

Neural Networks with Complex Activations and Connection Weights

Miklos N. Szilagyi

*Department of Electrical and Computer Engineering,
University of Arizona, Tucson, AZ 85721, USA*

Boaz Salik

*Department of Electrical Engineering,
California Institute of Technology, Pasadena, CA 91125, USA*

Abstract. The concept of neural networks is generalized to include complex connections between complex units. A mathematical model is presented. An expression for the network's energy as well as a complex learning rule are proposed. This innovation may lead to new neural network paradigms, architectures, and applications, and may help to better understand biological nervous systems. The similarity between the dynamics of some linear complex networks and the quantum mechanical behavior of atomic systems is shown. The convergence properties of two-neuron complex networks are explored as extensions of the neural description of the Mandelbrot set, and are found to possess similar fractal properties.

1. Introduction

The dynamic behavior of neural networks can be fully described by systems of differential equations [1, 2]. A large amount of information can be extracted from both theoretical investigation and numerical solution of such systems. It is also possible, however, to explore the behavior of neural networks by means of suitable physical models whose behavior is described by similar systems of differential equations.

Complex functions are commonly used for the description of physical quantities. A well-known example is the treatment of electrical networks in terms of complex potentials, currents, and impedances. The dynamics of such networks are represented by systems of differential equations in which the complex impedance matrix represents connections between the different network elements. One can even introduce the notion of complex power as the inner product of the complex voltage and current vectors. The real and imaginary parts of this power are the active (dissipative) and reactive powers, respectively.

Since the nature of information processing in biological nervous systems is totally different from that in digital computers, and still mostly unknown, one should not *ex cathedra* exclude the existence of complex neural networks. Indeed, there are neuron models based on two-state variables, for example, phase and frequency (the first determines the short-term behavior of the network, the second reflects the effect of learning and shapes its long-term evolution). Thus, the activity in the network model can be viewed as two superimposed relaxation processes: one for phases and another for frequencies [3]. Therefore, although the connection weights in the *anatomy* of macroscopic neural networks may be real, there is no evidence that would exclude the existence of complex neural networks at the *functional* level that can be anatomically realized, for example, between clusters of neurons.

2. Complex network model

Let us now consider a complex neural network. For a fully interconnected network of N units the network dynamics is described by the system of differential equations [1]

$$\frac{du_j}{dt} = \sum_{k=1}^N w_{jk} f_k(u_k(t)) + I_j(t) \quad (1)$$

where t is time, $u_j(t)$ are the state variables that describe the time dependence of each unit's input activation, $f_j(u_j)$ are the output activation functions, w_{jk} is the connection weight between the j th and k th neurons, and I_j is the external input, $j = 1, 2, \dots, N$. (For simplicity, we neglect the contribution of leakage currents in the network.)

In our case u_j , f_j , w_{jk} , and I_j are all complex:

$$u_j(t) = x_j(t) + iy_j(t) \quad (2)$$

$$f_j(u_j) = g_j(x_j, y_j) + ih_j(x_j, y_j) \quad (3)$$

$$w_{jk}(t) = p_{jk}(t) + iq_{jk}(t) \quad (4)$$

$$I_j(t) = J_j(t) + iK_j(t) \quad (5)$$

where x_j , y_j , g_j , h_j , p_{jk} , q_{jk} , J_j , and K_j are all real, $i = \sqrt{-1}$ is the imaginary unit. Then equations (1) can be transformed into

$$\frac{dx_j}{dt} = \sum_{k=1}^N (p_{jk}g_k - q_{jk}h_k) + J_j \quad (6)$$

$$\frac{dy_j}{dt} = \sum_{k=1}^N (p_{jk}h_k + q_{jk}g_k) + K_j \quad (7)$$

($j = 1, 2, \dots, N$). Thus we have described the network of N complex neurons by a system of $2N$ coupled differential equations in real variables; such behavior cannot be realized by a real Hopfield network of any size.

