*Neural Networks*, Vol. 1, pp. 17–61, 1988
Printed in the USA. All rights reserved.

0893-6080/88 $3.00 + .00

*ORIGINAL CONTRIBUTION*

# Nonlinear Neural Networks: Principles, Mechanisms, and Architectures

STEPHEN GROSSBERG

Boston University

(*Received and accepted* 1 *September* 1987)

**Abstract**—*An historical discussion is provided of the intellectual trends that caused nineteenth century interdisciplinary studies of physics and psychobiology by leading scientists such as Helmholtz, Maxwell, and Mach to splinter into separate twentieth-century scientific movements. The nonlinear, nonstationary, and nonlocal nature of behavioral and brain data are emphasized. Three sources of contemporary neural network research—the binary, linear, and continuous-nonlinear models—are noted. The remainder of the article describes results about continuous-nonlinear models: Many models of content-addressable memory are shown to be special cases of the Cohen-Grossberg model and global Liapunov function, including the additive, brain-state-in-a-box, McCulloch-Pitts, Boltzmann machine, Hartline-Ratliff-Miller, shunting, maskingfield, bidirectional associative memory, Volterra-Lotka, Gilpin-Ayala, and Eigen-Schuster models. A Liapunov functional method is described for proving global limit or oscillation theorems for nonlinear competitive systems when their decision schemes are globally consistent or inconsistent, respectively. The former case is illustrated by a model of a globally stable economic market, and the latter case is illustrated by a model of the voting paradox. Key properties of shunting competitive feedback networks are summarized, including the role of sigmoid signalling, automatic gain control, competitive choice and quantization, tunable filtering, total activity normalization, and noise suppression in pattern transformation and memory storage applications. Connections to models of competitive learning, vector quantization, and categorical perception are noted. Adaptive resonance theory (ART) models for self-stabilizing adaptive pattern recognition in response to complex real-time nonstationary input environments are compared with off-line models such as autoassociators, the Boltzmann machine, and back propagation. Special attention is paid to the stability and capacity of these models, and to the role of top-down expectations and attentional processing in the active regulation of both learning and fast information processing. Models whose performance and learning are regulated by internal gating and matching signals, or by external environmentally generated error signals, are contrasted with models whose learning is regulated by external teacher signals that have no analog in natural real-time environments. Examples from sensory-motor control of adaptive vector encoders, adaptive coordinate transformations, adaptive gain control by visual error signals, and automatic generation of synchronous multijoint movement trajectories illustrate the former model types. Internal matching processes are shown capable of discovering several different types of invariant environmental properties. These include ART mechanisms which discover recognition invariants, adaptive vector encoder mechanisms which discover movement invariants, and autoreceptive associative mechanisms which discover invariants of self-regulating target position maps.*

## TABLE OF CONTENTS

---

Based upon a lecture given on October 7, 1986 at the NSF meeting on Neural Networks and Neuromorphic Systems, Woburn, Massachusetts.

Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-86-C-0037 and AFOSR F49620-87-C-0018) and the National Science Foundation (NSF IRI-84-17756).

The author thanks Cynthia Suchta and Carol Yanakakis for their valuable assistance in the preparation of the manuscript and illustrations.

Requests for reprints should be sent to Stephen Grossberg, Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston, MA 02215.

> "The foundations of science as a whole, and of physics in particular, await their next great elucidations from the side of biology, and especially from the analysis of the sensations . . . psychological observation on the one side and physical observation on the other may make such progress that they will ultimately come into contact, and that in this way new facts may be brought to light. The result of this investigation will not be a dualism but rather a science which, embracing both the organic and the inorganic, shall interpret the facts that are common to the two departments." (Mach, 1914)

## 1. INTRODUCTION

The physical and mathematical theory of neural networks has been developing rapidly during the past 25 years. It is a theory whose diversity and complexity reflect the multifaceted organization of the brain processes that it sets out to explain. In this article, I will summarize some of the unifying principles, mechanisms, and mathematical methods that arise in this theory, as well as some of the specialized neural architectures which are important both in physical analyses of behavioral and brain data and in the development of novel technologies.

I will begin this article with some historical remarks that may clarify the complex and often confusing sociological milieu in which these exciting intellectual developments have been taking place.

## 2. INTERDISCIPLINARY STUDIES DURING THE NINETEENTH CENTURY: HELMHOLTZ, MAXWELL, AND MACH

Interdisciplinary studies flourished during the nineteenth century. In addition to pursuing their great work in physics, scientists such as Helmholtz, Maxwell, and Mach also made seminal contributions to psychology and neurobiology (Boring, 1950; Campbell & Garnett, 1882; Glazebrook, 1905; Koenigsberger, 1906; Ratliff, 1965). Their interests in the structure of physical space-time were balanced by a fascination with psychological space-time. Thus their contributions to understanding the observed world developed side-by-side their analysis of the observer.

