LA-UR -86-1449

CONF-8604173--2

RECEIVED BY OSTI MAY 12 1986

TITLE NEURONS WITH HYSTERESIS FORM A NETWORK THAT CAN LEARN WITHOUT ANY CHANGES IN SYNAPTIC CONNECTION STRENGTHS

AUTHOR(S) Geoffrey W. Hoffmann
Maurice W. Benson

LA-UR--86-1449

DE86 010205

SUBMITTED TO To be published by the AIP in the Proceedings of the conference on Neural Networks for Computing

Snowbird, Utah, April 13-16, 1986

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

NEURONS WITH HYSTERESIS FORM A NETWORK THAT CAN LEARN WITHOUT ANY CHANGES IN SYNAPTIC CONNECTION STRENGTHS

Geoffrey W Hoffmann
Theoretical Biology and Biophysics, Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, U S A ,
and
*Departments of Physics and Microbiology, University of British Columbia,
Vancouver, B C., Canada V6T 2A6

Maurice W Benson
Center for Nonlinear Studies, Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, U S A.,
and
*Department of Mathematics, Lakehead University,
Thunderbay, Ontario, Canada P7B 5E1

ABSTRACT

A neural network concept derived from an analogy between the immune system and the central nervous system is outlined The theory is based on a neuron that is slightly more complicated than the conventional McCullogh-Pitts type of neuron, in that it exhibits hysteresis at the single cell level This added complication is compensated by the fact that a network of such neurons is able to learn without the necessity for any changes in synaptic connection strengths The learning occurs as a natural consequence of interactions between the network and its environment, with environmental stimuli moving the system around in an N-dimensional phase space, until a point in phase space is reached such that the system's responses are appropriate for dealing with the stimuli Due to the hysteresis associated with each neuron, the system tends to stay in the region of phase space where it is located The theory includes a role for sleep in learning

INTRODUCTION

It is difficult to overstate the importance of analogy in the development of theories An example from physics is the development of the theory of electromagnetism in the last century, which involved the use of complicated mechanical analogies [1] An example from biology is Burnet's clonal selection theory of the immune response, a theory that is modelled on Darwin's theory of evolution, and describes a survival of the fittest process at the level of

*Permanent address

single cells within an animal.[2] The precursor of the clonal selection theory was Jerne's natural selection theory of antibody formation[3], which marked the transition from older instructionist ideas to Darwinian selectionist theories Analogies clearly played a similarly crucial role when Jerne subsequently postulated that the immune system functions as a network; this time he was impressed by analogies between the structures and the functions of the immune system and the central nervous system.[4,5]

In order to explore a variety of ideas on how the brain could work, we need to utilize whatever analogies we can think of. In the early 1970's, Little formulated a neural network theory[6] based on the analogy between an Ising spin system and a neural network. Klopf[7] and Barto[8] have described neural network models in which they assume that the selfish behaviour of neural networks might be a reflection of analogous selfishness at the level of the neuronal building blocks. It would be interesting to know what analogies with other physical system, if any, aided Hopfield in the formulation of his model,[9] which has had such a profound influence on this field during the last four years. His treatment of the effect of symmetric synaptic coefficients, T_{ij}, was probably inspired by familiarity with many physical systems in which symmetric couplings lead to interesting properties.

As mentioned above, similarities between the central nervous system and the immune system led to a new way of viewing the immune system. An exciting possibility is that we can now reverse the process, and apply insights gained in the study of the immune system network to the central nervous system problem. This possibility has been considered also by Edelman and Reeke.[10] The immune system is also a complex system, but in the last few years we have learned an enormous amount about how it functions. We now have at least the outline of an immune system network theory that seems to work quite well.[11-16] Since many of the similarities between the immune system and the central nervous system are at a high level,[17] we considered the possibility that the same kind of mathematical model could be applicable to both systems. That idea led to a theory that we here review briefly, and that has been developed in more detail elsewhere.[17,18]

At one level our theory is slightly more complex than other theories, in that we invoke hysteresis at the level of single neurons. This added complexity is compensated by a new simplicity at the level of the network; the network can learn without any changes in the synaptic connection strengths. Learned information is associated solely with a state vector; memory is a consequence of the fact that due to the hysteresis associated with each neuron, the system tends to stay in the region of an N-dimensional phase space to which its experiences have taken it. Its stimulus-response behaviour is determined by its location in that space.

