large positive or negative argument. The maximum slope β_i that appears in the stability criterion (5.11) is also indicated. (b) An example of a nonlinear function F (identical for all i) which meets the less general conditions assumed for the associative memory phase diagrams, Figs. 5.3 and 5.4. These conditions are given at the beginning of § 5.4.

The continuous-time version of Eq. (5.1), given by

$$dx_i(t)/dt = -x_i(t) + F_i\left(\sum_j T_{ij} x_j(t) + I_i\right) , \tag{5.4}$$

has the same fixed points as the iterated-map system (5.1), but not the same stability properties: For T_{ij} symmetric and all F_i monotonic, Eq. (5.4) will always converge to a fixed point, regardless of neuron gain. Finally, we note that the continuous-time system (5.4) is equivalent to the electronic circuit equations used in chapters 3 and 4,

$$C_i\, du_i(t')/dt' = -u_i(t')/R_i + \sum_j T_{ij} f_j\left(u_j(t')\right) + I_i \quad . \tag{5.5}$$

when the time constants R_iC_i are equal for all i. Equation (5.4) can be transformed into (5.5) by the change of variables (5.3), plus the rescaling of time, $t' \equiv \left(R_iC_i\right) t$.

We will now prove that all attractors of the iterated-map network (5.1) - or, equivalently, all attractors of (5.2) - are either fixed points or period-2 limit cycles [Marcus and Westervelt, 1989c]. The proof consists of showing that a function $E(t)$, defined as

$$\begin{aligned} E(t) \;=\; & -\sum_{i,j} T_{ij}\, x_i(t)\, x_j(t-1) \; - \sum_i I_i\left[x_i(t) + x_i(t-1)\right] \\ & + \sum_i \left[G_i\left(x_i(t)\right) + G_i\left(x_i(t-1)\right)\right] , \end{aligned} \tag{5.6a}$$

where

$$G_i(x_i) \equiv \int_0^{x_i} F_i^{-1}(z)dz \qquad , \tag{5.6b}$$

is a Liapunov function for the iterated-map network, Eq. (5.1), and that the minima of $E(t)$ are at either fixed points or period-2 limit cycles of Eq. (5.1).

The change in $E(t)$ between times t and $t+1$, defined as $\Delta E(t) \equiv E(t+1) - E(t)$, can be found from (5.6a) and (5.1) and the symmetry $T_{ij} = T_{ji}$, and is given by

$$\Delta E(t) = -\sum_i F_i^{-1}\big(x_i(t+1)\big)\Delta_2 x_i(t) + \sum_i \Big[G_i\big(x_i(t+1)\big) - G_i\big(x_i(t-1)\big)\Big] \qquad (5.7)$$

where $\Delta_2 x_i(t) \equiv x_i(t+1) - x_i(t-1)$ is the change in $x_i(t)$ over *two* time steps. For $G_i(x)$ concave up at all values of its argument x, we can write the following inequality (see Fig. 5.2):

$$G_i\big(x_i(t+1)\big) - G_i\big(x_i(t-1)\big) \;\le\; G_i'\big(x_i(t+1)\big)\Delta_2 x_i(t) \quad , \qquad (5.8)$$

where $G_i'\big(x_i(t+1)\big)$ is the derivative of $G_i(x)$ at the point $x = x_i(t+1)$. The case of equality in (5.8) only occurs when $\Delta_2 x_i(t) = 0$. The requirement that $G_i(x)$ be concave up is not very restrictive: it is satisfied as long as F_i is a single-valued, invertible, and increasing function. Inserting the inequality (5.8) into (5.7) gives

$$\Delta E(t) \;\le\; \sum_i \Big[G_i'\big(x_i(t+1)\big) - F_i^{-1}\big(x_i(t+1)\big)\Big]\,\Delta_2 x_i(t) \quad . \qquad (5.9)$$

The difference in the square bracket equals zero by Eq. (5.6b) giving the result:

$$\Delta E(t) \le 0 \qquad (5.10a)$$

$$\Delta E(t) = 0 \;\Rightarrow\; \Delta_2 x_i(t) = 0 \qquad (5.10a)$$

Thus $E(t)$ is a Liapunov function for the iterated-map network (5.1) and all attractors of (5.1) - where $\Delta E(t) = 0$ - satisfy $\Delta_2 x_i(t) = 0$, or $x_i(t+1) = x_i(t-1)$ for all i. Therefore all attractors of (5.1) are either fixed points or period-2 limit cycles.

5.3. A GLOBAL STABILITY CRITERION

In this section we show that all period-2 limit cycles of the iterated-map network (5.1) can be eliminated, leaving only fixed points, by lowering the neuron gains β_i to satisfy the stability criterion:

$$\frac{1}{\beta_i} > -\lambda_{min} \quad \text{for all } i\,, \tag{5.11}$$

where β_i (> 0) is the maximum slope of F_i and λ_{min} is the minimum eigenvalue of the connection matrix T_{ij} [Marcus and Westervelt, 1989c]. This criterion applies for any distribution of the (real) eigenvalues of T_{ij}. When T_{ij} has negative eigenvalues, $\lambda_{min}\left(T_{ij}\right)$ refers to the most negative eigenvalue. Assumptions made about the network in proving this result are the same as already used in § 5.2. As a reminder: The connection matrix T_{ij} is symmetric and the neuron transfer functions F_i are single-valued, monotonic and increase in magnitude slower than linear at large positive or negative argument.

The stability criterion (5.11) is derived by showing that the function $L(t)$, defined as

$$L(t) = -\frac{1}{2}\sum_{i,j} T_{ij} x_i(t) x_j(t) - \sum_i I_i x_i(t) + \sum_i G_i\left(x_i(t)\right), \tag{5.12}$$

with G_i given by Eq. (5.6b), is a Liapunov function of (5.1) *when the stability criterion*

is obeyed, and that the minima of $L(t)$ are at fixed points of (5.1).

From (5.1), (5.12) and the symmetry $T_{ij} = T_{ji}$, the change in $L(t)$ between times t and $t+1$, defined as $\Delta L(t) \equiv L(t+1) - L(t)$, can be written

$$\Delta L(t) = -\frac{1}{2}\sum_{i,j} T_{ij}\ \Delta x_i(t)\,\Delta x_j(t) - \sum_i F_i^{-1}\big(x_i(t+1)\big)\Delta x_i(t) + \sum_i \Big[G_i\big(x_i(t+1)\big) - G_i\big(x_i(t)\big)\Big], \tag{5.13}$$

where $\Delta x_i(t) \equiv x_i(t+1) - x_i(t)$. Note that $\Delta x_i(t)$ is the change in $x_i(t)$ in *one* time step.

We now construct an inequality similar to (5.8), but including a quadratic term along with the linear term. Choosing the coefficient of the quadratic term to be the minimum curvature of $G_i(x_i)$,

$$\min_{x_i}\left(d^2G_i/dx_i^2\right) = \beta_i^{-1}\ , \tag{5.14}$$

yields the following inequality, as illustrated in Fig. 5.2:

$$G_i\big(x_i(t+1)\big) - G_i\big(x_i(t)\big) \ \le\ G_i'\big(x_i(t+1)\big)\,\Delta x_i(t) - \tfrac{1}{2}\,\beta_i^{-1}\,\big(\Delta x_i(t)\big)^2\ . \tag{5.15}$$

Equations (5.13) and (5.15), and the equality $G'_i(x_i) = F_i^{-1}(x_i)$ from (5.6b) yield

$$\Delta L(t) \ \le\ -\frac{1}{2}\sum_{i,j}\left[T_{ij} + \delta_{ij}\beta_i^{-1}\right]\Delta x_i(t)\,\Delta x_j(t) \tag{5.16}$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. As long as the matrix $(\mathbf{T} + \mathbf{B}^{-1})$ - which appears in component form in the square brackets of Eq. (5.16) - is positive definite, then the right side of (5.16) is negative, and, *in that case*, we can immediately write

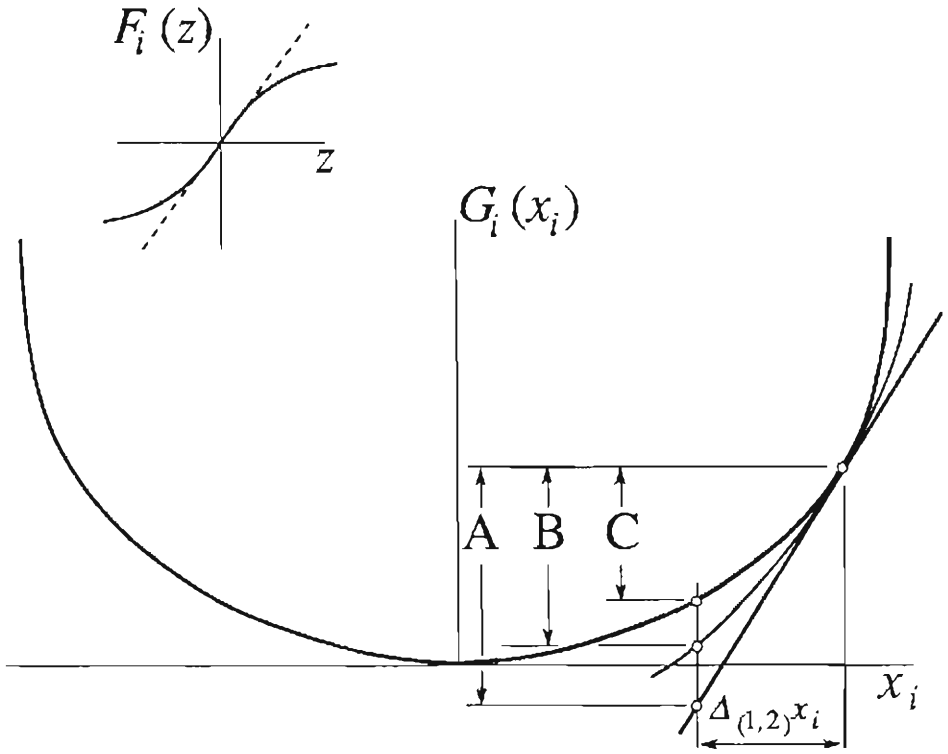


Fig. 5.2. Graphical representation of inequalities (5.8) and (5.15) for a typical sigmoid nonlinearity $F_i(z)$ with maximum slope β_i (inset). The concave-up function $G_i(x_i)$ is defined in Eq. (5.6b). A line and a parabola with second derivative $\beta_i^{-1} = \min_x [d^2G(x)/dx^2]$ are tangent to the curve $G_i(x_i)$ at the point $[x_i(t+1), G_i(x_i(t+1))]$. Eq. (5.15) is represented by the inequality $C \leq B$; Equation (5.8) is represented by the inequality $C \leq A$ where the case of equality, $C = A$, implies $\Delta_2 x_i(t) = 0$.

$$\Delta L(t) \leq 0 \quad , \tag{5.17a}$$

$$\Delta L(t) = 0 \Rightarrow \Delta x_i(t) = 0. \tag{5.17b}$$

Thus, as long as $(\mathbf{T} + \mathbf{B}^{-1})$ is a positive definite matrix, $L(t)$ is a Liapunov function for the iterated-map system (5.1). At the minima of $L(t)$ the condition $\Delta x_i(t) = 0$ holds for all i, thus all attractors are fixed points. A sufficient condition for $(\mathbf{T} + \mathbf{B}^{-1})$ to be positive definite is $\beta_i^{-1} > -\lambda_{min}$ for all i. This condition is therefore sufficient to guarantee that $L(t)$ is a Liapunov function and that all minima of (5.1) are fixed points, which gives the stability criterion (5.11).

5.4. ANALOG ASSOCIATIVE MEMORY

We now apply the iterated-map neural network to the problem of associative memory [Marcus *et al.*, 1990]. In this section we assume a less general form for the network, one in which $I_i = 0$ for all i and the nonlinear functions F_i are single-valued, odd functions and are the same for all i. We also assume the function F (now dropping the index i) has its maximum slope at zero input, $F'(0) = \beta$, and that the slope of F is a non-increasing function of the magnitude of the argument. As before, the maximum slope β will be referred to as the *gain* of the neurons. Possible forms for F include, but are not limited to, *tanh*-like functions including the transfer function for the electronic network, Eq. (3.3). As in § 5.3, we do not require that F saturate at large argument though it must increase in magnitude slower than linear at large positive or negative argument. We normalize the amplitude of F so that the accessible state space ("the hypercube") is of length O(1) on a side; that is, nonzero solutions of $m^* = F(m^*)$ are typically O(1). Fig. 5.1(b) shows a function which meets the conditions assumed in this section. Under these assumptions, the associative memory network is given by the set of

coupled maps

$$x_i(t+1) = F\left(\sum_j T_{ij} x_j(t)\right) \quad , \quad i = 1, ..., N. \tag{5.18}$$

We will consider connection matrices T_{ij} for two learning rules, the Hebb rule and the pseudo-inverse rule, storing random unbiased memory patterns, $\xi_i^{\mu} = \pm 1$. For the Hebb rule,

$$T_{ij} = \frac{1}{N} \sum_{\mu=1}^{\alpha N} \xi_i^{\mu} \xi_j^{\mu} \quad ; \quad T_{ii} = 0 \tag{5.19}$$

where αN is the number of stored memory patterns. For the pseudo-inverse rule

$$T_{ij} = \frac{1}{N} \sum_{\mu,\nu=1}^{\alpha N} \xi_i^{\mu} \left(C^{-1}\right)_{\mu\nu} \xi_j^{\nu} \quad ; \qquad T_{ii} = 0 \tag{5.20a}$$

where C^{-1} is the inverse of the correlation matrix,

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{\mu} \xi_i^{\nu} \quad . \tag{5.20b}$$

Notice that we are considering the modified pseudo-inverse rule with $T_{ii} = 0$ studied by Kanter and Sompolinsky [1987]. These authors showed that this modification increases the basins of attraction for the memories without sacrificing error-free recall. The analysis in the following two subsections assumes $\beta > 0$, $0 < \alpha < 1$ and $N >> 1$.

5.4.1. Hebb rule

A phase diagram for the Hebb rule, showing four distinct regions in the parameter space of analog gain β and the storage ratio α, is presented in Fig. 5.3. The four regions are characterized as follows: In the region marked 'origin' a fixed point at the origin, $x_i = 0$ for all i, is the global attractor. In the region marked 'spin glass' the origin is no longer an attractor, but neither are the memory recall states. In this region, the network converges to a fixed point with small [$O(N^{-1/2})$] overlap with all memories. In the region marked 'recall', fixed points having large overlaps with memory patterns exist and have large basins of attraction. In the 'recall' region the iterated-map network operates well as an associative memory. The boundary separating 'recall' from 'spin glass' is shown in Fig. 5.3 for the particular choice $F(z) = tanh(\beta z)$. With this choice of nonlinearity, this boundary agrees with the ferromagnetic transition curve found by Amit *et al.* [1985b; 1987] for the Ising-model associative memory at finite temperature. The present analysis leading to this curve, however, is not restricted to case $F(z) = tanh(\beta z)$. Details are given in appendix 5A. In the region marked 'oscillation' the stability criterion (5.11) is no longer obeyed and convergence to a fixed point is not guaranteed. Numerically, we find that limit cycles are quite abundant in this region, especially for larger values of β and α (see § 5.5).

The stability of the origin can be determined by linearizing Eq. (5.18) about the point $x_i = 0$, which gives N decoupled linear iterated maps: $\varphi_i(t+1) = \beta\lambda_i\varphi_i(t)$ for evolution along the i^{th} eigenvector of the matrix T_{ij},with associated eigenvalue λ_i. (cf. Eq. (4.5), the corresponding equation for the delay-differential system). For $|\beta\lambda_i| < 1$ for all i, the origin is stable, and because of the form of F, it is also the global attractor of Eq. (5.18). (The proof of this is based on a contraction mapping theorem. See: Ortega and Rheinboldt [1970], Thm 12.1.2].) Notice that when the eigenvalue spectrum

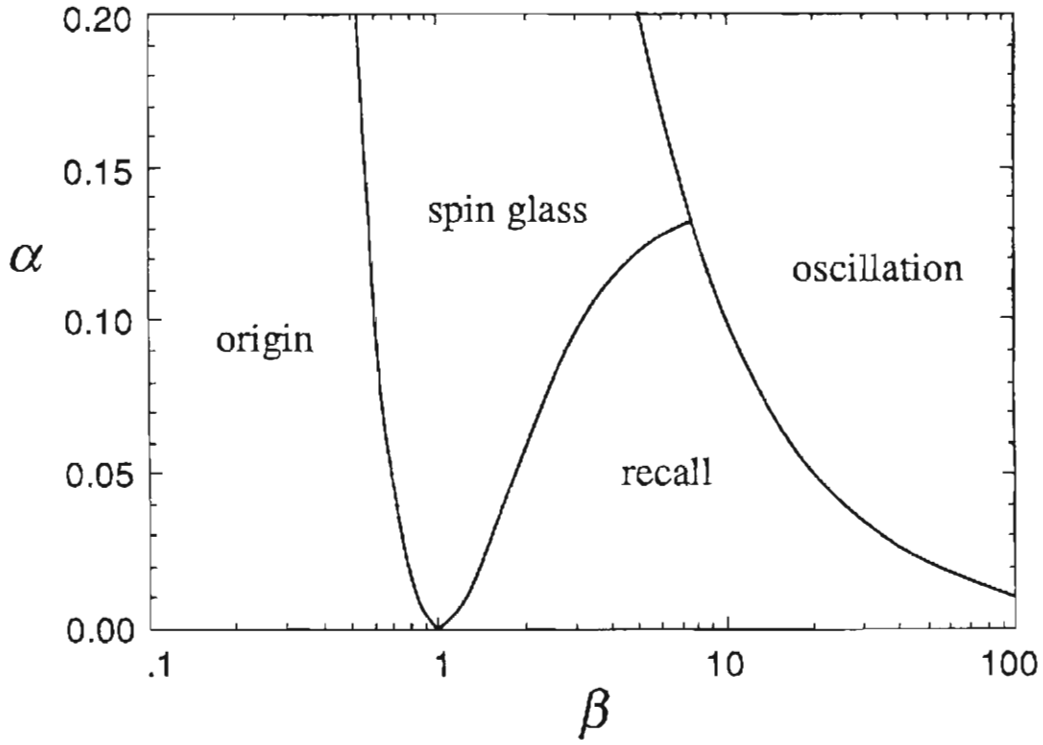


Fig. 5.3. Phase diagram for the Hebb rule associative memory with neuron transfer function $F(z) = tanh(\beta z)$. The parameter β is the neuron gain, and α is the number of stored patterns divided by the number of neurons N. All borders separating the regions are based on analysis at large N, as described in the text.

is "skewed negative" ($0 < \lambda_{max} < -\lambda_{min}$), the stability condition for the origin is identical to the global stability criterion (5.11).

The minimum and maximum eigenvalues for the Hebb matrix (5.19) with $\alpha < 1$ in the large-N limit are

$$\lambda_{min} = -\alpha, \qquad [N(1-\alpha) \text{ - fold degenerate}] \tag{5.21a}$$

$$\lambda_{max} = 1 + 2\sqrt{\alpha}\,, \qquad [\text{edge of continuous distribution}] \tag{5.21b}$$

[Geman, 1980; Silverstein, 1985; Crisanti and Sompolinsky, 1987], thus for $\alpha < 1$ the boundary where the origin loses stability is given by the condition $\beta = 1/\left(1+2\sqrt{\alpha}\right)$. From the value of λ_{min} in (5.21a) we can also identify the border of the oscillatory region as $\beta = 1/\alpha$. Crossing the 'origin'-'spin glass' line corresponds to a forward pitchfork bifurcation of the origin, analogous to a second order transition in thermodynamics. Note that this transition occurs along a different curve from the corresponding paramagnet-spin glass transition in the Ising model associative memory [Amit *et al.*, 1985b; 1987].

Crossing the border from the 'recall' region into the 'spin glass' region marks the disappearance of a fixed point having a large overlap with a single memory. As in the case of the Ising model network, this transition is due to the random overlaps of the state of the network with patterns other than the one being recalled. These overlaps generate an effective noise term which destabilizes the fixed point near the recalled pattern. Because our system has no reaction field, the analysis is somewhat simpler than either the replica [Amit *et al.*, 1985b; 1987] or cavity [Mezard *et al.*, 1987; Domany *et al.*, 1988] approaches used to analyze the thermodynamic Ising model network. In appendix 5A we derive a set of four self-consistent equations that determine the border between the 'recall' and 'spin glass' regions assuming random, unbiased memory patterns:

$$m^1 = \frac{1}{\sqrt{2\pi}} \int dy \exp\left(-y^2/2\right) F\left(\sigma y + m^1\right) \tag{5.22a}$$

$$C = \frac{1}{\sqrt{2\pi}} \int dy \exp\left(-y^2/2\right) F'\left(\sigma y + m^1\right) \tag{5.22b}$$

$$q = \frac{1}{\sqrt{2\pi}} \int dy \exp\left(-y^2/2\right) F^2\left(\sigma y + m^1\right) \tag{5.22c}$$

$$\sigma = \frac{\sqrt{\alpha q}}{1 - C} \tag{5.22d}$$

where $F'(z) \equiv dF(z)/dz$. The quantity m^1 in Eq. (5.22a) is the overlap of the network state vector with a single memory pattern, arbitrarily chosen to be pattern 1. In the recall state, these equations have a self-consistent solution with $m^1 \sim 1$. For the particular choice $F(z) = tanh(\beta z)$, the quantities C and q obey the usual Fischer relation $C=\beta(1-q)$ [Fischer, 1976].

5.4.2. Pseudo-inverse rule

The pseudo-inverse learning rule, Eq. (5.20), offers several advantages over the Hebb rule, chiefly a greater storage capacity, error-free recall states and the ability to store correlated patterns [Personnaz *et al.*,1985; Kanter and Sompolinsky, 1987]. Its primary disadvantage is that it is nonlocal, meaning that a given element of the connection matrix, T_{ij}, cannot be determined from the ith and jth elements of the memory patterns, but depends on all components of all memories. However, iterative learning algorithms have been presented which are local and which converge to the pseudo-inverse rule [Diederich and Opper, 1987].

A phase diagram for the pseudo-inverse rule showing three distinct regions depending on analog gain β and storage ratio α is shown in Fig. 5.4. This phase diagram differs from that of the Hebb rule in three distinctive ways: First, there is no

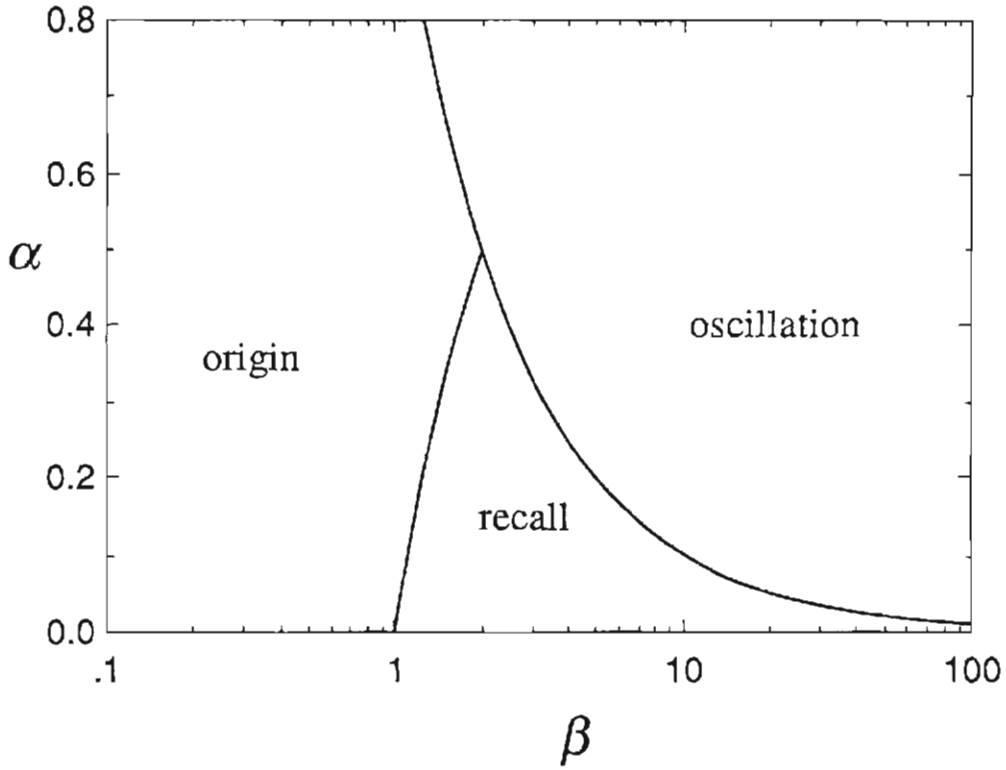


Fig. 5.4. Phase diagram for the pseudo-inverse rule (diagonal elements = 0) with sigmoidal neuron transfer function as described at the beginning of § 5.4. The parameter β is the neuron gain, and α is the number of stored patterns divided by the number of neurons N. All borders separating the regions are based on analysis at large N, as described in the text. Note that the pseudo-inverse rule does not possess a spin glass phase for $\alpha < 1$.

'spin glass' phase. This does not imply that the pseudo-inverse rule does not possess spurious attractors; just as for the Hebb rule, there are many spurious fixed point attractors within the recall and oscillatory regions which have small overlap with all memories. Unlike the Hebb rule, however, there is no region of the pseudo-inverse phase diagram where *only* spurious fixed-point attractors are found. The second difference is that the recall region is much larger, extending to $\alpha = 0.5$ for $\beta = 2$. Above this point, and for higher gain, recall states still exist, but convergence to a fixed point is not guaranteed. The third distinctive feature is the adjacency of the 'origin' and 'oscillation' regions at larger values of α. Crossing the border between these two regions, say by increasing β, constitutes a multiple flip bifurcation [Guckenheimer and Holmes, 1983] in which $N(1-\alpha)$ eigendirections about the origin simultaneously lose stability giving rise to period-2 limit cycles in the subspace orthogonal to all memories.

As in the Hebb rule phase diagram, the region marked 'origin' for the pseudo-inverse phase diagram satisfies $|\beta\lambda_i| < 1$ for all i, where λ_i are the N eigenvalues of the pseudo-inverse matrix (5.20). For $T_{ii} = 0$, the extremal eigenvalues in the limit of large N are given by

$$\lambda_{min} = -\alpha, \qquad [N(1-\alpha)\text{- fold degenerate}] \tag{5.23a}$$

$$\lambda_{max} = 1-\alpha\ . \qquad [N\alpha\text{ - fold degenerate}] \tag{5.23b}$$

[Kanter and Somplinsky, 1987]. Below $\alpha = 0.5$ the origin loses stability at gain $\beta = 1/(1-\alpha)$. This condition defines the border between the regions marked 'origin' and 'recall.' In appendix 5B we show that stable recall states appear as soon as this bifurcation occurs. From the stability criterion (5.11) and Eq. (5.23a), convergence to a fixed point is not guaranteed for $\beta > 1/\alpha$, which defines the region marked 'oscillation' in Fig. 5.4.

Adding a positive diagonal element $T_{ii} = \gamma > 0$ to the connection matrix shifts the

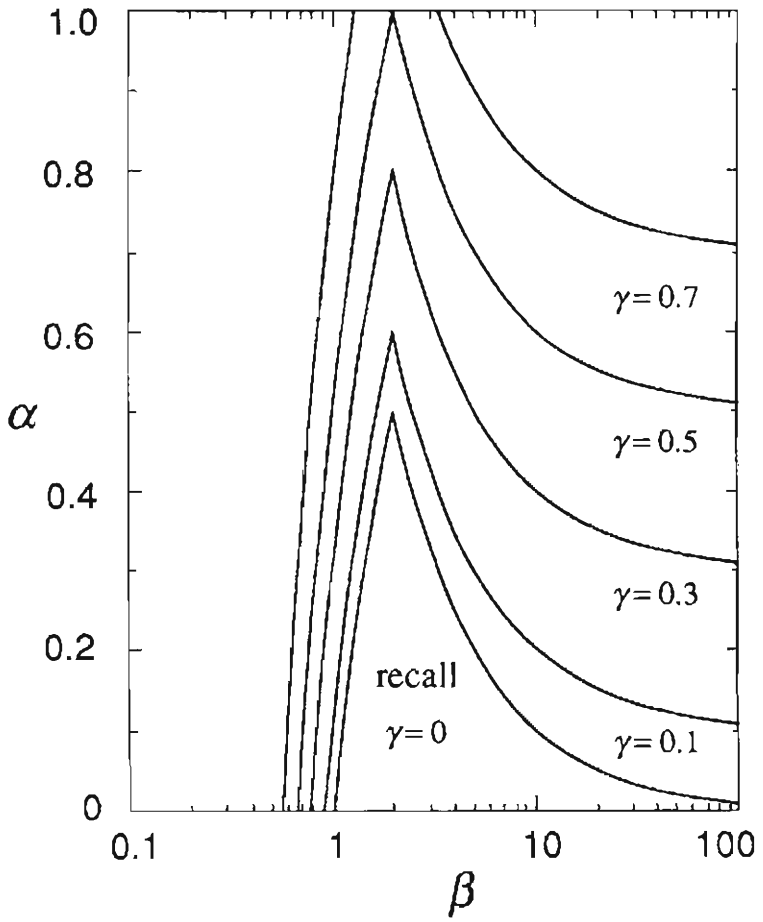


Fig. 5.5. The recall region for the pseudo-inverse rule for various values of diagonal element γ. Note that the maximum capacity in the recall region is for analog gain $\beta = 2$, regardless of γ. Although the recall region is expanded for positive diagonal element, too large a diagonal will greatly reduce the basins of attraction for the recall states, as discussed by Kanter and Sompolinsky [1987].

eigenvalues to $\lambda_{min} = -\alpha + \gamma$ and $\lambda_{max} = 1 - \alpha + \gamma$ and increases the maximum storage capacity in the recall region to $\alpha_{max} = 1/2 + \gamma$. [1] The recall region for several values of positive self-coupling are shown in Fig. 5.5. Note that the maximum always occurs at $\beta = 2$. Recently, Krauth *et al.* [1988] have demonstrated that using a small positive diagonal element with the pseudo-inverse rule in an Ising network (at zero temperature) increases the radius of attraction for the recall states.[2] For example, they find numerically that for α = 0.5, using a diagonal term of ~0.075 instead of zero increases the basins of attraction by about 50%. Too large of a diagonal term, however, greatly reduces the basins of attraction for the recall states [Kanter and Sompolinsky, 1987; Krauth *et al.*, 1988].

5.5. NUMERICAL RESULTS

5.5.1. Numerical verification of the phase diagrams

In this section, phase diagrams for the Hebb rule and pseudo-inverse rule are investigated numerically for networks of size N = 100 with $F(z) = tanh(\beta z)$ and random, unbiased memory patterns. The data in Figs. 5.6 and 5.7 show, as a function of analog gain β, the fraction of randomly chosen initial states which converged to a particular type of attractor - either the origin, a memory pattern (or its inverse), a spurious fixed point, or a period-2 limit cycle. These attractor types are the only possibilities. Each panel in these figures is for a fixed value of α, so each represents a horizontal slice through the phase diagrams for the Hebb rule (Fig. 5.3) or the pseudo-inverse rule

[1] This definition of γ differs from the one used in chapter 4: Here, the matrix elements T_{ii} are equal to γ. In chapter 4, the value γ is the diagonal matrix element *before* normalization. See, for example, Eq. (4.18).

[2] Krauth *et al.* [1988] use yet another definition of γ. For them, $T_{ii} = \gamma(1-\alpha)$ for large N.

(Fig. 5.4).

The data in each panel were generated as follows: For each of 38 values of β, ranging from $\beta \sim 0.3$ to $\beta \sim 90$, twenty T_{ij} matrices were generated using random, unbiased patterns, $\xi_i^\mu = \pm 1$. For each matrix, 50 initial states located at random corners of the state space ($x_i(0) = \pm 1$, $i = 1,...,100$) were chosen and the attractor for each was found by iterating the map, Eq. (5.18). The condition for convergence was $\|\vec{x}(t) - \vec{x}(t-2)\| < 10^{-6}$, where distances are defined by $\|\vec{z}\| \equiv (1/2N)\sum_i |z_i|$. Though the initial states were located at the corners of the hypercubic state space, all attractors were real-valued N-vectors located away from the corners of the state space. Plotted in each panel are the fractions of the 20x50 = 1000 runs for each value of β which converged to each of the four attractor types. A fixed point $\vec{x}$ * was counted as a recall state if, for any μ, $\left\| sgn(\vec{x}\,*) \pm \vec{\xi}^\mu \right\| < 0.05$; similar criteria were used to recognize the other attractor types.

Along the top of each panel in Figs. 5.6 and 5.7 is a strip marked 'orig.' , 'recall', etc. These strips show the regions of the theoretical phase diagram (from Figs. 5.3 and 5.4) for the particular value of α in that panel. The appearance of the various attractor types corresponds very closely to the theoretical regions in these slices, giving strong numerical support to the phase diagrams. Furthermore, the data indicate that the basins of attraction for limit cycles in the 'oscillation' region do occupy a significant part of state space as soon as the stability criterion is violated. That is, the 'oscillation' region is more than just the region where convergence to a fixed point is not guaranteed by the stability criterion, it is in fact the region where oscillatory modes are quite abundant.

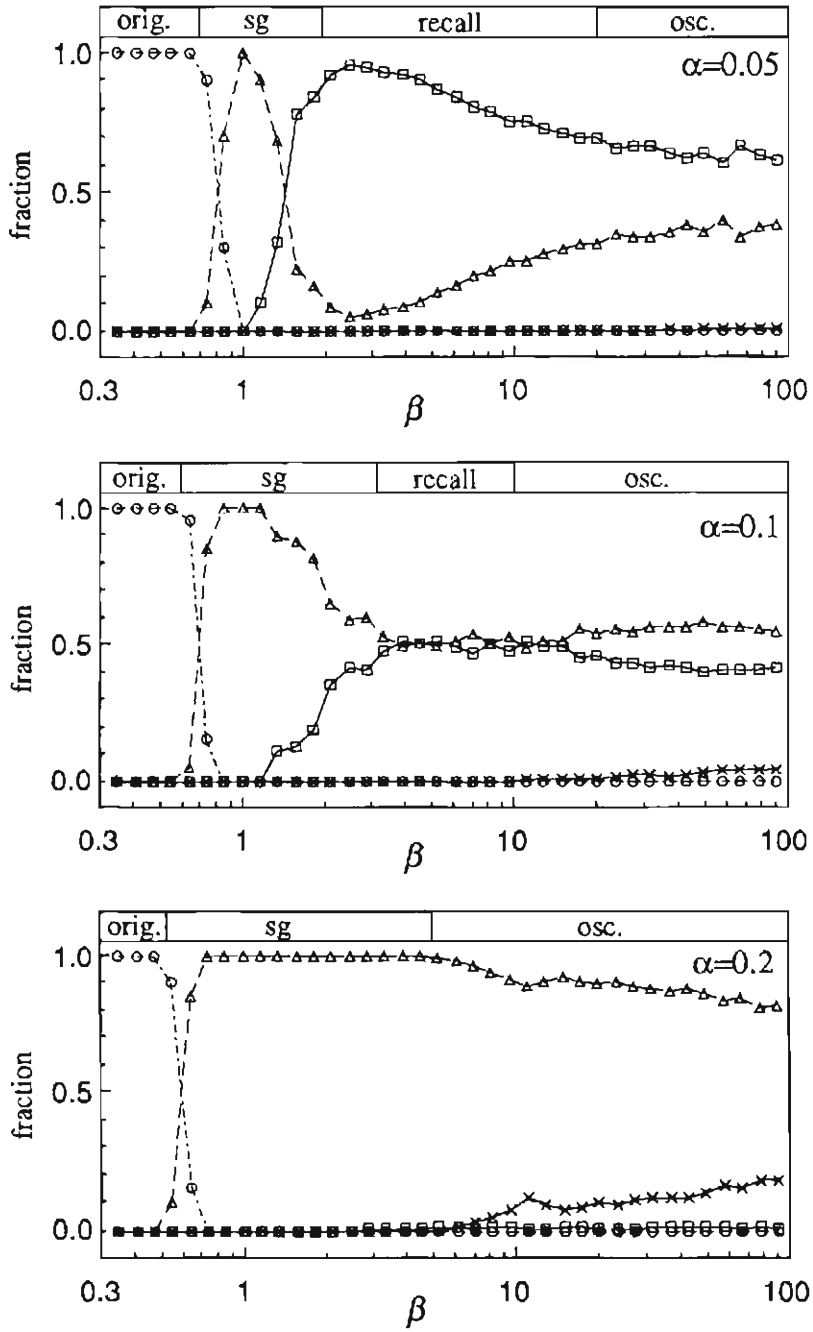


Fig. 5.6. Numerical data for the Hebb rule showing the fraction of random initial states which lead to the four types of attractors: the origin (circle), a memory pattern or its inverse (square), a spurious fixed point (triangle), or a period-2 limit cycle (cross), as a function of neuron gain β. Each data point represents a total of 1000 initial states from 20 matrices constructed from random, unbiased memory patterns with $N = 100$. The three panels are for $\alpha N = 5$, 10 and 20 patterns, and the strip along the top indicates the regions of the phase diagram, Fig. 5.3, for that value of α.

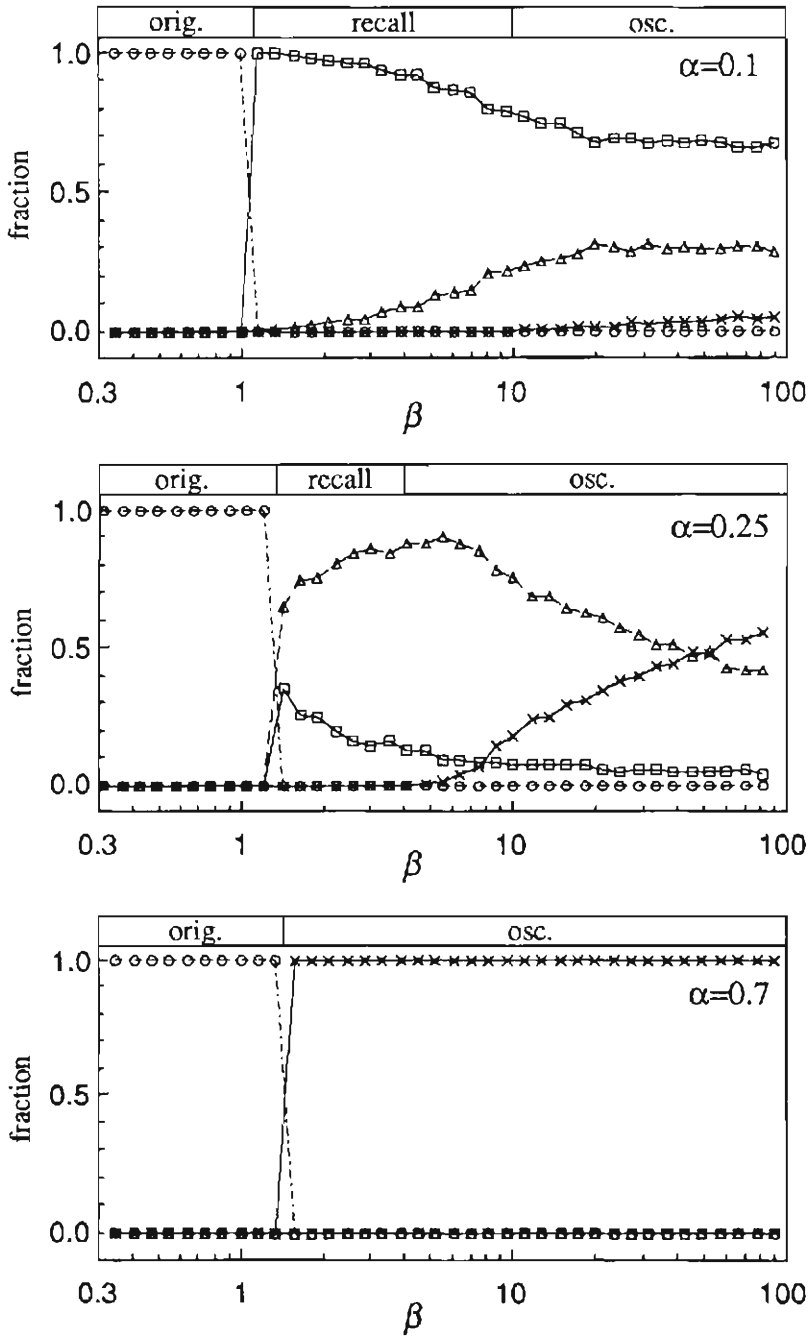


Fig. 5.7. Numerical data for the pseudo-inverse rule showing the fraction of random initial states which lead to the four types of attractors: the origin (circle), a memory pattern or its inverse (square), a spurious fixed point (triangle), or a period-2 limit cycle (cross), as a function of neuron gain β. Each data point represents a total of 1000 initial states from 20 matrices constructed from random, unbiased memory patterns with N = 100. The three panels are for αN = 10, 25 and 70 patterns, and the strip along the top indicates the regions of the phase diagram, Fig. 5.4, for that value of α.