For example, every physicist knows about the Mach numbers and about the influence of Mach's ideas upon Einstein's thinking during the development of relativity theory. Mach is also famous, however, for his investigations of the Mach bands in vision. Surprisingly few scientists have studied both types of contributions in school. In a similar way, every physicist knows about Maxwell's fundamental contributions to electromagnetic theory and to the molecular theory of gases. Maxwell is equally well known, however, for his work on developing trichromatic color theory.

Helmholtz's life is an inspiration to us all. Trained as an M.D., his experiments on the velocity of electrical signals in nerve axons led him to help discover the principle of conservation of energy, which is one of the cornerstones of nineteenth-century physics. He made fundamental contributions to optics, which served as a foundation for his classical contributions to vision. His work in acoustics likewise supported his major contributions to hearing.

Thus during the last half of the nineteenth century, a number of great scientists functioned successfully in an interdisciplinary research mode and made lasting contributions to both the physical and psychobiological sciences.

## 3. THE SCHISM BETWEEN PHYSICS AND PSYCHOLOGY

It is often accepted as a truism that success breeds success, just as money makes money. Likewise, the great interdisciplinary successes of Helmholtz, Maxwell, and Mach might have been expected to breed droves of dedicated interdisciplinary disciples. This did not, however, occur. In the next generation of physicists, Einstein himself, in a letter to his friend Queen Elizabeth of Belgium in 1933, wrote: "Most of us prefer to look outside rather than inside ourselves; for in the latter case we see but a dark hole, which means: nothing at all" (Nathan & Norden, 1960, p. 567).

Thus a schism of major scientific importance occurred towards the end of the nineteenth century. Scientists whose work was previously greatly energized by interdisciplinary investigations of physics and psychology were rapidly replaced by scientists who rarely had even a rudimentary knowledge of the other field. Although the explosion of scientific knowledge during the twentieth century, with its attendant requirement to specialize, surely contributed to this schism, deeper intellectual factors exacerbated this schism. An understanding of these factors is useful for appreciating the scientific climate in which neural network research has been carried out during the past few decades.

## 4. THE NONLINEAR, NONLOCAL, AND NONSTATIONARY PHENOMENA OF MIND AND BRAIN

Basic causes of this schism emerged from the scientific work of the very pioneers, such as Helmholtz, Maxwell, and Mach, whose interdisciplinary careers we have been considering. Two examples from Helmholtz's work on visual perception are illustrative.

### A. Color Theory

In the classical Newtonian approach to color theory, white light is defined by an energy spectrum that is locally measurable at each point in space. In contrast, Helmholtz realized that, during visual perception, the average color of a whole scene tends to look white (Beck, 1972; Helmholtz, 1962). Thus, instead of being reduceable to local measurements at each location, the analysis of how humans perceive white light at each location necessitates an investigation of long-range (or *nonlocal*) interactions across a network of locations. Such investigations disclosed the role of these interactions in "discounting the illuminant," or enabling humans and other species to detect the actual reflectances of visible surfaces under a wide variety of illumination conditions. In addition to being nonlocal, the network interactions which discount the illuminant, being sensitive to image reflectances, are also *nonlinear*. The neural processes whereby illuminants are discounted are still the subject of intensive experimental and theoretical investigation (Arend, Buehler, & Lockhead, 1971; Cornsweet, 1970; Hurvich, 1981; Land, 1977; Mollon & Sharpe, 1983) and only recently have a large number of paradoxical brightness and color phenomena been analysed in a unified way using a real-time neural network model (Cohen & Grossberg, 1984; Grossberg, 1987a, 1987b; Grossberg & Mingolla, 1985a, 1985b; Grossberg & Todorović, in press).

### B. Top-Down Learning, Expectation, and Matching

Helmholtz faced another barrier when he attempted to conceptualize the process of visual perception itself. His conception is known as the doctrine of *unconscious inference* (Boring, 1950). This doctrine held that a raw sensory datum, or perzeption, is modified by previous experience via a learned imaginal increment, or vorstellung, before it becomes a true perception, or anschauung. Thus Helmholtz realized that we perceive, in part, what we *expect* to perceive based upon past *learning.*

Helmholtz's doctrine can be recast in modern terminology as follows (Figure 1). Bottom-up environmentally-activated input signals trigger the read-out of learned top-down expectations. These bottom-up and top-down data cooperate and compete through a matching process until they generate an emergent consensus which is the final percept. Such a cooperative-competitive network interaction also requires nonlinear and nonlocal interactions. In addition, the learning of top-down expectations requires a *nonstationary* process. Thus Helmholtz's experimental discoveries about visual perception led to the realization that theoretical understanding of these phenomena would require the discovery of appropriate nonlinear, nonlocal, and nonstationary mathematics, which are now being developed on multiple fronts.