A NEURON WITH HYSTERESIS

Neurophysiologists tell me that we cannot, on the basis of presently available data, exclude the possibility that at least some neurons exhibit hysteresis at the single cell level. A simple, plausible biochemical model that could lead to single cell hysteresis is as follows. Let X be a key neural metabolite that controls the rate of firing of the cell, such that the rate of firing is (say) proportional to the concentration of X. The concentration of X is assumed to be affected by a second substance Y that reflects the net level of inhibitory signals the cell receives. We consider the following simple reaction:

$A \rightarrow X$

$X \rightarrow B$ (slow)

$X \rightarrow B$ (catalysed by Y)

Excess X inhibits Y

X is produced at a constant rate from A. X is broken down by two processes, one of which is slow and is independent of Y. The second breakdown process involves Y as an enzyme or the activator of an enzyme. High levels of X inhibit the enzymatic breakdown of X by Y. This phenomenon, known to biochemists as substrate inhibition of enzymes, can occur when an enzyme has two binding sites for the substrate. If we denote the concentrations of X and Y in neuron i by x_i and y_i, respectively, an appropriate differential equation for x_i as a function of time can have the form

$$\dot{x}_i = 1 - x_i - \frac{x_i y_i}{1 + \alpha x_i^2} \tag{1}$$

where α is a constant. y_i is given by

$$y_i = \sum_{j=1}^{N} \beta_{ij} x_j \tag{2}$$

β_{ij} is the synaptic connection strength from neuron j to neuron i. The β_{ij} correspond to the T_{ij} of Hopfield, except that learning can occur with the β_{ij} fixed (see below).

There can be three steady-state values for each x_i, two of which are stable (A and C, Fig. 1), and one of which is unstable (B). With N neurons, each of which can be in either a high x_i or a low x_i steady state, the number of attractors can be almost $2.^N$ The number of unstable steady states in the system can be close to $3.^N$

THE DYNAMICS OF AN N-DIMENSIONAL SYSTEM DISPLAYED ON A PHASE PLANE

It is convenient to consider a simple variant of the system (1) that also has about 2^N attractors:

$$\dot{x}_i = (1 - |x_i|)x_i - y_i \tag{3}$$

When the system (3) is at equilibrium, all the neurons are located on the reverse S-shaped curve

$$y_i = (1 - |x_i|)x_i$$

which is shown in Figure 2. Figure 2a shows trajectories following a small transient stimulus. The same stimulus with a larger amplitude results in a qualitatively identical perturbation of the network (Figure 2b). A still larger stimulus causes the x_i value of one or more neurons to change sign, and the other neurons then relax back to slightly different positions on the equilibrium curve (Figure 2c). The system can make a large number of different responses to a correspondingly large number of different stimuli and unless the amplitudes of the stimuli exceed a certain threshold level, the system stays in the same region of phase space.

The network has been shown to be capable of producing arbitrary outputs, providing N is sufficiently large.[17]

A particular output is produced if the system has an initial condition that can lead to that output. The hysteresis associated with each neuron tends to keep the system as a whole in a restricted region of the N-dimensional phase space, and this can account for memory. The particular region in that space where the system is located at any instant depends on the set of stimuli to which it has been subjected. Hence, in contrast to conventional neural network models, the formation of memories does not need to involve changes in the synaptic strengths, β_{ij}.

TEACHING

These networks can be taught to exhibit prespecified stimulus-response behavior. We systematically perturb the network until it reaches a region of phase space where it gives the desired stimulus-response behavior. Since the system tends to stay in a particular region of the phase space in the absence of systematic perturbations, the system is able to retain trained stimulus-response behavior or memories. The idea is that a brain is moved around in the N-dimensional phase space by the stimuli that it receives, until it reaches a point in that phase space such that its responses to the environment are such that it is not strongly perturbed by the environment.

We simply apply stimuli and observe responses. There is no "twiddling" of the synaptic connection strengths. Only the various stimuli themselves are used to move the system around in the N-dimensional phase space until a satisfactory region in that space is found. If at some point a satisfactory response to a stimulus is not obtained, the stimulus is reapplied and can

intensify. A wrong response means the system was in an unsatisfactory region of the phase space. Applying the same stimulus again takes the system further away from the region of phase space that did not give the correct response. If a good response is obtained, we pass on to another stimulus. This algorithm presumes to mimic the way real brains learn by experience. In an abstract sense there needs to be some complementarity between the stimulus and the response. If we are exposed to a stimulus and respond to it appropriately, our response can lead to the elimination of, or escape from, the stimulus. If we do not respond appropriately, the stimulus can persist, and possibly intensify, independently of whether the stimulus is being provided by a teacher or aspects of the non-human or inanimate environment. In a recent series of experiments we trained networks consisting of 20 neurons to respond to stimuli in prespecified ways using an algorithm based on these ideas.[18] The correct responses were defined in a binary fashion, so that untrained networks had a fifty percent rate of making correct responses. In one series of experiments, we achieved an average accuracy of 73% when we were training for a single stimulus, 70% when there were two stimuli, 66% with 3 stimuli, and 59% when there were 4 stimulus-response pairs being taught.