5.5.2. Improved recall at low gain: deterministic annealing

Figures 5.6 and 5.7 show that the probability of recall is greater at lower values of analog gain within the 'recall' region. This phenomenon suggests a potentially powerful technique for annealing a deterministic analog neural network to a good (low energy) solution [Hopfield and Tank, 1985]. Annealing by varying the analog gain is not only useful as a fast numerical technique, but can be easily implemented in analog electronics, eliminating the need for electronic noise generators to perform stochastic annealing.

As with standard simulated annealing [Kirkpatrick *et al.*, 1983], convergence times at reduced gain can be quite long. To speed convergence, the gain should follow an annealing schedule, starting at the low-gain border of the 'recall' phase, and ending at the high gain border. The phase diagrams, Figs. 5.3 and 5.4, can be used to find the range of gains over which annealing should take place. Note that annealing range depends strongly on the storage ratio α. The surprising fact that the performance of an associative memory can be improved by using analog neurons will be considered in more detail in Ch. 7.

5.6. DISCUSSION

In this chapter we studied the dynamics and associative-recall properties of an analog network with parallel dynamics. We found that using analog neurons has two important benefits: First, analog networks can be updated in parallel with guaranteed convergence to a fixed point as long as the stability criterion (5.11) is satisfied. For the associative memories considered, the iterated-map network has rather large regions in the space of neuron gain β and storage ratio α where recall states exist and the stability criterion is satisfied. The second benefit, seen numerically in Figs. 5.6 and 5.7, is that using a

reduced neuron gain improved the chances that a random initial state would make it to a memory state without getting caught in a spurious attractor.

The usefulness of analog dynamics goes beyond the stability and improved recall properties studied here. By taking advantage of the generality of the stability results of § 5.3, one can design stable networks of neurons having *nonsigmoidal* transfer functions with computationally useful properties. As an example, the stability results apply to three-state (+1, 0 ,-1) neurons [Yedidia, 1989; Meunier, *et al.*, 1989] generalized to a smooth 'staircase' analog transfer function. Networks of three-state analog neurons bear a strong resemblance to the mean field spin-1 Ising model at finite temperature [Blume, 1966; Capel, 1966], with regions of parameter space where both the origin and recall states are locally stable. Such systems might be used to allow an 'I don't know' state of the network, such that initial states with insufficient overlap with any pattern will converge to the origin. In a numerical investigation[3] of networks made of three-state analog neurons, it was found that the attractors included not only the recall states and the origin, but also new mixture states in which a pattern was partially recalled, with some neurons converging to the zero-output state.

Another generalization of the iterated-map associative memory is the deliberate inclusion of limit cycles as recall states. Several techniques for storing and recalling limit cycles have been explored in continuous-time systems with delay [Grossberg, 1970; Kleinfeld, 1986; Sompolinsky and Kanter 1986; Gutfreund and Mezard, 1988; Riedel, *et al.*, 1988; Herz, *et al.*, 1988; Kühn, *et al.*, 1989] and in discrete-time systems, both for sequential [Buhmann and Schulten, 1987; Nishimori *et al.*, 1990] and parallel dynamics [Dehaene, *et al.*, 1987; Guyon, *et al.*, 1988]. Because these models use asymmetric connections, little is known analytically about their stability or the types of attractors they can produce. On the other hand, it is possible to store 2-cycle attractors in

[3] This work was done by Fred Waugh.

the iterated-map network using a symmetric connection matrix. This can be done most easily with a generalized Hebb rule in which a weighted Hebb matrix for the oscillatory directions $\vec{\zeta}^{\nu}$ is *subtracted* from a Hebb matrix for the fixed-point patterns $\vec{\xi}^{\mu}$:

$$T_{ij} = \frac{1}{N}\left[\sum_{\mu=1}^{p_{fp}} \xi_i^{\mu}\xi_j^{\mu} - \Lambda \sum_{\nu=1}^{p_{osc}} \zeta_i^{\nu}\zeta_j^{\nu}\right]. \tag{5.24}$$

The weighting factor Λ can be used to cause fixed point patterns and 2-cycle patterns to appear at different values of analog gain. A detailed analysis of such an analog network, yielding for example the combined storage capacity of limit cycles as well as fixed points, remains an open problem.

APPENDIX 5A: STORAGE CAPACITY FOR THE HEBB RULE

In this appendix we find the border separating the 'spin glass' region from the 'recall' region in the phase diagram for the Hebb rule, Fig. 5.3. The derivation is a slight generalization of a cavity method approach presented by Domany *et al.* [1989], but is somewhat simpler because of the absence of the reaction field [Marcus *et al.*,1990]. The form assumed for the nonlinear function F (taken to be identical for all i) is described at the beginning of § 5.4. We also assume all memory patterns, $\xi_i^{\mu} = \pm 1$, to be uncorrelated, and we set $I_i = 0$ for all i. For the special choice $F(z) = tanh(\beta z)$, the border we obtain is the same as that obtained for the Ising model network at temperature $1/\beta$ [Amit *et al.*, 1985b; Amit *et al.*, 1987; Mezard *et al.*, 1987; Domany *et al.*, 1989]. Throughout this appendix and appendix 5B, sums over roman indices ($i, j, k, ...$) run from 1 to N; sums over greek indices ($\mu, \nu, \rho, ...$) run from 1 to αN.

A recall state is characterized by the existence of a fixed point of the iterated map,

which satisfies

$$x_i = F\left(\sum_j T_{ij} x_j \right) \quad , \quad i = 1,...,N \tag{5A.1}$$

and which has a large (O(1)) overlap with a single memory pattern, where the overlaps m^μ are defined

$$m^\mu = \frac{1}{N} \sum_i \xi_i^\mu x_i \quad . \tag{5A.2}$$

For the Hebb matrix, Eq. (5.19), the input h_i to neuron i can be written in terms of the m^μ as

$$h_i = \sum_j T_{ij} x_j = \sum_\mu \xi_i^\mu m^\mu \tag{5A.3}$$

which gives a set of αN fixed-point equations for the overlaps

$$m^\mu = \frac{1}{N} \sum_i \xi_i^\mu F(h_i) \quad , \quad \mu = 1,...,\alpha N \, . \tag{5A.4}$$

For F odd and $\xi_i^\mu = \pm 1$, these equations can be written

$$m^\mu = \frac{1}{N} \sum_i F\left(\xi_i^\mu h_i\right) = \frac{1}{N} \sum_i F\left(H_i^\mu\right) = \left\langle F\left(H^\mu\right) \right\rangle \tag{5A.5}$$

where $H_i^\mu \equiv \xi_i^\mu h_i$. Borrowing spin glass terminology, H_i^μ will be referred to as a local field for memory μ. The brackets in (5A.5) denote an average over the index i:

$\langle z \rangle \equiv 1/N \sum_i z_i$. In the large-N limit, this average can be written as an integral over the distribution of local fields $P(H^\mu)$:

$$m^\mu = \int dH^\mu P\left(H^\mu\right) F\left(H^\mu\right). \tag{5A.6}$$

We now seek a self-consistent expression for the distribution function $P(H^1)$ when $m^1 \sim 1$ and $m^\mu \sim O(N^{-1/2})$ for $\mu>1$. The local field for pattern 1,

$$H_i^1 = \xi_i^1 \sum_\nu \xi_i^\nu m^\nu , \tag{5A.7}$$

can be split into two parts,

$$H_i^1 = m^1 + \xi_i^1 \sum_{\nu>1} \xi_i^\nu m^\nu . \tag{5A.8}$$

For $\alpha \sim O(1)$, the second term on the right side of (5A.8) acts as a noise term which we take to be gaussian distributed with zero mean and variance σ^2 given by

$$\sigma^2 = \sum_{\nu>1} \left(m^\nu\right)^2 . \tag{5A.9}$$

To evaluate the sum of squares in (5A.9) we first write the overlaps m^ν with the uncondensed patterns using (5A.3) and (5A.4):

$$m^\nu = \frac{1}{N} \sum_i \xi_i^\nu F\left(\sum_\rho \xi_i^\rho m^\rho \right) . \tag{5A.10}$$

Notice that the right side of (5A.10) is of the form $\sum_i A_i B_i$. A sum of this form with

uncorrelated random variables A_i and B_i has an expected square of $\sum_i A_i^2 B_i^2$. In (5A.10) however, the two factors in the sum over i are *correlated* through the $\rho = \nu$ term in the argument of F, and this term must be treated separately before squaring. Writing the correlated term separately,

$$m^\nu = \frac{1}{N}\sum_i \xi_i^\nu F\left(\sum_{\rho\neq\nu} \xi_i^\rho m^\rho + \xi_i^\nu m^\nu\right), \tag{5A.11}$$

and noting that the single term ($\rho = \nu$) is small compared to the sum over all the rest ($\rho \neq \nu$), we expand F to first order in m^ν giving

$$m^\nu = \frac{1}{N}\sum_i \xi_i^\nu \left[F\left(\sum_{\rho\neq\nu} \xi_i^\rho m^\rho\right) + \xi_i^\nu m^\nu F'\left(\sum_{\rho\neq\nu} \xi_i^\rho m^\rho\right)\right] \tag{5A.12}$$

where F' is the derivative of the function F. The missing $\rho = \nu$ term in the argument of F' only affects the value of F' to order O(1/N) which we neglect by taking the argument to be the whole h_i. We now define the quantity C,

$$C \equiv \langle F'(h)\rangle = \frac{1}{N}\sum_i F'(h_i) \tag{5A.13}$$

and write (5A.12) as

$$m^\nu(1-C) = \frac{1}{N}\sum_i \xi_i^\nu F\left(\sum_{\rho\neq\nu} \xi_i^\rho m^\rho\right) \tag{5A.14}$$

With the $\rho = \nu$ term removed from the argument of F, the two factors in the sum over i on the right side of (5A.14) are now uncorrelated and can be squared to yield an

expected value of

$$\left([1-C]m^{\nu}\right)^2 = \frac{1}{N}\sum_i F^2\left(\sum_{\rho\neq\nu}\xi_i^{\rho}m^{\rho}\right) \approx \frac{1}{N}\sum_i F^2(h_i) \tag{5A.15}$$

where, again, the O(1/N) error in the value of F^2 from the $\rho = \nu$ term is ignored. Next, we define the quantity q in analogy with the Edwards-Anderson order parameter,

$$q \equiv \left\langle F^2(h)\right\rangle = \frac{1}{N}\sum_i F^2(h_i), \tag{5A.16}$$

and write (5A.15) as

$$\left(m^{\nu}\right)^2 = q/N(1-C)^2 \ . \tag{5A.17}$$

From (5A.9) and (5A.17), the variance σ^2 of the local field distribution is given in terms of the quantities C and q by

$$\sigma^2 = \alpha q/(1-C)^2 . \tag{5A.18}$$

Because F' and $F2$ are both even functions, we can multiply their arguments by ±1 without changing their values. This allows us to write the averages in Eqs. (5A.13) and (5A.16) in terms of H_i^1 rather than h_i, and finally as integrals over the distribution of local fields $P(H^1)$, given by the normalized gaussian distribution

$$P\left(H^1\right) = \frac{1}{\sqrt{2\pi}\sigma}\, exp\left[\frac{-\left(H^1 - m^1\right)^2}{2\sigma^2}\right], \tag{5A.19}$$

where the variance σ^2 is given by (5A.18). Together with Eq. (5A.6), the self-consistent equations for quantities m^1, C and q are given by the following integrals:

$$m^1 = \int dH^1 P(H^1) F(H^1) \tag{5A.20a}$$

$$C = \int dH^1 P(H^1) F'(H^1) \tag{5A.20b}$$

$$q = \int dH^1 P(H^1) F^2(H^1) \tag{5A.20c}$$

After a change of variables, $y \equiv (H^1 - m^1)/\sigma$, Eqs. (5A.18) - (5A.20) yield the self-consistent set of equations (5.22a)-(5.22d) in § 5.4.1.

APPENDIX 5B: RECALL STATES FOR THE PSEUDO-INVERSE RULE

In this appendix we show that for the pseudo-inverse learning rule, stable recall states exist whenever $\alpha < 1$ and $\beta > 1/(1-\alpha)$ [Marcus *et al.*,1990]. This implies that there is no spin glass phase for the pseudo-inverse rule in the iterated-map network, in contrast to the thermodynamic Ising-spin network with the same learning rule [Kanter and Sompolinsky, 1987]. The analysis below closely follows Kanter and Sompolinsky [1987].

As described in appendix 5A, a recall state is defined as a fixed point which has a large overlap with a single pattern (again, taken to be pattern 1). For large N, the single large overlap m^1 can be written as an integral over the distribution of local fields

$$m^1 = \frac{1}{N}\sum_i F(H_i^1) \xrightarrow{N\to\infty} \int dH^1 P(H^1) F(H^1). \tag{5B.1}$$

where $P(H^1)$ is a gaussian distribution whose mean and variance must be found self-consistently. The local field for memory pattern 1,

$$H_i^1 = \xi_i^1 \sum_{j \neq i} T_{ij} x_j \tag{5B.2}$$

with the pseudo-inverse matrix

$$T_{ij} = \frac{1}{N} \sum_{\mu,\nu} \xi_i^\mu \left(C^{-1}\right)_{\mu\nu} \xi_j^\nu \tag{5B.3}$$

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu \tag{5B.4}$$

is given by

$$H_i^1 = \xi_i^1 \left\{ \sum_{\mu,\nu} \xi_i^\mu \left[\left(C^{-1}\right)_{\mu\nu} m^\nu - \alpha x_i \right] \right\}. \tag{5B.5}$$

The $-\alpha x_i$ term explicitly takes care of setting the diagonals to zero since the T_{ii} as defined by (5B.3) are narrowly peaked around α at large N. The state vector x_i, $i = 1,..., N$ can be written as a weighted sum of the pattern vectors, with real-valued weights a^μ, plus a vector χ_i, $i = 1,..., N$, which is perpendicular to the subspace spanned by the patterns

$$x_i = \sum_\mu a^\mu \xi_i^\mu + \chi_i \ . \tag{5B.6}$$

From (5A.2), (5B.4) and (5B.6), the weights a^μ are related to the overlaps m^μ through

the inverse correlation matrix

$$a^{\mu} = \sum_{\nu} \left(C^{-1}\right)_{\mu\nu} m^{\nu} \ . \tag{5B.7}$$

Writing the local field H_i^1 in terms of the a^{μ},

$$H_i^1 = (1-\alpha)a^1 + \xi_i^1(1-\alpha)\left[\sum_{\mu>1} \xi_i^{\mu} a^{\mu}\right] - \alpha\chi_i \ , \tag{5B.8}$$

reveals a similar structure to the Hebb rule (compare (5B.8) to (5A.8)), with a 'signal' term proportional to a^1 and a 'noise' term due to the other patterns. The third term on the right causes the state to relax towards the subspace spanned by the memories, and does not add any destabilizing 'noise.' Comparing Eqs. (5B.8) and (5A.8) also reveals why the pseudo-inverse rule allows perfect recall with an extensive number of patterns and the Hebb rule does not: for the pseudo-inverse rule, the variance of the gaussian noise due to the other patterns is given by

$$\sigma_{PI}^2 = (1-\alpha)^2 \sum_{\mu>1} \left(a^{\mu}\right)^2 \tag{5B.9}$$

whereas for the Hebb rule, the variance is

$$\sigma_H^2 = \sum_{\mu>1} \left(m^{\mu}\right)^2 \ . \tag{5B.10}$$

When the state of the network is fully aligned with, say, pattern 1, then all a^{μ}, $\mu > 1$ vanish. On the other hand, the overlaps m^{μ}, $\mu > 1$ do not vanish, even when the state is perfectly aligned with a pattern, unless all memories are orthogonal. Therefore the

'noise' term for the Hebb rule is in general always non-zero.

In a recall state (for pattern 1), $a^1 = m^1$ and $a^\mu = 0$ for $\mu > 1$, giving a delta function distribution for the local fields

$$P\left(H^1\right) = \delta\left(H^1 - (1-\alpha)m^1\right). \tag{5B.11}$$

Inserting this distribution into (5B.1) gives the self-consistent solution for the overlap with pattern 1,

$$m^1 = F\left((1-\alpha)m^1\right). \tag{5B.12}$$

When the function F is tanh-like with maximum slope β, there is a non-zero m^1 given by (5B.12) whenever $\alpha < 1$ and $\beta > 1/(1-\alpha)$. The value of m^1 grows continuously from zero at the transition. In analogy with thermodynamics, the appearance of recall states is therefore a second order transition. As mentioned above, the behavior of the analog network with the pseudo-inverse rule for the particular choice $F(z) = tanh(\beta z)$ is *not the same* as the corresponding Ising-spin network at finite temperature $1/\beta$: as shown by Kanter and Sompolinsky [1987], the recall states for the Ising model appear at a value of β significantly above $1/(1-\alpha)$ and the transition to the recall state is first order. These differences can be attributed to the absence of a reaction field in our system.

Chapter 6

THE ANALOG MULTISTEP NETWORK

6.1. INTRODUCTION

In Ch. 5, we showed that analog neural networks offer several computational advantages over networks of binary neurons, including the property that convergence to a fixed point under parallel dynamics can be assured by a global stability criterion for networks with symmetric connections. The purpose of this chapter is to extend these stability results to networks with an updating rule based on multiple previous time steps, and apply the new stability criterion to the problem of associative memory.

The significant result which emerges from this analysis is that the criterion for assuring convergence to a fixed point allows a larger neuron gain in proportion to the number of time steps used in the updating rule, while other properties, including the storage capacity, are independent of the number of steps used. We emphasize that even when many previous states are used in the updating rule, parallel is still parallel: Thinking in terms of electronic hardware, one can imagine using a tapped analog delay line[1] at the input of each neuron; upon each time step, the local states at each neuron are *simultaneously* advanced one position in the delay line. In this scheme, previous time steps are local and only need to be evaluated once.

A further motivation for studying multiple-time-step networks is to provide a first

[1]Delay-line devices are described in Mead [1989]. Commercial tapped delay lines are available, for example, from EG&G Reticon Corp. (see: EG&G Reticon Application Note 105, *A Tapped Analog Delay for Sampled Data Signal Processing.*)

step in the application of global stability analysis to networks that make explicit use of the time domain as part of the computation, including recently proposed models for storing and generating sequences of patterns [Grossberg, 1970; Kleinfeld, 1986; Somplinsky and Kanter 1986; Dehaene, *et al.*, 1987; Gutfreund and Mezard, 1988; Riedel, *et al.*, 1988; Herz, *et al.*, 1988; Guyon, *et al.*, 1988; Kühn, *et al.*, 1989]. Typically, sequence-generating networks sample multiple previous states - for example by using time delay - to determine their evolution. Numerical studies [Riedel, *et al.*, 1988; Babcock and Westervelt, 1986; Aihara *et al.*, 1990] well as experiments using analog circuits [Marcus and Westervelt, 1988] (see § 4.6) show that the neural networks with time delay can be chaotic, and very few analytical results on their stability and convergence are known [Marcus and Westervelt, 1989a] (see § 4.3 and § 4.4). The analysis presented here is greatly facilitated by considering a relatively simple discrete-time multistep system with symmetric connections. In this sense, our results apply to sequence-generating networks when they are configured to retrieve fixed points only.

The dynamical system we will study is defined by a set of N coupled iterated maps

$$x_i(t+1) = F_i\left[\sum_{j=1}^{N} T_{ij}\, z_j(t) + I_i\right], \qquad i = 1,\ldots,N \tag{6.1a}$$

where $z_j(t)$ is the output of the j^{th} neuron time-averaged over M previous time steps:

$$z_j(t) \;=\; \frac{1}{M}\left\{\sum_{\tau=0}^{M-1} x_j(t-\tau)\right\}, \qquad j = 1,\ldots,N \;;\; M \in \{1,2,3,\ldots\} \tag{6.1b}$$

This system will be referred to as a *multistep neural network*. Updating of the state variables $x_i(t)$ as well as the $z_i(t)$ is done in *parallel* (i.e. synchronously) and is *fully deterministic*. We assume throughout that the connection matrix T_{ij} is *symmetric*. The

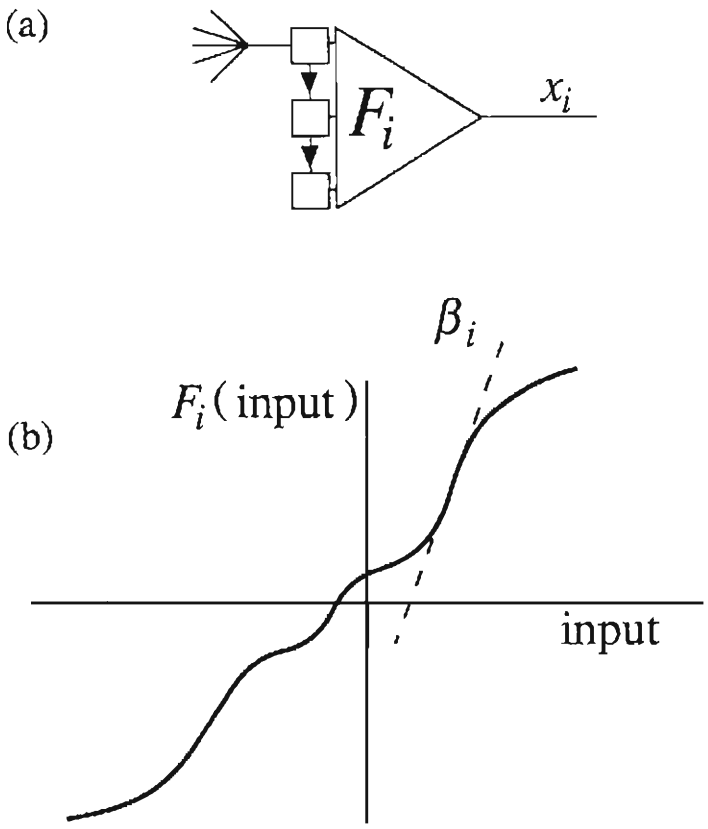


Fig. 6.1. The analog multistep neuron. (a) Schematic representation of a multistep neuron with $M = 3$. Electronic implementation could use a tapped delay line as part of the input (or output) circuitry. (b) An example of a neuron transfer function F_i (solid line) that satisfies the conditions required for the analysis: F_i is monotonic, single-valued, and increases in magnitude slower than linear at large argument. The maximum slope of F_i (dashed line) is defined as the neuron gain β_i.

nonlinear neuron transfer functions F_i must obey the same constraints as in Ch. 5: All F_i are *monotonic* (Without loss of generality, we can choose all F_i to be monotonically increasing) and *single-valued.* Also, the F_i may be locally concave up or concave down, and do not need to saturate, but must eventually increase in magnitude slower than linear for large positive and negative argument. An example of a function F_i that satisfies these conditions is shown in Fig. 6.1(b). The maximum slope of each F_i is defined as the *gain* β_i for that neuron. Results of § 6.2 and § 6.3 can be applied to networks of binary (Ising) neurons by letting all $\beta_i \to \infty$.

Equation (6.1) with $M = 1$ is the standard analog iterated-map neural network discussed in Ch. 5 [Marcus and Westervelt, 1989c]. The $M = 2$ case of (6.1) with all $F_i = Sgn$ was investigated numerically by Kanter and Sompolinsky [1987] for an associative memory neural network based on the pseudo-inverse learning rule. These authors found that the basins of attraction for the recall states are considerably larger for $M = 2$ than for $M = 1$ (the basins for recall states under sequential updating are larger than for either of these parallel schemes). Kanter and Sompolinsky [1987] offered the following explanation for this observation: First, the use of two previous states in the updating rule adds an effective momentum to the dynamics, allowing the network to "coast" over shallow local minima; Second, spurious *oscillatory* attractors for this network are 3-cycles, which, they argued, are rarer than the 2-cycles found in the $M = 1$ network. Below, we will prove (for general F_i) the claim of Kanter and Sompolinsky [1987] that the only attractors for $M = 2$ besides fixed points are period-3 limit cycles. On the other hand, numerical evidence for the $M = 2$ Ising spin glass presented in Ch. 7 suggests that spurious 3–cycles may be quite abundant for multistep networks.

The rest of the chapter is organized as follows. In § 6.2 we derive a global stability criterion which guarantees that the multistep analog network will always converge to a

fixed-point attractor as long as the maximum neuron gain $\beta \equiv max_i(\beta_i)$ does not exceed a critical value which is proportional to the number of time steps M in the update rule and inversely proportional to the minimum eigenvalue of the connection matrix. For the particular case $M = 2$ we also prove that the only other possible attractors (i.e. when the stability criterion is violated) besides fixed points are period-3 limit cycles. In § 6.3, we apply these results to multistep associative memory networks and give a simple stability criterion for the Hebb rule [Hebb, 1949, Hopfield, 1982] and pseudo-inverse rule [Personnaz, *et al.*, 1985; Kanter and Sompolinsky, 1987]. This criterion depends on M, β, the ratio α of stored memories to neurons and the self-coupling γ. In § 6.4, we show that the convergence time of the multistep network increases proportional to M, but that in some instances, optimal choices for both β and M give faster convergence with increasing M. Finally, conclusions and open problems are discussed in § 6.5.

6.2. LIAPUNOV FUNCTIONS FOR MULTISTEP ANALOG NETWORKS

6.2.1. Global stability criterion for general M

In this section we prove that the analog multistep network, Eq. (6.1), will have only fixed point attractors whenever

$$\frac{1}{\beta} > -\frac{\lambda_{min}(T_{ij})}{M} \tag{6.2}$$

where $\beta \equiv max_i(\beta_i) > 0$ is the maximum neuron gain, $M \in \{1,2,3,...\}$ is the number of time steps in the update rule and $\lambda_{min}(T_{ij})$ is the minimum eigenvalue of the symmetric connection matrix T_{ij}. This criterion applies for any distribution of the (real) eigenvalues of T_{ij}. In particular, when T_{ij} has both negative and positive eigenvalues, $\lambda_{min}(T_{ij})$

refers to the most negative eigenvalue. This result should be compared to the corresponding result for the iterated-map network (the case $M = 1$) presented in Ch. 5, which is the stability criterion (5.11).

Following a similar approach to Ch. 5, we consider the discrete-time evolution of the real scalar function $L(t)$ defined

$$L(t) \;=\; -\frac{1}{2}\sum_{i,j} T_{ij}\;\; z_i(t)\;\; z_j(t) \;+\; \sum_i \frac{1}{M}\sum_{\tau=0}^{M-1}\Big[G_i\big(x_i(t-\tau)\big) - I_i x_i(t-\tau)\Big] \qquad (6.3)$$

where

$$G_i(x_i) \;\equiv\; \int_0^{x_i} F^{-1}(z)dz \;\; . \qquad (6.4)$$

The requirement that F_i change in magnitude slower than linear at large argument insures that the function $L(t)$ is bounded below.

The change in $L(t)$ in one time step, defined as $\Delta L(t) \equiv L(t+1) - L(t)$, is

$$\begin{aligned}\Delta L(t) = \quad & -\frac{1}{2}\sum_{i,j} T_{ij}\Big[z_i(t+1)z_j(t+1) - z_i(t)z_j(t)\Big] \\ & +\sum_i \frac{1}{M}\Big[G_i\big(x_i(t+1)\big) - G_i\big(x_i(t-M+1)\big)\Big] \\ & -\sum_i \frac{1}{M} I_i\Big[x_i(t+1) - x_i(t-M+1)\Big] \;. \end{aligned} \qquad (6.5)$$

This can be simplified by defining the change in $z_i(t)$:

$$\Delta z_i(t) \equiv \big[z_i(t+1) - z_i(t)\big] = \frac{1}{M}\big[x_i(t+1) - x_i(t-M+1)\big] \qquad (6.6)$$

and using the symmetry of T_{ij}, giving

$$\Delta L(t) = -\frac{1}{2}\sum_{i,j} T_{ij}\,\Delta z_i(t)\,\Delta z_j(t) \;-\sum_i \Delta z_i(t)\left[\sum_j J_{ij} z_j(t) + I_i\right] + \sum_i \frac{1}{M}\left[G_i\left(x_i(t+1)\right) - G_i\left(x_i(t-M+1)\right)\right] . \tag{6.7}$$

Expanding the last term in (6.7) in a two-term Taylor series about the point $x_i(t+1)$ and replacing the coefficient of the quadratic term with the smallest value that it can take, which is

$$\min_{x_i}\left(d^2 G_i / dx_i^2\right) = \beta_i^{-1} \quad , \tag{6.8}$$

gives the following inequality [see Fig. 6.2]:

$$G_i\left(x_i(t+1)\right) - G_i\left(x_i(t-M+1)\right) \;\le\; G_i'\left(x_i(t+1)\right)\left[M\,\Delta z_i(t)\right] - \frac{1}{2}\beta_i^{-1}\left[M\,\Delta z_i(t)\right]^2 . \tag{6.9}$$

where G_i' is the derivative of G_i with respect to x_i. From Eqs. (6.1a) and (6.4), G_i' can be written

$$G_i'\left(x_i(t+1)\right) = F_i^{-1}\left(x_i(t+1)\right) = \sum_j T_{ij} z_j(t) + I_i \quad , \tag{6.10}$$

leading to an inequality for $\Delta L(t)$:

$$\Delta L(t) \;\le\; -\frac{1}{2}\sum_{i,j} T_{ij}\,\Delta z_i(t)\,\Delta z_j(t) \;-\frac{1}{2}\sum_i M\beta_i^{-1}\left[\Delta z_i(t)\right]^2 \quad . \tag{6.11}$$

By defining a matrix K_{ij} as

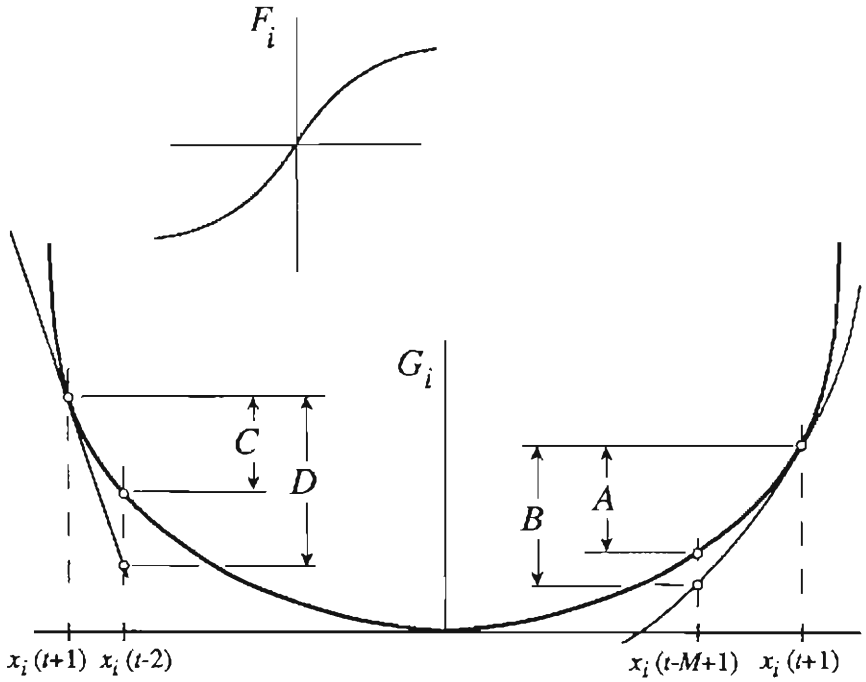


Fig. 6.2. Inequalities (6.9) and (6.20), illustrated for a particular transfer function F_i (inset) and its corresponding G_i, defined by Eq. (6.4). The curve tangent to G_i on the right side is a parabola with second derivative β_i^{-1}, the line on the left is tangent to G_i. Equation (6.9) is the statement $A \leq B$. ; Eq. (6.20) is the statement $C \leq D$, and $C = D$ only when $x_i(t+1) = x_i(t-2)$.

$$K_{ij} \equiv \frac{1}{2}\left[T_{ij} + \delta_{ij} M \beta_i^{-1}\right] = \frac{1}{2}\begin{pmatrix} T_{11} + M\beta_1^{-1} & T_{12} & \cdots & T_{1N} \\ T_{21} & T_{22} + M\beta_2^{-1} & & \\ \vdots & & \ddots & \\ T_{N1} & & & T_{NN} + M\beta_N^{-1} \end{pmatrix}, \quad (6.12)$$

the inequality (6.11) can be rewritten in the simple form

$$\Delta L(t) \;\leq\; -\sum_{i,j} K_{ij}\, \Delta z_i(t)\, \Delta z_j(t)\,. \quad (6.13)$$

From (6.13) it is clear that if K_{ij} is a positive definite matrix, then $\Delta L(t) \leq 0$, and that the case of equality ($\Delta L(t) = 0$) only can occur when all $\Delta z_i(t) = 0$. Writing the difference of Eq. (6.1) at subsequent times as

$$\sum_j T_{ij}\, \Delta z_j(t) \;=\; F_i^{-1}\big(x_i(t+1)\big) - F_i^{-1}\big(x_i(t)\big) \quad (6.14)$$

indicates that the condition $\Delta z_i(t) = 0$ for all i (and therefore the condition $\Delta L(t) = 0$) further implies that $x_i(t+1) = x_i(t)$ for all i; that is, that the system must be at a fixed point.

A sufficient condition for K_{ij} to be positive definite is $M\beta_i^{-1} > -\lambda_{min}(T_{ij})$ for all i, where $\lambda_{min}(T_{ij})$ is the minimum eigenvalue of the matrix T_{ij}. (This holds for any value for λ_{min}.) In terms of the maximum neuron gain $\beta \equiv max_i(\beta_i)$, this sufficient condition can be stated

$$\frac{1}{\beta} > -\frac{\lambda_{min}\left(T_{ij}\right)}{M} \quad \Rightarrow \quad K_{ij} \text{ positive definite}\,. \quad (6.15)$$

Equations (6.13) - (6.15) and the arguments in the preceding paragraphs can be summarized as follows:

Stability Criterion:

$$\frac{1}{\beta} > -\frac{\lambda_{min}(T_{ij})}{M} \quad \Rightarrow \quad \begin{array}{l} (1)\ L(t) \text{ is a Liapunov function of system (6.1),} \\ (2)\ \text{All attractors of (6.1) are fixed points.} \end{array} \qquad (6.16)$$

As mentioned above, this criterion applies for any distribution of the (real) eigenvalues of T_{ij}. When T_{ij} has both negative and positive eigenvalues, $\lambda_{min}(T_{ij})$ refers to the most negative eigenvalue. The stability criterion (6.16) is immediately satisfied if the connection matrix T_{ij} is itself positive definite (since M and β are both strictly positive), although this is usually not the case in most currently-used network applications (see § 6.3).

What can be said about the attractors of (6.1) when the stability criterion (6.16) is *not* satisfied? For $M = 1$, we showed in § 5.2 that the only new kind of attractor that can appear when (6.16) is violated is a limit cycle of period 2. For $M = 2$, we will prove in the following subsection that the only possible attractors besides fixed points are period-3 limit cycles. For $M > 2$, we know of no results - analytical or numerical - that limit the possible types of attractors of (6.1) when (6.16) is not satisfied. However, the trend for $M = 1, 2$ suggests that for general M, attractors of (6.1) (with symmetric T_{ij}) might be restricted to limit cycles with periods of $(M+1)$ and its divisors.

6.2.2. The case $M = 2$: Only fixed points and 3-cycles

The simplest multistep extension of the standard parallel update rule is the $M = 2$ case of (6.1). An Ising-model version of this network was studied by Kanter and Sompolinsky [1987] for the pseudo-inverse rule associative memory. These authors found numerically that the $M = 2$ network has improved recall over the corresponding standard ($M = 1$) Ising-model network. As part of their explanation for the improved performance, they also point out that the only non-fixed-point attractors for their network were period-3 limit cycles.

In this subsection we prove a generalization of the statement of Kanter and Sompolinsky [1987]: We show that all attractors of the multistep network (6.1) with $M = 2$ are either fixed points or period-3 limit cycles for general nonlinearities F_i as defined in § 6.1. Again, this result assumes a symmetric T_{ij}.

Consider the time evolution of the function $E(t)$, defined

$$\begin{aligned} E(t) &= -\sum_{i,j} T_{ij}\left[z_i(t)z_j(t-1)+\frac{1}{4}x_i(t-1)x_j(t-1)\right] \\ &\quad + \frac{1}{2}\sum_i \sum_{\tau=0}^{2}\left[G_i\big(x_i(t-\tau)\big)-I_i x_i(t-\tau)\right] \end{aligned} \tag{6.17}$$

where the function G_i is given by Eq. (6.4). As with $L(t)$, the requirement that all F_i change in magnitude slower than linear at large argument insures that the function $E(t)$ is bounded below. The exact form of Eq. (6.17) was not derived, but was found via guesswork and some intuition, though it bears an obvious resemblance to $L(t)$ as well as to other previously discovered Liapunov functions [Cohen and Grossberg, 1983; Hopfield, 1984; Goles-Chacc *et al.*, 1985; Golden, 1986]. The change in $E(t)$ in one time step, defined $\Delta E(t) \equiv E(t+1) - E(t)$, is

$$\Delta E(t) = -\sum_{i,j} T_{ij}\left[z_i(t+1)z_j(t) - z_i(t)z_j(t-1)\right] + \frac{1}{4}\sum_{i,j} T_{ij}\left[x_i(t)x_j(t) - x_i(t-1)x_j(t-1)\right] + \frac{1}{2}\sum_i \left[G_i(x_i(t+1)) - G_i(x_i(t-2)) - I_i x_i(t+1) + I_i x_i(t-2)\right]. \quad (6.18)$$

Expressing the first term on the right of Eq. (6.18) in terms of x_i's and x_j's (from Eq. (6.1b)) and using the symmetry of T_{ij}, the change $\Delta E(t)$ can be written as

$$\Delta E(t) = -\frac{1}{2}\sum_i \left[x_i(t+1) - x_i(t-2)\right]\left(\sum_j T_{ij} z_j(t) + I_i\right) + \frac{1}{2}\sum_i \left[G_i(x_i(t+1)) - G_i(x_i(t-2))\right]. \quad (6.19)$$

We now expand G_i to first order about the point $x_i(t+1)$ and use the following inequality [see Fig. 6.2]:

$$\left[G_i(x_i(t+1)) - G_i(x_i(t-2))\right] \le G_i'(x_i(t+1))\left[x_i(t+1) - x_i(t-2)\right]. \quad (6.20)$$

Equation (6.20) differs from the second-order expansion of G_i used above (Eq. (6.9)) in that the only case of equality for Eq. (6.20) is when $x_i(t+1) = x_i(t-2)$. Combining Eqs. (6.10), (6.19) and (6.20) yields

$$\Delta E(t) \le -\frac{1}{2}\sum_i \left[F_i^{-1}(x_i(t+1)) - G_i'(x_i(t+1))\right]\left\{x_i(t+1) - (x_i(t-2))\right\}. \quad (6.21)$$

From Eq. (6.10), the difference in square brackets equals zero, and therefore

$$\Delta E(t) \le 0; \tag{6.22a}$$

$$\Delta E(t) = 0 \quad \Rightarrow \quad x_i(t+1) - x_i(t-2) = 0 \text{ for all } i. \tag{6.22b}$$

Thus, $E(t)$ is a Liapunov function for the system (6.1) with $M = 2$, and all attractors (the minima of $E(t)$) satisfy the condition $x_i(t+1) = x_i(t-2)$ for all i. That is, all attractors are either period-1 (fixed points) or period-3 limit cycles.