In contrast, much of the mathematics available for physical theorizing during the nineteenth century was linear, local, and stationary mathematics. Thus the experimental discoveries about mind and brain by workers like Helmholtz, Maxwell, and Mach clarified that the available mathematics were not sufficient for supporting a sustained theoretical penetration of mind and brain mechanisms. Since theoretical scientists rely upon appropriate mathematical bread-and-butter techniques to express and develop their deepest intuitive ideas, the mismatch between psychological phenomena and nineteenth-century mathematics created an intellectual crisis for all theorists who might have wished to study mind and brain.

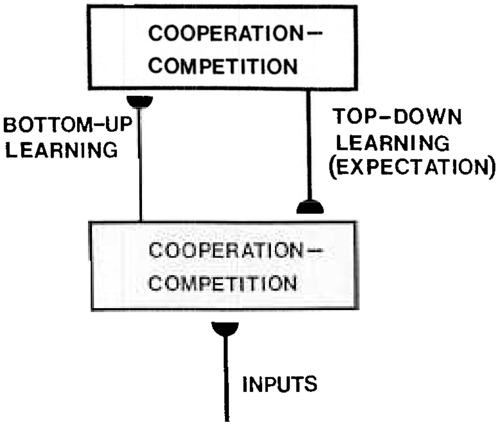


FIGURE 1. Bottom-up inputs and learned top-down expectations interact via a co-operative-competitive matching process until they generate an emergent consensus which represents the final, or resonant, percept.

This schism was exacerbated by the fact that the major revolutions of twentieth-century physics could be supported by nineteenth-century mathematics. For example, when Einstein finally realized that he needed a certain type of mathematics to express the general relativity theory, his burden was significantly lightened by the fact that nineteenth-century Riemannian geometry provided a perfect tool. As the early quantum mechanicians struggled towards expressing their intuitive insights using matrix theory and linear operator theory, they too were greatly aided by strong nineteenth-century mathematical traditions.

A major approach-avoidance paradigm was hereby established in the practice of theoretical science. Theoretical physicists abandoned psychology and neurobiology to rapidly fashion theories about the external world that could be quantitatively supported by available mathematical concepts and methods. Psychologists and neurobiologists returned the favor by abandoning physical concepts and mathematics that seemed irrelevant to their data and, over time, by also eschewing and even denegrating theoretical and mathematical training in general. This bifurcation was already apparent during the unfolding of Helmholtz's scientific life. Beginning his career as an M.D., he ended it as the first President of the new Physico-technical Institute in Berlin (Koenigsberger, 1906).

## 5. THE NATURE OF AN ENDURING SYNTHESIS

Left without an appropriate framework of concepts and mathematical techniques for interpreting and unifying their experiments, psychologists and neurobiologists nonetheless went about accumulating one of the largest and most sophisticated sets of data bases in the history of science. Remarkably, they accomplished this feat during a century of controversy that was spawned by the unavailability of a unifying theoretical and mathematical framework for explaining their data. As Hilgard and Bower (1975) have noted in their important textbook about theories of learning "Psychology seems to be constantly in a state of ferment and change, if not of turmoil and revolution" (p. 2).

While most mind and brain experimentalists ignored theory and most theorists looked for more hospitable frontiers, there arose the widespread tendency to interpret brain function in terms of whatever technological development happened to be current. The ever-expanding list of technological metaphors to which the brain has been compared includes telegraph circuits, hydraulic systems, information processing channels, digital computers, linear control systems, catastrophies, holograms, and spin glasses. All of these metaphors have been unable to explain a substantial data base about brain and behavior, as well they might, since none of them arose from a sustained analysis of behavioral or brain data.

The schism between physics and psychology encouraged theorists trained in the physics tradition to believe that no theories of behavior and brain exist. An inquiry about available theories by an interested physicist more often than not would confirm this impression, because the schism has prevented most psychologists and neurobiologists from getting the training necessary to understand the theories that have begun to cope with the nonlinear, nonlocal, and nonstationary nature of behavioral and brain data. Thus the theories which hold the greatest promise have been the ones that have been most difficult to evaluate in the social climate spawned by the great schism.