The energy of the complex network may be defined as

$$E = -\frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N f_j^* w_{jk} f_k - \operatorname{Re} \left(\sum_{j=1}^N f_j^* I_j \right) \quad (8)$$

where the asterisk represents the complex conjugate. It can be shown [4] that this energy is a nonincreasing function of time if the following three conditions are satisfied:

$$\frac{\partial g_j}{\partial x_j} \geq 0 \quad \text{and} \quad \frac{\partial h_j}{\partial y_j} \geq 0, \quad (9)$$

$$\frac{\partial g_j}{\partial y_j} = -\frac{\partial h_j}{\partial x_j} \quad (10)$$

and

$$w_{jk} = w_{kj}^* \quad (11)$$

(the connection matrix must be hermitian). Thus, when these conditions are satisfied, the network must evolve toward local energy minima. However, the network state may “jump” from the neighborhood of one minimum to another, provided the jump decreases its energy. Therefore these attractors are not Liapunov stable. Next we consider how to train the network so that it has the attractors we desire.

3. Learning

To make the network learn some patterns, its energy surface has to be shaped according to those patterns. Complex patterns can be quite convenient to represent images by forming complex vectors as sets of complex coordinates of pixels of the given image. The complex equivalent of the Hebbian learning rule is

$$\Delta w_{jk} = \eta f_j f_k^* \quad (12)$$

where η is the learning rate. It can be easily verified that such a learning rule cannot increase the energy of the network.

We have performed some simple experiments with content-addressable memory, utilizing this learning rule [4]. The weight matrix is constructed analytically by using the formula

$$w_{jk} = \sum_{s=1}^p V_j(s) V_k(s)^* \quad (13)$$

where $V(s)$ is the training pattern and p is the number of patterns to be learned. A network of $N = 24$ complex neurons could learn $p = 3$ randomly selected patterns quite successfully (in some cases even vectors with more than 50% of entries altered could be reconstructed).

Finding a suitable activation function is not easy. The extension of the usual sigmoid-type activation functions to the complex plane through analytic continuation creates a problem. According to Liouville's Theorem, the only bounded and continuously differentiable function in the complex plane is the constant function. The analytic continuations of the well-known sigmoid functions all have singularities and do not satisfy our conditions (9) and (10). The question is, therefore, whether to give up analyticity or boundedness. If we choose the former, we can utilize simple bounded functions without analytic continuation. For example, in our experiments we have used the simple activation function

$$f_j(u_j) = \tanh(x_j) + i \tanh(y_j). \quad (14)$$

Our search continues for workable activation functions that satisfy our requirements.

On the other hand, the emerging singularities of analytically continued activation functions may result in unexpected and interesting properties. Systematic investigations of small complex networks reveal patterns in this behavior that can be used to define new architectures and applications (see section 4).

It is natural that nearly all existing neural network paradigms can be generalized to include complex activations and connections. Some of them may not be practical but others will be useful. It is expected that a complex network of given size can perform more complicated functions than a real-valued one.

4. Dynamic properties

Discrete complex dynamical systems constitute interesting special cases of complex neural networks. It has been shown [11] that a single complex neuron is capable of determining membership in a Mandelbrot or Julia set. Chaotic behavior in more complex networks has been documented. For example, a general neural model of an oscillating cortex was shown to exhibit distinct nonlinearities [12]. A more detailed model of the olfactory system has clearly displayed chaotic behavior, and it was proposed that this dynamic complexity actually enhances the system's function [13]. Extending Clarke's model, we investigated the convergence properties of a two-neuron complex network.

Discretization of the Hopfield model yields:

$$u_j(t+1) - u_j(t) = \sum_{k=1}^N w_{jk} f_k(u_k(t)) + I_j(t). \quad (15)$$

As in the one-neuron network, convergence was plotted as a function of the initial (complex) value of one of the neurons. However, several dimensions are introduced here that are not present in the single-neuron case, namely the output function and initial value of the second neuron and the value of each connection matrix element. For simplicity, the output functions $f_k(u_k)$ were assumed to be the same and the inputs $I_j(t)$ were set to zero.