6.3 MULTISTEP ASSOCIATIVE MEMORIES

To illustrate some of the benefits of using a multistep update rule, we now consider associative memory networks based on the two learning algorithms studied in Ch. 5: the Hebb rule and the pseudo-inverse rule. The connection matrices T_{ij} to be inserted into the multistep network (6.1) in order to store a set of p fixed-point memory patterns ξ_i^μ ($\xi_i^\mu = \pm 1$; $i = 1, ..., N$; $\mu = 1, ..., p$) are specified by the following rules:

Hebb Rule:

$$T_{ij} = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{\mu=1}^{p} \xi_i^\mu \, \xi_j^\mu & (i \neq j) \\ \gamma & (i = j) \end{cases} \tag{6.23}$$

Pseudo-inverse Rule:

$$T_{ij} = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{\mu,\nu=1}^{p} \xi_i^\mu \left(C^{-1}\right)_{\mu\nu} \xi_j^\nu & (i \neq j) \\ \gamma & (i = j) \end{cases} \tag{6.24a}$$

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \, \xi_i^\nu \quad . \tag{6.24b}$$

We take the self-coupling term γ to have an adjustable value in both learning rules. The influence of self-coupling γ on the recall performance [Kanter and Somplinsky, 1987; Krauth *et al.*, 1988; Fontanari and Köberle, 1988a,b,c] and stability [Jeffery and Rosner, 1986a; Denker, 1986c; Fontanari and Köberle, 1988a; Marcus and Westervelt, 1990] on various associative memory models has been discussed previously.

Mean field analysis [Amit *et al.*, 1985b, 1987; Marcus and Westervelt, 1990] in the large-N limit suggests that overloading these two associative memories results in the disappearance - not merely the destabilizing - of fixed points close to the stored patterns. That is, storage capacities for these networks seem to depend only on the presence or absence of fixed points, not on the details of the dynamics. Because the fixed point condition for (6.1) is independent of M, *storage capacities for these associative memories are also independent of* M.

To apply the stability criterion (6.16) to these associative memories, we need the minimum eigenvalues of the connection matrices defined by (6.23) and (6.24). For the Hebb rule,

$$\lambda_{min}\left(T_{ij}^{Hebb}\right) = \gamma - \alpha \qquad [\alpha < 1] \qquad (6.25);$$

for all values of the network size N [Crisanti and Sompolinsky, 1987].[2] For the pseudo-inverse rule, the minimum eigenvalue asymptotically equals the Hebb-rule value,

$$\lambda_{min}\left(T_{ij}^{p-i}\right) \rightarrow \gamma - \alpha \qquad [\alpha < 1 \,;\, N >> 1] \qquad (6.26)$$

and is slightly [$O(N^{-1/2})$] more negative for finite N [Kanter and Sompolinsky, 1987].

[2]Equation (6.25) can be proved by first showing that the outer product matrix $\xi\,\xi^T$ is non-negative definite and has rank less than or equal to the number of patterns p. Thus for $p < N$, $\lambda_{min}(\xi\xi^T) = 0$. Then, because the diagonal of $(1/N)\xi\,\xi^T$ is α, setting the diagonal in (6.23) to γ immediately gives (6.25).

We assume $N >> 1$ and take $\lambda_{min} = \gamma - \alpha$ for both rules. We emphasize: this value of λ_{min} is exact for the Hebb rule and is valid as $N \to \infty$ for the pseudo-inverse rule. For both rules, it is valid for $0 < \alpha < 1$. We have not placed any restrictions on biases or correlations among memories, although (6.25) and (6.26) are *not* valid if the connection strengths are clipped or diluted [Marcus and Westervelt, 1989a].

From Eqs. (6.16), (6.25) and (6.26) we can immediately write the main result of this section: An M-step analog associative memory based on the Hebb or pseudo-inverse rule with self-coupling γ and all neuron gains less than or equal to β will always converge to a fixed point attractor when the condition

$$\frac{1}{\beta} > \left(\frac{\alpha - \gamma}{M}\right) \tag{6.27}$$

is satisfied.

This is a remarkably simple result. To illustrate its usefulness, consider a pseudo-inverse rule network loaded to $\alpha = 0.8$ and no self-coupling, $\gamma = 0$. From (6.27), the $M = 1$ network can oscillate whenever $\beta > 1.25$. On the other hand, it can also be shown using a multistep generalization of the contraction mapping theorem [Weinitschke, 1964; Baudet, 1978] that the pseudo-inverse network has a single, global attractor whenever the maximum gain satisfies

$$\beta < (1 - \alpha + \gamma)^{-1} \quad , \tag{6.28}$$

independent of M. In this low-gain state, all initial states evolve to the same fixed point and the network is not useful as an associative memory. For the present example, the condition (6.28) requires $\beta > 5$ to create recall states, which is unfortunately at odds with the stability condition $\beta < 1.25$. To simultaneously satisfy both requirements, one

could add positive self-coupling $\gamma > 0$, but this has the detrimental side effect of reducing the size of the basins of attraction for recall states [Kanter and Sompolinsky, 1987; Krauth *et al.*, 1988]. An alternative is to use a multistep updating rule: From (6.27) and (6.28), the two desired conditions - guaranteed convergence to a fixed point and existence of recall states - can be simultaneously satisfied in the pseudo-inverse rule when

$$\alpha < \left[\left(\frac{M}{M+1}\right) + \gamma\right] \tag{6.29}$$

Thus, in our example, the $M = 1$ network must have $\gamma > 0.3$ to simultaneously provide guaranteed convergence and the existence of recall states; however, a multistep network with $M > 3$ can satisfy both conditions without the use of positive self-coupling.

6.4 CONVERGENCE TIME

In this section we present a simple analysis of the convergence time for the multistep network, and show that the convergence time τ_M for the M-step network is greater than for the 1-step network by a factor τ_M/τ_1 where

$$(M+1)/2 < \tau_M/\tau_1 < M. \tag{6.30}$$

In the context of discrete-time dynamics, the expression "time" means "number of iterations," and is not equivalent to the real time taken to perform the updating, which depends on the details of the implementation. For example, in a multiprocessor implementation of the multistep network, each processor (one for each neuron) must read and sum the N states of the other neurons in order to determine its local field. Thus the

update time in this implementation is roughly N times the processor's cycle time. Note that the real time for an update does not scale with M, however, since local fields can be stored in an array and used M times. By this same argument, the real time taken to update all neurons *sequentially* is proportional to N^2.

Convergence times for associative memories have been studied previously for binary neurons and discrete-time parallel dynamics with $M = 1$ [Kanter, 1989] as well as continuous-time dynamics with time delay [Kerszberg and Zippelius, 1989]. An important result reported by Kanter [1989] is that the convergence time for binary networks under parallel dynamics increases in proportion to the logarithm of the network size for the Hebb rule, but appears to reach a size-independent limit for the pseudo-inverse rule.

To analyze the convergence time in the analog multistep network, we consider evolution in the vicinity of any attracting fixed point, which may be a memory recall state, a spurious fixed point, or, for very low gain ($\beta|\lambda(T_{ij})| < 1$ for all λ, independent of M), the single, globally attracting fixed point. Close to the fixed point $\vec{\mathbf{x}}^*$, the time evolution of the deviation $\vec{\delta}(t) \equiv (\vec{\mathbf{x}}(t) - \vec{\mathbf{x}}^*) << 1$ can be described by a linearized version of (6.1):

$$\delta_i(t+1) = \sum_{j=1}^{N} D_{ij}\left(\frac{1}{M}\sum_{\tau=0}^{M-1}\delta_j(t-\tau)\right) \tag{6.31a}$$

where D_{ij} is the Jacobian matrix,

$$D_{ij} = \left[\frac{d}{dx_j}\left(F_i\left(\sum_j T_{ij}x_j\right)\right)\right]_{\vec{\mathbf{x}}^*} \tag{6.31b}$$

In general, D_{ij} is not symmetric, but it can be shown that all of its eigenvalues are real,

due to the monotonicity of the F_i. We now further assume that the system has had sufficient time for fast modes to relax, allowing the evolution to be characterized by a single characteristic multiplier - that is, the system has reached the slow manifold. In this case, the approach to the fixed point can be described by the *linear*, *scalar* iterated map

$$\delta(t+1) = \Lambda_M \, \delta(t) \,, \tag{6.32}$$

where the characteristic multiplier Λ_M is real-valued and $|\Lambda_M| < 1$. We emphasize that even though (6.32) is a single step equation, this form applies for any value of M.

For $M = 1$, (6.31) is a simple N- dimensional linear map, and the value of Λ_1 is just the eigenvalue of D_{ij} in the slow manifold. For $M > 1$, the equation for $\delta(t)$ near $\vec{x}^*$ can be written in terms of Λ_1 as

$$\delta(t+1) = \frac{\Lambda_1}{M}\left[\delta(t) + \delta(t-1) + \cdots + \delta(t-M+1)\right] \tag{6.33}$$

By repeated application of Eq. (6.32), the multistep equation (6.33) can be cast in the form of the single step equation

$$\delta(t+1) = \left(\frac{\Lambda_1}{M}\left[1 + {\Lambda_M}^{-1} + \cdots + {\Lambda_M}^{-(M-1)}\right]\right)\delta(t) \quad . \tag{6.34}$$

For consistency with the definition of Λ_M in (6.32), we require

$$\Lambda_M = \frac{\Lambda_1}{M}\left[1 + {\Lambda_M}^{-1} + \cdots + {\Lambda_M}^{-(M-1)}\right] . \tag{6.35}$$

Summing the partial series in (6.35) gives a self-consistent expression relating the

characteristic multipliers Λ_1 and Λ_M:

$$\Lambda_1 = \frac{M(\Lambda_M)^M (1-\Lambda_M)}{1-(\Lambda_M)^M} . \tag{6.36}$$

We now define a characteristic time $\tau_M > 0$ as the number of time steps needed to reduce an initial distance $\delta_o << 1$ by a factor of $1/e$. From (6.32), the distance from the fixed point decreases according to $\delta(t) = \delta_o [\Lambda_M]^t$. This yields an equation for converting characteristic multipliers into characteristic times: $\tau_M = -\left[ln|\Lambda_M|\right]^{-1}$. The ratio τ_M/τ_1, which indicates how much slower the multistep network is compared to the single-step network, is therefore given by

$$\tau_M/\tau_1 = ln|\Lambda_1| / ln|\Lambda_M| . \tag{6.37}$$

Values for τ_M/τ_1 as a function of τ_1 for M = 1 through 4 are plotted in Fig. 6.3. For large and small values of τ_M, the limiting values of the ratio τ_M/τ_1 are

$$\lim_{\tau_M \to 0} \left[\tau_M/\tau_1\right] = M \tag{6.38a}$$

$$\lim_{\tau_M \to \infty} \left[\tau_M/\tau_1\right] = \frac{M+1}{2} \tag{6.38b}$$

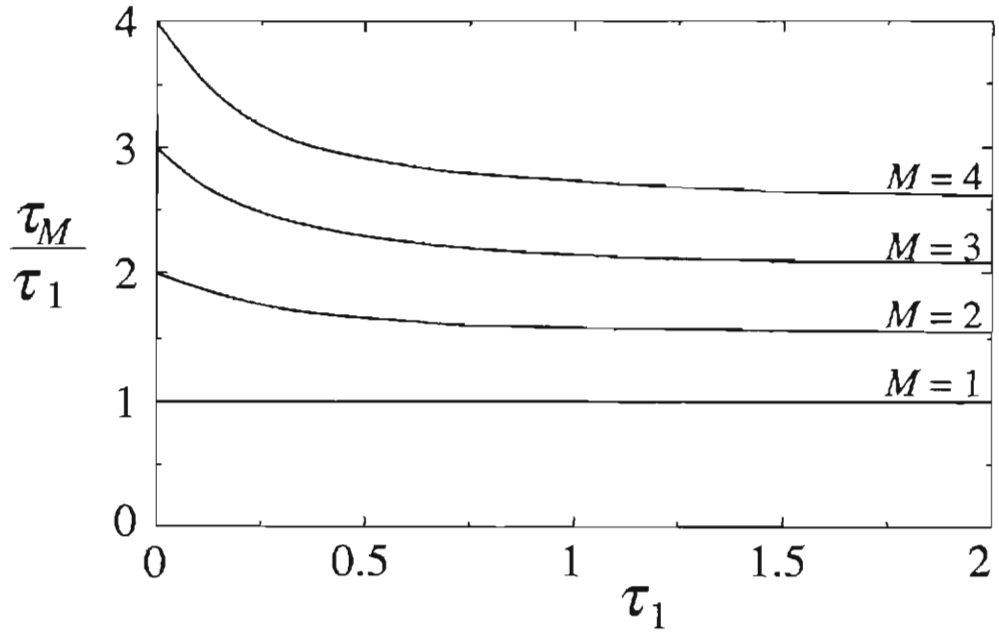


Fig. 6.3. The characteristic time τ_M for the M-step network, normalized by the single-step characteristic time τ_1, as a function of τ_1 for M = 1 - 4. Curves are from Eqs. (6.36) and (6.37), and are based on linear analysis in the vicinity of an attracting fixed point.

As a practical example of the convergence time result, we consider the multistep associative memories discussed in § 6.2, taking $F_i(z) = tanh(\beta z)$ and $I_i = 0$ for all i. In the vicinity of a recall state (or in the vicinity of the origin ($\vec{x} = 0$), when it is the unique attractor) the overlap $m(t)$ of the state of the network with a memory pattern evolves according the scalar iterated map

$$m(t+1) = tanh(b\,m(t)) \tag{6.39}$$

where

$$b = (1 - \alpha + \gamma)\beta \quad \text{[pseudo-inverse rule]}, \tag{6.40a}$$

$$b = (1 + \gamma)\beta \qquad \text{[Hebb rule; } \textit{analysis is only valid for } \alpha \to 0\text{]}. \tag{6.40b}$$

For these networks, the characteristic time for the single-step network is

$$\tau_1 = \frac{-1}{ln\left[b\ \mathrm{sech}^2(b\,m^*)\right]}\ , \tag{6.41}$$

where m^* is the stable fixed point of (6.39): $m^* = tanh(b\,m^*)$. A plot of τ_1 as a function of b from Eq. (6.41) is shown in Fig. 6.4. Values for the characteristic time τ_M for $M > 1$ can be found by multiplying τ_M/τ_1 (from Eqs. (6.36) - (6.38) or Fig. 6.3) by the value of τ_1 (from Eq. (6.41) or Fig. 6.4).

Although a given network configuration takes longer to converge as M increases (with other parameters fixed), it is possible in some instances to optimize both the neuron gain β and M to satisfy the stability criterion (6.16) with a resulting *reduction* in the convergence time for larger M. For example, in the case of the pseudo-inverse network

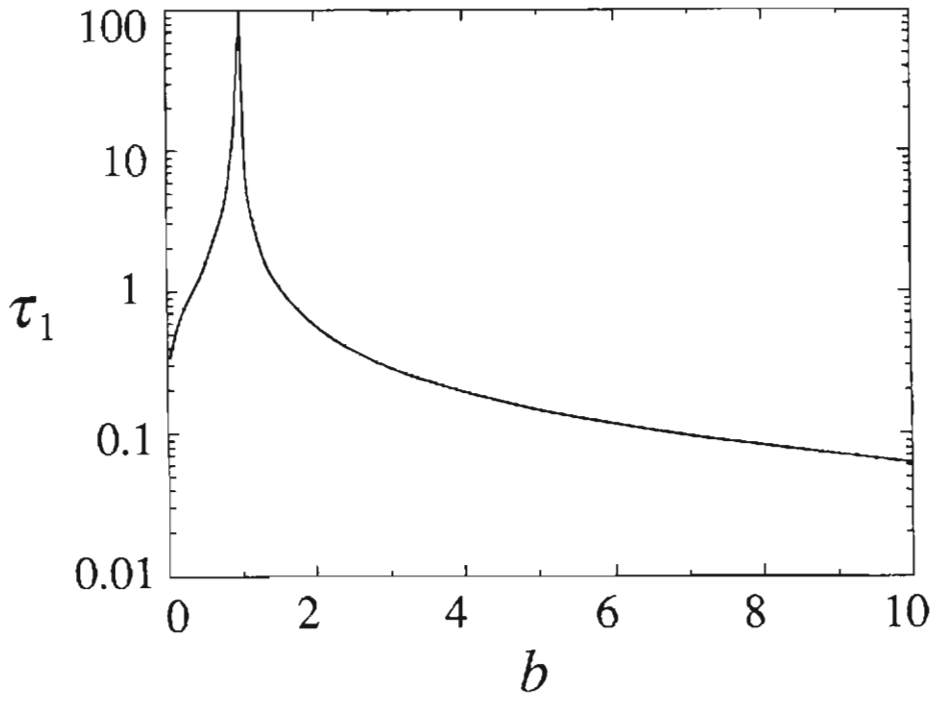


Fig. 6.4. The characteristic time τ_1 for the one-dimensional map $m(t+1) = tanh(b\,m(t))$ as a function of b, from Eq. (6.41). For $b<1$ the map converges to $m = 0$; for $b>1$ the map converges to one of two non-zero fixed points. This map describes the evolution of the overlap of the state vector with a memory pattern in the vicinity of a recall state for the Hebb rule in the $\alpha \to 0$ limit, and for the pseudo-inverse rule with $\alpha < 1$. Values for b are given by Eq. (6.40).

described above with α= 0.4 and $\gamma = 0$, the stability criterion (6.16) places an upper limit on the neuron gain that depends on M. For $M = 1, 2$ the optimal values are

$$M = 1\text{: } \beta < (1/0.4) = 2.5 \qquad [b < 1.5] \qquad \Rightarrow \qquad \tau_1 > [1.07 \times 1.00] = 1.07\,, \quad (6.42a)$$

$$M = 2\text{: } \beta < (2/0.4) = 5.0 \qquad [b < 3.0] \qquad \Rightarrow \qquad \tau_2 > [0.29 \times 1.72] = 0.49. \quad (6.42b)$$

In this example, using the maximum safe gain for the particular value of M gives a convergence time for the M = 2 network that is roughly half that of the M = 1 network.

6.5 CONCLUSIONS AND OPEN PROBLEMS

In this chapter we studied the dynamics of a symmetric analog neural network with a parallel update rule that averages over M previous time steps. We have shown that convergence to a fixed point attractor can be guaranteed by a simple stability criterion, Eq. (6.16), which limits the maximum neuron gain to a value proportional to M. The global analysis leading to this result is based on a new Liapunov function given in Eq. (6.3). For the system we have considered, certain aspects of the dynamics do not depend on M; these invariant properties include the associative memory storage capacity and the value of neuron gain needed to create fixed points away from the origin . The results were applied to multistep associative memories based on the Hebb and pseudo-inverse learning rules, giving the stability criterion (6.27).

In general, the multistep updating scheme is useful when (1) parallel dynamics is desired - for example, to take advantage of multiple processors; (2) connections are symmetric, and convergence to a fixed point is desired; and (3) the connection matrix has a negative eigenvalue of sufficient magnitude to render the stability criterion for the

single-step network overly restrictive (see: § 5.3) . For example, if $\lambda_{min}/\lambda_{max} < -1$, the only way to prevent oscillation in the $M = 1$ network is to lower all neuron gains until there is only a single, globally attracting fixed point (at which point the dynamics are computationally uninteresting). Increasing the number of time steps M allows the gain to be (safely) increased (as per (6.16)) to a sufficiently large value to create multiple fixed points.

As a quick (and final) example of where a multistep updating would be particularly useful, consider an analog network with a spin-glass connection matrix: the symmetric matrix T has random elements picked from a gaussian distribution with zero mean and variance J^2/N. This system has been studied [Soukoulis *et al.*, 1982; 1983; Ling *et al.*, 1983] as an approximation to the TAP mean-field approach [Thouless *et al.*, 1977] and yields reasonable predictions, comparable to Monte-Carlo techniques. [See, however: Reger *et al.*, 1984]. The numerical work of Soukoulis *et al.* was done using sequential update, and using a parallel update rule (for instance, using a multiprocessor computer) for such a system is problematic, since the large-N eigenvalue spectrum of T is symmetric about 0 (i.e. $\lambda_{max} = -\lambda_{min}$): as soon as spin glass states appear, period-2 limit cycles also appear. Figure 6.5 shows a phase diagram for the fully connected (SK-like) analog spin glass. The phase diagram shows that the parallel updating problem can be cured by going to a multistep updating rule. As long as $M > 1$, there is a range of neuron gain β for which spin glass states - but *not* oscillatory states - can be found.

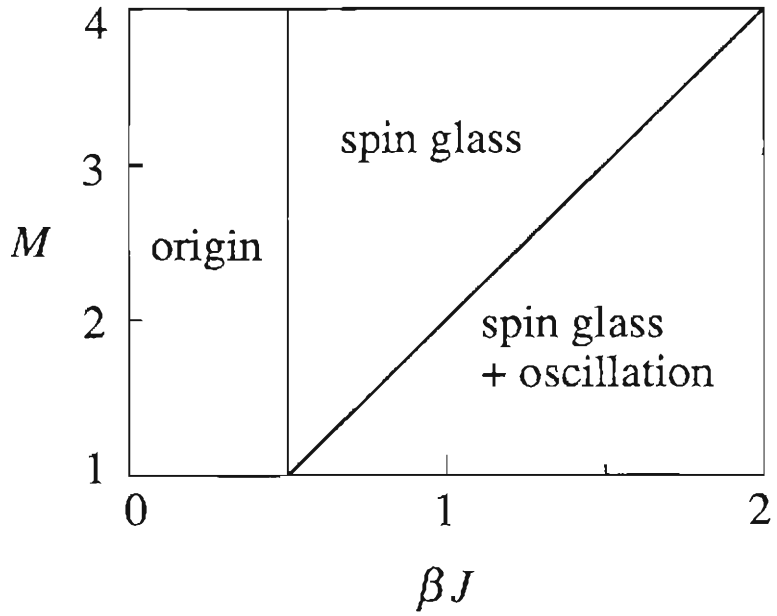


Fig. 6.5. Phase diagram for the M-step analog spin glass in the large-N limit. Connection matrix elements are random gaussian distributed with zero mean and variance J^2/N. The gain of each neuron is β. Notice that oscillation-free parallel updating requires $M > 1$.

We have also presented a simple analysis of convergence times and found that the number of iterations required for a multistep network to converge to a fixed point increases proportional to M when all other networks parameters held fixed. However, because an increase in the value of M allows the gain to be safely increased, in some instances using a larger M can reduce the convergence time when the gain is also optimally adjusted.

Two important extensions of the present results, which remain open problems, are the inclusion of (1) *weighted* averages over previous time; and (2) *nonsymmetric* matrices, especially those used to generate desired cyclic attractors. Another open problem is to prove the conjecture that the only attractors of (6.1) (with symmetric connections) for arbitrary M are limit cycles of period $M+1$ and all of its divisors (including 1, i.e. fixed points).

Chapter 7

COUNTING ATTRACTORS IN ANALOG SPIN GLASSES AND NEURAL NETWORKS

7.1. INTRODUCTION: DETERMINISTIC ANNEALING

The data in Figs. 5.6 and 5.7, showing the fraction of random initial states that settled onto various types of attractors in an analog associative memory, reveal a remarkable and useful property of analog neural networks: over a broad range of neuron gain, the chances of correctly finding a memory pattern increase as the gain is reduced. This property has been observed and discussed by several authors in a variety of applications [Hopfield and Tank, 1985, 1986; Koch *et al.*, 1986; Blake and Zisserman; 1987; Durbin and Willshaw, 1987], and may well be the primary motivation for developing parallel computation in analog. The benefits of analog computation were emphasized by Hopfield and Tank [1985, 1986], who explained the improved performance at lower gain by way of a comparison to the stochastic dynamics of simulated annealing [Kirkpatrick *et al.*, 1983]. In his review of neural networks in *Physics Today*, Sompolinsky [1988] assessed the situation from a slightly different perspective:

> What is the reason for the improved performance of the analog circuits? Obviously, there is nothing in the circuit's dynamics, which is the same as gradient descent, that prevents convergence to a local minimum. Apparently, the introduction of continuous degrees of freedom smooths the energy surface, thereby eliminating many of the shallow local minima.

Thus Sompolinsky makes a keen distinction between stochastic (or Monte-Carlo) dynamics on a rough energy landscape and deterministic, gradient-descent dynamics on a smooth energy landscape.[1] The results of the two schemes may be similar, but the dynamical process is quite different. The hypothesized smoothing of the energy landscape is illustrated schematically[2] in Fig. 7.1.

In this chapter we explore the structure of the energy landscape in an analog neural network by counting - analytically and numerically - the expected number of local minima in the landscape of a typical realization of a Hebb-rule associative memory. The analysis adapts techniques previously developed to count fixed points in the mean-field spin glass model of Thouless *et al.*[1977] (TAP) [Bray and Moore, 1980] and in neural networks with binary neurons [Gardner, 1986; Treves and Amit, 1988; Kepler, 1989]. The result provides a quantitative demonstration that using analog neurons dramatically reduces the number of local minima in the energy landscape [Waugh *et al.*, 1990].

We find that the expected number of local minima in the energy landscape $\langle N_{fp} \rangle_{\mathbf{T}}$, averaged over realizations of the connection matrix **T**, increases exponentially with the number of neurons N,

$$\langle N_{fp} \rangle_{\mathbf{T}} \sim exp[N a(\alpha,\beta)] \tag{7.1}$$

---

[1] "Energy" here means any Liapunov function appropriate to the particular network dynamics. In physical systems - or, more generally, in systems obeying detailed balance - the free energy behaves as a Liapunov function in the thermodynamic limit [for a proof, see Amit, 1989, §3.6.3], hence this expression. The "energy landscape" describes how a Liapunov function changes as one moves around in state space.

[2] These schematic energy landscapes are ubiquitous in the spin glass literature and seem to have a great influence on the way the community visualizes the complex state space of spin glasses and neural networks. Reducing state space to one dimension can create false intuitions if embraced too literally. Not only does this representation give a distorted sense of adjacency and distance, but it also creates the impression that fixed points are all either maxima or minima, with necessarily equal numbers of each. In high-dimensional space, this is not the case: First of all, there can be saddle points in addition to maxima and minima. Second, fixed points need not be evenly divided among the various types. For example, in the discrete state space of a binary-neuron network with deterministic sequential dynamics, *all* fixed points are minima.

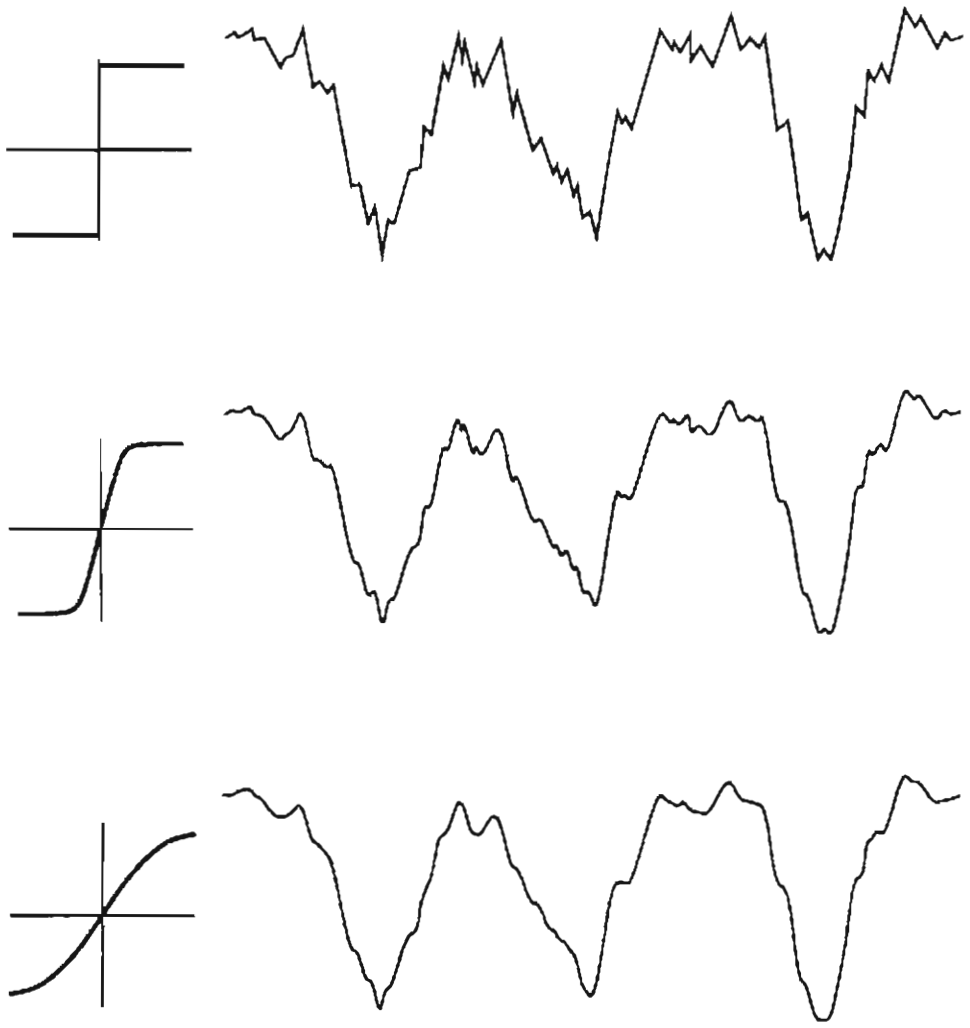


Fig. 7.1. Schematic energy landscapes illustrate how reducing the gain of the analog neurons smooths the landscape, thereby reducing the number of local minima.

with a scaling exponent $a(\alpha,\beta)$ which is an increasing function of both the neuron gain β and the ratio α of memory patterns to neurons.

For the case of binary neurons, Gardner [1986] found that fixed points in a Hebb rule network are not distributed evenly throughout state space, but tend to be correlated with the memory patterns, as seen in Fig. 7.2. For low storage ($p/N < 0.11$) there is a separate clump of fixed points very near each memory pattern. At a critical storage fraction, this highly correlated clump disappears, suggesting the disappearance of the recall state. This analysis provides an interesting alternative method for finding the storage capacity of a network. Although the result for the storage capacity ($p/N \sim 0.11$) is not as accurate as that found by Amit *et al.* [1985b; 1987], Gardner's technique is more versatile, being applicable, for example, to nonsymmetric matrices [Treves and Amit, 1988; Kepler, 1989; Fukai, 1990 (Fukai considers networks obeying Dale's rule, the physiologically-motivated requirement that a neuron be either purely excitatory or inhibitory)].

Despite the correlation of the fixed points with the memory patterns seen in Fig. 7.2, the vast majority of local minima are uncorrelated with all memory patterns. That is, nearly all fixed points are truly spurious states, having a vanishing overlap with the stored patterns. In fact, as Gardner points out, the distribution of the fixed points having overlap m with a stored pattern is very narrowly peaked about $m = 0$ in large systems; counting *all* fixed points or only counting the uncorrelated fixed points gives the same result in the large-N analysis.

In the limit of large neuron gain β, our scaling function $a(\alpha,\infty)$ from Eq. (7.1) agrees numerically with Gardner's result for the total number of fixed points in a Hebb rule network with binary neurons. In the limit where both $\alpha,\beta \to \infty$, we recover the familiar result for the zero-temperature Ising spin glass [Tanaka and Edwards, 1980; De Dominicis *et al.*, 1980; Bray and Moore, 1980],

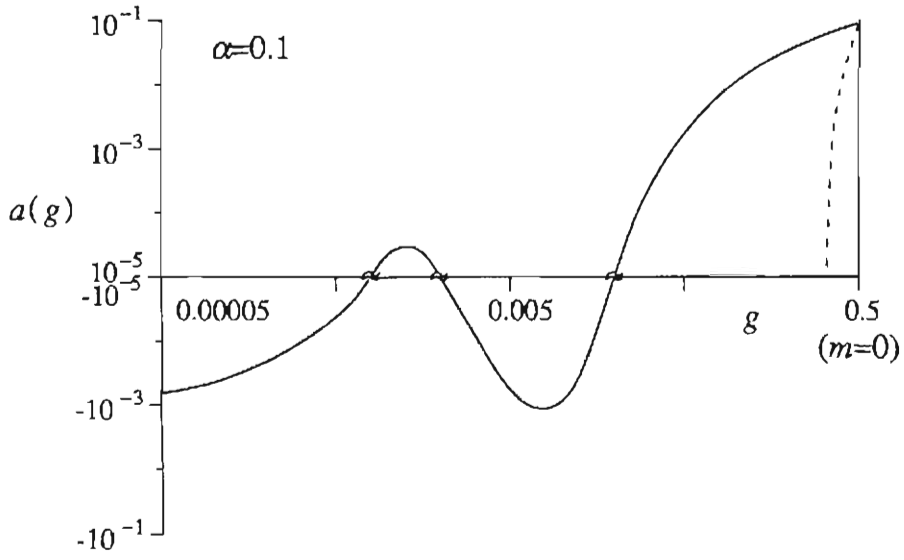


Fig. 7.2. Scaling exponent $a(g)$ for the number of fixed points a Hebb-rule neural network with binary neurons and storage ratio $\alpha = 0.1$. The expected number of fixed points at a Hamming distance Ng from a stored pattern in a typical realization is $\langle N_{fp}(N,g)\rangle = exp[a(g)N]$. Negative values indicate regions where no fixed points are expected. Fixed points at $g = 0.5$ (overlap with pattern $m = 0$) dominate at large N. The dashed line shows the scaling exponent if the fixed points were evenly distributed throughout state space. After Gardner [1986].

$$a(\infty,\infty) = 0.1992... \tag{7.2}$$

One might further suspect that for finite β, the limit $\alpha \to \infty$ would correspond to the finite-temperature spin glass result [Bray and Moore; 1980] with β playing the role of the inverse temperature in the TAP equations. This is not the case, however. The inclusion of a reaction field term in the TAP equations greatly affects the number of fixed points, thus the function $a(\infty,\beta) \equiv a(\beta)$ is *not equivalent* to the spin glass result of Bray and Moore [1980], as we will show below. The reaction field also makes the dynamical version of the TAP equations chaotic for finite β so that a direct numerical test of the theory for the TAP equations has not been possible [Bray and Moore, 1979].

The analytical results presented in § 7.2 are in good agreement with numerical counts of local minima, which are given in § 7.3 [Waugh *et al.*, 1990]. This agreement suggests that the analysis is correct despite its many approximations and assumptions.

The work presented in this chapter was done in collaboration with Fred Waugh, who led the way through most of the roughest terrain, both analytical and numerical.

7.2. COUNTING ATTRACTORS: ANALYSIS

7.2.1. Analog spin glass

Before considering the analog associative memory, we begin by solving a slightly easier problem. We will calculate the expected number of local minima in the energy landscape for the analog version of the SK spin glass [Sherrington and Kirkpatrick, 1975]. The results are interesting in their own right as a demonstration of landscape smoothing in a well-studied multi-attractor system. Also, the techniques used here will

appear again when we analyze the analog associative memory.

The fixed-point condition for the analog network is given by

$$x_i = F(h_i) = F\left(\sum_j T_{ij} x_j\right) \quad , \qquad i = 1, ..., N, \tag{7.3}$$

which can be written as

$$G_i \equiv g_i(x_i) - \beta \sum_j T_{ij} x_j = 0 \ ; \qquad i = 1, ..., N; \tag{7.4}$$

with the definition

$$g_i(x_i) \equiv \beta F^{-1}(x_i) \, . \tag{7.5}$$

The neuron transfer function F must be invertible[3] and the connection matrix $\mathbf{T} = \{T_{ij}\}$ is assumed symmetric with elements chosen from a gaussian distribution having zero mean and variance J^2/N. The normalized probability distribution for the T_{ij} is

$$P(T_{ij}) = \left(\frac{N}{2\pi J^2}\right)^{1/2} exp\left(\frac{-N T_{ij}^{\,2}}{2J^2}\right) \, , \qquad T_{ij} = T_{ji}. \tag{7.6}$$

The type of dynamics or update rule leading to the fixed point condition (7.4) is unspecified; the results apply to systems with continuous-time dynamics, discrete-time sequential dynamics and parallel dynamics as long as the stability criterion of § 5.3 is also satisfied.

We are most interested in counting the *stable* fixed points, that is, minima of the energy landscape, not saddles or maxima. To limit the count to include only stable fixed

---

[3] No other constraints on the nonlinear function are mandated by the analysis. However, the only nonlinearity for which the theory has been checked by numerical experiment is $F(h_i) = tanh(\beta h_i)$. How well the theory works with other nonlinear functions is unknown.

The squared sum on the right of (7.16) can be made linear using a Hubbard-Stratonovich transformation,

$$exp\left(\frac{\lambda a^2}{2}\right) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} dx\, exp\left[-\frac{\lambda x^2}{2} + a\lambda x\right] \tag{7.17}$$

(with $a \equiv (\beta J/N)\sum_i k_i x_i$ and $\lambda \equiv N$ in this case). This introduces a new integration variable V, which will eventually be evaluated by steepest descent. We now have

$$\left\langle N_{fp}\right\rangle_{\mathbf{T}} = \underset{V}{Max} \int_{-i\infty}^{i\infty} \prod_i \left(\frac{dk_i}{2\pi i}\right) \int \prod_i dx_i$$
$$exp\left[\frac{\beta^2 J^2}{2N}\left(\sum_i k_i^2\right)\left(\sum_i x_i^2\right) - \frac{NV^2}{2} + \sum_i \left(\beta J V x_i + g(x_i)\right)k_i\right] \left\langle \left| det\ \mathbf{A}\right|\right\rangle_{\mathbf{T}} \tag{7.18}$$

Next, we introduce the order parameter

$$q = \frac{1}{N}\sum_i x_i^2 \tag{7.19}$$

by multiplying (7.18) by an integral over a complex exponential that is equal to 1:

$$1 = \frac{N}{2\pi i}\int_{-i\infty}^{i\infty} d\lambda \int_{-\infty}^{\infty} dq\, exp\left[-\lambda\left(Nq - \sum_i x_i^2\right)\right]. \tag{7.20}$$

This adds two more variables, q and λ, to be determined by steepest descent. With this substitution, the integrals over k_i in (7.18) are now gaussian and can be integrated to give

$$\left\langle N_{fp}\right\rangle_{\mathbf{T}} = \underset{q,\lambda,V}{Max} \int \prod_i dx_i \, exp\left[-\frac{NV^2}{2} - \lambda N q \right] \left(\frac{1}{\sqrt{2\pi q}\,\beta J} \right)^N exp\left[-\sum_i \left(\frac{\left(\beta J V x_i + g(x_i)\right)^2}{2\beta^2 J^2 q} + \lambda {x_i}^2 \right) \right] \left\langle \left| det\ \mathbf{A} \right| \right\rangle_{\mathbf{T}} \tag{7.20}$$

Notice that the x_i's on different sites i are now decoupled. This allows the N integrals over the x_i's to be written as a product of N identical integrals, or as a single integral over x (no index) raised to the N^{th} power.

We next consider the $\left\langle \left| det\, \mathbf{A} \right| \right\rangle_{\mathbf{T}}$, which we evaluate using the following property of the multidimensional gaussian integral:

$$(det\ \mathbf{A}) \left\{ \prod_i \theta[\lambda_i(\mathbf{A})] \right\} = \left[\int_{-\infty}^{\infty} \prod_i \frac{d\rho_i}{\sqrt{2\pi}} exp\left(-\frac{1}{2} \sum_{i,j} \rho_i A_{ij} \rho_j \right) \right]^{-2} . \tag{7.22}$$

where λ_i, $i = 1, \ldots, N$ are the eigenvalues of $\mathbf{A}$, and θ is the step function, $\theta(z) = (1 + Sgn(z))/2$. Notice that the RHS of (7.22) equals $det\ \mathbf{A}$ as long as $\mathbf{A}$ is positive definite, otherwise it equals zero. This is just what we need to count stable fixed points: Recall that $\mathbf{A}$ is the Hessian of the Liapunov function for the dynamical systems discussed above. Thus replacing $|det\ \mathbf{A}|$ in (7.20) with $(det\ \mathbf{A}) \prod_i \theta[\lambda_i(\mathbf{A})]$ will pick out the minima of the energy landscape. The RHS of (7.22) can be averaged by introducing replicas [Bray and Moore, 1980], and eventually setting the number of replicas to -2. Dropping the step functions (and understanding that the count now includes only stable fixed points), we write

$$det\ \mathbf{A} = \lim_{m \to -2} \int_{-\infty}^{\infty} \prod_{i,\alpha} \frac{d\rho_{i\alpha}}{\sqrt{2\pi}} exp\left(-\frac{1}{2} \sum_{i,j} \sum_{\alpha=1}^{m} \rho_{i\alpha} A_{ij} \rho_{j\alpha} \right). \tag{7.23}$$

Calculating the average over realizations $\langle det\,\mathbf{A} \rangle_{\mathbf{T}}$ from (7.23) follows Bray and Moore [1980], and is presented in Appendix 7A. The result, which assumes replica symmetry, is

$$\langle det\,\mathbf{A} \rangle_{\mathbf{T}} = \min_{R} \prod_{i} \left(g'(x_i) - 2\beta JR \right) exp\left(2NR^2 \right). \tag{7.24}$$

The *Min* over R comes from a steepest descent integral. Having to minimize with respect to R (rather than maximize) is an artifact of the replica method, as explained in Appendix 7-A.