SLEEP

There is an automatic motion of each neuron from the central region (near $x_i = 0$ in the model (3)) towards one of the attractors (near $x_i = \pm 1$). The system can be strongly influenced by small signals when neurons are in the central region, while much larger stimuli are required when the system is at or near a system attractor. The system might therefore be much more easily taught if there were some means of keeping at least some of the neurons near the central regions. Perhaps during sleep the $\dot{x}_i$ versus x_i characteristic changes to, say,

$$\dot{x}_i = - ux_i$$

where u is a constant. The x_i of the system then relax back towards the origin during sleep, while maintaining their respective relative magnitudes. This part of our theory leads to the prediction that the variance in the rates of firing of neurons involved in memory should increase during waking hours and decrease during sleep.

PROSPECTS

The potential possibilities for systems of the above type has scarcely begun to be explored. In this final section I speculate about the directions that future research might take.

Since the β_{ij} matrix is fixed, special forms of that matrix that permit very rapid computation by digital or optical techniques can be used. As discussed elsewhere,18 the use of circulant matrices will permit computations

that are up to a factor of $N/\log N$ faster than calculations with random matrices, and such systems are also amenable to the application of extremely rapid analogue optical methods.

The brain is organized largely as groups of neurons that have high internal connectance, with lower connectance between groups. For artificial intelligence applications, it might be profitable to mimic this architecture, and thus obtain reliable behavior from unreliable components. If we couple together groups of our neurons that have been previously trained separately, it should be possible that the larger aggregates will give an improved fraction of "correct" responses than the smaller groups are able to do independently.

A scientific and perhaps philosophical mystery in neural science is the question of the neural bases of pain and joy. We are presently employing a teaching method that might be viewed as using "pain." (If the response is incorrect, we increase the magnitude of the stimulus.) Presumably the efficiency of our teaching could be much greater if we could add something to the algorithm equivalent to joy. A very speculative neural interpretation of pain and joy is in terms of familiar and unfamiliar regions of phase space. Perhaps pain is associated with stimuli that take one much away from a familiar region of phase space (as most completely random strong stimuli would do), and joy results from stimuli that are special in the sense that they move one back towards more familiar regions. New stimuli could move one back towards familiar regions of phase space if they are somehow correlated (in the context of the wiring diagram of our brains) with other stimuli we have received. If this speculation is correct, we might be able to simulate joy by making a network accustomed to a certain set of stimuli during its "ontogeny." Stimuli that are correlated with the early set might then be used to impart "joy."

ACKNOWLEDGEMENTS

This work was done while we were both on sabbatical leave at the Los Alamos National Laboratory. We thank George Bell and the other members of the Theoretical Biology group and David Campbell of the Center for Nonlinear Studies for their kind hospitality.

This research was financed by the Natural Sciences and Engineering Research Council of Canada, Grant No. A-6729, and the United States Department of Energy.

REFERENCES

1. F. Hund, Geschichte der physikalischen Begriffe (Bibliographisches Institut, Mannheim 1972).
2. F. M. Burnet, The Clonal Selection Theory of Acquired Immunity (Cambridge University Press, Cambridge, 1959).