The convergence criteria were

$$|\operatorname{Re}(u_1(t)) - \operatorname{Re}(u_1(t-1))| < 10^{-4} \quad (16)$$

and

$$|\operatorname{Im}(u_1(t)) - \operatorname{Im}(u_1(t-1))| < 10^{-4} \quad (17)$$

Divergence was defined as nonconvergence within 50 iterations or attainment of the divergence condition

$$|u_1(t)| > 10^6 \quad (18)$$

Note that, because convergence was plotted versus $u_1(0)$, these criteria were imposed on the first neuron. However, it is conceivable that in a complex two-neuron network, one neuron will converge and the other diverge, especially given our simplified criteria. Also, remember that the Mandelbrot and Julia sets are realizable as special cases of a two-neuron network by eliminating the extra dimensions, for example using a connection matrix with zero determinant. Thus, it is clear that, in general, the convergence properties of an N -neuron complex network are not realizable by an $(N-1)$ -neuron network, and it is natural to think of the former as a generalization of the latter. To that extent, our two-neuron convergence properties are generalizations of Julia sets, while properties of an N -neuron network, $N > 2$, are generalizations of ours. Even for the simple case of two neurons and a square-law output function, there is little to say analytically about the network's convergence. Therefore we present here several empirical observations.

For each initial value of neuron one $u_1(0)$, the initial value of neuron two $u_2(0)$ can be characterized by its absolute value and its phase (Ph) relative to $u_1(0)$. We found that for all connection matrices and output functions, preservation of the network's convergence requires that

$$\operatorname{Ph}(u_1(0)) - \operatorname{Ph}(u_2(0)) = \text{constant} \quad (19)$$

In other words, when convergence is plotted as a function of $u_1(0)$, an increase in $\operatorname{Ph}(u_2(0))$ causes a corresponding counter-clockwise rotation in the region of convergence (Figure 1).

Less intuitive results are attained by varying the magnitude of $u_2(0)$. In general, for a given output function and connection matrix W , an increase in $|u_2(0)|$ caused a decrease in the size (disintegration) of the region of convergence (ROC) for $u_1(0)$ (Figures 2–5). Additionally, increasing $|u_2(0)|$ caused the ROC to move in a constant radial direction outward from the origin before disappearing; this direction depends on the relative phase of $u_1(0)$ and $u_2(0)$ and on the connection matrix but does not seem to depend on the output function. Also note that for real-valued W , the ROC for $f_k(u_k) = u_k^2$ maintains an axial symmetry when disintegrating (a property attributable to the function's symmetry), while the ROC for $f_k(u_k) = u_k^3$ does not.

Much like initial values, a uniform rotation in the connection weights causes an opposite rotation of the ROC when the initial values are close to

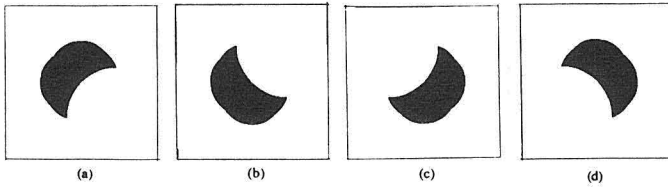


Figure 1: Effect of change in second neuron's initial phase on the network's convergence. Plots show convergence (black) for $-1 < \text{Re}(u_1(0)) < 1$, $-1 < \text{Im}(u_1(0)) < 1$. Initial phases of second neuron are (a) $\pi/4$, (b) $3\pi/4$, (c) $5\pi/4$, (d) $7\pi/4$; $|u_2(0)| = 1$; $f_i(u_i) = u_i^2$; $w_{ii} = 1$, $w_{ij} = 0.1$.

zero; for the symmetric output function $f_k(u_k) = u_k^2$, this rotation occurs even when both initial values are nonzero, although this is not generally the case. Furthermore, in general, an increase in the magnitude of the connection weights causes a decrease in the size of the ROC for a given output function, although there is no motion away from the origin.