In certain regions of state space $\{x_i\}$, the RHS of (7.24) becomes negative. We interpret a negative result in (7.24) as indicating that **A** is not positive definite in that region of state space, and that the replica symmetric treatment has failed to return a zero for the average, as it should. Because we are interested in counting the stable fixed points (where **A** *is* positive definite) *we will limit integrals over state space to the sub-region of the range of F where* $\left(g'(x) - 2\beta JR \right)$ *is positive*. This is indicated by a " + " marking such integrals. We note that limiting the integrals in this way was necessary to obtain meaningful results from the saddle-point equations below.

From (7.21), (7.24), and the change of variables: $B \equiv -2\beta JR$, $\Delta \equiv -\beta JV$, the average number of fixed points can now be written

$$\left\langle N_{fp}\right\rangle_{\mathrm{T}} = \underset{q,\lambda,\Delta}{Max}\ \underset{B}{Min}\left\{exp\left[N\left(\frac{B^2-\Delta^2}{2\beta^2J^2}-\lambda q+ln(I)\right)\right]\right\}$$

$$\equiv \underset{q,\lambda,\Delta}{Max}\ \underset{B}{Min}\left\{exp\left[N\,\tilde{a}(\beta,q,\lambda,B,\Delta)\right]\right\} \tag{7.25}$$

where

$$I = \frac{1}{\sqrt{2\pi q}\,\beta J}\int_{+} dx\,\left(g'(x)+B\right)exp\left[-\frac{\left(g(x)-\Delta x\right)^2}{2\beta^2J^2q}+\lambda x^2\right]. \tag{7.26}$$

and the + on the integral means "only integrate where $\left(g'(x)+B\right)>0$". Finding extrema of (7.25) is done by setting derivatives of $\tilde{a}(\beta,q,\lambda,B,\Delta)$ to zero. This leads to the following set of coupled equations

$$\begin{aligned}
\partial\tilde{a}/\partial\lambda = 0 \quad &\Rightarrow \quad q = \left\langle\left\langle x^2\right\rangle\right\rangle \\
\partial\tilde{a}/\partial\Delta = 0 \quad &\Rightarrow \quad \Delta = \frac{1}{2q}\left\langle\left\langle x\,g(x)\right\rangle\right\rangle \\
\partial\tilde{a}/\partial q = 0 \quad &\Rightarrow \quad \lambda = -\frac{1}{2q}\left(1-\frac{1}{\beta^2J^2q}\left\langle\left\langle\ \left(g(x)-\Delta x\right)^2\ \right\rangle\right\rangle\right) \\
\partial\tilde{a}/\partial B = 0 \quad &\Rightarrow \quad B = -\beta^2J^2\left\langle\left\langle\ \left(g'(x)+B\right)^{-1}\right\rangle\right\rangle
\end{aligned} \tag{7.27}$$

The double brackets in (7.27) indicate a weighted average, with the weight function given by the integrand of I from Eq.(7.26):

$$\left\langle\left\langle\ f(x)\ \right\rangle\right\rangle \equiv \frac{\int_{+} dx\ f(x)\,W(x)}{\int_{+} dx\ W(x)}\ ; \tag{7.28a}$$

$$W(x) \equiv \left(g'(x) + B\right) exp\left[-\frac{\left(g(x) - \Delta x\right)^2}{2\beta^2 J^2 q} + \lambda x^2\right]. \qquad (7.28b)$$

A self-consistent solution to the four equations in (7.27) can be found numerically. Resulting values of q, λ, B, Δ, and I are inserted into (7.25) to yield a value of $\left\langle N_{fp} \right\rangle_T \equiv exp\left[N\, a(\beta)\right]$. The resulting numerical values for $a(\beta)$ are shown in Fig. 7.3 for the particular nonlinearity $F(h_i) = tanh(\beta h_i)$ along with the corresponding value for the TAP spin-glass result of Bray and Moore [1980]. For all finite values of β, the analog spin-glass with neuron gain β has more local mimima than the TAP equations with inverse temperature β. In the limit of large gain, both the TAP result and the analog spin-glass result approach the zero-temperature Ising spin-glass value $a(\beta \to \infty) = 0.1992$..[Bray and Moore, 1980].

A slight complication: Numerically evaluating the integrals in Eq. (7.26) and (7.27) is difficult for saturating nonlinearities, for example $F(h_i) = tanh(\beta h_i)$ because the integrands diverge at the endpoints while the integral itself remains finite. To evaluate these integrals, it was necessary to expand the integrands near the endpoints and evaluate the integrals analytically in these regions. The relevant formulas are given in Appendix 7B.

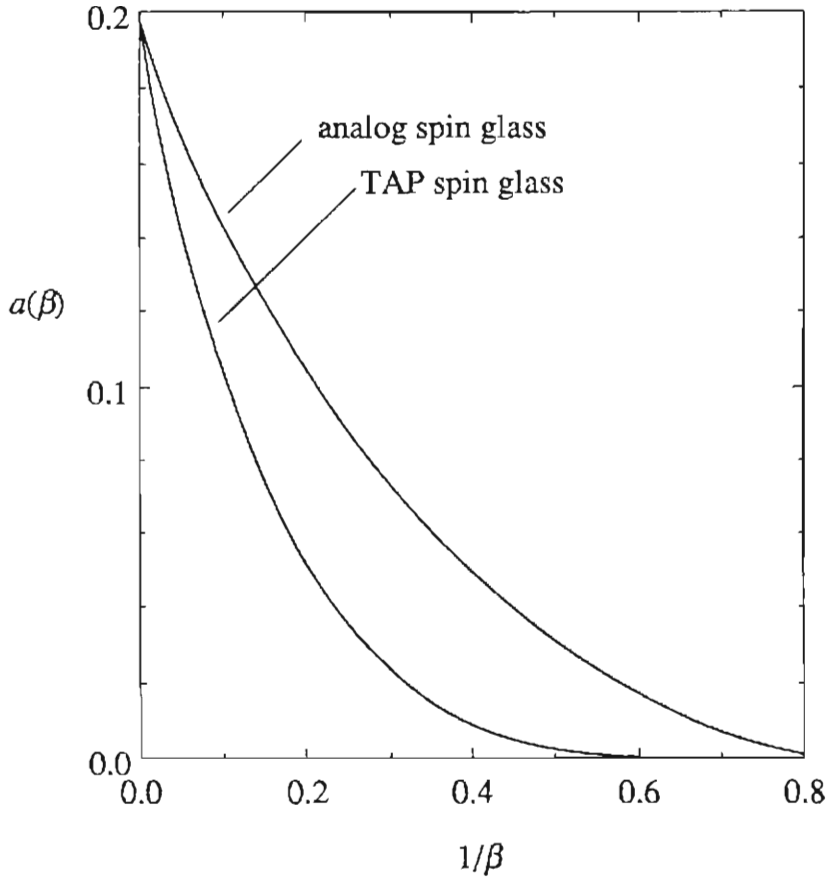


Fig. 7.3. Theoretical result for the scaling exponent $a(\beta)$ as a function of inverse neuron gain $1/\beta$ for the analog spin glass. The expected number of local minima is $\langle N_{fp} \rangle \approx exp[a(\beta)N]$. Also shown for comparison is the number of solutions of the TAP equations at temperature $1/\beta$ from Bray and Moore [1980].

7.2.2. Analog neural network

In this section, we calculate the average number of local minima in the energy landscape for an analog neural network as a function of neuron gain β and the ratio α of stored memories to neurons. This calculation [Waugh *et al.*, 1990] is a hybrid of the methods used in the previous section - which were originally developed by Bray and Moore [1980] for the TAP spin-glass [Thouless *et al.*, 1977] - and the analysis by Gardner *et al.* [Gardner, 1986; Bruce *et al.*, 1987] of the number and distribution of fixed points in a Hebb-rule neural network with binary (Ising) neurons. We will not consider how the fixed points are distributed in state space, but will instead count their total number. Based on Gardner's results, we expect that most of the fixed points have a vanishing overlap with any of the memory patterns as $N \to \infty$. That is, most of the fixed points are completely spurious states.

The analysis begins just as in the previous section, except that now the interconnection matrix $\mathbf{T} = \{T_{ij}\}$ is given by the Hebb rule:

$$T_{ij} = \frac{1}{N\sqrt{\alpha}} \sum_{\mu=1}^{\alpha N} \xi_i^\mu \xi_j^\mu ; \quad T_{ii} = 0 \qquad i, j = 1, ..., N. \tag{7.29}$$

where each $\xi_i^\mu = \pm 1$ at random with equal probability and αN is the number of stored patterns. The normalization $(N\sqrt{\alpha})^{-1}$ is chosen to make $\sum_j T_{ij} \sim 1$, independent of N and α.

As in the previous section, the expected number of fixed points is found by integrating a product of delta functions on G_i over state space,

$$\left\langle N_{fp}\right\rangle_{\xi} = \left\langle \int \prod_i dx_i \prod_i \delta\left[g(x_i) - \beta \sum_j T_{ij} x_j \right] \left| det\ \mathbf{A} \right| \right\rangle_{\xi} \tag{7.30}$$

with $g(x_i)$ defined in (7.5) and **A**, the Hessian of the Liapunov function, defined in (7.8). We use an integral representation of the delta function, this time taking k_i real, which gives

$$\left\langle N_{fp}\right\rangle_{\xi} = \left\langle \int_{-\infty}^{\infty} \prod_i \left(\frac{dk_i}{2\pi}\right) \int \prod_i dx_i \ \times \right.$$

$$\left. exp\left[i \sum_i k_i g(x_i) - \frac{i\beta}{N\sqrt{\alpha}} \sum_{i,j,\mu} k_i \xi_i^{\mu} \xi_j^{\mu} x_j + i\beta\sqrt{\alpha} \sum_i k_i x_i \right] \left| det\ \mathbf{A} \right| \right\rangle_{\xi} , \tag{7.31}$$

We now make the approximation that *det* **A** can be averaged separately from the rest of the integral in (7.31). That is, we set the average of a product equal to the product of the averages: $\langle X \times det\,\mathbf{A} \rangle = \langle X \rangle \times \langle det\,\mathbf{A} \rangle$. Physically, this procedure assumes that the vast majority of local minima have identical local curvatures, so that a single value - the average - can be used as the multiplier for each. This assumption is reasonable in light of the fact that other features of the local minima, such as their energies and overlaps with the stored patterns, behave in this way, i.e. their distributions are dominated by a single value at large N. The only term in (7.31) which depends on the ξ_i^{μ} - besides *det* **A** - is the second term in the exponential, which we write as

$$\frac{-i\beta}{N\sqrt{\alpha}} \sum_{i,j,\mu} k_i \xi_i^{\mu} \xi_j^{\mu} x_j = \sum_{\mu} -i \left(\left(\frac{\beta}{N\sqrt{\alpha}} \right)^{1/2} \sum_i k_i \xi_i^{\mu} \right) \left(\left(\frac{\beta}{N\sqrt{\alpha}} \right)^{1/2} \sum_i x_i \xi_i^{\mu} \right), \tag{7.32}$$

This term can be further "simplified" by separating the product on the right of (7.32) and

introducing two new integration variables for each μ. This replacement uses the double Fourier integral:

$$exp\left[-i(AB)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} da\, db\, exp\left[i(ab - Aa - Bb)\right]. \tag{7.33}$$

Rearranging slightly, (7.31) can be written

$$\left\langle N_{fp}\right\rangle_{\xi} = \int_{-\infty}^{\infty}\prod_{i}\left(\frac{dk_i}{2\pi}\right)\int\prod_{i}dx_i\int_{-\infty}^{\infty}\prod_{\mu}\left(\frac{da_\mu\, db_\mu}{2\pi}\right)\times$$

$$exp\left[i\sum_{i}k_i\left(g(x_i)+\beta\sqrt{\alpha}x_i\right)+i\sum_{\mu}a_\mu b_\mu\right]\times$$

$$\left\langle exp\left[-i\left(\frac{\beta}{N\sqrt{\alpha}}\right)^{1/2}\sum_{i,\mu}\left(a_\mu k_i + b_\mu x_i\right)\xi_i^\mu\right]\right\rangle_{\xi}\left\langle\left|det\ \mathbf{A}\right|\right\rangle_{\xi}. \tag{7.34}$$

For $\xi_i^\mu = \pm 1$ with equal probability, the average over ξ_i^μ can be done easily,

$$\left\langle exp\left(-ia\xi_i^\mu\right)\right\rangle_{\xi} = cos(a) \xrightarrow[a<<1]{} exp\left(-\frac{a^2}{2}\right), \tag{7.35}$$

where the term corresponding to a in (7.34) is << 1 for large N. Applying (7.35) to (7.34) gives

$$\left\langle N_{fp} \right\rangle_{\xi} = \int_{-\infty}^{\infty} \prod_i \left(\frac{dk_i}{2\pi} \right) \int \prod_i dx_i \int_{-\infty}^{\infty} \prod_\mu \left(\frac{da_\mu \, db_\mu}{2\pi} \right) \times$$

$$exp\left[i \sum_i k_i \left(g(x_i) + \beta \sqrt{\alpha}\, x_i \right) + i \sum_\mu a_\mu b_\mu \right] \times$$

$$exp\left[-\frac{\beta}{2N\sqrt{\alpha}} \left(\sum_i k_i^2 \sum_\mu a_\mu^2 + \sum_i x_i^2 \sum_\mu b_\mu^2 + 2 \sum_\mu a_\mu b_\mu \sum_i k_i x_i \right) \right] \times$$

$$\left\langle |det\ \mathrm{A}| \right\rangle_{\xi} \tag{7.36}$$

We now define three order parameters, q, s, and t, and their conjugate fields Q, S, and T via integral definitions of 1:

$$q = \frac{1}{N} \sum_i x_i^2 \quad \rightarrow \quad 1 = (N/2\pi i) \iint dq\, dQ\ exp\left[Q \left(Nq - \sum_i x_i^2 \right) \right] \tag{7.37}$$

$$s = \frac{1}{N} \sum_i k_i^2 \quad \rightarrow \quad 1 = (N/2\pi i) \iint ds\, dS\ exp\left[S \left(Ns - \sum_i k_i^2 \right) \right] \tag{7.38}$$

$$t = \frac{i}{N} \sum_i x_i k_i \quad \rightarrow \quad 1 = (N/2\pi i) \iint dt\, dT\ exp\left[T \left(Nt - i \sum_i x_i k_i \right) \right] \tag{7.39}$$

This allows the integrals over a_μ and b_μ in (7.36) to be evaluated by straightforward gaussian integration, giving

$$\left\langle N_{fp} \right\rangle_{\xi} = \left(\frac{N}{2\pi i}\right)^3 \int dq\, dQ\, ds\, dS\, dt\, dT\ exp[NG] \times$$

$$\int_{-\infty}^{\infty} \prod_i \left(\frac{dk_i}{2\pi}\right) \int \prod_i dx_i\ exp\left[i \sum_i k_i \left(g(x_i) + \left(\beta\sqrt{\alpha} - T\right) x_i \right) \right] \times$$

$$exp\left[-Q \sum_i x_i^2 - S \sum_i k_i^2 \right] \left\langle |det\ \mathbf{A}| \right\rangle_{\xi} \qquad (7.40)$$

where

$$G \equiv qQ + sT + tT - \frac{\alpha}{2} ln\left[\left(\beta^2 qs/\alpha\right) + \left(1 + \beta t/\sqrt{\alpha}\right)^2 \right]. \qquad (7.41)$$

Because q, s, and t now only appear in the $exp[NG]$ term, integrals over these variables can be evaluated easily using steepest descent by setting partial derivatives of G equal to zero,

$$\frac{\partial G}{\partial q} = \frac{\partial G}{\partial s} = \frac{\partial G}{\partial t} = 0\,, \qquad (7.42)$$

which yields solutions

$$q = \frac{2S\alpha}{4QS + T^2}\ ;\ \ s = \frac{2Q\alpha}{4QS + T^2}\ ;\ \ t = \frac{T\alpha}{4QS + T^2} - \frac{\sqrt{\alpha}}{\beta}\ ; \qquad (7.43)$$

$$G = \alpha - \frac{T\sqrt{\alpha}}{\beta} - \frac{\alpha}{2} ln\left[\frac{\beta^2 \alpha}{4QS + T^2} \right]. \qquad (7.44)$$

The integrals over k_i are now in gaussian form and can be evaluated using (7.17), to

give

$$\left\langle N_{fp}\right\rangle_{\xi} = \underset{Q,S,T}{Max}\left\{\left(\frac{1}{\sqrt{4\pi S}}\right)^{N} exp\left[N\left(\alpha-\frac{T\sqrt{\alpha}}{\beta}-\frac{\alpha}{2}ln\left(\frac{\beta^{2}\alpha}{4QS+T^{2}}\right)\right)\right]\times\right.$$

$$\left.\int\prod_{i}dx_{i}\ exp\left[\sum_{i}-Qx_{i}^{2}-\frac{\left(g(x_{i})+\left(\beta\sqrt{\alpha}-T\right)x_{i}\right)^{2}}{4S}\right]\right\}\left\langle\left|det\ \mathbf{A}\right|\right\rangle_{\xi} \qquad (7.45)$$

where *Max* indicates that the integrals over Q, S, and T are evaluated by steepest descent by numerically maximizing the expression in curly brackets.

Calculating $\left\langle\left|det\ \mathbf{A}\right|\right\rangle_{\xi}$ is done in a similar manner as for the analog spin glass: We start with (7.23), which picks out only the stable fixed points (the subspace where **A** is positive definite) and average using the replica method over realizations of the Hebb matrix. Again, we drop the absolute value brackets, as (7.23) is only applicable where $det\ \mathbf{A} > 0$. This rest of the calculation is shown in Appendix 7C. The replica symmetric solution is

$$\left\langle det\ \mathbf{A}\right\rangle_{\xi} = \underset{R}{Min}\left\{exp\left[N\left(\frac{2R\sqrt{\alpha}}{\beta}-\alpha+\alpha\ ln\left(\frac{\beta\sqrt{\alpha}}{R}\right)-\alpha\ ln2\right)\right]\times\right.$$

$$\left.\prod_{i}\left(g'(x_{i})+\beta\sqrt{\alpha}-2R\right)\right\} \qquad (7.46)$$

where $g'(x_i)$ is the derivative of $g(x_i)$ with respect to x_i. As we discovered for the analog spin glass, the RHS of Eq. (7.46) becomes negative in certain regions of state space. Again, we exclude such regions from all integrations over state space. This provides an approximate way of counting only the stable fixed points, as described in the previous section. Restricting the domain of integration in this way was also necessary in

order to obtain meaningful results from the saddle-point equations below. Substituting (7.46) into (7.45), and changing the variables Q, R, S, and T to a new set: λ, B, q, and Δ, by the definitions

$$\begin{aligned} B &\equiv \beta\sqrt{\alpha} - 2R \\ \lambda &\equiv -Q \\ q &\equiv 2S/\beta^2 \\ \Delta &\equiv T - \beta\sqrt{\alpha} \end{aligned} \tag{7.47}$$

gives

$$\begin{aligned} \left\langle N_{fp} \right\rangle_{\xi} &= \underset{q,\lambda,\Delta}{Max}\ \underset{B}{Min} \left\{ exp\left[N\left(-\frac{\sqrt{\alpha}}{\beta}(B+\Delta) + \frac{\alpha}{2}\, ln\left[\frac{\left(\Delta + \sqrt{\alpha}\beta\right)^2 - 2\lambda\beta^2 q}{\left(B - \sqrt{\alpha}\beta\right)^2} \right] + ln\left(\hat{I}\right) \right) \right] \right\} \\ &\equiv \underset{q,\lambda,\Delta}{Max}\ \underset{B}{Min} \left\{ exp\left[N\,\tilde{a}(\alpha,\beta,q,\lambda,B,\Delta) \right] \right\} \end{aligned} \tag{7.48}$$

where

$$\hat{I} = \frac{1}{\sqrt{2\pi q}\,\beta} \int_{+} dx\, \left(g'(x) + B\right) exp\left[-\frac{\left(g(x) - \Delta x\right)^2}{2\beta^2 q} + \lambda x^2 \right] \tag{7.49}$$

and the "+" on the integral means that the region of integration is restricted to the range of F where $\left(g'(x) + B\right) > 0$. Notice that $\hat{I}$ from (7.49) is identical to I from (7.26) upon setting J = 1. The rest of the present solution, Eq. (7.48), also bears a strong

resemblance to the corresponding analog spin-glass solution, (7.25), but has some significant differences. We note, however, that as $\alpha \to \infty$, $\tilde{a}(\alpha,\beta,q,\lambda,B,\Delta)$ can be expanded to the leading order in α and becomes identical to $\tilde{a}(\beta,q,\lambda,B,\Delta)$ from Eq. (7.25). That is, in the limit $\alpha \to \infty$ the analog Hebb-rule network is formally equivalent to the analog spin glass. This correspondence has been noted already for the case of binary neurons [Gardner, 1986; Treves and Amit, 1988].

To find extrema of (7.48) with respect to q, λ, B, and Δ, we set partial derivatives of $\tilde{a}(\alpha,\beta,q,\lambda,B,\Delta)$ equal to zero,

$$\frac{\partial \tilde{a}}{\partial q} = \frac{\partial \tilde{a}}{\partial \lambda} = \frac{\partial \tilde{a}}{\partial B} = \frac{\partial \tilde{a}}{\partial \Delta} = 0 \tag{7.50}$$

which gives a set of four integral equations:

$$\begin{aligned}
q &= \frac{\left(\Delta + \sqrt{\alpha}\beta\right)^2 - 2\lambda\beta^2 q}{\alpha\beta^2} \left\langle\!\left\langle x^2 \right\rangle\!\right\rangle \\
0 &= \alpha\beta^2 + \left(\left(\Delta + \sqrt{\alpha}\beta\right)^2 - 2\lambda\beta^2 q\right)\left[\frac{\left\langle\!\left\langle\, x\, g(x)\,\right\rangle\!\right\rangle}{\beta\sqrt{\alpha}\, q} - 1\right] \\
B &= \left(\frac{\beta B}{\sqrt{\alpha}} - \beta^2\right)\left\langle\!\left\langle\, \left(g'(x) + B\right)^{-1}\,\right\rangle\!\right\rangle \\
\lambda &= -\left[\frac{\left(\Delta + \sqrt{\alpha}\beta\right)^2 - 2\lambda\beta^2 q}{2\alpha\beta^2 q}\right]\left(1 - \frac{1}{\beta^2 q}\left\langle\!\left\langle\, \left(g(x) - \Delta x\right)^2\,\right\rangle\!\right\rangle\right)
\end{aligned} \tag{7.51}$$

where, as above, the double brackets are a weighted average with the weight function given by the integrand of $\hat{I}$. The weighted average is exactly that given in Eq. (7.28), setting $J = 1$ in (7.28b). For given values of α and β, a set of solutions for q, λ, B, and Δ are found by numerically solving (7.51). These four values are then inserted into (7.48) to yield a numerical value for the quantity of interest, $a(\alpha,\beta)$, defined as

$$\left\langle N_{fp} \right\rangle_{\xi} = exp\left[N\, a(\alpha,\beta)\right]. \tag{7.52}$$

Notice that the terms within the double brackets in (7.51) are identical to those appearing in the spin glass analysis, Eq. (7.27). Thus to evaluate the integrals implied by the double brackets we again must use the expansion of Appendix 7B.

Values for $a(\alpha,\beta)$ are plotted in Fig. 7.4 for the particular case $F(h_i) = tanh(\beta h_i)$. The data show that for any value of the storage ratio α, the number of fixed points in the energy landscape is reduced as the neuron gain is lowered. In the large-gain limit, $a(\alpha,\infty)$ is found numerically to agree with Gardner's [1986] result for binary neurons.

The form of (7.52) indicates that the number of fixed points is dramatically affected by even small changes in $a(\alpha,\beta)$, especially for large N. As an example, consider the effect of lowering the neuron gain from $\beta = 100$ to $\beta = 10$ in a network with storage ratio $\alpha = 0.1$: Using the values $a(0.1,100) = 0.059$ and $a(0.1,10) = 0.040$, we expect that the average number of fixed points will be reduced by ~ 97% for N=200 and by eight orders of magnitude for N=1000.

7.3. COUNTING ATTRACTORS: NUMERICAL RESULTS

7.3.1. Technique for counting fixed points

The analytical results of § 7.2, summarized in Figs. 7.3 and 7.4, have been tested for the standard sigmoidal nonlinearity $F(h_i) = tanh(\beta h_i)$ by directly counting the stable fixed points in small computer-generated analog spin glasses and neural networks. Numerical data were obtained by the following procedure: At several values of β for the spin glass, or pairs of values (α,β) for the neural network, 20 random connection

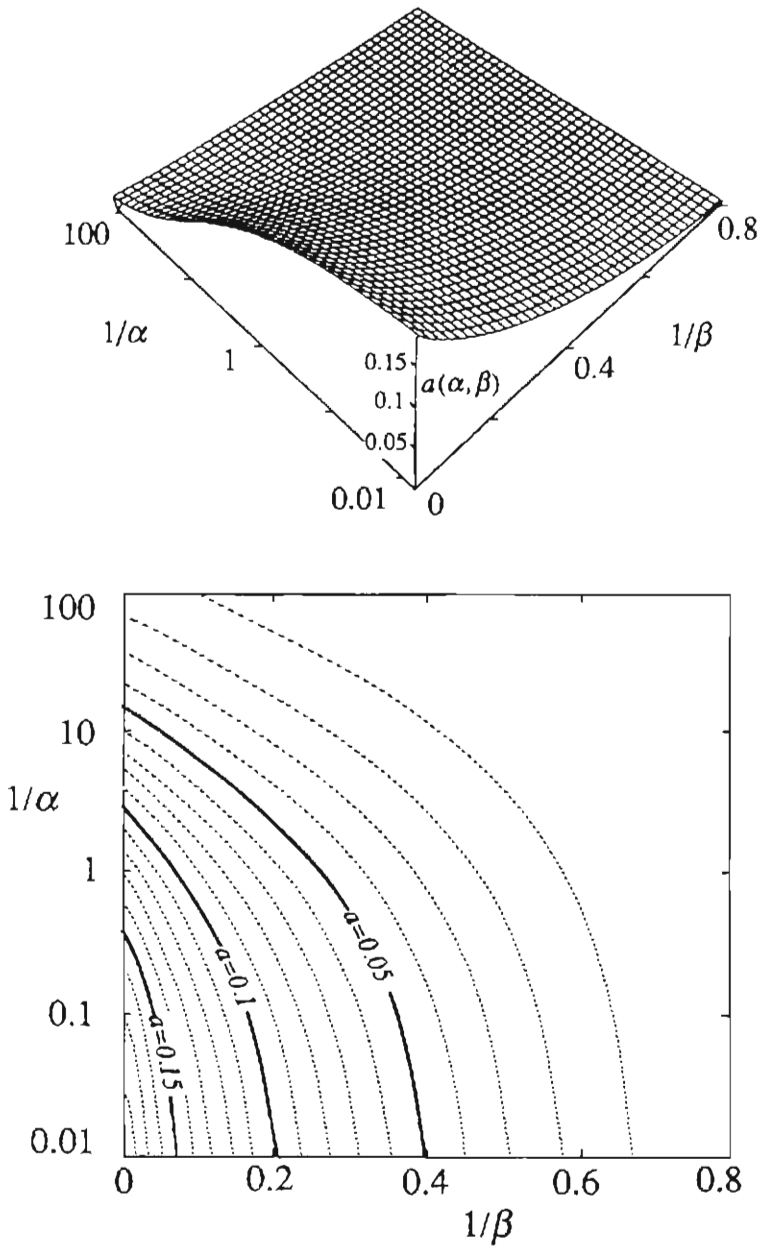


Fig. 7.4. Theoretical result for the scaling exponent $a(\alpha,\beta)$ as a function of inverse neuron gain $1/\beta$ and inverse storage ratio $1/\alpha$ for the analog Hebb-rule associative memory. The expected number of fixed points is $\langle N_{fp} \rangle \approx exp[a(\alpha,\beta)N]$.

matrices were generated for each of 6 values of N. For the spin glass, matrices were symmetric with gaussian distributed elements having zero mean and variance $1/N$. For the neural network, the matrices were constructed using the Hebb rule, Eq. (7.29) with $\xi_i^\mu = \pm 1$ at random. The values of N were chosen so that the number of fixed points was roughly in the range 20 to 400.

The number of fixed points in each network was counted by choosing random initial conditions $x_i(0) = \pm 1$ and iterating the map

$$x_i(t+1) = tanh(\beta h_i(t)) \tag{7.53}$$

$$h_i(t) = \sum_{j<i} T_{ij} x_j(t+1) + \sum_{j>i} T_{ij} x_j(t) \tag{7.54}$$

(sequential updating) until convergence to a fixed point was reached. Recall that under sequential updating of state variables, a network with symmetric connections (and zero diagonals) will converge to a fixed point for all values of gain β.

For each realization, the search for new fixed points was terminated after 10^5 initial conditions or when no new fixed points had been found for 10^4 consecutive initial conditions and for every fixed point found, the inverse point ($x_i \rightarrow -x_i$ for all i) had also been found. Then, for each set of parameters (α,β,N), the mean $\overline{N_{fp}}$ and the variance of the observed number of fixed points (for the 20 realizations) were computed, and an experimental value for a, defined by the line

$$ln\left(\overline{N_{fp}}(N)\right) = aN + const., \tag{7.55}$$

was found by a weighted least-square fit. Note that we average the number of observed fixed points in each realization rather than averaging the logarithm of the number of fixed

points. This coincides with the analysis above: recall that in § 7.2, we calculated $ln\langle N_{fp}\rangle$, not $\langle ln(N_{fp})\rangle$. As we have argued, the two types of averages are expected to agree in the large-N limit for the fully connected network.

7.3.2. Numerical results for analog spin glass

Numerical results for the analog spin glass, obtained using the above procedure, are shown in Figs. 7.5 and 7.6. Notice that the data in Fig. 7.6 agree quite well with the analytical result for the analog spin glass, but are quite different from the TAP result of Bray and Moore [1980]. To our knowledge, there are no comparable data verifying the Bray and Moore curve away from $T = 1/\beta = 0$. This is due, in part, to the chaotic dynamics exhibited by the dynamical version of the TAP equations [Bray and Moore, 1979]. In particular, since few initial conditions terminate at fixed points for the TAP equations at finite β, it is extremely difficult to count fixed points numerically in even a single realization[5].

7.3.3. Numerical results for the neural network

Fixed point counts for the Hebb-rule neural network were performed at six points in the (α,β) plane. Results are shown in Figs. 7.7 and 7.8. In Fig. 7.8, the numerical data are presented along with the analytical results for α = 10, 1, and 0.1. The agreement is very good at larger values of α and β, and reasonably good - though outside the range of the error bars - for smaller α and β.

It is not clear why theory and numerics disagree for $a(\alpha,\beta) <\sim 0.05$. At small α,

[5] Nemoto and Takayama [1985] claimed to be investigating just this problem using a variation of the TAP equations which is guaranteed to converge and which has as its solutions a superset of the solutions of the TAP equations. Apparently, this work has not been published.

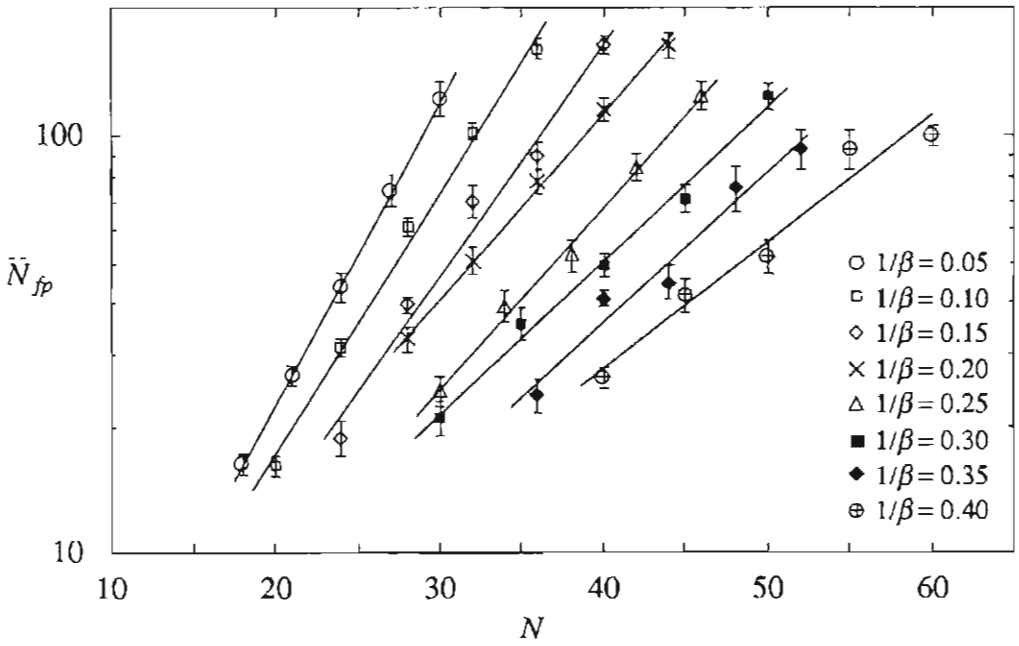


Fig. 7.5. Numerical counts of stable fixed points for the analog spin glass at several values of gain β, as a function of the system size N. Lines are weighted exponential fits to the data. Numerical values for the scaling exponent $a(\beta)$ are given by the slopes of the lines (using log-log scale).

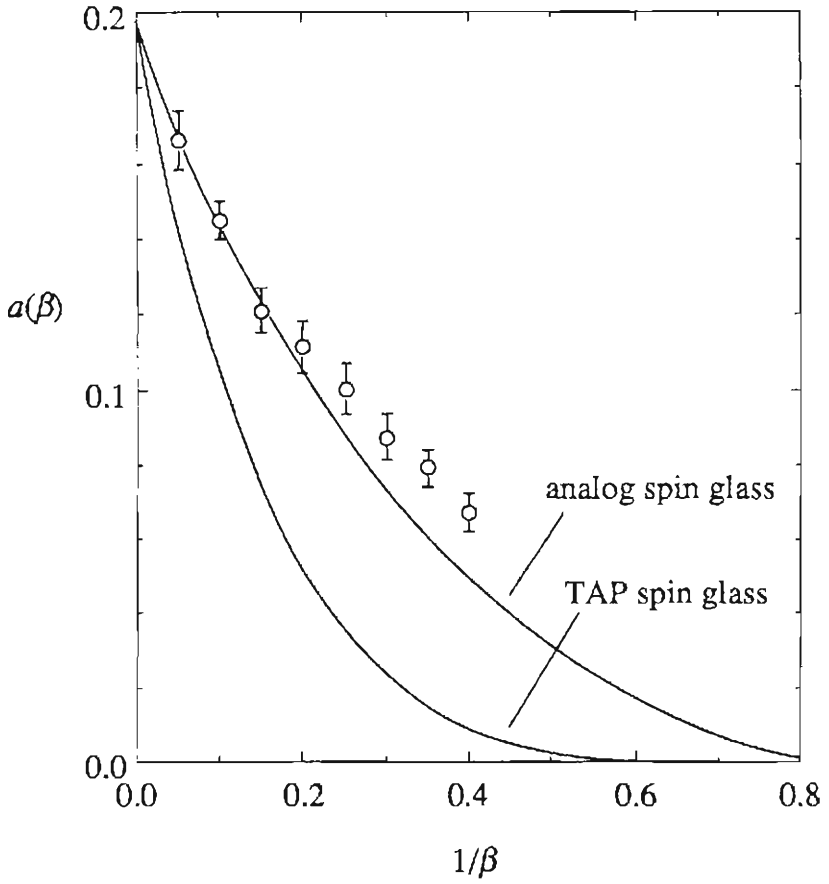


Fig. 7.6. Comparison of theoretical and numerical results for the scaling exponent $a(\beta)$ in the analog spin glass, as a function of the inverse neuron gain $1/\beta$. Note that the agreement is good, and that the numerical values clearly differ from the corresponding result for the TAP equations.

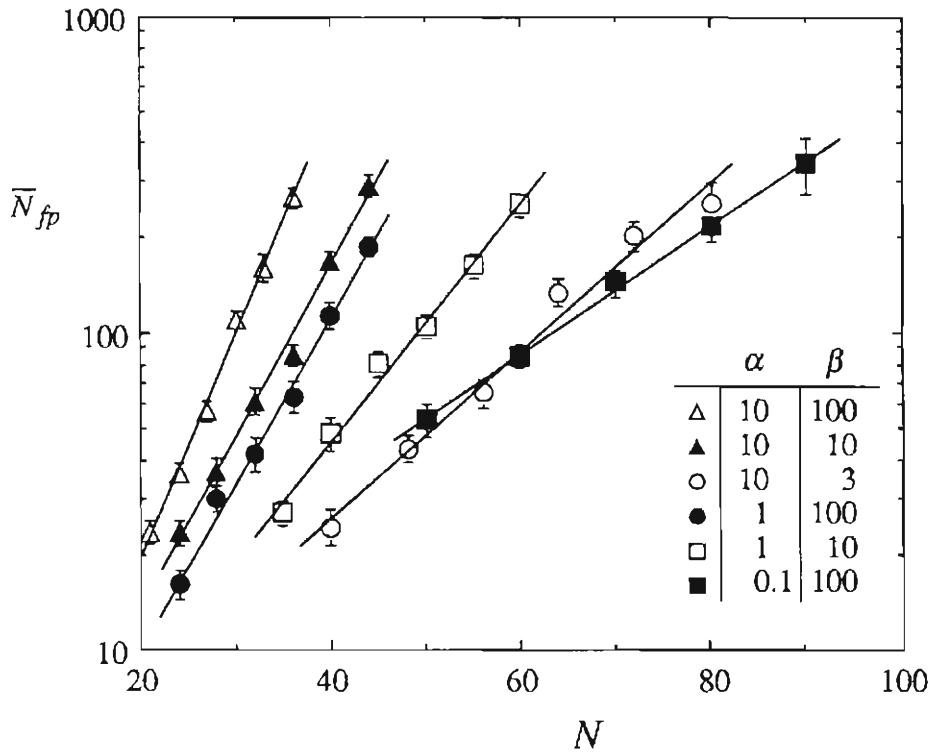


Fig. 7.7 Numerical counts of stable fixed points for the analog neural network for several values of α and β, as a function of system size N. Lines are weighted exponential fits to the data. Numerical values for the scaling exponent $a(\alpha,\beta)$ are given by the slopes of the lines (using log-log scale).

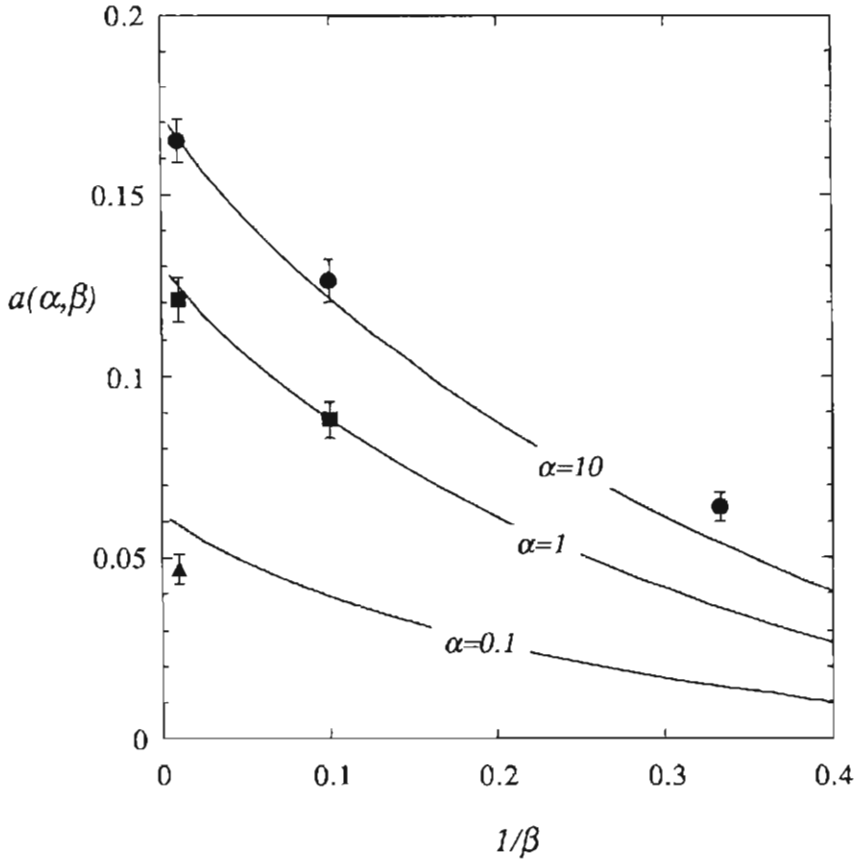


Fig. 7.8. Comparison of theoretical and numerical results for the scaling exponent $a(\alpha,\beta)$ in the analog associative memory. The agreement is good, especially at larger values of α and β.

the problem can likely be traced to the assumption in the analysis that $N\sqrt{\alpha} >> 1$, while numerical data for $\alpha = 0.1$ are limited to the range $15 < N\sqrt{\alpha} < 30$. Why smaller values of β should also cause problems - even for large α and for the analog spin glass - is unknown.