3. N. K. Jerne, Proc. Nat. Acad. Sci. (USA) 41 849-857 (1955)
4. N. K. Jerne, Sci. Amer. 229 (1), 52-60, (1973).
5. N. K. Jerne, Ann. Immunol. (Inst. Pasteur) 125C 373-389 (1974).
6. W. A. Little, Math. Biosci. 19 101-120 (1974).
7. A. H. Klopf, The Hedonistic Neuron: A Theory of Memory, Learning and Intelligence (Hemisphere, Washington, D.C., 1982).
8. A. G. Barto, Human Neurobiol. 4 229-256 (1985).
9. J. J. Hopfield, Proc. Nat. Acad. Sci. (USA) 79 2554-2558 (1982).
10. G. M. Edelman and G. N. Reeke Jr., Proc. Nat. Acad. Sci., 79 2091-2095 (1982).
11. G. W. Hoffmann, Eur. J. Immunol. 5 638-647 (1975).
12. G. W. Hoffmann, in Theoretical Biology, G. I. Bell, A. S. Perelson and G. H. Pimbley, (eds.), Marcel Dekkar (New York, 1978), pp. 571-602.
13. G. W. Hoffmann, in Contemp. Topics in Immunobiol., vol. XI, 185-226 (1980).
14. N. Gunther and G. W. Hoffmann, J. Theoret. Biol. 94 815-855 (1982).
15. G. W. Hoffmann, in Regulation of Immune Response Dynamics, C. DeLisi and J. Hiernaux, (eds.), (CRC Press, Boca Raton, Florida, 1982), pp. 137-162.
16. G. W. Hoffmann, A. Cooper-Willis and M. Chow, in Lecture Notes in Biomathematics, vol. 36 (Springer-Verlag, 1986), p. 15.
17. G. W. Hoffmann, J. Theor. Biol., in press.
18. G. W. Hoffmann, M. W. Benson, G. M. Bree and P. E. Kinahan, Physica D, in press.

FIGURE CAPTIONS

Fig. 1. The rate of change in the rate of firing, $\dot{x}_i$, as a function of x_i for three different values of the input y_i: low ($y_i = y_L$), intermediate ($y_i = y_I$), and high ($y_i = y_H$). For $y_i = y_I$ the neuron has two stable steady state rates of firing, $x_i = x_A$ and $x_i = x_C$.

Fig. 2. The dynamics of a network with 20 neurons portrayed in the y_i, x_i phase plane. The system is at equilibrium when the y_i, x_i coordinates of every neuron is on the reverse S shaped curve. (a) A transient small perturbation causes the neurons to leave the curve, and then relax back to their original positions. (b) A larger stimulus causes a qualitatively similar response. (c) A still larger stimulus causes one neuron to switch from one branch of the curve to the other, and the equilibrium values of each of the other neurons changes somewhat. For these plots we used a network with 20 neurons with a random β_{ij} matrix, connectance 0.5.