We can recognize in this sociological milieu touches of irony when we acknowledge that a key scientific issue in understanding behavior and brain is to explain how humans rapidly and spontaneously adapt to noisy and complex environments whose rules may change unexpectedly, or in William James' engaging phrase: How do we cope with the "blooming buzzing confusion" of every day? It remains to be seen how the several scientific communities now converging with enthusiasm but vastly different training and goals upon the interdisciplinary study of mind and brain will assimilate the noisy and unexpected constraints imposed by each others' existence, notably by the fact that, despite the extra burden of difficult sociological conditions, relevant theories of mind and brain have been developing rapidly during the past few decades.

## 6. SOURCES OF NEURAL NETWORK RESEARCH: BINARY, LINEAR, CONTINUOUS-NONLINEAR

### A. Binary

At least three sources of neural network research can be identified which have had a substantial influence on contemporary research. The streams of research generated by the sources have intersected in complex ways through the years and have tended to converge during the past several years. The present brief review merely sets the stage for the article's later discussions. Due to the sheer size and complexity of the neural network literature, this review must necessarily be selective. Other recent collections of classical and current neural network results include both articles (Carpenter & Grossberg, 1987b; Grossberg, 1987e; Hecht-Nielsen, 1986; Hestenes, 1987; Levine, 1983; Szu, 1986) and books (Amari & Arbib, 1982; Denker, 1986; Grossberg, 1982, 1987c, 1987d, 1988; Grossberg & Kuperstein, 1986; Hinton & Anderson, 1981; Kohonen, 1977, 1984; McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986).

Table 1 indicates several of the contributions that initiated or illustrate significant research developments. The stream of *binary* neural networks was initiated by the classical article of McCulloch and Pitts (1943). This article investigated threshold logic systems of the form

$$x_i(t + 1) = \mathrm{sgn}[\sum A_{ij}x_j(t) - B_j], \tag{1}$$

where sgn $(w) = +1$ if $w > 0$, 0 if $w = 0$, and $-1$ if $w < 0$. Such binary systems were inspired in part by neurophysiological observations showing that neural signals between many cells are carried by all-or-none spikes. The variables $x_i$ in Equation (1) are often called short term memory (STM) traces, or activations. Caianiello (1961) used a binary STM equation of the form

$$x_i(t + \tau) = 1[\sum_{j=1}^{n} \sum_{k=0}^{l(m)} A_{ij}^{(k)}x_j(t - k\tau) - B_i] \tag{2}$$

where $1(w) = 1$ if $w > 0$ and 0 if $w \leq 0$. Rosenblatt (1962) used an STM equation of the form

$$\frac{d}{dt}x_i = A_{x_i} + \sum_{j=1}^{n} \phi(B_j + x_j)C_{ij} \tag{3}$$

where $\phi(w) = 1$ if $w \geq \theta$ and 0 if $w < \theta$. Mueller, Martin, and Putzrath (1962) designed circuits which used the McCulloch-Pitts logical operations and also extended their analysis to analog circuits for applications to acoustic pattern recognition.

The binary, discrete-time approach to neural modeling was encouraged by the technical liberation which the use of oscilloscopes brought to neurophysiology. After years of heroic efforts to measure the tiny electrical signals in nerves, each spike could at last be easily

TABLE 1

| Binary | Linear | Continuous and Nonlinear |
|---|---|---|
| McCulloch-Pitts (1943) | | Hartline-Ratliff-Miller (1963) |
| Caianiello (1961) | | Grossberg (1967, 1968) |
| Rosenblatt (1962) | | Sperling-Sondhi (1968) |
| | | Wilson-Cowan (1972) |

amplified until it filled the whole oscilloscope screen. The all-or-none property of the individual spike was celebrated by making each spike much bigger than life. Although the oscilloscope provided a way for people to look at spikes, this representation did not necessarily correspond to what the cells which received the spikes were measuring. Cell body potentials may vary slowly and continuously relative to the time scale of a single spike. Thus neurons may process the frequencies or other statistical properties of spike sequences through time. If, for example, the spiking activity of a visual cortical feature detector is amplified by a microphone instead of by an oscilloscope, and if an object to which the detector is sensitive is brought in and out of its receptive field, one hears a continuous waxing and waning of the sound of the cell's spike discharges through time. If the potentials of the cells receiving these spike sequences fluctuate slowly enough to average across clusters of spikes, then such cells will be better modeled by continuous than binary dynamics.

Both Caianiello (1961) and Rosenblatt (1962) also introduced equations to change the weights $A_{ij}^{(k)}$ in (2) and $C_{ij}$ in (3) through learning. Such adaptive weights are often called long term memory (LTM) traces. Both workers decoupled the interactions between STM traces and LTM traces in order to partially analyze their nonlinear equations. These LTM equations also had a digital aspect. The equations of Caianiello (1961) increased or decreased at constant rates until they hit finite upper or lower bounds. Those of Rosenblatt (1962) were used to classify patterns into two distinct classes, as in the Perceptron Learning Theorem.