7.4 DISCUSSION

7.4.1 Asymmetry - An alternate way to eliminate spurious attractors

We have shown analytically and numerically that lowering the gain of the neuron transfer function in an analog spin glass or neural network greatly reduces the number of local minima in the energy landscape. This phenomenon provides one mechanism for the observed improvement of performance in analog neural networks compared to their binary-neuron counterparts. Because the present method of smoothing the landscape is fully deterministic, it can be implemented in electronic hardware much more easily than stochastic methods such as simulated annealing.

An alternate strategy for deterministically eliminating the spurious ("glassy") fixed points in a neural network is to add some degree of (quenched) asymmetry to the connection matrix. This idea has been considered by several authors [Parisi, 1986; Hertz *et al.*, 1987; Crisanti and Sompolinsky, 1987; Treves and Amit, 1988; Kepler, 1989]. In a deterministic Ising spin glass, for example, asymmetry reduces the number of fixed points (Fig. 7.9a) as well as the total number of attractors (Fig.7.9b) [Gutfreund *et al.*, 1988]. For the fully asymmetric spin glass (k = 1 in Fig. 7.9), the number of fixed points is no longer exponentially increasing with N, and the total number of attractors is minimized.

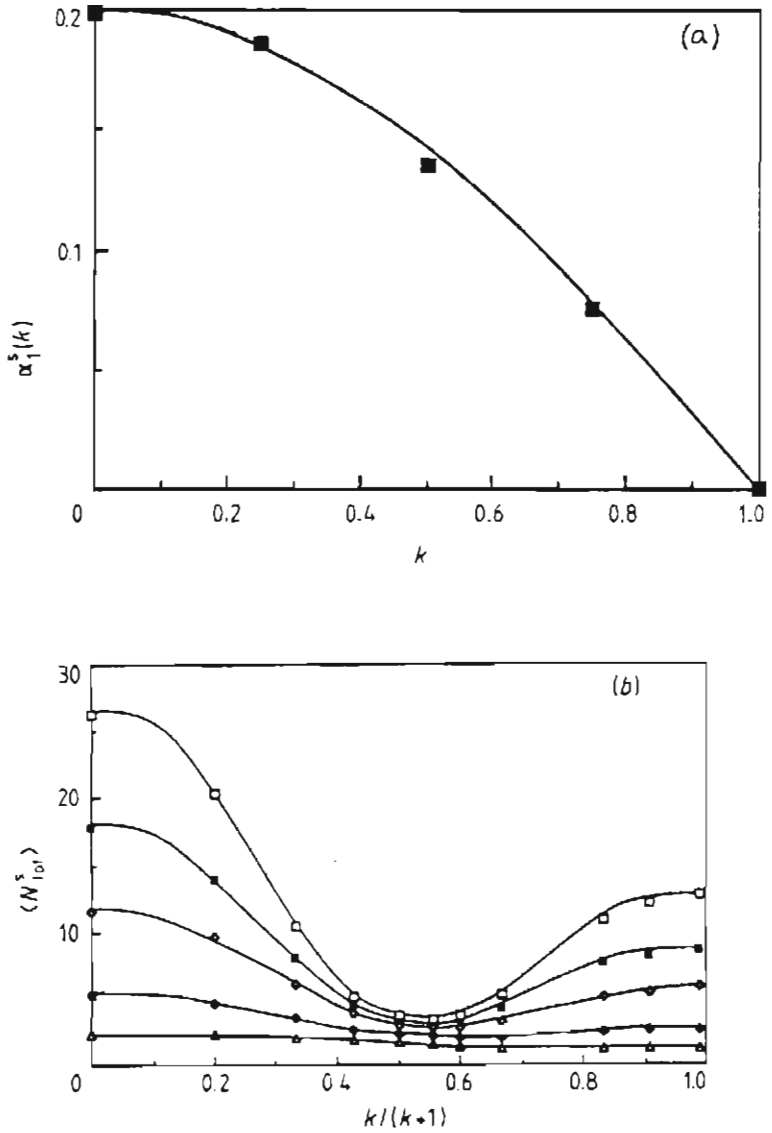


Fig. 7.9. (a) Scaling exponent $a(k)$ for fixed points and as a function of asymmetry parameter k for asymmetric spin glass with binary (Ising) state variables and sequential dynamics. Connection matrix is composed of gaussian random symmetric and antisymmetric parts, $T_{ij} = T_{ij}^S + kT_{ij}^A$. Line is theory, squares are from numerical counts. (b) Numerical counts of the total number of attractors as a function of asymmetry for various size systems. Both the number of fixed points and the total number of attractors can be reduced by introducing asymmetry, but this also introduces non-fixed-point attractors. After Gutfreund *et al.*[1988].

In the limit $N \to \infty$, all attractors of finite period disappear entirely; the only dynamical state for the fully asymmetric analog spin glass (besides the origin at low gain) is chaos [Sompolinsky *et al.*, 1988].

Treves and Amit [1989] have studied the distribution of fixed points in Hebb-rule neural network (with binary neurons) for arbitrary symmetry and dilution. They find that, in contrast to the spin glass, a neural network at finite α has an exponential number of fixed-point attractors for *all* values of asymmetry - including full asymmetry. Kepler [1989] has taken an additional step to eliminate these remaining spurious fixed points by adding a self-inhibition term $T_{ii} < 0$, which eliminates fixed points, and (for parallel dynamics) creates in their place period-2 limit cycles. Kepler's result, as well as Fig. 7.9, highlights an important drawback of smoothing with asymmetry: *Using asymmetry to reduce the number of fixed point attractors necessarily creates new, non-fixed-point attractors*. In contrast, smoothing the landscape using analog neurons eliminates fixed point attractors without introducing any new attractors.

7.4.2. A short discussion of attractors in multistep systems

Finally, we will briefly discuss the nature of the attractors in the multistep updating rule defined in Ch. 6. We will restrict our attention to the binary (Ising) spin glass (T_{ij} gaussian random symmetric), and compare a sequential update scheme to the $M = 2$ updating rule, where state variables are updated in parallel based on the average of two previous time steps. We expect that our observations will be qualitatively correct for analog systems and neural networks.

First - to check the validity of our numerical method -we reproduce the well-known result that the Ising spin glass under single-time-step sequential dynamics always converges to a fixed point and that the expected number of fixed points is

$\langle N_{fp} \rangle = \exp[aN]$ with $a = 0.1992...$[Tanaka and Edwards, 1980; De Dominicis *et al.*, 1980; Bray and Moore, 1980; Gutfreund *et al.*, 1988]. This result is shown in Fig. 7.10. The numerical method is similar to the one described above: Here, 40 random gaussian matrices were generated for each value of N with $8 \leq N \leq 18$. For each matrix, random initial states (random N-vectors of ± 1's) were generated and their associated attractors were found using sequential dynamics:

$$x_i(t+1) = Sgn\left[\sum_{j<i} T_{ij} x_j(t+1) + \sum_{j>i} T_{ij} x_j(t)\right]. \tag{7.56}$$

Fixed points for each matrix were counted and tabulated. A particular matrix was considered fully mined for attractors when 500 consecutive initial states were tested without finding a new attractor. (This occurred after anywhere from 502 to several thousand initial states had been examined). The numbers of fixed points for each of the 400 matrices were plotted and a direct, least-square exponential fit to all 400 points was made (using KaleidaGraph 2.0) to find the scaling exponent a. This method of averaging is not the same as finding the average number of fixed points for each value of N first, and then doing an exponential fit. The two methods, however, yield very similar results. The least-square fit gives $a = 0.2030$, within 2% of the theoretical value.

Now we turn to the $M = 2$ multistep update rule. Recall that in § 6.2.2 it was proved that all attractors for this update rule (with symmetric connections) are either fixed points or period-3 limit cycles. This result is supported by the present numerical investigation. Counts of fixed points and 3-cycles are shown in Fig. 7.11. The fixed points and 3-cycle were counted by the same method described above (40 matrices for each N, $8 \leq N \leq 18$ and 4 matrices for $N = 19$). In this case, initial states were generated until no new attractors of *either* type had been found for 500 consecutive initial conditions.

Because there were typically a large number of 3-cycles, this often meant checking tens of thousands of initial conditions per matrix. Care was taken not to overcount 3-cycles that are identical under cyclic permutation, though noncyclic permutations were counted as separate attractors. We point out three important features of the data in Fig. 7.11:

(1) The scaling exponent for the number of stable fixed points using 2-step dynamics is significantly below 0.1992. The measured value is $a_{fp} = 0.175$.

(2) The scaling exponent for stable 3-cycles is larger than that of the fixed points (for either type of dynamics), and is measured to be $a_{3\text{-}cyc} = 0.237$. Thus 3-cycles are quite abundant in the 2-step network - much more so than fixed points. On the other hand, the 3-cycles are not as abundant as the 2-cycles generated by standard (M = 1) parallel updating, which are known to have a scaling exponent of $a_{2\text{-}cyc} = 2(0.1992) = 0.3984$ [Gutfreund *et al.*, 1988, Cabasino *et al.*, 1988].

(3) The number of 3-cycles seems *not* to be self-averaging. For a given system size, the number of 3-cycles found in particular realizations varies by up to two orders of magnitude, with $\left(N_{fp}^{max} - N_{fp}^{min}\right) / \overline{N}_{fp} \sim O(1)$.

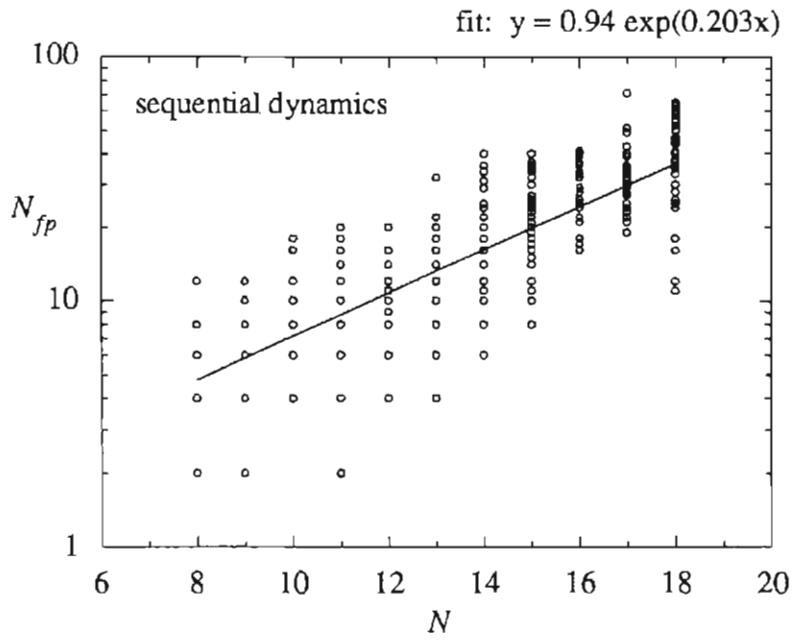


Fig. 7.10. Number of fixed points N_{fp} as a function of system size N for the Ising spin glass with single-step sequential dynamics. See text for details of the numerical method. The direct exponential fit to all 400 points (line) gives a scaling exponent of $a = 0.203$, in good agreement with the theoretical value of 0.1992.

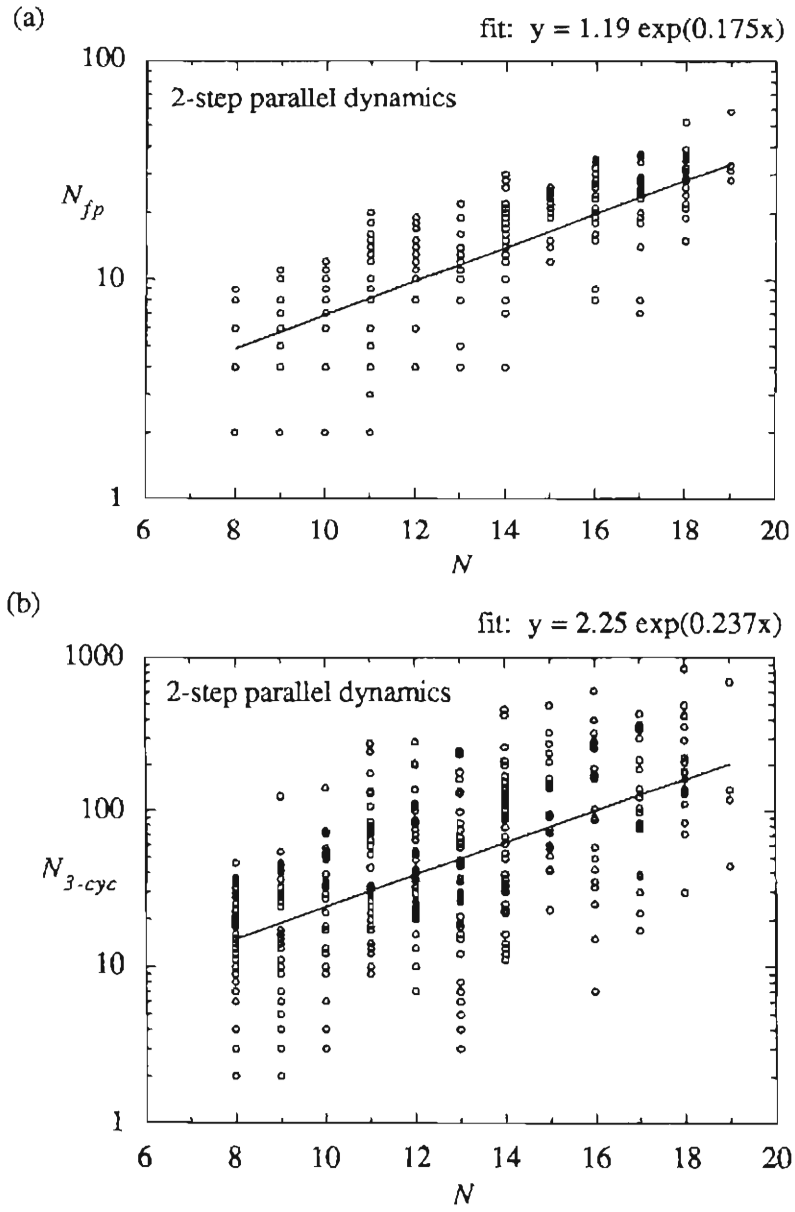


Fig. 7.11. Number of fixed points N_{fp} and distinct 3-cycles $N_{3\text{-}cyc}$ as a function of system size N for the Ising spin glass with 2-step parallel dynamics (see § 6.6.2). See text for details of the numerical method. (a) The numerically determined fixed-point scaling exponent is a_{fp}= 0.175. This value is significantly less than the sequential update value of 0.1992 (see Fig. 7.10). (b) The numerically determined 3-cycle scaling exponent is $a_{3-cyc} = 0.237$. This value is larger than a_{fp}, so at large N the vast majority of attractors for the 2-step spin glass are 3-cycles.

APPENDIX 7A: $\langle det\ \mathbf{A}\rangle_{\mathbf{T}}$ FOR THE ANALOG SPIN GLASS

In this appendix we calculate the average (over realizations) of the determinant of the Hessian matrix **A**, defined in Eq. (7.8). The method follows Bray and Moore [1980]. Replicas were introduced in the main text; we pick up the calculation starting at Eq. (7.23).

From Eq. (7.23), the average over realizations is given by

$$\langle det\ \mathbf{A}\rangle_{\mathbf{T}} = \lim_{m\to -2} \left\langle \int_{-\infty}^{\infty} \prod_{i,\alpha} \frac{d\rho_{i\alpha}}{\sqrt{2\pi}} exp\left(-\frac{1}{2} \sum_{i,j} \sum_{\alpha=1}^{m} \rho_{i\alpha} A_{ij} \rho_{j\alpha} \right) \right\rangle_{\mathbf{T}} . \tag{7A.1}$$

Writing the average in (7A.1) as an integral over the gaussian distribution of matrix elements (7.6),

$$\langle det\ \mathbf{A}\rangle_{\mathbf{T}} = \lim_{m\to -2} \left\{ \int_{-\infty}^{\infty} \prod_{i,\alpha} \frac{d\rho_{i\alpha}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \prod_{(ij)} dT_{ij} \left(\frac{N}{2\pi J^2} \right)^{1/2} exp\left[\sum_{(ij)} -\left(\frac{N}{2J^2} \right) T_{ij}^{\,2} \right] \times \right.$$

$$\left. exp\left[-\frac{1}{2} \sum_{i,j} \sum_{\alpha}^{m} \rho_{i\alpha} \left(g'(x_i)\delta_{ij} - \beta T_{ij} \right) \rho_{j\alpha} \right] \right\} , \tag{7A.2}$$

with (ij) indicating distinct pairs. Integrals over T_{ij} are gaussian and can be integrated to give

$$\langle det\ \mathrm{A}\rangle_{\mathrm{T}} = \underset{m\to -2}{Lim} \left\{ \int_{-\infty}^{\infty} \prod_{i,\alpha} \frac{d\rho_{i\alpha}}{\sqrt{2\pi}}\ exp\left[\left(\frac{J^2\beta^2}{4N} \right) \sum_{i,j} \left(\sum_{\alpha} \rho_{i\alpha}\rho_{j\alpha} \right)^2 \right] \times \right.$$

$$\left. exp\left[-\frac{1}{2} \sum_{i,\alpha} g'(x_i){\rho_{i\alpha}}^2 \right] \right\} \qquad (7A.3)$$

where now the double sums over i and j are unrestricted. The squared sum can be reduced using a Hubbard-Stratonovich transformation (7.17), which introduces two new order parameters, an m-component vector R_α and an $m \times m$ matrix $M_{\alpha\beta}$,

$$exp\left[\left(\frac{J^2\beta^2}{4N} \right) \sum_{i,j} \left(\sum_{\alpha} \rho_{i\alpha}\rho_{j\alpha} \right)^2 \right]$$

$$= exp\left[\frac{J^2\beta^2}{4N} \sum_{\alpha} \left(\sum_{i} {\rho_{i\alpha}}^2 \right)^2 \right]\ exp\left[\frac{J^2\beta^2}{4N} \sum_{\alpha,\beta<\alpha} \left(\sum_{i} \rho_{i\alpha}\rho_{i\beta} \right)^2 \right]$$

$$= (term\ A)(term\ B) \qquad (7A.4)$$

where

$$(term\ A) = \left(\frac{N}{\pi} \right)^{1/2} \int_{-\infty}^{\infty} \prod_{\alpha} dR_\alpha\ exp\left[-N \sum_{\alpha} {R_\alpha}^2 + J\beta \sum_{\alpha} \left(\sum_{i} {\rho_{i\alpha}}^2 \right) R_\alpha \right] \qquad (7A.5)$$

$$(term\ B) = \left(\frac{N}{2\pi} \right)^{1/2} \int_{-\infty}^{\infty} \prod_{\alpha,\beta<\alpha} dM_{\alpha\beta}\ exp\left[-\frac{N}{2} \sum_{\alpha,\beta<\alpha} {M_{\alpha\beta}}^2 + J\beta \sum_{\alpha,\beta<\alpha} \left(\sum_{i} \rho_{i\alpha}\rho_{i\beta} \right) M_{\alpha\beta} \right] \qquad (7A.6)$$

Following Bray and Moore [1980], we adopt the replica symmetric solutions $R_\alpha = R$ for

all α, and $M_{\alpha\beta} = 0$ for all (α,β), which reduces (7A.5) and (7A.6) to

$$\begin{aligned} (term\,A) &= \left(\frac{N}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} dR\, exp\left(-NmR^2\right) exp\left[J\beta R \sum_i {\rho_{i\alpha}}^2 \right]^m, \\ (term\,B) &= 1. \end{aligned} \tag{7A.7}$$

Inserting (7A.7) into (7A.3) gives

$$\begin{aligned} \langle det\ \mathbf{A} \rangle_{\mathbf{T}} = \underset{m\to -2}{Lim} \Bigg\{ \int_{-\infty}^{\infty} \prod_{i,\alpha} \frac{d\rho_{i\alpha}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dR \left(\frac{N}{\pi}\right)^{1/2} exp\left[-NmR^2\right] \times \\ exp\left[-\frac{1}{2} \sum_i \left(g'(x_i) - 2\beta JR \right) \sum_\alpha {\rho_{i\alpha}}^2 \right] \Bigg\}. \end{aligned} \tag{7A.8}$$

Integrals over $\rho_{i\alpha}$ are now gaussian and can be integrated to give

$$\begin{aligned} \langle det\ \mathbf{A} \rangle_{\mathbf{T}} &= \underset{m\to -2}{Lim} \left\{ \int_{-\infty}^{\infty} dR \left(\frac{N}{\pi}\right)^{1/2} exp\left[-NmR^2\right] \left[\prod_i \left(g'(x_i) - 2\beta JR \right)^{-1/2} \right]^m \right\} \\ &= \int_{-\infty}^{\infty} dR \left(\frac{N}{\pi}\right)^{1/2} exp\left[2NR^2\right] \prod_i \left(g'(x_i) - 2\beta JR \right). \end{aligned} \tag{7A.9}$$

The integral over R in (7A.9) is eventually done by steepest descent. Because the number of replicas has been set to -2, the correct solution of this steepest decent integral turns out to be a minimum over R, rather than a maximum. This is also the case in the analysis of Bray and Moore [1980] as they use the solution $B = 0$, which is likewise a minimum of *det* **A**. Neglecting numerical prefactors of O(1), we obtain the following result:

$$\langle det\ \mathbf{A} \rangle_{\mathrm{T}} = \underset{R}{Min} \left[exp\left[2NR^2 \right] \prod_i \left(g'(x_i) - 2\beta JR \right) \right], \tag{7A.10}$$

which appears as Eq. (7.24).

APPENDIX 7B: EXPANSIONS FOR STEEPEST DESCENT INTEGRALS

Finding a solution of the saddle-point equations for the analog spin glass (7.27) or the analog neural network (7.51) requires solving a set of four coupled equations, each of which contains an integral in the form of the double angle brackets defined by (7.28). These integrals, five in total, must be evaluated numerically. For the analog transfer function $F(h_i) = tanh(\beta h_i)$, four out of the five integrands diverge at the endpoints of the domain of integration, ± 1, causing fatal problems for the numerical integration package used (NAGLIB D01AHF). To get around this problem, we split the domain of integration into three regions:

$$\int_{-1}^{1} = \int_{-1}^{-1+\varepsilon} + \int_{-1+\varepsilon}^{1-\varepsilon} + \int_{1-\varepsilon}^{1} \quad , \tag{7B.1}$$

which can be evaluated as

$$\int_{-1}^{1} = \int_{-1+\varepsilon}^{1-\varepsilon} + 2\int_{1-\varepsilon}^{1} \tag{7B.2}$$

by virtue of the (even) symmetry of all of the integrands. The integrals with limits at $\pm(1-\varepsilon)$ no longer have divergent integrands over this reduced domain, and can be evaluated accurately using the NAGLIB integration package. The remaining parts, extending over $[1-\varepsilon,1]$, can be approximated for $\varepsilon << 1$, keeping only those terms

which diverge as $x \to 1$. Once nonleading terms are dropped, the $[1-\varepsilon,1]$ integrals can be evaluated in closed form, giving expressions which depend on ε. The procedure is straightforward, and we give only the results. The approximations were checked by comparing several values of ε, ranging from 10^{-4} to 10^{-8}, and confirming that the *sum* of the two integrals on the right of (7B.2) was insensitive to the choice of ε, though the individual parts of each integral did depend on ε.

We have suppressed the "+" markers on the integrals indicating that the range of integration is limited to a sub-region where $\langle det\, \mathbf{A} \rangle > 0$. The excluded region is in the center of state space - covered by the "easy" integral over [-1+ε, 1-ε] which is done numerically.

First, we consider the integral $\tilde{I}$ of the weight function $W(x)$ defined in (7.28b) for the case of the neuron transfer function $F(h_i) = tanh(\beta h_i)$:

$$\tilde{I} = \int_{-1}^{1} dx\, W(x) \quad , \tag{7B.3}$$

$$W(x) = \left(\frac{1}{1-x^2} + B \right) exp\left[-\frac{\left(tanh^{-1}(x) - \Delta x \right)^2}{2\beta^2 J^2 q} + \lambda x^2 \right] \quad . \tag{7B.4}$$

Notice that $\tilde{I}$ is proportional to the I's defined previously: $\tilde{I} = \left[\sqrt{2\pi q}\, \beta J \right] I$ from (7.26) and $\tilde{I} = \left[\sqrt{2\pi q}\, \beta \right] \hat{I}$ from (7.49). Expanding (7B.3) near ± 1 as descibed above gives

$$\tilde{I} \cong \int_{-1+\varepsilon}^{1-\varepsilon} dx\, W(x) + \frac{b\sqrt{\pi}}{2} e^{\lambda}\, erfc[a/b] \, , \tag{7B.5}$$

where

$$a \equiv ln(2/\varepsilon) - 2\Delta\,, \tag{7B.6a}$$

$$b \equiv 2\beta J\sqrt{2q}\,, \tag{7B.6b}$$

and *erfc* is the complementary error function,

$$erfc(z) \;=\; \frac{2}{\sqrt{\pi}}\int_z^\infty dt\; e^{-t^2}\;. \tag{7B.7}$$

Using these same definitions of a and b, the integrals in double brackets from (7.27) and (7.51) have the following expansions:

$$\begin{aligned}\left\langle\!\left\langle\, x^2 \,\right\rangle\!\right\rangle &= \frac{1}{\bar{I}}\int_{-1}^{1} dx\, x^2\, W(x) \\ &\cong \frac{1}{\bar{I}}\left[\int_{-1+\varepsilon}^{1-\varepsilon} dx\, x^2\, W(x) \;+\; \frac{b\sqrt{\pi}}{2} e^{\lambda}\; erfc[a/b]\right]\,,\end{aligned} \tag{7B.8}$$

$$\begin{aligned}\left\langle\!\left\langle\, x\, g(x) \,\right\rangle\!\right\rangle &= \frac{1}{\bar{I}}\int_{-1}^{1} dx\left(x \tanh^{-1}(x)\right) W(x) \\ &\cong \frac{1}{\bar{I}}\left[\int_{-1+\varepsilon}^{1-\varepsilon} dx\left(x \tanh^{-1}(x)\right) W(x)\right. \\ &\qquad \left. +\; \frac{e^{\lambda b}}{2}\left(\frac{b}{2} exp\left(-a^2/b^2\right) + \Delta\sqrt{\pi}\; erfc(a/b)\right)\right]\,,\end{aligned} \tag{7B.9}$$

$$\begin{aligned}\left\langle\!\left\langle\, g(x) - \Delta x \,\right\rangle\!\right\rangle &= \frac{1}{\bar{I}}\int_{-1}^{1} dx\left(\tanh^{-1}(x) - \Delta x\right) W(x) \\ &\cong \frac{1}{\bar{I}}\left[\int_{-1+\varepsilon}^{1-\varepsilon} dx\left(\tanh^{-1}(x) - \Delta x\right) W(x)\right. \\ &\qquad \left. +\; \left(\frac{b}{2}\right)^3 e^{\lambda}\left(\left(\frac{a}{b}\right) exp\left(-a^2/b^2\right) + \frac{\sqrt{\pi}}{2}\; erfc(a/b)\right)\right]\,.\end{aligned} \tag{7B.10}$$

Finally, the integrand in $\left\langle\!\left\langle \left(g'(x)+B\right)^{-1} \right\rangle\!\right\rangle$ does not blow up at ±1 for $F(h_i)$ = $tanh(\beta h_i)$, so it is not necessary to expand the integrand in order to numerically evaluate this integral.

APPENDIX 7C: $\langle det\ \mathbf{A}\rangle_\xi$ FOR THE ANALOG NEURAL NETWORK

In this appendix, we derive the expression (7.46) for $\langle det\ \mathbf{A}\rangle_\xi$, the average determinant of the Hessian matrix A, defined in (7.8). The average of *det* A is taken over realizations of Hebb matrices, defined in Eq.(7.29), each storing αN unbiased and uncorrelated random memory patterns. That is, each ξ_i^μ, i = 1, ..., N; μ= 1, ..., αN in (7.29) equals ±1 at random.

We start with the identity (7.22),

$$(det\ \mathbf{A})\left\{\prod_i \theta[\lambda_i(\mathbf{A})]\right\} = \left[\int_{-\infty}^{\infty}\prod_i \frac{d\rho_i}{\sqrt{2\pi}} exp\left(-\frac{1}{2}\sum_{i,j}\rho_i A_{ij}\rho_j\right)\right]^{-2} \tag{7C.1}$$

and introduce replicas, indexed by γ, with the number of replicas eventually set to -2,

$$det\ \mathbf{A} = \lim_{m\to -2}\int_{-\infty}^{\infty}\prod_{i,\gamma}\frac{d\rho_{i\gamma}}{\sqrt{2\pi}} exp\left(-\frac{1}{2}\sum_{i,j}\sum_{\gamma=1}^{m}\rho_{i\gamma}A_{ij}\rho_{j\gamma}\right). \tag{7C.2}$$

As with the analog spin glass, we drop the absolute value brackets around the determinant, recognizing that (7C.1) is nonzero only when A is positive definite, which implies *det* A > 0. The claim that (7C.2) picks out only the stable fixed points after averaging is only valid insofar as replica symmetry is valid. The validity of replica symmetry in this problem will not be studied.

We write the average over ξ_i^μ as

$$\langle det\,\mathbf{A}\rangle_\xi = \underset{m\to -2}{Lim} \left\langle \int_{-\infty}^{\infty} \prod_{i,\gamma} \frac{d\rho_{i\gamma}}{\sqrt{2\pi}} exp\left(-\frac{1}{2} \sum_{i,j} \sum_{\gamma=1}^{m} \rho_{i\gamma} A_{ij} \rho_{j\gamma} \right) \right\rangle_\xi \tag{7C.3}$$

where the angle brackets denote an average over all $2^{\alpha N^2}$ states of ξ_i^μ.
Note in (7C.3) that averaging is done before the number of replicas is set to -2. Inserting A_{ij} from (7.8) and T_{ij} from (7.29) into (7C.3) gives

$$\langle det\ \mathbf{A}\rangle_\xi = \underset{m\to -2}{Lim} \left\langle \int_{-\infty}^{\infty} \prod_{i,\gamma} \frac{d\rho_{i\gamma}}{\sqrt{2\pi}} exp\left[-\frac{1}{2} \sum_{i,\gamma} \rho_{i\gamma}{}^2 \left(g'(x_i) + \beta\sqrt{\alpha} \right) \right.\right.$$
$$\left.\left. + \frac{1}{2} \frac{\beta}{N\sqrt{\alpha}} \sum_{\gamma,\mu} \left(\sum_i \rho_{i\gamma}\ \xi_i^\mu \right)^2 \right] \right\rangle_\xi \tag{7C.4}$$

The square in the last term of (7C.4) can be reduced to a linear form via a Hubbard-Stratonovich transformation,

$$exp\left(\frac{\lambda a^2}{2} \right) = \left(\frac{\lambda}{2\pi} \right)^{1/2} \int_{-\infty}^{\infty} dx\, exp\left[-\frac{\lambda x^2}{2} + a\lambda x \right] \tag{7C.5}$$

[with λ= 1 and $a = \left(\beta/N\sqrt{\alpha}\right)^{1/2} \sum_i \xi_i^\mu \rho_{i\gamma}$ in this case]. This introduces a new set of integration variables, $\sigma_{\gamma\mu}$ (γ= 1, ..., m; μ = 1, ..., αN) and gives

$$\langle det\,\mathbf{A}\rangle_\xi = \lim_{m\to -2} \int_{-\infty}^{\infty} \prod_{\gamma,\mu} \frac{d\sigma_{\gamma\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \prod_{i,\gamma} \frac{d\rho_{i\gamma}}{\sqrt{2\pi}} \times$$

$$exp\left[-\frac{1}{2} \sum_{i,\gamma} \rho_{i\gamma}{}^2 \left(g'(x_i) + \beta\sqrt{\alpha} \right) - \frac{1}{2} \sum_{\gamma,\mu} \sigma_{\gamma\mu}{}^2 \right] \times$$

$$\left\langle exp\left[\left(\frac{\beta}{N\sqrt{\alpha}} \right)^{1/2} \sum_{i,\gamma,\mu} \rho_{i\gamma}\, \sigma_{\gamma\mu}\, \xi_i^\mu \right] \right\rangle_\xi . \qquad (7C.6)$$

Averaging over the ξ_i^μ can now be done immediately using the relation

$$\left\langle exp\left(a\,\xi_i^\mu \right) \right\rangle_\xi = cosh(a) \xrightarrow[a<<1]{} exp\left(a^2/2 \right) + O\left(a^4 \right) \qquad (7C.7)$$

The term corresponding to a in (7C.6) is small for large N, so that the $O(a^4)$ terms can be dropped:

$$\left\langle exp\left[\left(\frac{\beta}{N\sqrt{\alpha}} \right)^{1/2} \sum_{i,\gamma,\mu} \rho_{i\gamma}\, \sigma_{\gamma\mu}\, \xi_i^\mu \right] \right\rangle_\xi \rightarrow exp\left[\frac{\beta}{2N\sqrt{\alpha}} \sum_{i,\mu} \left(\sum_\gamma \rho_{i\gamma}\, \sigma_{\gamma\mu} \right)^2 \right] . \qquad (7C.8)$$

We now assume replica symmetry by setting $\sigma_{\gamma\mu} = \sigma_\mu$ and $\rho_{i\gamma} = \rho_i$ for all γ. This allows (7C.6) to be written as a single-site integral (in replica space) raised to the power m:

$$\langle det\,\mathbf{A}\rangle_\xi = \lim_{m\to -2} \left\{ \int_{-\infty}^{\infty} \prod_{\mu} \frac{d\sigma_\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \prod_{i} \frac{d\rho_i}{\sqrt{2\pi}}\; exp\left[-\frac{1}{2} \sum_i \rho_i{}^2 \left(g'(x_i) + \beta\sqrt{\alpha} \right) \right] \times \right.$$

$$\left. exp\left[-\frac{1}{2} \sum_\mu \sigma_\mu{}^2 + \frac{\beta}{2\sqrt{\alpha}} \left(\frac{1}{N} \sum_i \rho_i{}^2 \right) \left(\sum_\mu \sigma_\mu{}^2 \right) \right] \right\}^m . \qquad (7C.9)$$

Next, we introduce the order parameter

$$r = \frac{1}{N}\sum_i \rho_i{}^2 \tag{7C.10}$$

and its conjugate field R via an integral definition of 1:

$$1 = \frac{N}{2\pi i}\iint dR\,dr\ exp\left[R\left(\sum_i \rho_i{}^2 - Nr\right)\right]. \tag{7C.11}$$

Inserting (7C.10) and (7C.11) into (7C.9) gives

$$\langle det\,\mathbf{A}\rangle_\xi = \lim_{m\to -2}\left\{\left(\frac{N}{2\pi i}\right)\iint dr\,dR\int_{-\infty}^{\infty}\prod_\mu \frac{d\sigma_\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\prod_i \frac{d\rho_i}{\sqrt{2\pi}}\ \ exp[-NRr]\times\right.$$

$$\left. exp\left[-\frac{1}{2}\sum_i \rho_i{}^2\left(g'(x_i)+\beta\sqrt{\alpha}-2R\right)\right]\ exp\left[-\frac{1}{2}\left(1-\frac{\beta r}{\sqrt{\alpha}}\right)\sum_\mu \sigma_\mu{}^2\right]\right\}^m, \tag{7C.12}$$

which, after gaussian integration of σ_μ and ρ_i, yields

$$\langle det\,\mathbf{A}\rangle_\xi = \lim_{m\to -2}\left\{\left(\frac{N}{2\pi i}\right)\iint dr\,dR\ exp\left[N\left(-rR-\frac{\alpha}{2}ln\left(1-\frac{\beta r}{\sqrt{\alpha}}\right)\right)\right]\times\right.$$

$$\left.\prod_i\left(g'(x_i)+\beta\sqrt{\alpha}-2R\right)^{-1/2}\right\}^m. \tag{7C.13}$$

The integrals over r and R are evaluated by steepest descent, which is justified for large

N. Each integral produces a factor $(2\pi/N)^{1/2}$ to cancel the existing prefactor proportional to N. Other numerical prefactors of O(1) will be ignored. The steepest descent integral over r can be done explicitly by setting

$$\frac{\partial}{\partial r}\left[rR+\frac{\alpha}{2}ln\left(1-\beta r/\sqrt{\alpha}\right)\right]=0 \tag{7C.14}$$

which gives

$$r=\left(\frac{\sqrt{\alpha}}{\beta}-\frac{\alpha}{2R}\right). \tag{7C.15}$$

The steepest descent integral over R will eventually be done numerically. Setting the number of replicas to -2 *before* performing the saddle point integral makes the minimum (not the maximum) with respect to R the valid solution. Inserting (7C.15) into (7C.13) and setting m = -2 yields the desired result,

$$\langle det\,\mathbf{A}\rangle_{\xi}=\underset{R}{Min}\left\{exp\left[N\left(\frac{2\sqrt{\alpha}R}{\beta}-\alpha+\alpha\,ln\left(\frac{\beta\sqrt{\alpha}}{R}\right)-\alpha\,ln(2)\right)\right]\times\right.$$

$$\left.\prod_i\left(g'(x_i)+\beta\sqrt{\alpha}-2R\right)\right\} \tag{7C.16}$$

which appears as Eq. (7.46).

Chapter 8

THE DISTRIBUTION OF BASIN SIZES IN THE SK SPIN GLASS

8.1. INTRODUCTION: BACK TO BASINS

In this chapter, we return to the problem considered in Ch. 3, namely, the structure of the basins of attraction in systems with a large number of attractors. We restrict our attention to the zero-temperature SK spin glass [Sherrington and Kirkpatrick, 1975], and study the distribution of basin sizes, averaged over the ensemble of (gaussian) random connection matrices.

The dynamical system we consider is the SK model with deterministic, discrete-time (sequential) dynamics:

$$S_i(t+1) = Sgn\left[\sum_{j<i} T_{ij} S_j(t+1) + \sum_{j>i} T_{ij} S_j(t)\right] \qquad i = 1, ..., N. \qquad (8.1)$$

The state space of this system is discrete, $S_i = \pm 1$, equivalent to the corners of an N-dimensional hypercube. Later in the chapter (§ 8.5) we will consider an analog version of Eq. (8.1). The connection matrix $\mathbf{T} = \{T_{ij}\}$ is taken to be symmetric ($T_{ij} = T_{ji}$) with off-diagonal elements drawn at random from a gaussian distribution $P(T_{ij})$ with zero mean and variance $1/N$,

$$P(T_{ij}) = \left(\frac{N}{2\pi}\right)^{1/2} exp\left[-\frac{N}{2}T_{ij}^{2}\right], \qquad (8.2)$$

and all $T_{ii} = 0$.

The main result of this chapter is that over a wide range of basin sizes W, the numerically-measured distribution $f(W)$ of basin sizes for Eq. (8.1) is roughly described by a power law, $f(W) = KW^{-\gamma}$, with $\gamma \sim 3/2$. (These quantities are defined precisely below, see (8.4) and (8.7)). After exploring some of the immediate consequences of such a power-law distribution (§ 8.3), we will compare this result to known basin-size distributions for other systems, and to a closely related quantity defined for the SK spin glass, the distribution of *cluster weights* (defined in § 8.4). Finally, we show that using analog state variables selectively eliminates fixed points with small basins as analog gain is lowered (§ 8.5). A discussion and open questions are presented at the end (§ 8.6).