The implications of these results for network architecture, assuming they are generalizable to larger networks, are rather straightforward. A complex

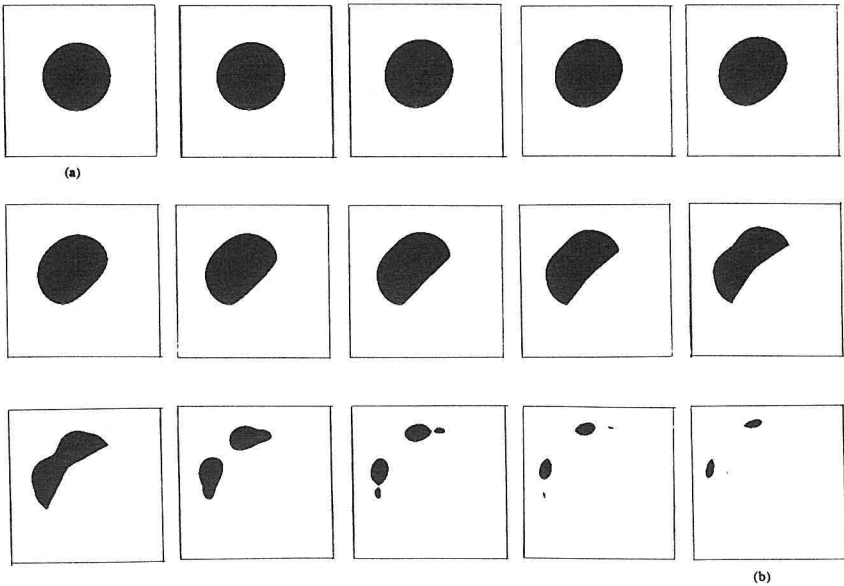


Figure 2: Disintegration of ROC with increasing $|u_2(0)|$ for real connections w_{ij} and symmetric f_i . $|u_2(0)|$ is incremented by steps of .1 from (a) 0 to (b) 1.4; $\text{Ph}(u_2(0))$ has the constant value $\pi/4$. $f_i(u_i) = u_i^2$; $w_{ii} = 1$, $w_{ij} = 0.4$. Region shown is $-1 < \text{Re}(u_1(0)) < 1$, $-1 < \text{Im}(u_1(0)) < 1$.