The historical importance of the binary McCulloch-Pitts (1943) model cannot be overestimated. For example, in addition to its seminal influence on neural modelling per se, it also was very much in the thoughts of von Neumann as he developed his ideas for the modern digital computer. In fact, a number of brain-inspired developments have found spin-offs over the years into other technologies.

## B. Linear

Concepts from linear system theory have provided a classical source of models for representing some of the continuous aspects of neural dynamics. Solutions of simultaneous linear equations $Y = AX$ using matrix theory and concepts about cross-correlation have been among the useful tools.

Inspired by an interest in brain modeling, Widrow (1962) developed his classical gradient descent *Adeline* adaptive pattern recognition machine before using this background to make his major contributions to the theory of adaptive antennas. Anderson (1968) initially described his intuitions about neural pattern recognition using the spatial cross-correlation function

$$\phi_{12}(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_1(i, j) f_2(i + x, j + y). \tag{4}$$

Kohonen (1971) made his transition from linear algebra concepts such as the Moore-Penrose pseudoinverse to more biologically motivated studies which he has summarized in his influential books (Kohonen, 1977, 1984). These workers thus began to develop their intuitions within a mathematically familiar engineering framework which was progressively developed to include more biologically motivated nonlinear interactions.

## C. Continuous-Nonlinear

Continuous-nonlinear network laws typically arose from a direct analysis of behavioral or neural data. One distinguished modeling tradition can be traced directly to the influence of Mach (Ratliff, 1965). This tradition set out to model data taken from the lateral eye of the *Limulus*, or horseshoe crab, and led to the award of a Nobel prize to H. K. Hartline.

The basic model from this tradition is the steady state Hartline-Ratliff model

$$r_i = e_i - \sum_{j=1}^{n} k_{ij}[r_j - r_{ij}]^+ \tag{5}$$

where $[w]^+ = \max(w, 0)$. This model describes how cellular excitations $e_i$ are transformed into net responses $r_i$ due to inhibitory feedback interactions governed by threshold-linear signals $-k_{ij}[r_j - r_{ij}]^+$. Thus, the Hartline-Ratliff model is a type of continuous threshold-logic system. Ratliff, Hartline, and Miller (1963) extended this steady-state model to a dynamical model of the form

$$r_i(t) = e_i(t) - \sum_{j=1}^{n} k_{ij}\left[\frac{1}{\tau}\int_0^t e^{-(t-s)/\tau} r_j(s)ds - r_{ij}\right]^+, \tag{6}$$

which also behaves linearly in the suprathreshold range. This model is a precursor of the additive model that is described below.

Another classical tradition arose from the analysis of how the excitable membrane of a single neuron can generate electrical spikes capable of rapidly and nondecrementally traversing the axon, or pathway, of the cell. The original experimental and modeling work on the squid giant axon by Hodgkin and Huxley (1952) also led to the award of a Nobel prize. Since this work focused on individual cells rather than networks of cells, it will not be further discussed herein except to note that it provides the foundation for the shunting model that is described below. The Hodgkin-Huxley model and some of its variations are reviewed elsewhere (Carpenter, 1981; Hodgkin, 1964; Hodgson, 1983; Katz, 1966; Plonsey & Fleming, 1969; Ricciardi & Scott, 1982; Scott, 1977).

Another source of continuous-nonlinear network models arose through a study of adaptive behavior, rather than of neural mechanisms per se, in Grossberg (1964, 1967, 1968a, 1968b). Its primary concern was to understand how the behavior of individuals adapts stably in real-time to complex and changing environmental contingencies. In order to analyze adaptive behavior, it is necessary to characterize the functional level on which a system's behavioral success is defined and achieved, as well as the computational units that are manipulated by this level. This behavioral analysis led to the derivation of continuous neural networks defined by nonlinearly coupled STM and LTM traces, and to the mathematical proof that the computational units of these networks are not individual STM and LTM variables, but are rather distributed spatial patterns of STM and LTM variables (Grossberg, 1968b, 1969a, 1969b, 1970a). Thus, neural networks describe a proper level for an analysis of adaptive behavior because the functional units which govern behavioral success are emergent properties due to interactions on the network level.

As in the tradition of binary models, this continuous-nonlinear approach defined laws for STM traces and LTM traces (Figure 2). The two primary versions of the STM equation which were introduced through this approach have been used in many applications since the 1960s and have received increasing experimental support.