With all of the attention that has been paid to the SK spin glass over the past fifteen years [Binder and Young, 1986; Mezard *et al.*, 1987], it is surprising that a direct measurement of basin sizes has not been presented previously.[1] Two explanations for this lacuna seem likely: First, there is the well-known "universality" of systems with multivalley energy landscapes [Derrida and Flyvbjerg, 1987a, 1987b; Gutfreund, 1988; Derrida, 1988b]. This universality has prompted comparisons of different quantities in different systems, all of which characterize state space in some way. For some models - such as the Kauffman model (§ 8.4.2) - the quantity used for comparison is in fact the

[1]N. Parga and G. Parisi have studied a related distribution in the T=0 SK model using a numerical technique very similar to ours. Rather than looking at the distribution of basin sizes, they measured the fraction of initial states that terminate at a fixed point with energy between E and $E + \Delta E$, as a function of E. They found that as the system becomes large, most initial states flow to fixed points with $E/N \sim 0.7$, with a rather narrow peak. Apparently, this work has not been published except as a preprint [Triest preprint IC/85/133 (1985)]. Some aspects of the work are discussed in Parga [1987], and a figure from the preprint appears in the book by Chowdhury [1986, p. 82-83].

distribution of basin sizes. For the SK spin glass, it is another quantity, the distribution of cluster weights which is generally used for comparison. An exact expression for the distribution of cluster weights has been presented and analyzed thoroughly [Mezard *et al.*, 1984a; 1984b]. Our results for the SK model suggest that besides the important conceptual difference between the distribution of cluster weights and the distribution of basin sizes, there are also fundamental qualitative differences between these two distributions. Our conclusion is that perhaps such distributions are not so universal, and that care must be taken in making comparisons between them.

The second reason why this problem has not received more attention is that basins of attraction can only be strictly defined for deterministic systems. Furthermore, their shape can depend on the details of the dynamics. It is only recently that spin glasses have been treated as dynamical systems in their own right, and that the standard dynamical-systems type questions have begun to be addressed [see, for example: Gutfreund *et al.*, 1988; Cabasino *et al.*, 1988; Sompolinsky, 1988; Kanter, 1990].

8.2. PROBABILISTIC BASIN MEASUREMENT

First, we note that all attractors of (8.1) are fixed points (this is not true for asymmetric connections or parallel dynamics). This fact can be established by showing that the total energy, E (i.e. the spin-glass Hamiltonian)

$$E = \sum_i E_i = -\frac{1}{2}\sum_{i,j} T_{ij} S_i S_j \tag{8.3}$$

is a Liapunov function of (8.1). In Eq. (8.3), E_i is the energy contributed by site i, equivalent to the local field at site i times $-S_i/2$. We will also refer to the average energy per site, defined $\epsilon \equiv E/N$. Intuitively, we expect that the most stable attractors (those

with the most negative energy) will have the largest basins of attraction. Indeed, this idea is observed to hold for the recall states in associative memory models [Forrest, 1988; Kepler and Abbott, 1988; Opper *et al.*, 1989], and is an important principle for developing robust learning algorithms [Krauth *et al.*, 1988; Abbott, 1990]. The numerically measured relationship between basin size and average energy per spin ϵ for the SK model (8.1) is shown for $N = 20$ in Fig. 8.1. This figure shows ~ 11,000 fixed points from 200 realizations, and confirms the intuition that more stable fixed points have larger basins of attraction. It is significant, however, that the dependence of basin size on energy is quite weak: Fixed points with identical energies have basin sizes ranging over two orders of magnitude.

We now explain how Fig. 8.1 was made. Define the size W_s of the basin of attraction of the s^{th} attractor as

$$W_s \equiv \frac{\text{number of initial states leading to attractor } s}{\text{total number of initial states}} \quad . \tag{8.4}$$

This definition satisfies the normalization

$$\sum_s W_s = 1 \quad . \tag{8.5}$$

Having only fixed point attractors greatly simplifies the task of measuring basins, since it is always clear to which attractor a particular initial condition has flowed. Still, because of the large state space (2^N states for a system of size N) a complete enumeration of basin sizes is prohibitively time-consuming for all but the smallest systems (such an approach was used by Gutfreund *et al.*, [1988]). The problem is compounded by the need to average over large numbers of realizations in order to obtain reliable statistics.

Instead, we compute basin sizes W_s by the following probabilistic method. For each

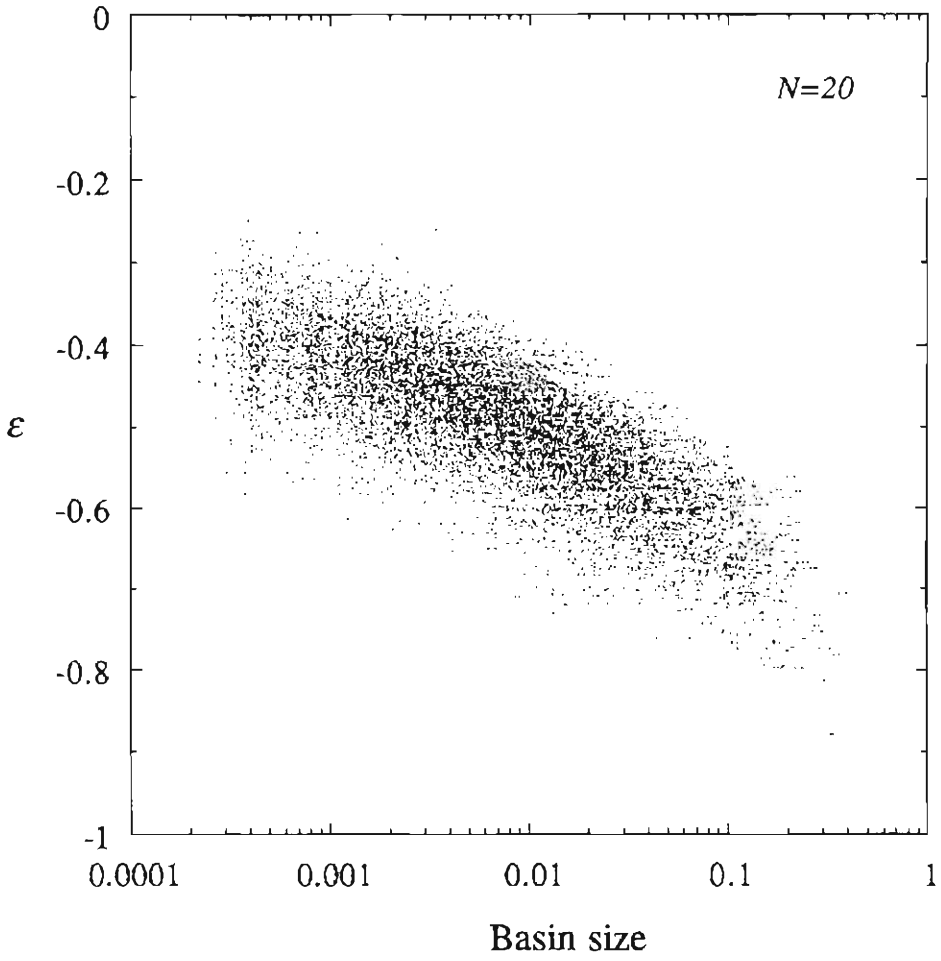


Fig. 8.1. Basin size versus energy per site $\varepsilon = E/N$ for fixed points in 200 realizations of the deterministic SK spin glass, Eq. (8.1) and (8.2). The number of data points, equal to the number of fixed points found using the statistical technique, is 11,032. The relation $N_{fp} = exp[aN]$ gives a = 0.200 for this data, in good agreement with the theoretical value 0.199. Each realization was checked for fixed points until 500 consecutive initial states failed to produce a new fixed point (see text). Note the general trend that the most stable fixed points (i.e. fixed points with the most negative ε) have the large basins of attraction. Also note that the trend is rather weak.

realization of T_{ij}, random initial states (random, unbiased strings of ± 1's) are generated, and the attractor for each initial state is found under the dynamics of (8.1). Convergence is fast, typically requiring fewer than 5 updates per site. A list of the distinct attractors is kept along with the total number of initial states that flowed to each attractor. For each initial state, the attractor found is compared to the list of those previously found, and if there is a match, the basin size of that match is incremented. If no match is found after checking the entire list, the attractor found must be new. The new attractor is then added to the list of attractors with its basin size initialized to 1. For each realization of T_{ij}, initial states are generated until a quitting condition is reached. The quitting condition is that a specified number of *consecutive* initial conditions have been generated without finding a new attractor. Typically this number is set to 500 for $N \leq 20$ and 800 for $N > 20$. This quitting condition is superior to simply using a large, fixed number of initial conditions to test each realization; It is efficient over a wide range of possible numbers of attractors without requiring a good "guess" (i.e. a pre-inserted theory) for how long to sample. After reaching the quitting condition, basin sizes for each of the found attractors are given by the number of initial states that flowed to that attractor divided by the total number of initial states tested. Statistics are accumulated over a large number of realizations (typically 200 - 500).

There is a check which tells us if we have sampled long enough. We know the total number of fixed points that we should find. The expected number of fixed points N_{fp} for (8.1) is

$$\left\langle N_{fp} \right\rangle_{\mathrm{T}} = Ae^{aN} \tag{8.6}$$

with $A \sim 1$ and $a = 0.1992...$ [Tanaka and Edwards, 1980; De Dominicis *et al.*, 1980; Bray and Moore, 1980]. (Henceforth we will drop A, calling it 1, and drop the brackets

on N_{fp} indicating an average over realizations). The result (8.6) was originally calculated assuming large N, but it is remarkably accurate for N as small as $N = 4$ (!) when averaging is done over a large number of realizations [Gutfreund *et al.*, 1988].

The distribution of basin sizes can be found numerically by setting up a histogram of basin sizes, and collecting data over a large number of realizations. A histogram of basin sizes for the 11,000 points of Fig. 8.1 is shown in Fig. 8.2(a). Figure 8.2(b) shows a histogram of energies per site ε for this same data set.

8.3 THE DISTRIBUTION OF BASIN SIZES

8.3.1. Definitions

The distribution of basin sizes, averaged over realizations, can be written as a continuous function[2]

$$f(W) = \left\langle \sum_s \delta(W - W_s) \right\rangle_{\mathbf{T}} \quad . \tag{8.7}$$

The 0^{th} and 1^{st} moments of the distribution $f(W)$ must satisfy the following constraints:

$$\int f(W)\, dW = N_{fp} \ , \tag{8.8}$$

$$\int f(W) W\, dW = 1 \ . \tag{8.9}$$

---

[2] The distribution $g(E)$ studied by Parga and Parisi [Chowdhury 1986, p. 82-83; see footnote 1] is $g(E) = \left\langle \sum_s W_s \ \delta(E - E_s) \right\rangle_{\mathbf{T}}$.

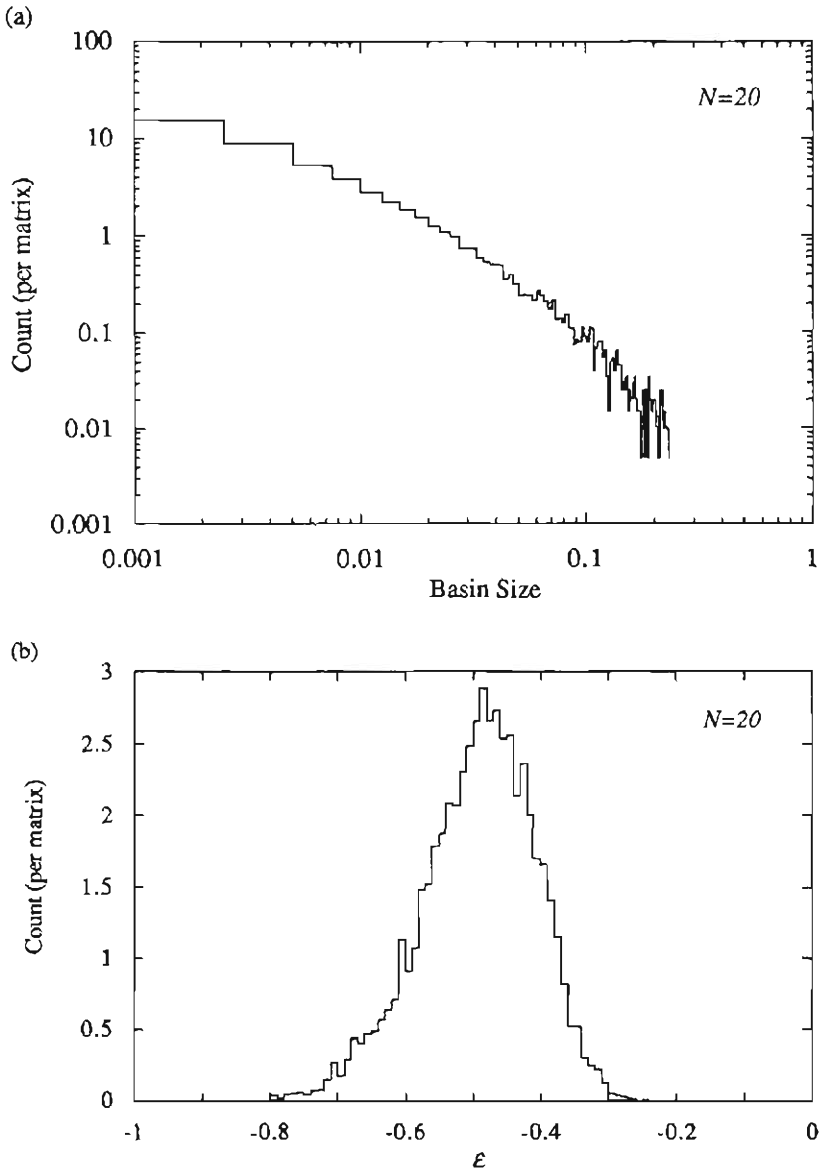


Fig. 8.2. Histograms of the data of Fig. 8.1, normalized by the number of realizations = 200. (a) Histogram of basin sizes, showing the characteristic power law behavior with rounding at large and small basin size. Bin width $\Delta W = 0.002$. (b) Histogram of energy per site $\varepsilon = E/N$, with E defined by (8.3). Bin width $\Delta\varepsilon = 0.01$. Distribution of energies is in good agreement with theory of Tanaka and Edwards [1980]. A least square fit of the energy histogram to the gaussian $f(\varepsilon) = A\,exp\left[-N(\varepsilon-\bar{\varepsilon})/2\sigma^2\right]$ gives $\bar{\varepsilon}$ = –0.49 and $\sigma = 0.36$; theoretical values are $\bar{\varepsilon}$ = -0.50 and $\sigma \cong 0.31$ [Tanaka and Edwards, 1980].

Other moments of $f(W)$, defined generally as

$$y_m = \int f(W) W^m \, dW \quad , \tag{8.10}$$

will also be of interest , especially for comparing our results to results for other dynamical systems and to numerical data already in the literature.

As a first example, consider the case where all basins are the same size. The constraints (8.8) and (8.9) then require the distribution to have the following form:

$$f(W) = N_{fp} \, \delta\left(W - 1/N_{fp}\right) \qquad \text{[equal-sized basins].} \tag{8.11}$$

We recover the obvious result that if all basins were the same size, that size would be $1/N_{fp}$ ($= e^{-0.1992N}$). Figures. 8.1 and 8.2(a) show that it is clearly *not* the case that all basins are the same size.

The range of possible basin sizes is limited by the dynamics. On the small end, a fixed point of (8.1) will always be stable to the flipping of a single state, thus no fewer than N states will will flow to any fixed point. This automatically gives a minimum basin size W_{min} of

$$W_{min} = \frac{(N+1)}{2^N} \, . \tag{8.12}$$

On the large end, the maximum basin size consistent with the invariance of (8.1) to the global inversion $S_i \rightarrow -S_i$ for all i, is

$$W_{max} = \frac{1}{2} \quad . \tag{8.13}$$

That is, there will always be at least two fixed points dividing state space into equal halves, so the largest possible basin size is 1/2. Even if we demand that the number of fixed points in every realization equal exactly N_{fp} - and we should, since (8.6) is self-averaging as $N \to \infty$ - we still find that the largest possible basin size is ~1/2 to a very good approximation. For example, with $N = 20$, a maximum basin size of $W_{max} = .4995$ still allows room for $(e^{0.1992*20} - 2) \cong 52$ other attractors, each of minimal basin size. This approximation, $W_{max} \cong 1/2$, improves for larger N.

8.3.2. Numerically observed power-law behavior of $f(W)$

Figure 8.3 shows the main observation of this chapter. *Over a broad range of basin sizes, $f(W)$ is approximately described by a power-law*

$$f(W) = K \; W^{-\gamma} \tag{8.14}$$

with $\gamma \sim 3/2$ (± 0.2). The data in Fig. 8.3 were collected from 330 realizations of (8.1) with $N = 28$. The average number of basins found per matrix was 232.1, corresponding to a scaling exponent (8.6) of $a = 0.195$, which is in reasonably good agreement with the theoretical value of 0.199. This suggests that most of the basins were counted. This result is consistent with the histogram in Fig. 8.2(b), but contains more data.

The value of γ is found to be *independent of N*, though as we will discuss in the following subsection, there is a cutoff for small basins which does vary with N.

We emphasize that the observed power law is an empirical result. Below, in § 8.4.1, we will discuss a related theoretical result which supports this observation. However, we do not yet have a theory which directly explains the power law, let alone the exponent. We also point out that the distribution we observe is not a perfect power law, but is

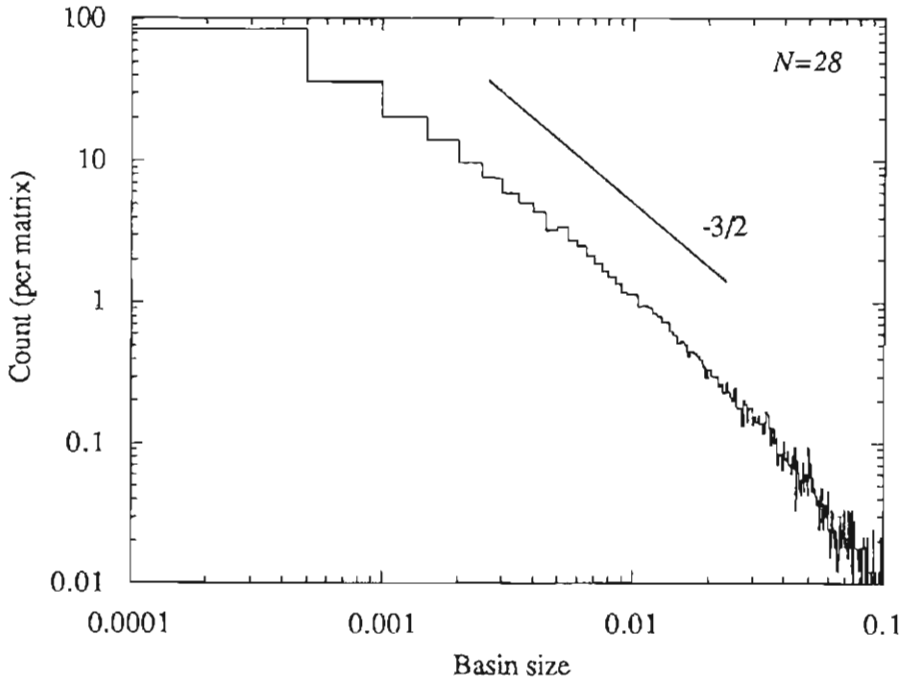


Fig. 8.3. Normalized histogram of basin sizes for $N = 28$ based on 330 realizations. Count (per realization) corresponds to $f(W)\Delta W$, where $f(W)$ is the distribution of basin sizes and ΔW is the bin width. Here, $\Delta W = 0.0005$. Each realization was checked for fixed points until 800 consecutive initial states failed to produce a new fixed point (see text). Line on right indicates slope of -3/2, for comparison.

concave down for both large and small W. Such deviations might be explained as numerical artifacts; on the other hand, there is no compelling reason to insist on a perfect power law for $f(W)$. At our present level of knowledge, the fairest statement is this: The assumption that $f(W)$ is a power law allows us to easily calculate some consequences of the observed basin-size distribution. Hopefully, these conclusions do not depend critically on the value of γ, or on slight deviations from a perfect power law. Lacking a theoretical model, this seems to be a reasonable place to start.

8.3.3. Consequences of a power law distribution of basin sizes

We now explore some consequences of a distribution of basin sizes given by $f(W) = K\ W^{-\gamma}$, focusing on the case $\gamma = 3/2$. The normalization conditions (8.8) and (8.9) imply that there is a cutoff $W_{cutoff} > 0$ which sets the scale of the smallest basin size. Taking the maximum basin size $W_{max} = 1/2$ gives the pair of equations

$$\int_{W_{cutoff}}^{1/2} K\ W^{-\gamma}\, dW = N_{fp}\ , \tag{8.15}$$

$$\int_{W_{cutoff}}^{1/2} K\ W^{1-\gamma}\, dW = 1\ . \tag{8.16}$$

Eqs. (8.15) and (8.16) together determine values for K and W_{cutoff} which depend only on γ. These normalization equations imply that γ is between 1 and 2. This is consistent with our observation $\gamma \sim 3/2$. For the case $\gamma = 3/2$, Eqs. (8.15) and (8.16) yield

$$K = \frac{\sqrt{2}}{2 - 4/N_{fp}} \; ; \quad \left[\lim_{N \to \infty} (K) = 1/\sqrt{2} \right] \tag{8.17}$$

$$W_{cutoff} = 2 N_{fp}^{-2} \; ; \left[\lim_{N \to \infty} \left(W_{cutoff} \right) = 0 \right] \tag{8.18}$$

Equations (8.17) and (8.18) already lead to three rather surprising conclusions:

(*i*) The value of K depends very weakly on N, especially for larger values of N. That means that the average (absolute) number of fixed points with basin size between W and $W+dW$ for any $W > W_{cutoff}$ is *independent of* N.

(*ii*) The cutoff of the power law, W_{cutoff}, tends to zero more slowly than the minimum basin size W_{min} as the size of the system N becomes large. Specifically, $W_{cutoff} \sim e^{-0.398N}$ while $W_{min} \sim e^{-0.693N}$. Below W_{cutoff}, the density of basins falls rapidly to zero. Thus the smallest (typical) basin size is much larger than W_{min} for large N.

(*iii*) W_{cutoff} is also the most common basin size, since $f(W)$ has its maximum at this value. The value of W_{cutoff} is different from the average basin size W_{ave}, which is $1/N_{fp}$ by definition. Thus as the size of the system increases, the most common basin size goes to zero faster than the average basin size.

Theoretical values for W_{ave}, W_{cutoff}, and W_{min} are shown in Fig. 8.4(a) along with numerical measurements of W_{cutoff} for several values of N. Numerical values of W_{cutoff} are taken to be the maxima of the histograms of basin sizes, as shown in Fig. 8.4(b) for the case $N = 22$.

Higher moments of $f(W)$ can also be calculated. Using $f(W) = K\ W^{-\gamma}$ with $\gamma = 3/2$ and values for K and W_{cutoff} from (8.16) and (8.17), we find

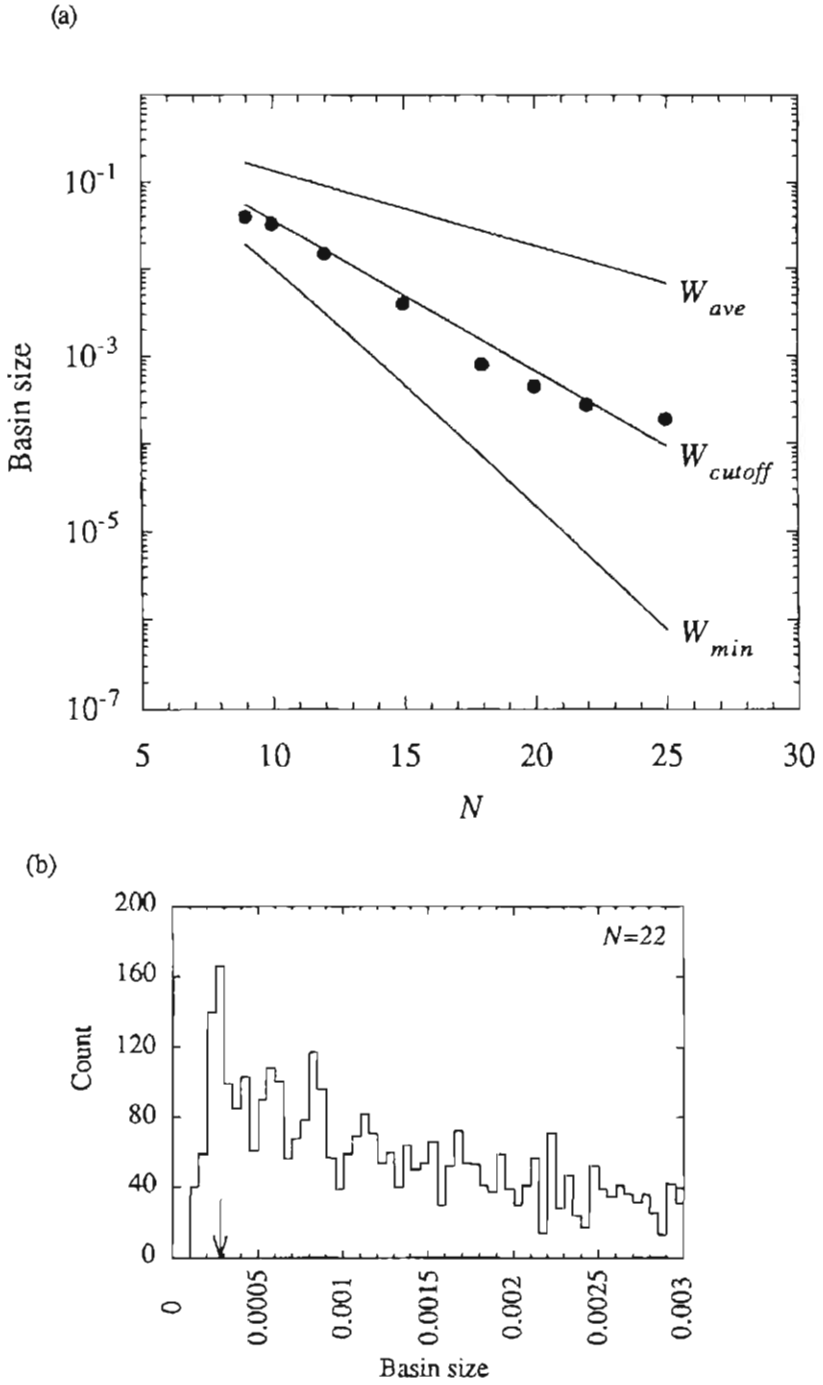


Fig. 8.4. The small-basin cutoff W_{cutoff} for the power law $f(W) = K\ W^{-\gamma}$ with $\gamma = 3/2$, as a function of system size N. From Eq. (8.18), $W_{cutoff} = 2N_{fp}^{-2} = 2e^{-.398N}$. Also shown are the average basin size $W_{ave} = 1/N_{fp}$ and minimum basin size $W_{min} = 2^{-N}(N+1)$, both of which are independent of the form of $f(W)$. Data (circles) are numerical values of W_{cutoff} for several values of N, defined as the maximum of the $f(W)$, as shown in (b) for the case $N = 22$.

$$y_m = \frac{2^{1-m} {N_{fp}}^{2m} - 2^m N_{fp}}{{N_{fp}}^{2m-1}\left(N_{fp}-2\right)(2m-1)} \quad , \; m = 0,1,2,3, \ldots \tag{8.19}$$

$$\lim_{N\to\infty}\left[y_m\right] = \frac{1}{(2m-1)2^{m-1}} \; , \quad m = 1,2,3, \ldots \tag{8.20}$$

Of particular interest is second moment y_2, which approaches the value 1/6 as the system size (and thus the number of fixed points) becomes large, according to (8.20). Numerical results of Gutfreund *et al.*[1988] for a system identical to (8.1), generalized to include asymmetric connections, showed that y_2 tends to zero as the system size increases for symmetric connections. The data of Gutfreund *et al.* were obtained using a different technique from ours, one which tested only two initial conditions per realization and sampled an extremely large number of realizations. Repeating these measurements using our probabilistic method also indicates that the numerical value of y_2 tends to zero , not 1/6, as N becomes large. This disagreement shows that a strict -3/2 power law does not provide a completely accurate description of $f(W)$. In particular, the -3/2 power law must exaggerate the number of large basins, leading to a finite second moment.

8.4. DISTRIBUTIONS FOR OTHER MODELS

There is theoretical and numerical evidence for universality in the way state space breaks into clusters - basins of attraction, for example - in dynamical systems with many attractors [Derrida and Flyvbjerg, 1986, 1987a, 1987b; Derrida, 1988b; Gutfreund *et al.*, 1988]. This section provides a brief summary of results for some of the systems which show these universal properties.

8.4.1. Clusters of states in the SK model

Mezard *et al.*[1984a, 1984b] supplied a crucial piece of the spin glass puzzle by describing the state space of the SK model in terms of a hierarchical or ultrametric geometry. Ultrametricity gave deep physical insight into the meaning of the Parisi order parameter $q(x)$, and also provided a satisfying picture of the state space of a spin glass as a hierarchy of valleys within valleys [Mezard *et al.*, 1987; for a review of ultrametricity, see: Rammal, 1986]. These papers [Mezard *et al.*, 1984a; 1984b] also presented an analytic form for the distribution of clusters of states[3], with clusters defined according to the mutual overlap q of states. We will present some of their results without providing details (see Mezard *et al.*, 1987, Ch. IV, for a very readable account).

Consider a division of the state space of the SK model into clusters, such that states with overlap larger than q belong to the same cluster. The overlap between a pair of states α and β is defined $q^{\alpha\beta} = N^{-1}\sum_i \langle S_i^{\alpha}\rangle\langle S_i^{\beta}\rangle$ with brackets denoting a thermodynamic expectation value. Define $\tilde{W}_s$ to be the size of the s^{th} cluster. When averaged over realizations, the cluster sizes will form a continuous distribution $\tilde{f}(\tilde{W})$, which was given by Mezard *et al.* [1984a; 1984b],

$$\tilde{f}(\tilde{W}) \equiv \left\langle \sum_s \delta(\tilde{W} - \tilde{W}_s) \right\rangle_{\mathbf{T}} = \frac{\tilde{W}^{y-2}(1-\tilde{W})^{-y}}{\Gamma(y)\Gamma(1-y)} \quad . \tag{8.21}$$

The quantity y in (8.21) depends on the choice of q used to set the cluster size and also on physical parameters such as temperature and applied magnetic field. The gamma

[3]Throughout this section the term "states" will refer to equilibrium states, or states separated by energy barriers that become infinite in the thermodynamic limit. At $T = 0$, equilibrium states of the SK model are equivalent to fixed points of the dynamical system (8.1).

functions in (8.21) normalize $\tilde{f}(\tilde{W})$ so that

$$\int \tilde{f}(\tilde{W})\tilde{W}\, d\tilde{W} = 1\,. \tag{8.22}$$

The derivation of $\tilde{f}(\tilde{W})$ takes into account the Boltzmann probability of a state being present in counting the number of states in a cluster. Setting q to its maximum value, which is q_{EA} (the Edwards-Anderson order parameter) reduces the cluster size such that each cluster contains just a single state (recall: $q_{EA} \equiv q^{\alpha\alpha}$ is the equilibrium overlap of a state with itself). For this choice of q, the function $\tilde{f}(\tilde{W})$ is just the average density of states with Boltzmann weight $\tilde{W}$. That is, for $q = q_{EA}$ the weight $W_s(y(q))$ of the s^{th} state is given by

$$\tilde{W}_s = \frac{e^{-\beta F_s}}{\sum_s e^{-\beta F_s}} \tag{8.23}$$

where $\beta = 1/kT$ and F_s is the free energy of the s^{th} state.

It can be shown that $y(q)$ for $q = q_{EA}$ is given by the length of the plateau of the Parisi order parameter $q(x)$ [Mezard *et al.*, 1984a, 1984b]. A value for $y(q_{EA})$ can be found using an approximation known as the PaT hypothesis [Vannimenus *et al.*, 1981], which assumes certain quantities to be independent of magnetic field and temperature. In the limit $T \to 0$ the PaT hypothesis gives

$$\lim_{T\to 0} y(q_{EA}) = \frac{1}{2}. \tag{8.24}$$

Inserting (8.24) into (8.21) gives the following encouraging result: For small clusters,

$$\tilde{f}(\tilde{W}) \sim \frac{1}{\pi}\tilde{W}^{-3/2} \quad , \quad \tilde{W} << 1, T \to 0. \tag{8.25}$$

Before celebrating the appearance of a -3/2 power law, we emphasize two points:

(*i*) Recall what we have found. $\tilde{f}(\tilde{W})d\tilde{W}$ is the number of states with Boltzmann weight between $\tilde{W}$ and $\tilde{W}+\mathrm{d}\tilde{W}$. It is not at all clear how this weight is related to the basin size, although Fig. 8.1 may offer some clues.

(*ii*) We have considered the solution as $T \to 0$ because in this limit, the fixed points of (8.1) correspond to thermodynamic equilibria of the SK model. However, another effect of the $T \to 0$ limit is to heavily weight the states with lowest free energy. Thus $\tilde{f}(\tilde{W};T \to 0)$ pertains only to the lowest energy states, not all equilibria. This explains why the distribution $\tilde{f}(\tilde{W})$ has a divergence at $\tilde{W} \to 1$: As $T \to 0$ the sum in the denominator of (8.23) is dominated by a single term, the ground state for that realization. When the state s in the numerator of (8.23) is the ground state, the numerator is nearly equal to the denominator, so $\tilde{W}_{gs} \sim 1$. Note that the divergence in $\tilde{f}(\tilde{W})$ as $\tilde{W} \to 1$ is not present in the distribution of basin sizes $f(W)$ which must vanish above $W = 1/2$.

Nevertheless, the form of (8.25) is tantalizing. At present, however, we do not have any satisfying way to relate this result to the observed power law for the distribution of basin sizes.

8.4.2. The Kauffman Model and the Random Map

Kauffman introduced a simple model of genetic mutation and adaptation which shares many features with spin glass models [Kauffman, 1969, 1984, 1990]. The model consists of N sites, representing individual genes, each of which is characterized by a binary state (0 or 1) indicating one of two possible alleles for that gene. Each site is affected by exactly K sites, selected at random from the N sites in the system. The

response of a site to its K inputs is a random Boolean function which is chosen independently for each site, and is fixed for all time (i.e. quenched) once chosen. The dynamics are deterministic and parallel, with each site following its random truth table (2^K input states assigned randomly to a 0 or 1 output state). Because the model is deterministic and has a discrete and finite state space, all attractors must be periodic, with periods ranging from 1 (fixed points) to 2^N.

The distribution of basin sizes in the Kauffman model has been studied numerically as a function of K [Derrida and Flyvbjerg, 1986; see also Kauffman, 1990 for basin size vs. energy plots similar to Fig. 8.3]. The numerical results of Derrida and Flyvbjerg [1986] suggest a surprising universality: moments of the distribution of basin sizes, plotted one against another, agree extremely well with analytical results relating the moments of $\tilde{f}(\tilde{W})$, the distribution of weights in the SK model. Such agreement indicates that certain statistical properties of multivalley state spaces are insensitive to the details of the underlying dynamical system.

In the limit $K \to \infty$, the Kauffman model is equivalent to a random mapping of an N-dimensional binary space onto itself [Derrida and Flyvbjerg, 1987a, 1987b]. (Note that the limit $K \to \infty$ can be taken even for finite N, since the connection rule does not restrict how many times a particular site may appear in a truth table.) In contrast to the Kauffman model with general K, the random map model is mathematically tractable, and many of its statistical properties are known analytically [Harris, 1960]. From these results, Derrida and Flyvbjerg [1987a, 1987b] derived the following exact expression for the distribution of basin sizes in the random map model:

$$f_{RM}(W) = \frac{1}{2} W^{-1} (1 - W)^{-1/2} . \tag{8.26}$$

Note the similarity between $f_{RM}(W)$ and the distribution of weights $\tilde{f}(\tilde{W})$ for the SK spin glass in Eq. (8.21). Of particular interest is the divergence at $W = 1$ common to both, but absent in the distribution of basin sizes for our system (8.1).

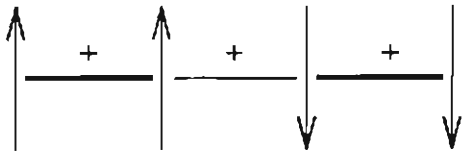
8.4.3. The 1-D spin glass

The one dimensional (1-D) spin glass at zero temperature is another example of a system with many attractors for which the distribution of basin sizes is known exactly (the other being the random map). In this case, the result is that all the basins are the same size, that is, $f(W)$ is given by Eq. (8.11) [Ettelaie and Moore, 1985; Derrida and Gardner, 1986].

Derrida and Gardner [1986] analyzed the metastable states of a 1-D chain of L Ising spins with symmetric nearest neighbor coupling and free ends. The dynamics they considered were single spin flip, with either random or deterministic order of updating. Their results do not depend on the details of the distribution $P(T_{ij})$ of the random connections, as long as it is symmetric ($P(T_{ij}) = P(-T_{ij})$) and contains no delta functions. They find that the typical number of metastable states in a realization of length L is $N_{fp} = 2^{L/3}$ and that the average number (over realizations) is $N_{fp}^{ave} = (4/\pi)^L$ [see also: Li, 1981; Ettelaie and Moore, 1985]. That the average and typical number of fixed points in a realization are not equal is characteristic of short range models; for the infinite-range SK model, $N_{fp} = N_{fp}^{ave}$.

It may seem surprising that the 1-D spin glass has many metastable states, since it is not frustrated in the sense of Toulouse [1977] (see § 4.3.2). Indeed, for zero external field - where the number of metastable states achieves its maximum [Derrida and Gardner, 1986] - the 1–D spin glass is equivalent to a purely ferromagnetic chain by a Mattis transformation [Mattis, 1976], albeit a ferromagnet with a distribution of

ferromagnetic coupling strengths. How then, can such a system possess an exponential number of metastable states? The answer is that single-spin-flip dynamics does not allow kinks at weak bonds to unkink themselves. This point is illustrated by a simple example with four spins and three ferromagnetic bonds, two strong ones on either side of a relatively weak one.



For this configuration, no single spin flip will allow the weak bond in the middle to become satisfied. The metastable state shown is also stable for parallel updating of spins. On the other hand, an update rule which checks for energy reduction upon simultaneously flipping pairs of spins will eliminate the metastable state shown. Moore [1987] has discussed the computational efficiency of update rules which use multiple spin flips to eliminate local energy minima. Finally, we note that recasting the 1-D chain in terms of analog state variables - by replacing $Sgn[h]$ with $tanh[\beta h]$ in (8.1), for instance - can also be used to reduce the number of metastable states. Continuing with the simple four spin arrangement above, with the particular connection strengths $|T_{12}| = |T_{34}| = (3/2)|T_{23}|$, setting $\beta < 1.44$ will destabilize the metastable states, while the ground states will remain stable for $\beta > 0.48$. This holds for sequential, parallel, or continuous-time analog dynamics.

8.5. BASIN SIZES IN AN ANALOG SPIN GLASS

The last example in the previous section showed how recasting a binary system into its analog counterpart can dramatically change the system's energy landscape. This idea was also discussed at length in Ch. 7. What effect will using analog state variables have on the basin structure? We address this question for an analog version of the SK spin glass,

$$x_i(t+1) = tanh\left[\beta\left(\sum_{j<i} T_{ij}x_j(t+1) + \sum_{j>i} T_{ij}x_j(t)\right)\right] \qquad i = 1, ..., N. \qquad (8.27)$$

where the states $x_i(t)$ now take on continuous values. As before, connection strengths T_{ij} are symmetric ($T_{ij} = T_{ji}$) and gaussian distributed according to (8.2). In the limit of large gain , $\beta \rightarrow \infty$, (8.27) reduces to (8.1).