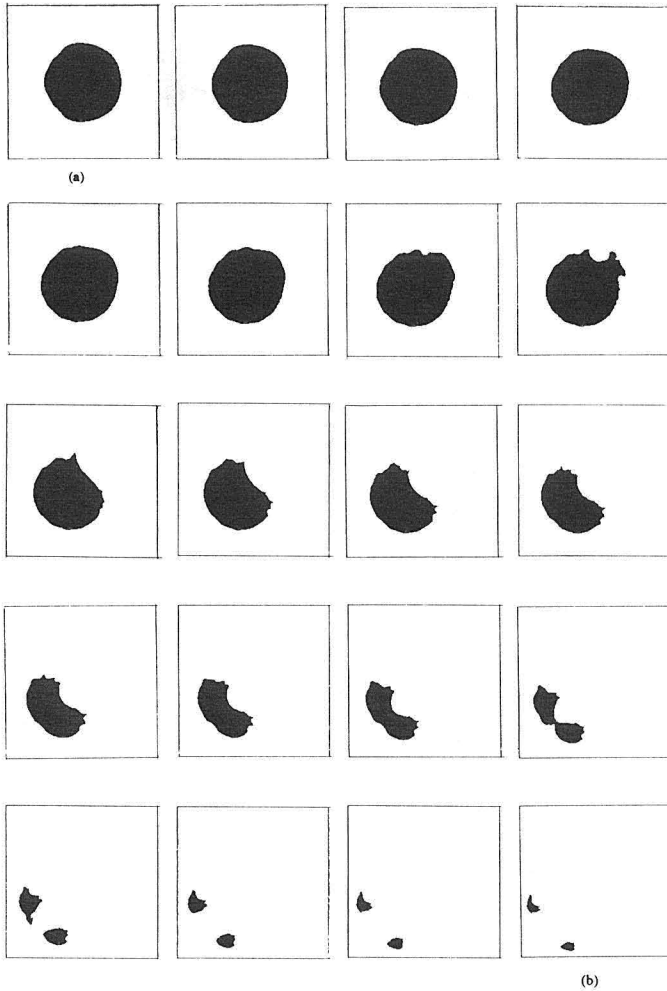


Figure 3: Disintegration of ROC with increasing $|u_2(0)|$ for complex connections w_{ij} and symmetric f_i . $|u_2(0)|$ is incremented by steps of .1 from (a) 0 to (b) 1.9; $\text{Ph}(u_2(0))$ has the constant value $\pi/4$. $f_i(u_i) = u_i^2$; $w_{12} = 0.5i$, $w_{21} = -0.5i$. Region shown is $-1 < \text{Re}(u_1(0)) < 1$, $-1 < \text{Im}(u_1(0)) < 1$.

network can use unbounded, analytic output functions and still converge under the right circumstances. Specifically, one with small connection weights and initial values will more likely converge (and do so faster) than one with larger connection weights or initial values; relative phases of initial values or connection weights will not affect the size of the ROC, but may affect its orientation. Finally, because of their fractal nature, it is impossible to know

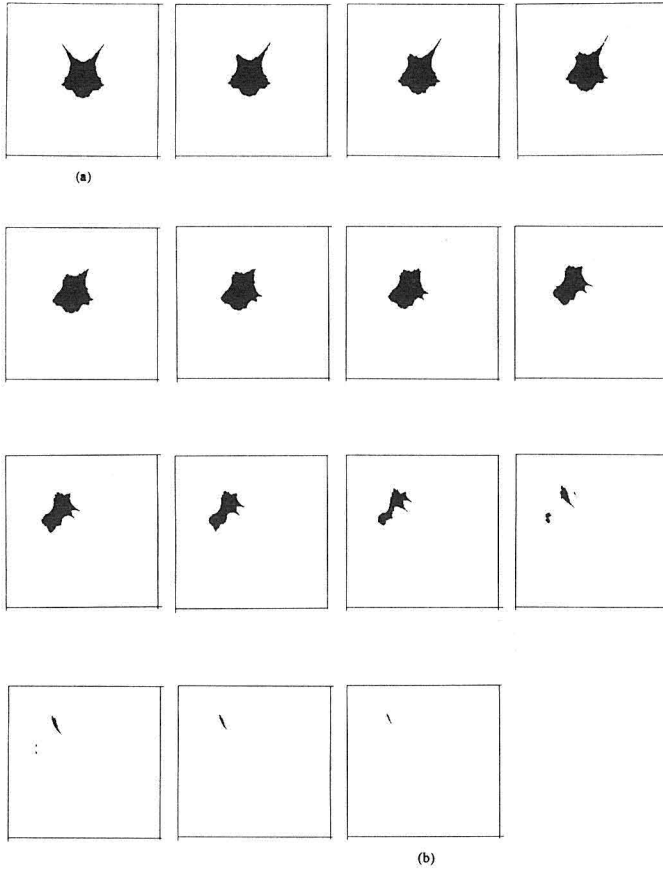


Figure 4: Disintegration of ROC with increasing $|u_2(0)|$ for real connections w_{ij} and asymmetric f_i . $|u_2(0)|$ is incremented by steps of .1 from (a) 0 to (b) 1.4; $\text{Ph}(u_2(0))$ has the constant value 0. $f_i(u_i) = u_i^3$; $w_{ii} = 1$, $w_{ij} = 1.5$. Region shown is $-1 < \text{Re}(u_1(0)) < 1$, $-1 < \text{Im}(u_1(0)) < 1$.

a priori whether a given network will diverge except in trivial cases. This presents an alternative to using bounded, non-analytic output functions, but sets limits on its effective use.