*Additive STM Equation*

$$\frac{d}{dt} x_i = -A_i x_i + \sum_{j=1}^{n} f_j(x_j) B_{ji} z_{ji}^{(+)} - \sum_{j=1}^{n} g_j(x_j) C_{ji} z_{ji}^{(-)} + I_i. \quad (7)$$

Equation (7) includes a term for passive decay ($-A_i x_i$), positive feedback ($\sum_{j=1}^{n} f_j(x_j) B_{ji} z_{ji}^{(+)}$), negative feedback ($-\sum_{j=1}^{n} g_j(x_j) C_{ji} z_{ji}^{(-)}$), and input ($I_i$). Each feedback term includes a state-dependent nonlinear signal ($f_j(x_j)$, $g_j(x_j)$), a connection, or path, strength ($B_{ji}$, $C_{ji}$), and an LTM trace ($z_{ji}^{(+)}$, $z_{ji}^{(-)}$). If the positive and negative feedback terms are lumped together and the connection strengths are lumped with the LTM traces, then the additive model may be written in the simpler form

$$\frac{d}{dt} x_i = -A_i x_i + \sum_{j=1}^{n} f_j(x_j) z_{ji} + I_i. \quad (8)$$

Early applications of the additive model included computational analyses in vision, associative pattern learning, pattern recognition, classical and instrumental conditioning, and the learning of temporal order in applications to speech and language behavior and to planned sensory-motor control (Grossberg, 1969a, 1969b, 1969c, 1970a, 1970b, 1971a, 1972a, 1972b, 1974; Grossberg & Pepe, 1971). The additive model has continued to be a cornerstone of neural network research to the present day; see, for example, Amari and Arbib (1982) and Grossberg (1982). Some physicists unfamiliar with the classical status of the additive model in neural network theory erroneously called it the Hopfield model after they became acquainted with Hopfield's first application of the additive model in Hopfield (1984); see Section 9A. The classical McCulloch-Pitts (1943) model in Equation (1) has also erroneously been called the Hopfield model by some physicists who became acquainted with the McCulloch-Pitts model in Hopfield (1982). These historical errors can ultimately be traced to the schism between physics and psychology that was described in Section 3.

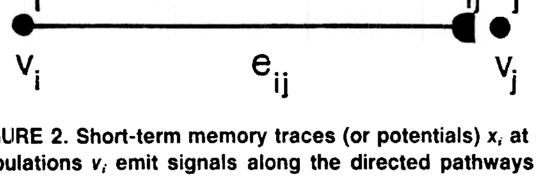


**FIGURE 2. Short-term memory traces (or potentials) $x_i$ at cell populations $v_i$ emit signals along the directed pathways (or axons) $e_{ij}$ which are gated by long-term memory traces $z_{ij}$ before they can perturb their target cells $v_j$.**

A related behaviorally derived STM equation was found to more adequately model the shunting dynamics of individual neurons (Hodgkin, 1964; Kandel & Schwartz, 1981; Katz, 1966; Plonsey & Fleming, 1969). In such a shunting equation, each STM trace is restricted to a bounded interval $[-D_i, B_i]$ and automatic gain control, instantiated by multiplicative shunting terms, interacts with balanced positive and negative feedback signals and inputs to maintain the sensitivity of each STM trace within its interval (see Section 15).

*Shunting STM Equation*

$$\frac{d}{dt} x_i = -A_i x_i + (B_i - x_i)\Big[\sum_{j=1}^{n} f_j(x_j) C_{ji} z_{ji}^{(+)} + I_i\Big]$$
$$- (x_i + D_i)\Big[\sum_{j=1}^{n} g_j(x_j) E_{ji} z_{ji}^{(-)} + J_i\Big]. \quad (9)$$

Several LTM equations have been useful in applications. Two particularly useful variations have been:

*Passive Decay LTM Equation*

$$\frac{d}{dt} z_{ij} = -F_{ij} z_{ij} + G_{ij} f_i(x_i) h_j(x_j) \quad (10)$$

and

*Gated Decay LTM Equation*

$$\frac{d}{dt} z_{ij} = h_j(x_j)[-F_{ij} z_{ij} + G_{ij} f_i(x_i)]. \quad (11)$$

In both equations, a nonlinear learning term $f_i(x_i) h_j(x_j)$, often called a Hebbian term after Hebb (1949), is balanced by a memory decay term. In (10), memory decays passively at a constant rate $-F_{ij}$. In (11), memory decay is gated on and off by one of the nonlinear signals. A key property of both equations is that the size of an LTM trace $z_{ij}$ can either increase or decrease due to learning. Neurophysiological support for an LTM equation of the form (11) has recently been reported (Levy, 1985; Levy, Brassel, & Moore, 1983; Levy &

Desmond, 1985; Rauschecker & Singer, 1979; Singer, 1983). Extensive computational analyses of these STM and LTM equations in a number of specialized circuits led gradually to the identification of a general class of networks for which one could prove invariant properties of associative spatio-temporal pattern learning and recognition (Grossberg, 1969a, 1971b, 1972c, 1982).