Figure 8.5 shows the distribution of basin sizes and energies for the analog spin glass with N = 20 for two values of neuron gain, β = 4 and β = 20, averaged over 100 realizations. Basin-size counts and energies are based on binary states generated by applying the *Sgn* function to an analog fixed point after convergence of (8.27). Initial states were random corners of the unit hypercube, as before. The data in Fig. 8.5 show that the reduction in the number of fixed points at lower gain (see § 7.2.1 for theory) is strongly biased toward eliminating fixed points with small basin of attraction. For example, we see from Fig. 8.5 that the average number of fixed points whose basins occupy 0.1 of the state space (10^5 initial states for N = 20) is essentially the same for β = 4 and β = 20, while the average number of fixed points with basins occupying 0.001 of the state space (10^3 initial states) is ~50 times smaller at β = 4 than at β = 20. The distribution of energies - calculated using (8.3) after applying the *Sgn* function to the analog fixed point - also shows that lowering the gain shifts the distribution to lower

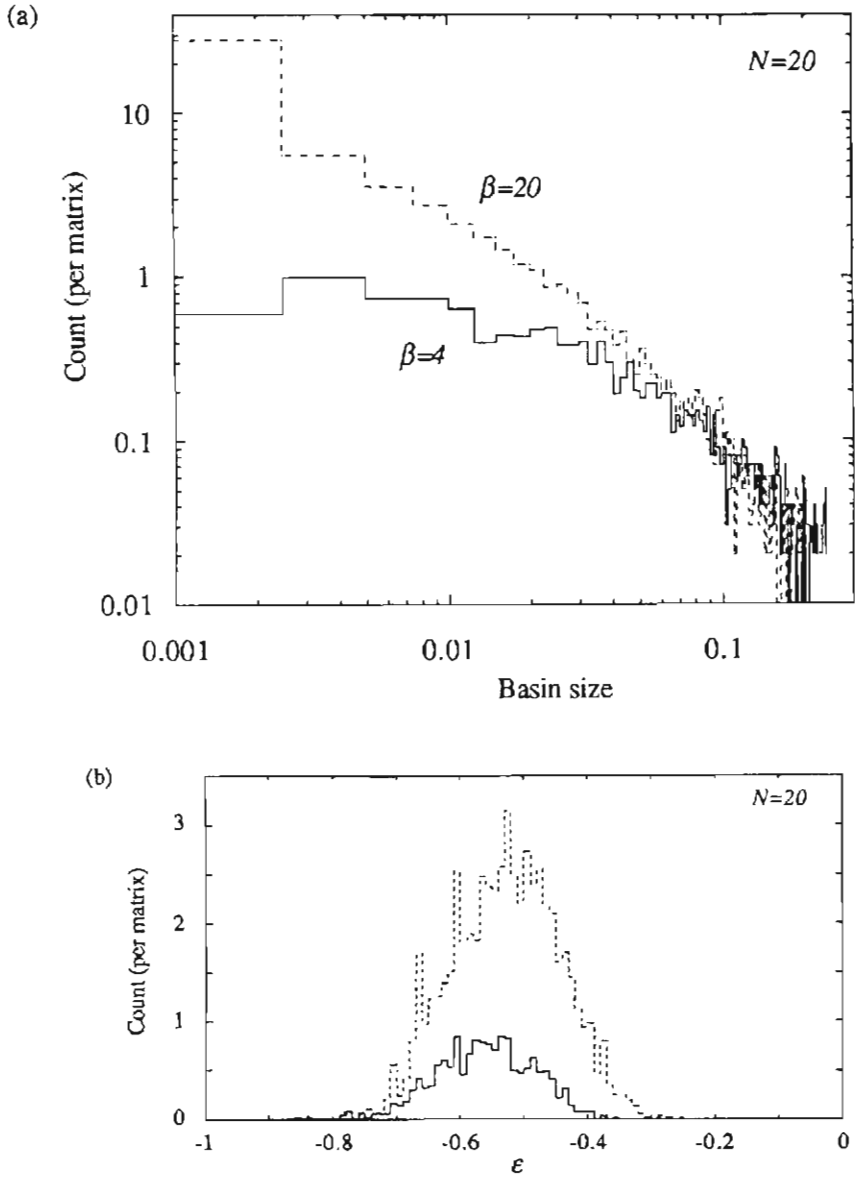


Fig. 8.5. Distribution of basin sizes and energies per site ε for the deterministic analog spin glass (defined in text) at two values of analog gain, $\beta = 4$ (solid line) and $\beta = 20$ (dashed line), for system size $N = 20$. 100 realizations were counted for each value of β. The critical value of gain where fixed points first appear away from the origin ($x = 0$) is $\beta = 1/2$, not $\beta = 1$, as it would be for the finite temperature spin glass. (a) Histogram of counts (per matrix) $=f(W)\Delta W$ for a bin width $\Delta W = 0.0025$. Data show that lowering the gain greatly reduces the number of small basins, while leaving most of the big ones. (b) Histogram ($=f(\varepsilon)\Delta\varepsilon$) of fixed point energies per site $\varepsilon = E/N$. Bin width $\Delta\varepsilon = 0.01$. Lowering β reduces the number of fixed points and also shifts $f(\varepsilon)$ to lower (more negative) energies.

energies. This effect, however, is weak compared to the shift in the basin-size distribution. From these observations, we conclude that using analog state variables is effective for selectively eliminating attractors with small basins, though not specifically the most shallow attractors.

8.6. DISCUSSION AND OPEN PROBLEMS

We have found numerically that the distribution $f(W)$ of basin sizes in the SK model is approximately descibed by a power law with exponent $\sim -3/2$. Thus, $f(W)$ appears to be quite different from the distribution $f_{RM}(W)$ of basin sizes in the random map model and different from the distribution $\tilde{f}(\tilde{W})$ of Boltzmann weights in the SK model. So what has become of the universality of multivalley landscapes discussed by Derrida? Gutfreund *et al.*[1988] have suggested that the general features of a multivalley landscape depend on whether the number of attractors grows exponentially with system size N. In the random map model for example, the average number of attractors $\langle A \rangle$ is linear in N: $\langle A \rangle = N \ln 2/2$. This is different from the SK model, where the average number of attractors is exponential in N. On the other hand, numerical data show that the number of attractors for the *fully asymmetric* spin glass does scale linearly with the size of the system, and also that the distribution of basin sizes for the asymmetric spin glass shares some common features with $f_{RM}(W)$ and $\tilde{f}(\tilde{W})$ [Gutfreund *et al.*, 1988]. Perhaps basin structure, and the general structure of state space, can be divided into universality classes depending on whether the number of attractors does or does not increase exponentially with N. At present, is not clear that such broad classes of dynamical systems exist, or even by what phenomenological criteria such distinctions ought to be made.

The results presented in this chapter raise many new questions. Here are some:

-- Will other techniques for measuring basin sizes yield the same results?

-- How do the the details of the dynamics affect basin structure? For example, how is state space shared between the fixed points and the 2-cycles in a parallel version of (8.1)?

-- Why does a power law - or an approximate power law - appear? How is it related to the ultrametric structure of state space?

-- How well do the results generalize to include (among other things) nongaussian connection distributions, nonsymmetric connection, correlated connections, and stochastic dynamics?

-- How can these results be applied to neural networks, where control over basins of attraction is of central importance? A first step in approaching such a question would be to study the distribution $f(W)$ in an associative memory as the number of stored patterns p goes from $p << N$ (the ferromagnetic limit), where presumably the basins for the stored patterns are all the same size, to $p >> N$, (the spin glass limit), where the present results should be recovered.

-- Is the weak correlation between the energy or depth of a fixed point and the size of its basin of attraction a general phenomenon? What about the correlation between basin size and the stability parameter κ, defined as the minimum of the distribution of local fields? For the associative memory, numerical and analytical evidence suggests that the correlation between κ and basin size is strong [Forrest, 1988; Abbott, 1990].

Chapter 9

CONCLUSIONS

In the introduction of this thesis, we discussed the idea that models of large nonlinear systems ought to be simple on a microscopic scale if one is to have any hope of understanding their large-scale dynamical behavior. At that point, no mention was made of how, given a physical system, one should go about separating wheat from chaff, discarding the unimportant microscopic details and keeping the essential ones. Making such distinctions *a priori* seems to be impossible; small changes often make big differences, and vise versa. This distinction, and, more generally, how a dynamical system's microscopic features influence its large-scale behavior, is at the heart of the controversy surrounding the whole neural network approach. One often hears that describing a neuron as a binary threshold element is a ridiculous oversimplification. If this is so, then why? *Specifically*, what global properties are lost, or even eroded, by making such an approximation?

In this thesis, we have studied how properties of analog neural networks *at the level of the individual neuron* affect the large-scale dynamics of the network, and how this effect is related to global network properties such as the spectrum of eigenvalues of the connection matrix. The neuron properties we have considered are especially relevant to designing fast, stable neural networks in electronic hardware. For example, an electronic neuron (or a biological neuron for that matter) does not have an infinite switching speed, and so, in principle, one must account for any delay in describing the overall dynamics of the network. Intuition tells us that when the delay is extremely small (compared to some characteristic time of the network), it can probably be ignored. In this case, a simple

model which assumes instantaneous response will suffice. But when is small no longer small? At some point the chaff becomes wheat, and the delay must be accounted for. The problem of delay-induced instability was discussed in detail in chapters 3 and 4. In this example and throughout the thesis, we found that the gain of the neuron (i.e. maximum slope of its transfer function) has a strong influence on the global dynamics of the network. In several instances, the influence of neuron gain could be reduced to a simple stability criterion for insuring convergence of the dynamics to a fixed point.

From a practical point of view, the most important conclusion of this thesis is that networks of analog neurons offer important computational advantages over networks of binary (Ising) elements. Those advantages are: (1) Symmetrically connected analog networks can be updated *in parallel* with guaranteed convergence to a fixed point. In general, networks with binary neurons must be be updated sequentially to prevent oscillation. Parallel updating is faster than sequential updating by a factor of $O(N)$ where N is the number of neurons in the network. (2) Lowering the gain of analog neurons shows many of the beneficial effects of using temperature to escape local minima. This was demonstrated numerically for associative memories in Ch. 5, and an explanation for this surprising effect was given in Ch. 7. Specifically, it was shown in Ch. 7 that lowering the neuron gain dramatically reduces the number of local minima in the network's energy landscape. (3) The analog networks we considered have deterministic dynamics, which means that they can be built using standard analog VLSI technology. Implementing a stochastic update rule in hardware is difficult because of the need for lots of random numbers. Implementing a stochastic algorithm with parallel dynamics in electronics would be even more difficult, as it would require N independent random numbers at each time step. Deterministic annealing could also be implemented easily in analog VLSI by changing all neuron gains simultaneously via a single control line.

One might also be lead to speculate that the analog character of biological neurons - that is, their well-known graded response - has similar computational significance, rather than being an artifact of evolution.

Another significant idea which has emerged in this work is that certain analytical techniques originally developed for Ising spin glasses, and later extended to treat binary neural network models, can also be successfully adapted to the study of analog neural networks. The analysis of storage capacity leading to the phase diagrams in Ch. 5 and the analysis of the number of local minima for the analog spin glass and associative memory in Ch. 7 are two places where techniques developed for discrete systems have been adapted to the analog problem.

The systems we have considered were not arbitrarily chosen by any means. One should not get the impression that relaxing some of the assumptions made will make the problem only slightly more difficult. Usually, things get much harder. The most restrictive assumption made throughout the thesis was that the coupling was symmetric. Neural networks with asymmetric connections have vastly richer dynamics but are correspondingly more difficult to approach analytically. The extension of the present results to include asymmetric networks is the first and most obvious direction in extending the present results. Be warned, this first step is a big one. Another example: The analysis of the multistep network in Ch. 6 does not generalize in any simple way to allow weighted averages of previous time steps. It is unclear whether qualitatively new dynamics would arise if different weights were allowed. One thing is certain: the present analytical approach quickly runs aground when the assumption of equal weights is relaxed.

Finally, we end with some interesting but unanswered questions:

-- Why is chaos so rare in finite-size networks? Can a "learning algorithm" be developed to train a network to be chaotic?

-- A related question is how to train a network to possess higher dimensional attractors, or chaotic attractors with a specific (noninteger) dimension. Measuring the dimension of an attractor is straightforward, so there ought to be a way to develop a training algorithm which yields an attractor of arbitrary dimension.

-- How can nonsigmoidal neuron transfer functions be used to advantage? One example, the smooth staircase function, is discussed in Ch. 5. What about nonmonotonic functions? Certainly from a dynamics point of view nonmonotonic neurons are more interesting. But interesting is not what one wants in a neural network. Boring and predictable make for robust computation.

-- What other sorts of problems, besides associative memory and a few *ad hoc* optimization problems, can take advantage of the extensive feedback of Hopfield-type networks? An answer to this question will determine whether this model is destined to become a valuable technology or an academic curiosity.

Chapter 10

APPENDIX:
REPRINTS OF CHARGE-DENSITY WAVE PAPERS

Reprinted with permission from the American Physical Society

REFERENCES

Abbott, L. F., (1990), Network **1**, 105.

Aguirregabiria, J. M., and J. R. Etxebarria, (1987), Phys. Lett. A **122**, 241.

Aihara, K., T. Takabe, and M. Toyoda, (1990), Phys. Lett. A **144**, 333.

Amari, S., (1971), Proc. IEEE, **59**, 35.

Amari, S., (1972), IEEE Trans. **SMC-2**, 643.

Amit, D. J., H. Gutfreund, and H. Sompolinsky, (1985a), Phys. Rev. A **32**, 1007.

Amit, D. J., H. Gutfreund, and H. Sompolinsky, (1985b), Phys. Rev. Lett. **55**, 1530.

Amit, D. J., H. Gutfreund, and H. Sompolinsky, (1987), Ann. Phys. NY **173**, 30.

Amit, D. J., (1989), *Modeling Brain Function: The World of Attractor Neural Networks*, (Cambridge University Press, Cambridge).

an der Heiden, U., (1979), J. Math. Biol. **8**, 345.

an der Heiden, U., (1980), *Analysis of Neural Networks*, Vol. 35 of *Lecture Notes in Biomathematics* (Springer, New York).

Anderson, D. Z., (1988), Ed., *Neural Information Processing Systems, Denver CO, 1987* (AIP, New York).

Ashcroft, N. W., and N. D. Mermin, (1976), *Solid State Physics*, (Saunders College, Philadelphia).

Babcock, K. L., and R. M. Westervelt, (1986a), Physica **23D**, 464.

Babcock, K. L., and R. M. Westervelt, (1986b), Physica **28D**, 305.

Barlow, R. B., and A. J. Fraioli, (1978), J. Gen. Physiol., **71**, 699.

Basharan, G., Y. Fu and P. W. Anderson, (1986), J. Stat. Phys. **45**, 1.

Baudet, G. M., (1978), J. Assoc. Comp. Mach. **25**, 226.

Bauer, M., and W. Martienssen, (1989), Europhys. Lett. **10**, 427.

Bellman, R., and K. L. Cooke, (1963), *Differential-Difference Equations* (Academic Press, New York).

Bergé, P., Y. Pomeau, and C. Vidal, (1984), *Order within Chaos*, (Hermann, Paris).

Bilbro G., R. Mann, T. Miller, W. Snyder, D. Van den Bout, and M. White, (1989), in *Advances in Neural Information Processing, Denver CO 1988*, edited by D. S. Touretzky (Morgan Kaufmann, San Mateo), p. 568.

Bilbro G., and W. Snyder, (1989), in *Advances in Neural Information Processing, Denver CO 1988*, edited by D. S. Touretzky, (Morgan Kaufmann, San Mateo), p. 594.

Binder, K., and A. P. Young, (1986), Rev. Mod. Phys. **58**, 801.

Blake, A., and A. Zisserman, (1987),*Visual Reconstruction* (MIT Press, Cambridge, MA).

Blume, M., (1966), Phys. Rev. **141**, 517.

Bray, A. J., and M. A. Moore, (1979), J. Phys. C **12**, L441.

Bray, A. J., and M. A. Moore, (1980), J. Phys. C **13**, L469.

Bray, A. J., and M. A. Moore, (1981), J. Phys. C **14**, 1313.

Brout, R., (1959), Phys. Rev. **115**, 824.

Brout, R., and H. Thomas, (1967), Physics **3**, 317.

Bruce, A. D., E. J. Gardner, and D. J. Wallace, (1987), J. Phys. A **20**, 2909.

Buhmann, J., and K. Schulten, (1987), Europhys. Lett. **4**, 1205.

Burgess, N., and M. A. Moore, (1989), J. Phys. A **22**, 4599.

Cabasino, S., E. Marinari, P. Paolucci, and G. Parisi, (1988), J. Phys. A **21**, 4201.

Capel, W., (1966), Physica **32**, 966.

Carpenter, G. A., M. A. Cohen, and S. Grossberg, (1987), Science **235**, 1226.

Chaffee, N., (1971), J. Math. Anal. and Appl. **35**, 312.

Choi, M. Y., and B. A. Huberman, (1983a), Phys. Rev. B **28**, 2547.

Choi , M. Y., and B. A. Huberman, (1983b), Phys. Rev. B **31**, 2862.

Chowdhury, D. (1986), *Spin Glasses and Other Frustrated Systems*, (Princeton University Press, Princeton, NJ).

Cohen, M. A., and S. Grossberg, (1983), IEEE Trans. **SMC-13**, 815.

Coleman, B. D., and G.H. Renninger, (1975), J. Theor. Biol. **51**, 243.

Coleman, B. D., and G.H. Renninger, (1976), SIAM J. Appl. Math. **31**, 111.

Coleman, B. D., and G.H. Renninger, (1978), Math. Biosc. **38**, 123.

Coolen, A. C. C., and C. C. A. M. Gielen, (1988), Europhys. Lett. **7**, 281.

Cragg, B. G., and H. N. V. Temperley, (1954), Electroenceph. Clin. Neuro. **6**, 85.

Crisanti, A., and H. Sompolinsky, (1987), Phys. Rev. A **36**, 4922.

Crisanti, A., and H. Sompolinsky, (1988), Phys. Rev. A **37**, 4865.

De Dominicis, C. , M. Gabay, T. Garel, and H. Orland, (1980), J. Physique **41**, 923.

Dehaene, S., J. P. Changeux, and J. P. Nadal, (1987), Proc. Nat. Acad. Sci. USA **84**, 2727.

Denker, J. S., (1986a), Physica **22D**, 216.

Denker, J. S., (1986b), Ed., *Neural Networks for Computing*, AIP Conf. Proc. 151 (American Institute of Physics, New York).

Denker, J. S., (1986c), in *Neural Networks for Computing*, edited by J. S. Denker, AIP Conf. Proc. 151 (American Institute of Physics, New York)., p. 121.

Derrida, B., (1988a), J. Phys. A **20**, L721.

Derrida, B., (1988b), in *Nonlinear Evolution and Chaotic Phenomena*, Proc. of NATO Adv. Study Workshop - Noto, Sicily, Italy, June 1987. (Plenum, NY), p. 213.

Derrida, B., and H. Flyvbjerg, (1986), J. Phys. A **19**, L1003.

Derrida, B., and H. Flyvbjerg, (1987a), J. Physique (Paris) **48**, 971.

Derrida, B., and H. Flyvbjerg, (1987b), J. Phys. A **20**, 5273.

Derrida, B., and E. Gardner, (1986), J. Physique (Paris) **47**, 959.

Diederich, S., and M. Opper, (1987), Phys. Rev. Lett. **58**, 949.

Domany, E., W. Kinzel, and R. Meir, (1989), J. Phys. A **22**, 2081.

Dowling, J. E., (1987), *The retina: An approachable part of the brain* (Harvard Univ. Press, Cambridge, MA).

Durbin R., and D. Willshaw, (1987), Nature **326**, 689.

Edwards, S. F., and R. C. Jones, (1976), J. Phys. A **9**, 1595.

Ettelaie, R., and M. A. Moore, (1985), J. Physique Lett. **46**, L893.

Farmer, J. D., (1982), Physica **4D**, 366.

Fischer, K. H., (1976), Solid State Commun., **18**, 1515.

Fleisher, M., (1988), in *Neural Information Processing Systems, Denver CO, 1987*, edited by D. Z. Anderson (AIP, New York), p. 278.

Fontanari, J. F., and R. Koberle, (1988a), J. Phys. (France) **49**, 13.

Fontanari, J. F., and R. Koberle, (1988b), J. Phys. A **21**, L259.

Fontanari, J. F., and R. Koberle, (1988c), J. Phys. A **21**, L667.

Forrest, B. M., (1988), J. Phys. A **21**, 245.

Fradkin, E., B. A. Huberman, and S. H. Shenker, (1978), Phys. Rev. **B18**, 4789.

Frumkin, A., and E. Moses, (1986), Phys. Rev. A **34**, 714.

Fu, Y., and P. W. Anderson, (1986), J. Phys A **19**, 1605.

Fukai, T., (1990), J. Phys. A **23**, 249.

Gardner, E. J., (1986), J. Phys. A **19**, L1047.

Gardner, E. J., (1988), J. Phys. A **21**, 257.

Geman, S., (1980), Ann. Prob. **8**, 252.

Glass, L. and M. C. Mackey, (1988), *From Clocks to Chaos - The Rhythms of Life*, (Princeton University Press).

Golden, R. M., (1986), J. Math. Psych. **30**, 73.

Goles, E., and G. Y. Vichniac, (1986), in *Neural Networks for Computing*, edited by J. S. Denker, AIP Conf. Proc. 151 (American Institute of Physics, New York), p.165.

Goles-Chacc, E., F. Fogelman-Soulie and D. Pellegrin, (1985), Disc. Appl. Math. **12**, 261.

Grinstein, G., C. Jayaprakash, and Y. He, (1985), Phys. Rev. Lett. **55**, 2527.

Grossberg, S., (1970), Stud. Appl. Math. **44**, 135.

Grossberg, S., (1988), Neural Networks **1**, 17.

Guckenheimer, J., and P. Holmes, (1983), *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, NY).

Gutfreund, H., and M. Mezard, (1988), Phys. Rev. Lett. **61**, 235.

Gutfreund, H., J. D. Reger, and A. P. Young, (1988), J. Phys. A, **21**, 2775.

Guyon, I., L. Personnaz, J. P. Nadal, and G. Dreyfus, (1988), Phys. Rev. A **38**, 6365.

Hadeler, K. P., and J. Tomiuk, (1977), Arch. Rat. Mech. Anal. **65**, 87.

Hale, J. K., and N. Sternberg, (1988), J. Comp. Phys. **77**, 221.

Harris, B., (1960), Ann. Math. Stat. **31**, 1045.

Hebb, D. O., (1949), *The Organization of Behavior* (Wiley, New York).

Hentschel, H. G. E., and A. Fine, (1989), Phys. Rev. A **40**, 3983.

Hertz, J. A., G. Grinstein, and S. A. Solla, (1987) in *Proc. Heidelberg Colloquium on Glassy Dynamics*, Vol. 275 of *Lecture Notes in Physics*, edited by L. J. van Hemmen and I. Morgenstern, (Springer, Berlin).

Herz, A., B. Sulzer, R. Kühn, and J. L. van Hemmen, (1988), Europhys. Lett. **7**, 663.

Hirsch, M. W., (1987), *Convergence in Neural Nets*, in *Proc. of the IEEE Conf. on Neural Networks, San Diego, CA*, edited by M. Caudill and C. Butler, (IEEE, New York).

Hirsch, M. W., (1989) , Neural Networks **2**, 331.

Hopfield, J. J., (1982), Proc. Nat. Acad. Sci. USA **79**, 2554.

Hopfield, J. J., (1984), Proc. Nat. Acad. Sci. USA **81**, 3008.

Hopfield, J. J., and D. W. Tank, (1985), Biol. Cybern. **52**, 141.

Hopfield, J. J., and D. W. Tank, (1986), Science **233**, 625.

Jeffrey, W., and R. Rosner, (1986a), Astr. J. **310**, 473.

Jeffrey, W., and R. Rosner, (1986b), in *Neural Networks for Computing*, edited by J. S. Denker, AIP Conf. Proc. 151 (American Institute of Physics, New York), p.241.

Kanter, I., (1989), Phys. Rev. A **40**, 2611.

Kanter, I., (1990), Europhys. Lett. **11**, 397.

Kanter, I., and H. Sompolinsky, (1987), Phys. Rev. A **35**, 380.

Kauffman, S. A. (1969), J. Theor. Biol. **22**, 437.

Kauffman, S. A. (1984), Physica **10D**, 145.

Kauffman, S. A. (1990), *Origins of Order: Self Organization in Evolution*, (Oxford University Press, Oxford).

Keeler , J. D., (1986), in *Neural Networks for Computing*, edited by J. S. Denker, AIP Conf. Proc. 151 (American Institute of Physics, New York), pg. 259.

Kepler, T. B., (1989), *Model Neural Networks*, Ph. D. dissertation, Brandeis University.

Kepler, T. B., and L. F. Abbott, (1988), J. Physique **49**, 1657.

Kepler, T. B., S. Datt. R. B. Meyer. and L. F. Abbott, (1989), *Chaos in a neural network circuit*, Brandies University preprint BRX-265.

Kerszberg, M., and A. Zippelius, (1989), *Synchronization in Neural Assemblies*, preprint .

Kirkpatrick, S., C. D. Gelatt, Jr. and M. P. Vecchi, (1983), Science **220**, 671.

Kirkpatrick, S., D. Sherrington, (1978), Phys. Rev. B **17**, 4384.

Kleinfeld, D., (1986), Proc. Nat. Acad. Sci. USA **83**, 9469.

Koch, C., J. Marroquin, and A. Yuille, (1986), Proc. Nat. Acad. Sci. USA **83**, 4263.

Kolmanovskii, V.B., and V.R. Nosov, (1986), *Stability of Functional Differential Equations* (Academic Press, New York).

Krauth, W., M. Mezard, and J. P. Nadal, (1988), Complex Systems **2**, 387.

Kühn, R., J. L. van Hemmen, and U. Riedel, (1989), J. Phys. A, **22**, 3123.

Kürten, K. E., (1988), Phys. Lett. **129A**, 157.

Kürten, K. E., and J. W. Clark, (1986), Phys. Lett. **114A**, 413.

Li, T., (1981), Phys. Rev. B **24**, 6579.

Ling, D. D., D. R. Bowman, and K. Levin, (1983), Phys. Rev. B **28**, 262.

Lippmann, R. P., (1987), IEEE ASSP, April issue, 4.

Little, W. A., (1974), Math. Biosci. **19**, 101.

Little, W. A., and G. L. Shaw, (1978), Math. Biosci. **39**, 281.

Lynch, G., (1986), *Synapses, Circuits, and the Beginnings of Memory*, (MIT Press, Cambridge, MA).

Mackey, M. C., and U. an der Heiden, (1984), J. Math. Biology, **19**, 221.

Mackey, M. C., and L. Glass, (1977), Science **197**, 287.

Maddox, J., (1987), *Modelling for its own sake*, Nature **328**, 571.

Marcus, C. M., and R. M.Westervelt, (1988), in *Neural Information Processing Systems, Denver CO, 1987*, edited by D. Z. Anderson, (AIP, New York), p. 524.

Marcus, C. M., and R. M.Westervelt, (1989a), Phys. Rev. A **39**, 347.

Marcus, C. M., and R. M.Westervelt, (1989b), in *Advances in Neural Information Processing, Denver CO 1988*, edited by D. S. Touretzky (Morgan Kaufmann, San Mateo), p. 568.

Marcus, C.M., and R. M. Westervelt, (1989c), Phys. Rev. A **40**, 501.

Marcus, C. M., and R. M.Westervelt, (1990), to appear in Phys. Rev. A.

Marcus, C. M., F. R. Waugh, and R. M.Westervelt, (1990), Phys. Rev. A, **41**, 3355.

Mattis, D. C., (1976), Phys. Lett. **56A**, 421.

May, R. M., (1974), *Stability and Complexity in Model Ecosystems, 2nd Ed.* (Princeton University Press, New Jersey).

May, R. M., (1976), Nature **261**, 459.

McCulloch, W., and W. Pitts, (1943), Bull. Math. Biophys. **5**, 115. Reprinted in *Brain Theory, Reprint Volume, Advanced Series in Neuroscience, Vol. 1*,ed. by G. L. Shaw and G. Palm (World Scientific, Singapore, 1988).

Mead, Carver A., (1989), *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, MA).

Meir, R., and E. Domany, (1987), Phys. Rev. Lett. **59**, 359.

Meir, R., and E. Domany, (1988), Phys. Rev. A **37**, 608.

Meunier, C., D. Hansel and A. Verga, (1989), J. Stat. Phys. **55**, 859.

Mezard, M., G. Parisi, N. Sourlas, G. Toulouse, and M. A. Virasoro, (1984a), Phys. Rev. Lett. **52**, 1156.

Mezard, M., G. Parisi, N. Sourlas, G. Toulouse, and M. A. Virasoro, (1984b), J. Physique (Paris) **45**, 843.

Mezard, M., G. Parisi, and M. A. Virasoro, (1987), *Spin Glass Theory and Beyond*, (World Scientific, Singapore).

Moore, M. A., (1987), Phys. Rev. Lett. **58**, 1703.

Mukherjee, A., (1985), *Introduction to nMOS and CMOS VLSI System Design*, (Prentice Hall, New Jersey).

Nemoto, K., and H. Takayama, (1985), J. Phys. C **18**, L529.

Nishimori, H., T. Nakamura, and M. Shiino, (1990) Phys. Rev. A **41**, 3346.

Ortega, J. M., and W. C. Rheinboldt, (1970), *Iterative solution of nonlinear equations in several variables*, (Academic Press, NY).

Opper, M., J. Kleinz, and W. Kinzel, (1989), J. Phys. A **22**, L407.

Parga, N., (1987), J. Physique (Paris) **48**, 499.

Parisi, G., (1986), J. Phys. A **19**, L675.

Peretto, P., (1984), Biol. Cybern. **50**, 51.

Personnaz, L., I. Guyon, and G. Dreyfus, (1985), J. Phys. Lett. (Paris) **46**, L359.

Rammal, R. G. Toulouse, and M. A. Viarsoro, (1986), Rev. Mod. Phys. **58**, 765.

Ratliff , F., (1965), *Mach Bands: Quantitative Studies on Neural Networks in the Retina*, (Holden-Day, San Francisco).

Ratliff, F., (1974), Ed., *Studies on Exitation and Inhibition in the Retina*, (Rockefeller University Press, New York).

Reger, J. D., and K. Binder, (1985), Z. Phys. B, **60**, 137.

Reger, J. D., K. Binder, and W. Kinzel, (1984), Phys. Rev. B **30**, 4028.

Renals, S., and R. Rohwer, (1990), J. Stat. Phys. **58**, 825.

Riedel, U., R. Kühn, and J. L. van Hemmen, (1988), Phys. Rev. A **38**, 1105.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group, (1986), *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol 1 and 2* (MIT Press, Cambridge, MA).

Shaw, G. L., and G. Palm, Eds., (1988), *Brain Theory, Reprint Volume, Advanced Series in Neuroscience, Vol. 1*,(World Scientific, Singapore, 1988).

Sherrington, D. and S. Kirkpatrick, (1975), Phys. Rev. Lett. **35**, 1792.

Shinomoto, S., (1986), Prog. Theor. Phys. **75**, 1313.

Silverstein, J. W., (1985), Ann. Prob. **13**, 1364.

Sommers, H. J., A. Crisanti, H. Sompolinsky, and Y. Stein, (1988), Phys. Rev. Lett. **60**, 1895.

Sompolinsky, H., (1986), Phys. Rev. A **34**, 2571.

Sompolinsky, H., (1988), Physics Today **41**, No. 12, 70.

Sompolinsky, H., A. Crisanti, and H. J. Sommers, (1988), Phys. Rev. Lett. **61**, 259.

Sompolinsky, H., and I. Kanter, (1986), Phys. Rev. Lett. **57**, 2861.

Soukoulis, C. M., K. Levin, and G. S. Grest, (1982), Phys. Rev. Lett. **48**, 1756.

Soukoulis, C. M., K. Levin, and G. S. Grest, (1983), Phys. Rev. B **28**, 1495.

Spitzner, P., and W. Kinzel, (1989), Z. Phys. B **77**, 511.

Sze, S. M., (1981), *Physics of Semiconductor Devices, 2e*, (John Wiley, New York).

Tanaka, F. and S. F. Edwards, (1980), J. Phys. F **10**, 2769.

Thouless, D. J., P. W. Anderson, and R. G. Palmer, (1977), Phil. Mag. **35**, 593.

Toulouse, G., (1977), Commun. Phys. **2**, 115.

Toulouse, G., (1989), J. Phys. A **22**, 1959.

Treves, A., and D. J. Amit, (1988), J. Phys. A **21**, 3155.

van Hemmen, L. J., (1987), Phys. Rev. A **36**, 1959.

van Hemmen, L. J., and I. Morgenstern, (1987), Eds., *Proc. Heidelberg Colloquium on Glassy Dynamics*, Vol. 275 of *Lecture Notes in Physics* (Springer, Berlin).

Vannimenus, J., G. Toulouse, and G. Parisi, (1981), J. Physique (Paris) **42**, 565.

Wannier, G. M., (1950), Phys. Rev. **79**, 357.

Waugh, F. R., C. M. Marcus, and R. M. Westervelt, (1990), Phys. Rev. Lett. **64**, 1986.

Weinitschke, H. J., (1964), Num. Math. **6**, 395.

Wigner, E. P., (1958), Ann. Math. **67**, 325.

Wyatt, Jr., J. L., and D. L. Stanley, (1988), in *Neural Information Processing Systems, Denver CO, 1987*, edited by D. Z. Anderson, (AIP, New York), p. 860.

Yedidia, J. S., (1989), J. Phys.A **22**, 2265.

Yuille, A., (1989), Biol. Cybern. **61**, 115.

Simple Model of Collective Transport with Phase Slippage

S. H. Strogatz,[a] C. M. Marcus, and R. M. Westervelt
Division of Applied Sciences and Department of Physics, Harvard University, Cambridge, Massachusetts 02138

R. E. Mirollo
Department of Mathematics, Brown University, Providence, Rhode Island 02912
(Received 29 April 1988)

We present a mean-field analysis of a many-body dynamical system which models charge-density-wave transport in the presence of random pinning impurities. Phase slip between charge-density-wave domains is modeled by a coupling term that is periodic in the phase differences. When driven by an external field, the system exhibits a first-order depinning transition, hysteresis, and switching between pinned and sliding states, and a delayed onset of sliding near threshold.

PACS numbers: 71.45.Lr, 03.20.+i, 05.45.+b, 72.15.Nj

Collective transport in coupled dynamical systems is a topic of considerable current interest.[1] An experimental example is the nonlinear conduction seen in charge-density-wave (CDW) samples.[2,3] When a sufficiently strong dc electric field is applied to a sample with a static CDW, the CDW depins from impurities in the lattice and begins to slide and carry current. Classical models of CDW transport[4-10] assume that the dynamics are dominated by competition between the internal elasticity of the CDW and the local potentials of randomly spaced impurities.

These models of CDW transport do not account for experimentally observed hysteresis, switching, and delayed conduction in "switching samples."[11-15] CDW inertia,[16] current noise,[17] and avalanche depinning[18] have been proposed to account for switching. More recently, switching and hysteresis have been ascribed to phase slippage in the CDW.[12-15] A physical model for a CDW in a switching sample is a collection of domains, each with a well-defined phase, separated by regions where the amplitude of the CDW is weak.[12,15] Phase slip can occur easily in these connecting regions. A rigorous theory of CDW transport for this case is very difficult, although a detailed analysis of a phenomenological model with a few degrees of freedom has been presented.[15] It is also uncertain which of the observed complex phenomena are intrinsic and which are properties of particular samples or experimental treatments.

In this Letter we present a simple model of collective transport which is applicable to CDW transport in switching samples. The model consists of many phases which represent the states of CDW domains, and phase slip due to amplitude collapse[15] is modeled by a weak-coupling term periodic in the phase differences. This is a simple modification of a well-understood model[6-8] with elastic coupling and no phase slip. As we will show, the periodic coupling gives rise to switching, hysteresis, and delayed conduction. Our approach is to analyze a simple model which may have some generality, rather than to make a detailed phenomenological treatment specific to charge-density waves.

The Hamiltonian is

$$H=\frac{J}{2N}\sum_{i,j}[1-\cos(\theta_i-\theta_j)]+b\sum_j[1-\cos(\theta_j-a_j)], \tag{1a}$$

and we assume zero temperature and relaxational dynamics with a driving field

$$\dot{\theta}_j=-\frac{\partial H}{\partial\theta_j}+E_0,\quad j=1,\ldots,N. \tag{1b}$$

The θ_j represent the phases of weakly coupled domains,[8,9] In other models,[6,7] θ_j represent the phase distortion of the CDW at the jth pinning site. In Eq. (1), J is the coupling strength, b is the pinning strength, a_j is a pinning phase randomly distributed on $[-\pi,\pi]$, and E_0 is an electric field applied along the CDW wave vector. The coupling term favors phase coherence, whereas the random fields try to pin each θ_j at a_j. For weakly coupled domains, the ratio $K=J/b$ is small. The infinite-range coupling in Eq. (1) corresponds to the mean-field approximation, also used for previous work.[6-8]

The model (1) is closely related to the system studied by Fisher.[6] The difference is that in the Hamiltonian Eq. (1a) we have assumed a periodic coupling $1-\cos(\theta_i-\theta_j)$ rather than a quadratic coupling $\frac{1}{2}(\theta_i-\theta_j)^2$. The periodic coupling in Eq. (1a) allows for phase-slip processes[6,11] and corresponds physically to the effects of CDW defects[1,19] or amplitude collapse[12,15] between coherent regions of the CDW. In particular, the model is appropriate in CDW systems with strong pinning centers that favor the formation of weakly coupled domains.[12,20] We have made the simplifying assumption that the argument of the periodic coupling is the phase

difference $\theta_i-\theta_j$ rather than a more general multiple $\lambda(\theta_i-\theta_j)$, where λ reflects the amount of polarization that can be built up before phase slip occurs. Additional metastable states[10] with different polarizations can exist below the depinning threshold for $\lambda\neq 1$; these are not present in our model.

We first consider the static configuration of the system when $E_0=0$. The phase coherence of the equilibrium configurations depends on the normalized coupling strength K. For instance, in the absence of coupling ($K=0$), the θ_j are pinned at α_j and are completely incoherent, whereas for $K\to\infty$, there is perfect coherence ($\theta_i=\theta_j$ for all i,j). To characterize the transition from incoherence to coherence, we define a complex order parameter

$$re^{i\phi}=N^{-1}\sum_j \exp(i\theta_j),$$

where r measures the coherence and ϕ is the average phase.

We now show analytically that there is a first-order transition in the model from the incoherent state ($r=0$) to the coherent state ($r\approx 1$) at $K=2$, when the domains are strongly coupled. This zero-field transition is an artifact of mean-field theory, but a related hysteretic transition occurs for nonzero E_0 in the physically relevant weak-coupling regime, as discussed below. The strategy of the analysis is to derive a self-consistent equation for r, by use of the fact that r determines the equilibrium phases θ_j and is in turn determined by them.

Equilibria of H satisfy $\partial H/\partial\theta_j=0$, i.e.,

$$\sin(\alpha_j-\theta_j)+\frac{K}{N}\sum_i \sin(\theta_i-\theta_j)=0.$$

Rewriting the sum in terms of the order parameter yields

$$\sin(\alpha_j-\theta_j)+Kr\sin(\phi-\theta_j)=0. \tag{2}$$

We may set $\phi=0$ because there is no absolute phase reference. This choice removes the rotational degeneracy. Solving Eq. (2) for θ_j yields

$$e^{i\theta_j}=\left[\frac{Kr+e^{i\alpha_j}}{Kr+e^{-i\alpha_j}}\right]^{1/2}. \tag{3}$$

Combining Eq. (3) with $r=N^{-1}\sum_j\exp(i\theta_j)$ and letting $N\to\infty$, we obtain the self-consistency relation for r:

$$r=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{Kr+\cos\alpha}{(1+2Kr\cos\alpha+K^2r^2)^{1/2}}\,d\alpha. \tag{4}$$

For each K, the values of r that satisfy Eq. (4) may be found as follows [Fig. 1(a)]. Let $u=Kr$ and let $f(u)$ denote the integral in Eq. (4), which may be expressed exactly as an elliptic integral.[21] Because $f(u)$ and u/K are both equal to r, the intersections of $f(u)$ and the line u/K yield solutions for the coherence r, given the normalized coupling strength K [Fig. 1(a)].