5. Physical applications

In the case of linear networks the activation function is trivial: $f_j(u_j) = u_j$. An interesting special case of a complex linear network is represented by the quantum mechanical Hamiltonian equations [5]. Schrodinger's wave equation

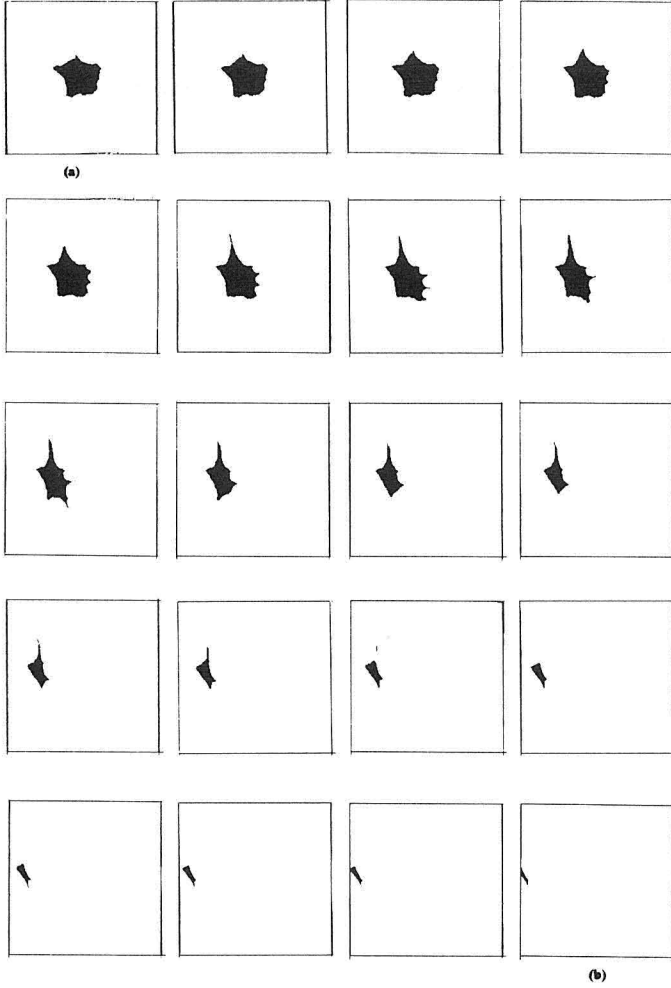


Figure 5: Disintegration of ROC with increasing $|u_2(0)|$ for complex connections w_{ij} and asymmetric f_i . $|u_2(0)|$ is incremented by steps of .1 from (a) 0 to (b) 1.9; $\text{Ph}(u_2(0))$ has the constant value $\pi/4$. $f_i(u_i) = u_i^3$; $w_{ii} = 1$, $w_{12} = 1.5(1 + i)/1.414$, $w_{21} = 1.5(1 - i)/1.414$. Region shown is $-1 < \text{Re}(u_1(0)) < 1$, $-1 < \text{Im}(u_1(0)) < 1$.

can be written in the form of the following system of N coupled differential equations [6]:

$$i\hbar \frac{dC_j}{dt} = \sum_{k=1}^N H_{jk} C_k \quad (j = 1, 2, \dots, N) \quad (20)$$

where N is the number of *base states*, $\hbar = h/2\pi$, h is Planck's constant, H_{jk}

is an element of the Hamiltonian matrix, and C_k is the probability amplitude of finding the state of the system in the k th base state at time t .

The system of differential equations (20) describes the time dependence of all these probability amplitudes, that is, the global dynamic behavior of the entire atomic system. The elements of the Hamiltonian matrix characterize the probability amplitudes of the generation of one base state from another, that is, the coupling between the system's different base states.

In general, the probability amplitudes as well as the elements of the Hamiltonian matrix are complex. However, since the Hamiltonian matrix is hermitian, its diagonal elements are all real.

Equation (20) is a special case of equation (1). The state variables are the probability amplitudes: $f_j = u_j = C_j$. We also have $I_j = 0$ and $w_{jk} = -(i/\hbar)H_{jk}$. Thus, the connection matrix is anti-hermitian, since all its diagonal elements are imaginary, and such a network does not satisfy our condition (11) and the energy formula (8) cannot be utilized. However, the system's dynamics is identical to that of a specific linear complex neural network. Of course, this does not imply that a classical model of quantum mechanics is possible, but the analogy may shed some light on the actual function of biological nervous systems.