Sperling and Sondhi (1968) utilized a shunting STM equation in an important contribution to visual psychophysics. Wilson and Cowan (1972) introduced a modified shunting STM equation of the form

$$\frac{d}{dt} x_i = -A_i x_i + (B_i - x_i) f_i\left(\sum_{j=1}^{n} x_j C_{ji}\right) \tag{12}$$

which replaces the sum $\sum_{j=1}^{n} f_j(x_j) C_{ji} z_{ji}^{(+)}$ of nonlinear signals in (9) with a nonlinear function of the sum. Equation (12) possesses one automatic gain control term $(B_i - x_i)$, whereas (9) possesses two. Consequently, the dynamics of (12) saturate in many situations where the dynamics of (9) remain sensitive to input fluctuations (see Section 15).

## 7. NONLINEAR FEEDBACK BETWEEN FAST DISTRIBUTED STM PROCESSING AND SLOW ASSOCIATIVE LTM PROCESSING

These dynamical equations incorporate two general types of nonlinear processes which explicate some of the themes that were already touched upon in Helmholtz's work. On the one hand, there are the cooperative-competitive nonlinear feedback processes which operate on a relatively fast time scale. These processes instantiate the distributed information processing and STM storage capabilities of the network. They can, for example, carry out matching of bottom-up data with top-down expectations (Figure 1) to generate the perceptual consensus discussed by Helmholtz.

Interacting with these fast STM interactions via nonlinear feedback are the more slowly varying LTM processes which instantiate associative learning. Such a learning process can, for example, adaptively tune the bottom-up filters and encode the learned top-down expectations (Figure 1) that were adumbrated in Helmholtz's concept of unconscious inference.

## 8. PRINCIPLES, MECHANISMS, AND ARCHITECTURES

Such STM and LTM equations were discovered through the analysis of two mutually supportive, but complementary, types of results.

On the one hand, a small number of general design principles and their mechanistic instantiations were discovered through a comparative analysis of several interdisciplinary data bases. For example, the functional importance of the shunting STM equation (9) became clear through analyses of data about perception, conditioning, and cognitive information processing. Such analyses led to the realization that a single type of network was needed that was capable of ratio processing, conservation or normalization of total activation (limited capacity), Weber law modulation, adaptation level processing, noise suppression, contrast enhancement, short term memory storage, energetic amplification of matched input patterns, and energetic suppression of mismatched input patterns. The discovery that these multiple constraints are all satisfied by a ubiquitous type of on-center off-surround network of cells which obey the membrane equations of neurophysiology created an irresistable intellectual pressure to study them exhaustively (see Sections 9F and 13–15).

In addition to such general laws, a growing number of specialized architectures have also been developed. Each architecture is a synthesis of several types of design principles and mechanisms in a carefully crafted circuit. The organization of the brain into functionally distinctive regions—such as cerebellum, hippocampus, retina, visual cortex, parietal cortex, frontal cortex, hypothalamus, septum, amygdala, and reticular formation—illustrates why a considerable number of specialized architectures need to be developed.

Due to the highly interactive nature of brain dynamics, the development of general organizational principles, mechanisms, and specialized architectures have proceeded hand-in-hand, each bootstrapping the scientific understanding of the others. Here is a research area where it is essential to keep the forest, the trees, and the individual branches simultaneously in view. In the remainder of the article, I will summarize several of the principles, mechanisms, and architectures whose further development is still engaging the efforts of many scientists.

## 9. CONTENT-ADDRESSABLE MEMORY STORAGE: A GENERAL STM MODEL AND LIAPUNOV METHOD

From a mathematical perspective, the question of content-addressable memory (CAM) in a neural network can be formulated as follows: Under what conditions does a neural network always approach an equilibrium point in response to an arbitrary, but sustained, input pattern? The equilibrium point represents the stored pattern in response to the input pattern. In a satisfactory analysis of this problem, the behavior of the network in response to arbitrary initial data, an arbitrary sustained input pattern, and an arbitrary choice of network parameters is provided. Also an account of how many equilibrium points exist and of how they are approached through time is desirable. Such a mathematical analysis is called a global analysis, to distinguish it from a local stability analysis around individual equilibrium points.