Figure 1(b) shows the first-order transition between incoherent and coherent states. The incoherent state

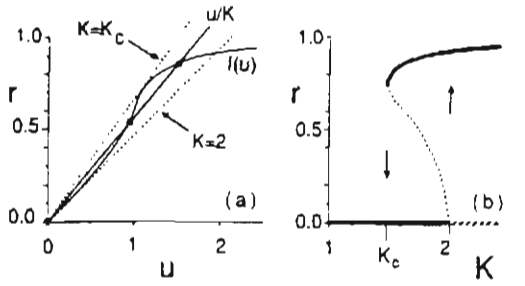


FIG. 1. (a) Solid lines indicate the integral $f(u)$ plotted from Eq. (4) together with the line u/K (see text). Equilibrium solutions for r occur where $f(u)$ intersects the line u/K. For the value of K shown, three solutions exist (filled circles). Dashed lines show u/K for the critical values $K=K_c$ and $K=2$. (b) Plot of the exact equilibrium solutions for r vs K: solid lines, locally stable equilibria; broken lines, unstable equilibria.

$r=0$ always solves Eq (4). An unstable second branch of the solutions bifurcates from $r=0$ at $K=2$, with $r\sim(2-K)^{1/2}$ as can be seen from Eq. (4) and the series expansion $f(u)=u/2+u^3/16+O(u^5)$, valid for $u<1$. We have also proven[21] that $r=0$ is locally stable for $K<2$ and unstable for $K>2$. A locally stable third branch of solutions, with $r\approx 1$, is created when u/K intersects $f(u)$ tangentially [Fig. 1(a)] at $K=K_c\approx 1.489$. Note that for K between K_c and 2, the system is bistable. We emphasize that this first-order transition is a consequence of the cosine coupling in Eq. (1a) and would not be seen if a quadratic coupling were assumed.

We turn now from statics to dynamics. In the presence of a driving field, the equations of motion from Eq. (1b) are

$$d\theta_j/dt=E+Kr\sin(\phi-\theta_j)+\sin(\alpha_j-\theta_j). \tag{5}$$

By letting $E=E_0/b$ and time $t\to bt$, we have set $b=1$ without loss of generality; as before, $K=J/b$. The second term on the right-hand side of Eq. (5) is the collective torque exerted on θ_j by all other phases. For $E=0$ and small K, the phase coherence $r=0$ and therefore the collective torque is zero. If r becomes nonzero, the collective torque provides a positive feedback which tries to increase r further by aligning each θ_j with the average phase ϕ. The physical consequences of this process are hysteresis and delayed conduction, as discussed below. In our model hysteresis and switching result from the transition to coherence of a randomly pinned state. Incoherence of the pinned state occurs naturally for a large number of random pinning phases α_j; systems as small as three phases show hysteresis and switching, but only when the α_j are chosen evenly spaced on $[-\pi,\pi]$. Thus in our model these phenomena are associated with many degrees of freedom.

Figure 2 plots the regions of stability of the pinned and sliding CDW states. The pinned state ($d\theta_j/dt = 0$) in this model can be analytically shown to be incoherent ($r = 0$). Using variational stability analysis about the pinned state, we have proven[21] that this state becomes unstable above the depinning threshold field $E_T = (1 - K^2/4)^{1/2}$ when $K < 2$, as shown in Fig. 2. For strong coupling, $K > 2$ where the model is not physically relevant, $E_T = 0$ and the CDW slides ($d\phi/dt > 0$) for any fields $E > 0$. This is an artifact of mean-field theory which also occurs in models[6-8] with elastic coupling. Numerical solutions of Eq. (5) show that the sliding state is always coherent ($r > 0$). The sliding state becomes pinned and incoherent below a separate pinning threshold E_P shown as the dashed line in Fig. 2, which was calculated numerically from the initial condition $r = 1$. This boundary extends from the critical value K_c found analytically for $E = 0$, also shown in Fig. 1. The solid and dashed lines in Fig. 2 bound a hysteretic region where both pinned and sliding solutions are stable; the final state reached depends on the initial conditions. The physically relevant weak-coupling region of Fig. 2 is for $K < K_c$, where E_T and E_P are nonzero.

The model predicts hysteresis and switching between pinned and sliding states as illustrated by the numerical solutions of Eq. (5) shown in Fig. 3. As E is increased slowly past E_T, the induced collective velocity $d\phi/dt$ corresponding to the CDW current jumps up discontinuously, then increases nearly linearly. If E is then decreased, the velocity $d\phi/dt$ decreases and then drops discontinuously to zero at the separate pinning threshold $E = E_P$ as shown in Fig. 3. When the CDW pins, the coherence r also drops discontinuously to zero. This loss of coherence is illustrated in the limit $E_P = 0$ for the analytical results in Fig. 1(b). Hysteretic current-voltage curves have been seen in low-temperature experiments on CDW samples with dilute impurities or irradiation-induced defects, which act as dilute, strong pinning sites.[11-14] The switching and hysteresis predicted by the model depend crucially on the periodic coupling in Eq. (1a); neither switching nor hysteresis are predicted for quadratic coupling.[6-8]

The model exhibits delayed conduction above the depinning threshold E_T when $E < 1$. Numerical solutions of Eq. (5) were used to study the evolution of the system from a random initial state. The system first rapidly reaches an incoherent configuration with $\theta_j \approx a_j + \sin^{-1}E$, then gradually develops coherence, and finally depins suddenly when r becomes appreciable.[21] The delay before depinning increases near the threshold E_T, as observed experimentally.[11,14] If $E > 1$, switching occurs immediately.

Numerical solutions of Eq. (5) show that the individual phases do not move with a constant velocity in the sliding state, although the collective velocity $d\phi/dt$ is constant. Near threshold, the motion of each phase is periodic, alternating between rapid advances by nearly 2π, and slow creep about its pinning phase. In this respect, Eq. (5) and other mean-field models[6-8] agree with results from more realistic short-range models,[10] and with recent experiments[12,20] which suggest a spatially nonuniform rate of CDW phase advance near threshold. An artifact of the mean-field approximation is that all the phases θ_j execute identical periodic motions shifted in time and phase.

We have also analyzed the dynamics of Eq. (5) far above the depinning threshold. For $E \gg E_j$, perturbation theory[21] yields $(d\phi/dt)/E = 1 - (1/2E^2) + O(E^{-4})$. Thus, the deviation from the limiting dc conductivity as $E \to \infty$ is proportional to E^{-n} with $n = 2$, in agreement with some CDW models[6,8,22] and in contrast to the value $n = \frac{1}{2}$ predicted by others.[5] The available data for high-field conductivity in CDW's[23] suggest $n \approx 1-2$.

Simplification of approximations in the model are infinite sample size N and infinite-range coupling. Solu-

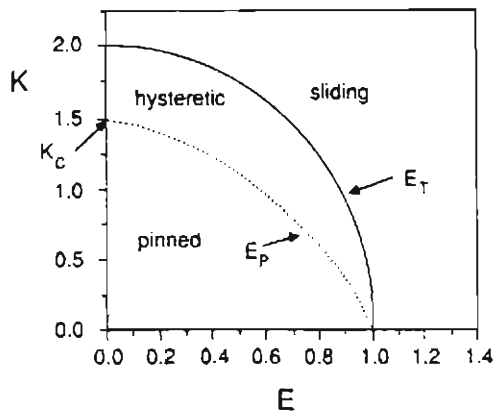


FIG. 2. Stability diagram for the model Eq. (5): solid line, depinning threshold $E_T = (1 - K^2/4)^{1/2}$ determined analytically; dashed line, pinning threshold E_P obtained by numerical integration of Eq. (5). Note the presence of hysteretic region.

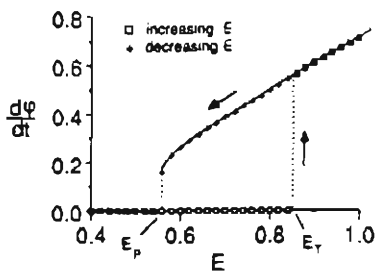


FIG. 3. Hysteresis and switching between pinned and sliding states. Data points obtained for $N = 300$ phases by numerical integration of Eq. (5) with $K = 1$, for which $E_T = (\frac{1}{4})^{1/2}$. The curve is a guide for the eye.

tions of the infinite-range model are relatively insensitive to N, and closely approximate the results presented here. To assess the effects of infinite-range coupling, we have numerically integrated Eq. (1) on a cubic lattice in three dimensions with nearest-neighbor coupling. The numerical solutions show hysteresis and switching,[21] though over a reduced range in E. For $N=216$ and $N=1000$ sites, the width of the hysteresis is approximately 20% and 15%, respectively, of the width predicted by the infinite-range model for the same values of N. Thus the qualitative behavior is similar to the mean-field theory, at least for finite sample sizes, but the thresholds are quantitatively different.

In summary, we have analyzed a dynamical system of many driven phases with random pinning and infinite-range coupling. The periodic coupling in the model gives rise to a first-order depinning transition, hysteresis, and switching between pinned and sliding states, and a time delay before the onset of sliding. These results demonstrate that some of the complex phenomena observed experimentally in strongly pinned charge-density-wave systems can be accounted for by a simple dynamical model.

We thank P. B. Littlewood, P. A. Lee, and B. I. Halperin for helpful discussions. One of us (S.H.S.) acknowledges financial support by the NSF, and another (C.M.M.) acknowledges financial support from AT&T Bell Laboratories. This research was supported in part by ONR Contract No. N00014-84-K-0465.

[a]Also at Department of Mathematics, Boston University, Boston, MA 02215.

[1]*Spatio-Temporal Coherence and Chaos in Physical Systems*, edited by A. R. Bishop, G. Grüner, and B. Nicolaenko (North-Holland, Amsterdam, 1986).

[2]*Charge Density Waves in Solids*, edited by G. Hutiray and J. Solyom, Lecture Notes in Physics, Vol. 217 (Springer-Verlag, Berlin, 1985).

[3]G. Grüner and A. Zettl, Phys. Rep. **119**, 117 (1985).

[4]H. Fukuyama and P. A. Lee, Phys. Rev. B **17**, 535 (1978); P. A. Lee and T. M. Rice, Phys. Rev. B **19**, 3970 (1979).

[5]L. Sneddon, M. C. Cross, and D. S. Fisher, Phys. Rev. Lett. **49**, 292 91982); P. B. Littlewood and C. M. Varma, Phys. Rev. B **36**, 480 (1987).

[6]D. S. Fisher, Phys. Rev. Lett. **50**, 1486 (1983), and Phys. Rev. B **31**, 1396 (1985).

[7]L. Sneddon, Phys. Rev. B **30**, 2974 (1984); P. Alstrom and R. K. Ritala, Phys. Rev. A **35**, 300 (1987).

[8]P. F. Tua and A. Zawadowski, Solid State Commun. **49**, 19 (1984); J. C. Gill, in Ref. 2, p. 377.

[9]M. Inui and S. Doniach, Phys. Rev. B **35**, 6244 (1987).

[10]P. B. Littlewood and T. M. Rice, Phys. Rev. Lett. **48**, 44 (1982); S. N. Coppersmith and P. B. Littlewood, Phys. Rev. B **31**, 4049 (1985); P. B. Littlewood, Physica (Amsterdam), **23D**, 45 (1986).

[11]A. Zettl and G. Grüner, Phys. Rev. B **26**, 2298 (1982).

[12]R. P. Hall, M. F. Hundley, and A. Zettl, Phys. Rev. Lett. **56**, 2399 (1986).

[13]J. Dumas, C. Schlenker, J. Marcus, and R. Buder, Phys. Rev. Lett. **50**, 757 (1983); H. Mutka, S. Bouffard, J. Dumas, and C. Schlenker, J. Phys. (Paris), Lett. **45**, L729 (1984); S. Bouffard, M. Sanquer, H. Mutka, J. Dumas, and C. Schlenker, in Ref. 2, p. 449; K. Tsutsumi, Physica (Amsterdam) **143B**, 129 (1986).

[14]G. Kriza, A. Janoczy, and G. Mihaly, in Ref. 2, p. 426.

[15]R. P. Hall, M. F. Hundley, and A. Zettl, Physica (Amsterdam) **143B**, 152 (1986); M. Inui, R. P. Hall, S. Doniach, and A. Zettl, Phys. Rev. B (to be published).

[16]R. P. Hall, M. Sherwin, and A. Zettl, Phys. Rev. B **29**, 7076 (1984).

[17]W. Wonneberger and H. J. Breymayer, Z. Phys. B **56**, 241 (1984).

[18]B. Joos and D. Murray, Phys. Rev. B **29**, 1094 (1984).

[19]N. P. Ong, G. Verma, and K. Maki, Phys. Rev. Lett. **52**, 663 (1984).

[20]J. R. Tucker, Phys. Rev. Lett. **60**, 1574 (1988), and references therein.

[21]S. H. Strogatz, C. M. Marcus, R. E. Mirollo, and R. M. Westervelt, to be published.

[22]G. Grüner, A. Zawadowski, and P. M. Chaikin, Phys. Rev. Lett. **46**, 511 (1981).

[23]M. Oda and M. Ido, Solid State Commun. **44**, 1535 (1982); N. P. Ong and X. J. Zhang, Physica (Amsterdam) **143B**, 3 (1986).

Delayed switching in a phase-slip model of charge-density-wave transport

C. M. Marcus, S. H. Strogatz, and R. M. Westervelt
Division of Applied Sciences and Department of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 30 March 1989)

We analyze the dynamics of the depinning transition in a many-body model of charge-density-wave (CDW) transport in switching samples. The model consists of driven massless phases with random pinning and a coupling term that is periodic in the phase difference, thus allowing phase slip. When the applied field in our model exceeds the depinning threshold by a small amount, there is a delay before the appearance of a coherent moving solution. This delay is also seen experimentally in switching CDW materials. We find that close to threshold the switching delay is approximately proportional to the inverse distance above threshold. Analytical results agree with numerical integration of the model equations. Results are also compared to available experimental data on delayed switching.

I. INTRODUCTION

The nonlinear conduction in charge-density-wave (CDW) materials has been extensively studied in a variety of quasi-one-dimensional metals and semiconductors, and a large number of theoretical models have been presented, each explaining some of the phenomena observed in these materials.[1,2] Classical models of CDW dynamics[3-6] which consider only the phase degrees of freedom of the CDW condensate have been quite successful at describing the behavior of both pinned and sliding CDW's in a regime where pinning forces are weak, phase distortions are small, and higher-energy amplitude modes are not excited. While many aspects of CDW dynamics are well described by a rigid phase model with only one degree of freedom,[6] dynamics near the depinning threshold are best treated by a model with many degrees of freedom, which allows for local distortion of the CDW in response to random pinning forces. For example, a mean-field discrete-site model of many coupled phases analyzed by Fisher[5] gives a continuous depinning transition with a concave-upward I-V curve at depinning, in agreement with experimental data, and in contrast to depinning of the corresponding single-phase model.

Recently, experimental and theoretical interest has turned to a class of CDW systems which have discontinuous depinning transitions and hysteresis between the pinned and sliding states.[7-23] This discontinuous depinning has been termed "switching." Several authors[11-17] (though not all[18-21]) have attributed switching to phase slip between coherent CDW domains occurring at ultrastrong pinning sites. Experimentally, it is known that switching can be induced by irradiating the sample, which creates strong pinning sites.

A very interesting phenomenon associated with switching, first reported by Zettl and Grüner[7] for switching samples of $NbSe_3$, and subsequently seen in other materials,[8-10] is a delayed onset of CDW conduction near threshold. When an electric field slightly larger than the depinning threshold is applied to a switching sample, there is a time delay before the pinned CDW begins to slide. The delay grows as threshold is approached from above and varies over several orders of magnitude from tenths to hundreds of microseconds.

In this paper, we analyze the dynamics of depinning for a many-body model of CDW transport in which phase slip is allowed. We show that, for an applied driving field slightly above the depinning threshold, switching from the pinned state to the coherent moving state is delayed, and that the delay grows roughly as the inverse of the distance above threshold. In Sec. II the phase-slip model is described and briefly compared to other models of CDW transport. The dynamics of the depinning transition in the model are then analyzed and an expression for the switching delay is derived. These results are shown to agree with direct numerical integration of the model. In Sec. III our results are compared with the available experimental data on switching delay.

II. PHASE-SLIP MODEL OF DELAYED SWITCHING

A. Mean-field model of switching CDW's

The dynamical system we will study is given by Eq. (1):

$$\frac{d\theta_j}{dt} = E + \sin(\alpha_j - \theta_j) + \frac{K}{N}\sum_{i=1}^{N}\sin(\theta_i - \theta_j), \quad j = 1, 2, \ldots, N . \qquad (1)$$

The system (1) is formally very similar to the model studied by Fisher[5] and Sneddon,[4] with the exception of the final term: the coupling between phases in (1) is periodic in the phase difference, rather than linear. Also, the physical interpretation of the phases θ_j is somewhat different than in these elastic-coupling models, as discussed below.

The phases θ_j in Eq. (1) represent the phases of CDW domains; $E > 0$ represents the applied dc field and $K > 0$ is the strength of coupling between domains. The α_j represent the preferred phase of each domain, taken to be randomly distributed on $[0, 2\pi]$. The strength of pinning is assumed to be constant for each domain, and has been

normalized to one. The values E and K thus represent the strengths of the coupling and applied field relative to pinning. The time scale similarly reflects this normalization. For all-to-all coupling in the large-N limit, a random distribution of pinning phases is equivalent to an evenly spaced distribution $\alpha_j = 2\pi j/N$, $j = 1, 2, \ldots, N$.

In the phase-slip model (1), the θ_j represent the phases of entire CDW domains, or subdomains,[17] separated by ultrastrong pinning sites. In this sense, our phase variables have a similar interpretation to those in the model of Tua and Zawadowski.[24] Physically, the dynamics of Eq. (1) represent a competition between the energy in the applied field, the large pinning energy, and the energy of CDW amplitude collapse at the pinning sites. Phase distortions within a single domain due to weak pinning are not included in this model.

By describing a phase-slip process at the pinning sites by a phase-only model, we have neglected the dynamics of amplitude collapse, except as it is reflected by a periodic coupling term. Inui *et al.*[16] recently presented a detailed analysis showing how phase slip (with amplitude collapse) can be implicitly included in a phase-only model of switching CDW's in the limit of fast amplitude-mode dynamics. Zettl and Grüner[7] suggested that phase slip in switching CDW's could be accounted for by a sinusoidal coupling term. The model presented here is a discrete-site mean-field version of the phase-slip process, in the spirit of Fisher's treatment.[5] By choosing a particularly simple form for the periodic coupling function—but one with the right overall behavior—we are able to analyze much of the model's dynamics.

Previously,[22,23] we have shown that the large-N, mean-field version of (1) has a discontinuous and hysteretic depinning transition as the applied field E is varied. The switching seen in this model is in contrast to the smooth, reversible depinning which occurs in the corresponding equations with elastic phase coupling.[4,5] The threshold field $E_T(K)$ where the pinned solution, $\theta_j = \alpha_j + \sin^{-1}(E)$, becomes unstable to the formation of a coherent moving solution was shown[22,23] to be $E_T(K) = (1 - K^2/4)^{1/2}$ for $K < 2$ and $E_T(K) = 0$ for $K > 2$. At this threshold, the pinned solution bifurcates from a stable node to a saddle point in configuration space.[23]

B. Delayed switching

We now consider the time evolution of the system (1) during depinning. In order to simulate the experimental procedure where delayed switching is seen, we "apply" a superthreshold field $E > E_T(K)$ at $t = 0$ to the system (1) starting in the $E = 0$ pinned state $\theta_j = \alpha_j$. The response of the system depends on the value of E: for $E > 1$ the phases quickly leave the $E = 0$ pinned state and organize into a coherent moving solution. There is no delayed switching in this case. For $E_T(K) < E < 1$ the phases leave the pinned state quickly, but come to a near standstill at the saddle point $\theta_j = \alpha_j + \sin^{-1}(E)$, where they linger for a long time before finally leaving—again, very quickly—along the unstable manifold of the saddle point to form a coherent moving solution. Close to threshold, the time spent in the vicinity of the saddle point is much longer than any other part of the depinning process, resulting in a long delay before a rapid switch to the moving state. In this subsection we analyze the dynamics of (1) near the saddle point and derive an expression for the switching delay.

To characterize the collective state of the phases, we define the complex order parameter $re^{i\Psi} \equiv (1/N)\sum_j e^{i\theta_j}$. In the limit $N \to \infty$, the distribution of pinning phases α_j becomes continuous on $[0, 2\pi]$ and the phase θ_j can be written as a continuous function θ_α parametrized by the pinning phase α. In this limit the order parameter is given by

$$re^{i\Psi} = \frac{1}{2\pi}\int_0^{2\pi} e^{i\theta_\alpha}\, d\alpha \ . \tag{2}$$

We find numerically that for evenly spaced pinning sites the model is quite insensitive to the choice of N for all $N \geq 3$. In the infinite-N limit, an evenly spaced distribution of pinning sites becomes equivalent to a random distribution, but at finite N, simulations with a random distribution of pinning sites showed a much stronger finite-size effect than those with evenly spaced pinning sites. The insensitivity to N for evenly spaced pinning provides a useful trick allowing us to numerically investigate the large-N behavior of Eq. (1) using relatively small simulations (typically $N = 300$). Simulations with evenly spaced pinning sites did show a slight size dependence, especially at very small initial coherence ($r_0 < 10^{-4}$), and care was taken in using a sufficiently large system to eliminate any measurable dependence on N. The excellent agreement between the simulations and the analysis shows that the dynamics of Eq. (1) are well approximated by the infinite-N limit. Analysis of Eq. (1) will henceforth treat the case $N \to \infty$. Justification for applying the large-N limit to real CDW systems will be discussed in Sec. III.

Physically, the magnitude of the order parameter r ($0 \leq r < 1$) characterizes phase coherence between CDW domains. In a pinned configuration, for example, where the phases of the domains are determined by a random locally preferred phase, there is no coherence among the domains; accordingly, all stable pinned solutions of (1) have $r = 0$. In the steady sliding state ($d\Psi/dt > 0$) the model shows a large coherence between domains; all stable sliding solutions of (1) have $r \sim 1$. The rate of change of the order-parameter phase, $d\Psi/dt$, corresponds to the current carried by the CDW. The delayed switching observed experimentally corresponds to a delay in the current carried by the CDW. In our model, the onset of a "current" ($d\Psi/dt > 0$) and the onset of coherence ($r \sim 1$) occur simultaneously, as seen in Fig. 1.

The slowest step in the depinning process for $E_T(K) < E < 1$ is the evolution near the saddle point $\theta_\alpha = \alpha + \sin^{-1}(E)$. As we have shown elsewhere,[23] an interesting feature of the dynamics of (1) is that *any* initially incoherent ($r \sim 0$) configuration will be funnelled towards this saddle, and from there coherence and steady rotation will evolve. Thus the delay before switching for any incoherent initial state is approximately given by

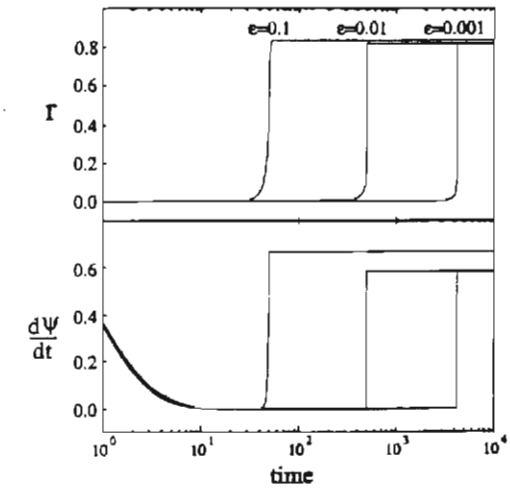


FIG. 1. Delayed onset of coherence r and current $d\Psi/dt$ for three normalized distances above threshold, $\varepsilon \equiv (E-E_T)/E_T = 0.1$, 0.01, and 0.0001. Note the logarithmic scale for the (dimensionless) time axis. Curves are from numerical integration of Eq. (1) with $N=300$, $r_0 \cong 5\times 10^{-5}$, and $K=1$, giving $E_T=(\frac{3}{4})^{1/2}$. The initial state of the system was the $E=0$ pinned state with a small amount of random jitter: $\theta_j=\alpha_j+\eta_j$ with $\eta_j \sim O(10^{-2})$. The large values of $d\Psi/dt$ at $t<10$ show the rapid evolution from the initial state to the saddle point $\theta_j=\alpha_j+\sin^{-1}(E)$. Note that for a given ε the order parameter r and the rotation rate $d\Psi/dt$ switch after the same delay.

the time for coherence to develop starting near $\theta_\alpha = \alpha - \sin^{-1}(E)$.

Because the depinning process with $E>0$ changes an initially static configuration ($d\Psi/dt=0$) into a uniformly rotating configuration ($d\Psi/dt>0$), the analysis is greatly simplified by working in a coordinate system that is corotating with Ψ. In terms of the corotating phase variables defined as $\phi=(\theta-\Psi)$ and $\gamma=(\alpha-\Psi)$, the depinning transition appears as a Landau-type symmetry-breaking transition from the unstable equilibrium at $r=0$ to a stable, *static* equilibrium at $r\sim 1$. Recasting the dynamical system (1) in terms of γ and ϕ_γ and writing the coupling term in (1) in terms of the order parameter gives the following mean-field equation:

$$\frac{d\Psi}{dt}\left[1-\frac{\partial\phi_\gamma}{\partial\gamma}\right]+\frac{dr}{dt}\frac{\partial\phi}{\partial r}=E+\sin(\gamma-\phi_\gamma)-Kr\sin(\phi_\gamma)\,. \tag{3}$$

In deriving Eq. (3) we have assumed that ϕ_γ only depends on r and γ. Assuming this dependence is equivalent to assuming that as the system leaves the saddle point $\phi_\gamma=\gamma+\sin^{-1}(E)$ it will not be free to visit all of state space, but is constrained to lie only in the unstable manifold of the saddle. Within the unstable manifold, two quantities are sufficient to characterize the state of the entire system: γ, which reflects the direction in which rotational symmetry has broken, and r, which reflects the position along the unstable manifold.

We now expand ϕ_γ about the saddle point $\phi_\gamma=\gamma+\sin^{-1}(E)$ in a Fourier series:

$$\phi_\gamma=\gamma+\sin^{-1}(E)+\sum_{k=1}^{\infty}A_k\sin(k\gamma)+\sum_{k=0}^{\infty}B_k\cos(k\gamma)\,, \tag{4}$$

where the Fourier coefficients A_k and B_k only depend on r. We assume that for small r each A_k and B_k can be expanded as a power series in r. We then solve Eq. (3) using a solution of the form of Eq. (4) with A_k, B_k, dr/dt, and $d\Psi/dt$ expanded in powers of r. This procedure gives a solution for ϕ_γ at each order of r. For self-consistency, these solutions must also satisfy the definition of the order parameter, which requires

$$\frac{1}{2\pi}\int_0^{2\pi}\cos(\phi_\gamma)d\gamma=r,\quad \frac{1}{2\pi}\int_0^{2\pi}\sin(\phi_\gamma)d\gamma=0\,. \tag{5}$$

Retaining terms to third order in r for $k\le 2$ uniquely determines (after much algebra) the evolution equation for the coherence:

$$\frac{dr}{dt}=\left[\frac{K-K_T}{2}\right]r+\left[\frac{6-K_T^2}{2K_T}\right]r^3+O(r^5)\,, \tag{6}$$

where $K_T\equiv 2(1-E^2)^{1/2}$. The form of the $O(r^3)$ coefficient in Eq. (6) was derived assuming that the system is close to threshold, that is, assuming $K-K_T \ll K_T$.

The value of r where the two terms on the right-hand side of Eq. (6) are equal, defined as $r^*\equiv[K_T(K-K_T)/(6-K_T^2)]^{1/2}$, marks a crossover point in the evolution of coherence. For $r(t)<r^*$, the cubic term is negligible and r grows as a slow exponential: $r(t)\cong r_0\exp(\sigma t)$, where $\sigma\equiv(K-K_T)/2$. Note that $\sigma\to 0$ as $E\to E_T$. After $r(t)$ reaches the value r^*, the cubic term in Eq. (6) dominates the linear term and r grows very rapidly. The rapid onset of coherence is accompanied by the simultaneous rapid growth of $d\Psi/dt$, as seen in Fig. 1. We identify the switching delay τ_{switch} as the time the system takes to evolve from r_0 to r^* by slow exponential growth: $r^*\equiv r_0\exp(\sigma\tau_{\text{switch}})$. Solving for τ_{switch} gives

$$\tau_{\text{switch}}=\frac{1}{K-K_T}\ln\left\{\frac{K_T(K-K_T)}{(6-K_T^2)r_0^2}\right\}\,. \tag{7}$$

The switching delay given by Eq. (7) agrees very well with numerical integration data for all values of $E_T<E<1$ and $r_0<r^*$. Figure 2 compares numerical data with Eq. (7) as a function of the normalized distance above threshold $\varepsilon\equiv(E-E_T)/E_T$. Numerical data were obtained using fourth-order Runge-Kutta numerical integration of Eq. (1). The pinning phases were evenly spaced on $[0,2\pi]$ and the initial phase configuration was the $E=0$ pinned state plus a repeatable random jitter: $\theta_j=\alpha_j+\eta_j$ with $\eta_j\sim O(10^{-2})$, giving $r_0\cong 1\times 10^{-4}$ at each ε. The random jitter η_j is introduced to break the symmetry of the unstable equilibrium at $r=0$, which is present for infinite N and also for finite N with evenly spaced pinning. In the absence of any initial jitter, r_0 is

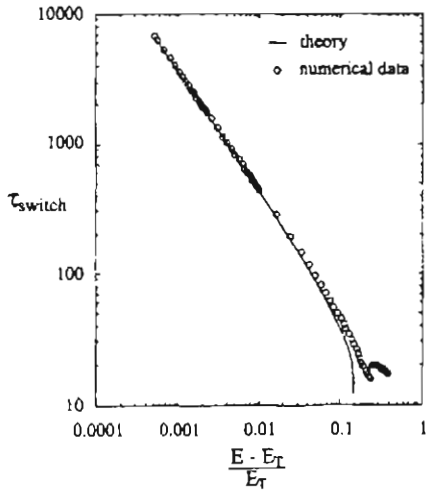


FIG. 2. Dependence of the delay (in dimensionless units) on the normalized distance above threshold $\varepsilon \equiv (E - E_T)/E_T$ for $K = 1$. Curve shows the theory from Eq. (7); circles are from numerical integration of Eq. (1) with $N = 300$. The initial state was the $E = 0$ pinned state with random jitter: $\theta_j = \alpha_j + \eta_j$ with $\eta_j \sim O(10^{-2})$. The same initial state was used for all values of ε. The value of r_0 used to calculate the theoretical curve was taken directly from the numerical data as the smallest value of r during its evolution; the minimum r depended very slightly on ε, and a single value of $r_0 = 1 \times 10^{-4}$ was used in Eq. (7). The disagreement between theory and numerics at $\varepsilon > 0.1$ is due to the small but finite time taken for the *other* parts of the depinning process besides the time spent lingering near the saddle point.

zero and the predicted switching time is infinite from Eq. (7). An interesting detail is that during the very early evolution, as the system evolves from the initial configuration $\theta_j = \alpha_j + \eta_j$ towards the saddle point (before the delay), r actually decreases—that is, the system becomes *less* coherent as it approaches the saddle. Because of this effect, the appropriate value of r_0 to use in Eq. (7)— and the value used for the theory in Fig. 2— is not the initial r, but the *minimum* r, which is slightly less than the initial value. The switching event for the numerical data was defined as the time where $r(t)$ reached 0.75. Because of the rapidity of the switch, any other reasonable definition of the switch time would have given nearly identical results.

III. DISCUSSION OF THE MODEL AND ITS PREDICTIONS

The phase-slip model of delayed switching presented here is a highly simplified treatment of CDW dynamics. Several approximations made in the interest of keeping the model analytically tractable are known to be physically unrealistic, including the all-to-all coupling of domains and the uniform coupling and pinning strengths. Possible justifications for these assumptions have been discussed elsewhere.[5,22]

An unphysical aspect of the model is the absence of multiple pinned configurations. Metastable pinned states are seen experimentally in both switching and nonswitching systems,[23] as well as in models with elastic coupling.[5,26] The absence of multiple pinned states in our model is a direct result of the periodic coupling, which does not allow large phase differences between domains to build up. Our assumption is not a necessary feature in modeling phase slip: an alternative phase-slip model that does allow a large buildup of phase difference before phase slip begins has been proposed by Hall *et al.*[12,14,16]

A subclass of switching samples, termed "type II" by Hundley and Zettl,[17] shows multiple depinning transitions as the applied field is swept. In contrast, type-I switching samples[17] show a single, hysteretic switch. Because the parameters in our model are uniform, we do not see multiple switching; our model always behaves like a type-I sample. The effects of distributed parameters in our model remains an interesting open problem.

Because we interpret phases as entire domains, our use of a large number of phases might be questioned in light of recent experiments identifying a small number of coherent domains separated by phase-slip centers.[11,13,17] We believe the large-N treatment is justified: Several experiments on switching samples indicate that even when a small number of coherent domains can be identified, these large domains have been formed by the collective depinning of many subdomains. The relevant dynamical process leading to switching and delayed conduction in this case is the *simultaneous depinning of many subdomains* within a single large domain. Experiments by Hundley and Zettl,[17] for example, show that a switching sample will depin smoothly when subdomains are forced to depin individually rather than collectively by applying a temperature gradient across the sample.

In spite of these limitations and the model's simplicity, we find that several aspects of delayed switching seen experimentally are produced by the mean-field phase-slip model. These similarities include the following.

(1) For $\varepsilon \equiv (E - E_T)/E_T \ll 1$ and $r_0 < r^*$, the switching delay in the model is approximately related to ε by the power law

$$\tau_{switch} \propto \varepsilon^{-\beta}, \quad \beta \sim 1 . \tag{8}$$

This behavior is clearly seen at small ε in Fig. 2. This dependence is different from that predicted by other models of delayed switching,[12,16–21] as will be discussed elsewhere.[27] Experimentally, power-law behavior with exponent $\beta \sim 1$ at small ε is consistent with the data of Maeda *et al.* (Fig. 5 of Ref. 9).

(2) Above a certain value of ε, defined as ε_0, the depinning transition in our model is not delayed. The value of ε_0 is defined by the condition $E = 1$. Similar behavior was seen experimentally by Zettl and Grüner,[7] who report that, for a bias current exceeding 1.25 times the threshold current, switching occurred without measurable delay. Measuring ε_0 in a switching sample will uniquely determine the appropriate value of K to be used in our model according to the formula $K = 2[1 - (\varepsilon_0 + 1)^{-2}]^{1/2}$.

(3) For values of ε slightly below ε_0, the switching delay in the model decreases more rapidly than the power

law (8), giving a concave-downward shape to a log-log plot of τ_{switch} versus ε, as seen in Fig. 2. This feature is also seen in the experimental data of Maeda *et al.* (Fig 5, Ref. 9).

(4) The switching delay in our model depends on an initial coherence. In Eq. (7) this dependence appears as the r_0 term in the logarithm, indicating that a sizable initial coherence will shorten the delay. For a sufficiently large initial coherence $r_0 > r^*$, Eq. (7) is not applicable because the cubic term in Eq. (6) will dominate throughout the evolution, leading to an extremely short switching delay. A reduction of the switching delay due to an organized initial state has been observed indirectly in the experiments of Kriza *et al.*[5] Using two closely spaced superthreshold pulses, Kriza *et al.*[8] were able to reduce the switching delay for the depinning which occurred during the second pulse; the shorter the spacing between the pulses, the shorter the observed switching delay.

(5) The delay in our model is a deterministic function of E, K, and r_0; we do not predict a scatter in the observed switching delay near threshold. Probabilistic models of depinning, for example, the model of Joos and Murray,[21] predict a scatter of delay times. A sizable scatter was reported by Zettl and Grüner;[7] more recently, Kriza *et al.*[8] attributed all scatter in the measured delay to "instrumental instability." It is not clear whether the scatter seen experimentally is an artifact or an intrinsic feature of delayed switching.

In conclusion, we have analyzed a very simple model of CDW domain dynamics with phase slip, and have found several features seen experimentally in switching samples. The results suggest that switching, hysteresis, and delayed onset of conduction are closely related phenomena which appear together when phase slip between domains occurs. Further experiments on delayed switching would be very useful to test in greater detail the predictive power of such a simple model.

ACKNOWLEDGMENTS

We thank Mark Sherwin for helpful discussions. One of us (C.M.M.) acknowledges partial support by AT&T Bell Laboratories, and another of us (S.H.S.) acknowledges financial support from the U.S. National Science Foundation (NSF). This research was supported in part by the U.S. Office of Naval Research (ONR) under Contracts No. N00014-84-K-0465 and No. N00014-84-K-0329.

[1] *Charge Density Waves in Solids*, Vol. 217 of *Lecture Notes in Physics*, edited by G. Hutiray and J. Sólyom (Springer, Berlin, 1985).

[2] For a recent review of charge-density-wave dynamics, see G. Grüner, Rev. Mod. Phys. **60**, 1129 (1988).

[3] H. Fukuyama and P. A. Lee, Phys. Rev. B **17**, 535 (1978); P. A. Lee and T. M. Rice, *ibid.* **19**, 3970 (1979); L. Sneddon, M. C. Cross, and D. S. Fisher, Phys. Rev. Lett. **49**, 292 (1982).

[4] L. Sneddon, Phys. Rev. B **30**, 2974 (1984).

[5] D. S. Fisher, Phys. Rev. Lett. **50**, 1486 (1983); Phys. Rev. B **31**, 1396 (1985).

[6] G. Grüner, A. Zawadowski, and P. M. Chaikin, Phys. Rev. Lett. **46**, 511 (1981).

[7] A. Zettl and G. Grüner, Phys. Rev. B **26**, 2298 (1982).

[8] G. Kriza, A. Jánossy, and G. Mihály, in *Charge Density Waves in Solids*, Ref. 1, p. 426.

[9] A. Maeda, T. Furuyama, and S. Tanaka, Solid State Commun. **55**, 951 (1985). Note that the graph for Fig. 5 is incorrectly printed as Fig. 3 in this paper.

[10] J. Dumas, C. Schlenker, J. Marcus, and R. Buder, Phys. Rev. Lett. **50**, 757 (1983); H. Mutka, S. Bouffard, J. Dumas, and C. Schlenker, J. Phys. (Paris) Lett. **45**, L729 (1984); S. Bouffard, M. Sanquer, H. Mutka, J. Dumas, and C. Schlenker, in *Charge Density Waves in Solids*, Ref. 1, p. 449.

[11] R. P. Hall, M. F. Hundley, and A. Zettl, Phys. Rev. Lett. **56**, 2399 (1986).

[12] R. P. Hall, M. F. Hundley, and A. Zettl, Physica B+C **143B**, 152 (1986).

[13] R. P. Hall, M. F. Hundley, and A. Zettl, Phys. Rev. B **38**, 13 002 (1988).

[14] R. P. Hall and A. Zettl, Phys. Rev. B **38**, 13 019 (1988).

[15] M. S. Sherwin, A. Zettl, and R. P. Hall, Phys. Rev. B **38**, 13 028 (1988).

[16] M. Inui, R. P. Hall, S. Doniach, and A. Zettl, Phys. Rev. B **38**, 13 047 (1988).

[17] M. F. Hundley and A. Zettl, Phys. Rev. B **37**, 8817 (1988).

[18] G. Mihály, G. Kriza, and A. Janossy, Phys. Rev. B **30**, 3578 (1984).

[19] A. Jánossy, G. Mihály, and L. Mihály, in *Charge Density Waves in Solids*, Ref. 1, p. 412.

[20] L. Mihaly, Ting Chen, and G. Grüner, Solid State Commun. **61**, 751 (1987).

[21] B. Joos and D. Murray, Phys. Rev. B **29**, 1094 (1984).

[22] S. H. Strogatz, C. M. Marcus, R. M. Westervelt, and R. E. Mirollo, Phys. Rev. Lett. **61**, 2380 (1988).

[23] S. H. Strogatz, C. M. Marcus, R. M. Westervelt, and R. E. Mirollo, Physica D **36**, 23 (1989).

[24] P. F. Tua and A. Zawadowski, Solid State Commun. **49**, 19 (1984).

[25] R. J. Cava, R. J. Fleming, E. A. Rietman, R. G. Dunn, and L. F. Schneemeyer, Phys. Rev. Lett. **53**, 1677 (1984); L. Mihaly, K.-B. Lee, and P. W. Stephens, Phys. Rev. B **36**, 1793 (1987).

[26] P. B. Littlewood, Phys. Rev. B **33**, 6694 (1986); Physica D **23**, 45 (1986).

[27] S. H. Strogatz and R. M. Westervelt, Phys. Rev. B (to be published).