For example, Penrose suggests that somewhere deep in the brain one may find cells of single quantum sensitivity (perhaps in the intercellular microtubules). If this is true, then quantum mechanics is indeed directly involved in brain activity [7].

In addition, we have previously shown that the response of biological neurons to stimuli can be represented as a measuring process, and there is evidence that the quantum mechanical theory of measurement can be applied to it. A simple form of this model takes into account two observables, the stimulus intensity and length, measured by individual neurons with only one threshold for each. An uncertainty relation can be derived between the possible accuracy of the measured intensity and length of time. The model is extended to measurements by neuron populations [8].

The idea of complex neural networks was first proposed by Szilagyi in 1988 [5]. An application of this idea has led to the holographic model of memory and learning [9], and even to the development of a commercial product [10]. Thanks to their ability to accept inputs with two real dimensions (amplitude and phase), complex networks are well suited for visual and auditory pattern recognition, the signal amplitude occupying one dimension and its frequency the other.

6. Conclusions

A Hopfield-like model for complex neural networks, including network energy and a Hebbian learning rule, has been presented. This model and others based on complex neurons and connection weights should lead to useful new architectures, paradigms, and applications, and may help to better understand biological nervous systems.

We generalized Clarke's neural description of the Mandelbrot and Julia sets to multiple neurons. The Mandelbrot and Julia sets can be produced as special cases of these networks (degenerate connection matrices). The effects of certain two-neuron network features on its convergence have been detailed, and their implications for network architectures outlined.

Acknowledgments

We are indebted to Professors J. D. Garcia, G. A. Korn, B. Mayoh, R. H. Parmenter, and T. Triffet for stimulating discussions, to Professor E. T. Paulsen for a valuable suggestion, and to D. Targovnik for careful editing of this paper.

References

- [1] J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," *Proceedings of the National Academy of Sciences*, **81** (1984) 3088–3092.
- [2] S. Grossberg, "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Neural Networks*, **1** (1988) 17–61.
- [3] C. T. Genis, in *Neural Computers*, edited by R. Eckmiller and C. Malsburg (Berlin: Springer, 1988).
- [4] M. Szilagyi, J. C. Mikkelsen, and K. H. Mortensen, *Some New Thoughts on Neural Networks with Complex Connection Matrices*, DAIMI PB-322 (Computer Science Department, Aarhus University, Aarhus, Denmark, 1990).
- [5] M. Szilagyi, *Neural Networks with Complex Connection Matrices*, DAIMI PB-260 (Computer Science Department, Aarhus University, Aarhus, Denmark, 1988).
- [6] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vol. 3 (Reading, MA: Addison-Wesley, 1965).
- [7] R. Penrose, *The Emperor's New Mind* (New York: Penguin Books, 1991).
- [8] F. Jolesz and M. Szilagyi, "A Possible New Model of Neurons and Neural Processes Based on the Quantum Mechanical Theory of Measurement," *Acta Cybernetica*, **4** (1978) 221–239.
- [9] J. G. Sutherland, "A Holographic Model of Memory, Learning and Expression," *International Journal of Neural Systems I*, **3** (1990) 259–267.
- [10] J. G. Sutherland, *HNeT Development System* (AND Corporation, Hamilton, Ontario, Canada, 1991).
- [11] T. L. Clarke, "Generalization of Neural Networks to the Complex Plane," pages 435–440 in *Proceedings of the International Joint Conference on Neural Networks*, vol. 2 (San Diego, CA, 1990).

- [12] B. Baird, "Associative Memory in a Simple Model of Oscillating Cortex," in *Neural Information Processing Systems II*, edited by D. S. Touretzky (Palo Alto, CA: Morgan Kaufmann, 1990).
- [13] B. Baird and W. J. Freeman, "Applications of Chaotic Neurodynamics in Pattern Recognition," *Proceedings of SPIE*, (1991) 1469.