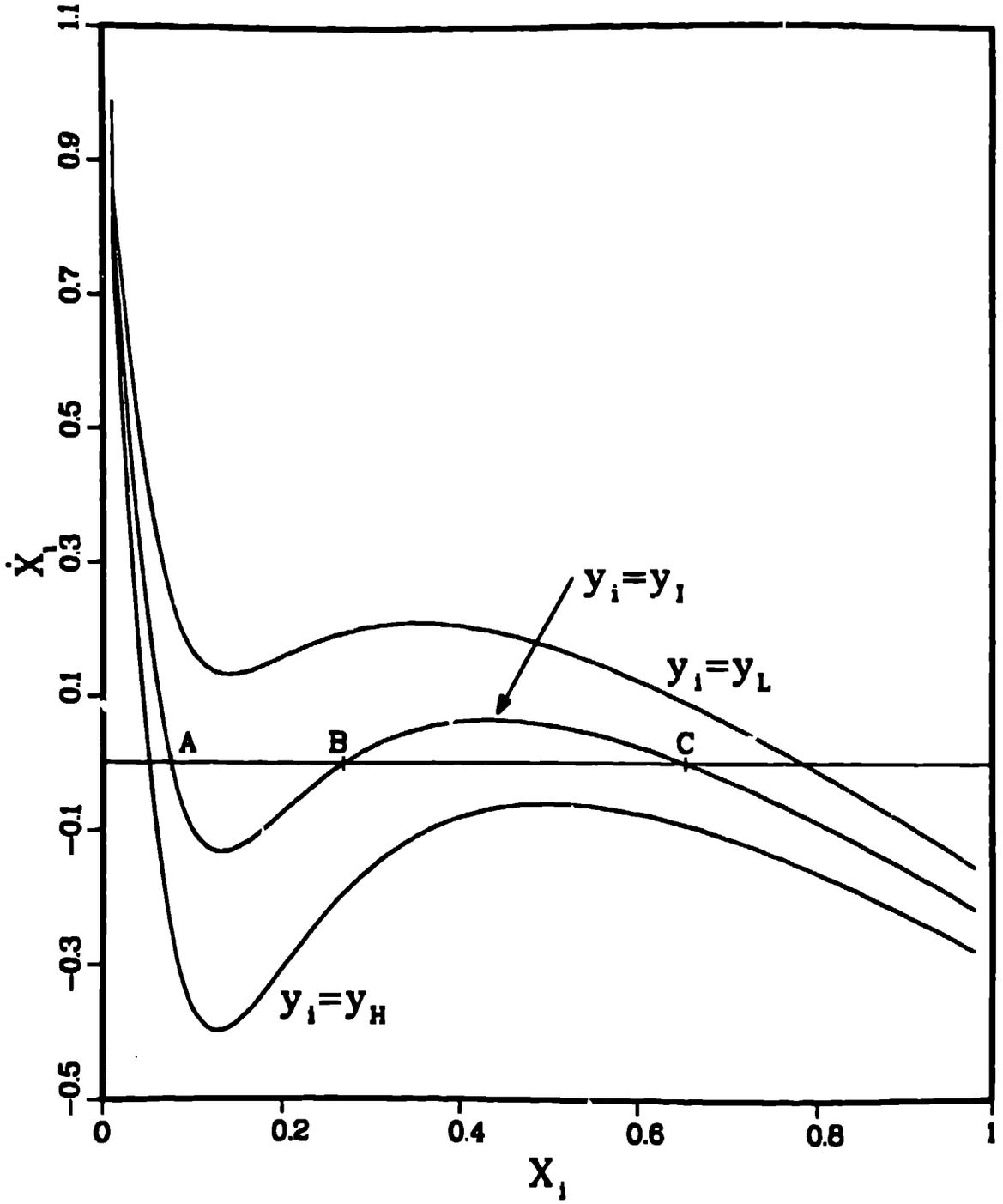
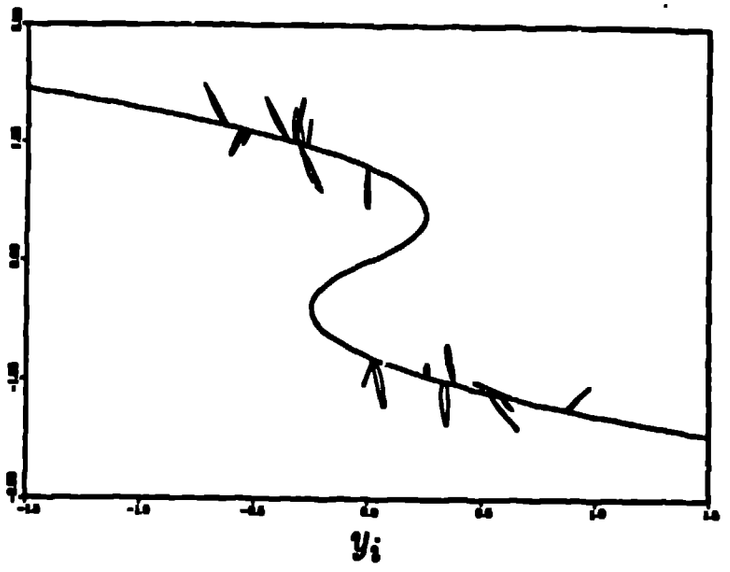

1.1
0.9
0.7
0.5
0.3
0.1
−0.1
−0.3
−0.5
$\dot{X}_1$
0
0.2
0.4
0.6
0.8
1
X_1
$y_i=y_I$
$y_i=y_L$
$y_i=y_H$
A
B
C

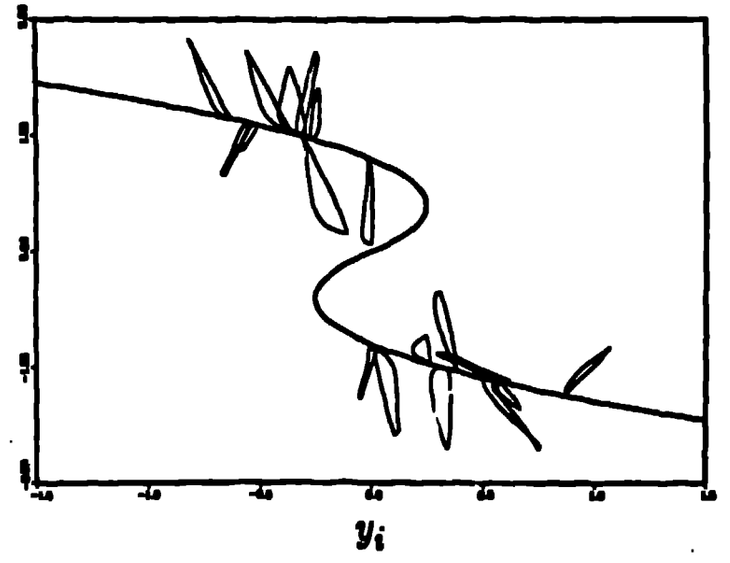
Fig. 1

Fig

2(a)



2(b)



2(c)