Amari and Arbib (1982) and Levine (1983) include a number of contributions to the local analysis of neural networks. Our concern herein is with global methods. A global analysis of equilibrium behavior is of importance for an understanding both of CAM and of the types of nonequilibrium behavior—such as traveling waves, bursts, standing waves, and chaos—which can be obtained by perturbing off systems which always approach equilibrium (Carpenter, 1977a, 1977b, 1979, 1981; Cohen & Grossberg, 1983; Ellias & Grossberg, 1975; Ermentrout & Cowan, 1979, 1980; Hastings, 1976, 1982; Hodgson, 1983; Kaczmarek & Babloyantz, 1977). A global mathematical analysis of nonlinear associative learning networks was begun in Grossberg (1967, 1968b). A global mathematical analysis of nonlinear shunting cooperative-competitive feedback networks was begun in Grossberg (1973). Some of the main articles in these series are brought together in Grossberg (1982).

One approach to the global approach to equilibrium which has attracted widespread interest is the use of global Liapunov, or energy, methods. Such global methods were introduced for the analysis of neural networks in the 1970s. Herein I summarize a general model of a nonlinear cooperative-competitive neural network for which a global Liapunov function has been explicitly constructed. I then show that a number of popular models are special cases of the general model, and thus are capable of CAM.

Cohen and Grossberg (1983) described a general principle for designing CAM networks by proving that models that can be written in the form

$$\frac{d}{dt}x_i = a_i(x_i)[b_i(x_i) - \sum_{j=1}^{n} c_{ij}d_j(x_j)] \tag{13}$$

admit the global Liapunov function

$$V = -\sum_{i=1}^{n} \int^{x_i} b_i(\xi_i)d'_i(\xi_i)d\xi_i + \frac{1}{2}\sum_{j,k=1}^{n} c_{jk}d_j(x_j)d_k(x_k) \tag{14}$$

if the coefficient matrix $C = \|c_{ij}\|$ and the functions $a_i$, $b_i$, and $d_j$ obey mild technical conditions, including

*Symmetry*:

$$c_{ij} = c_{ji}, \tag{15}$$

*Positivity*:

$$a_i(x_i) \geq 0, \tag{16}$$

*Monotonicity*:

$$d'_j(x_j) \geq 0. \tag{17}$$

Integrating $V$ along trajectories implies that

$$\frac{d}{dt}V = -\sum_{i=1}^{n} a_i d'_i[b_i - \sum_{j=1}^{n} c_{ij}d_j]^2. \tag{18}$$

If (16) and (17) hold, then $(d/dt)V \leq 0$ along trajectories. Once this basic property of a Liapunov function is in place, it is a technical matter to rigorously prove that every trajectory approaches one of a possibly large number of equilibrium points.

For expository vividness, the functions in the Cohen-Grossberg model (13) are called the *amplification* function $a_i$, the *self-signal* function $b_i$, and the *other-signal* functions $d_j$. Specialized models are characterized by particular choices of these functions.

### A. Additive STM Equation

Cohen and Grossberg (1983, p. 819) noted that "the simpler additive neural networks . . . are also included in our analysis." The additive equation (8) can be written using the coefficients of the standard electrical circuit interpretation (Plonsey & Fleming, 1969) as

$$C_i\frac{dx_i}{dt} = -\frac{1}{R_i}x_i + \sum_{j=1}^{n} f_j(x_j)z_{ji} + I_i. \tag{19}$$

Substitution into (13) shows that

$$a_i(x_i) = \frac{1}{C_i} \text{ (constant!)} \tag{20}$$

$$b_i(x_i) = -\frac{1}{R_i}x_i + I_i \text{ (linear!)} \tag{21}$$

$$c_{ij} = -T_{ij} \tag{22}$$

and

$$d_j(x_j) = f_j(x_j). \tag{23}$$

Thus in the additive case, the amplification function (20) is a positive constant, hence satisfying (16), and the self-signal term (21) is linear. Substitution of (20)–(23) into (14) leads directly to the equation

$$V = \sum_{i=1}^{n}\frac{1}{R_i}\int^{x_i} \xi_i f'_i(\xi_i)d\xi_i - \sum_{i=1}^{n} I_i f_i(x_i) - \frac{1}{2}\sum_{j,k=1}^{n} T_{jk}f_j(x_j)f_k(x_k). \tag{24}$$

This Liapunov function for the additive model was later published by Hopfield (1984). In Hopfield's treatment, $\xi_i$ is written as an inverse $f_i^{-1}(V_i)$. Cohen and Grossberg (1983) showed, however, that although $f_i(x_i)$ must be nondecreasing, as in (17), it need not have an inverse in order for (24) to be valid.

### B. Brain-State-in-a-Box Model: $S\,\Sigma$ Exchange

The BSB model was introduced in Anderson, Silverstein, Ritz, and Jones (1977). It is often described in discrete time by the equation

$$x_i(t+1) = S(x_i(t) + \alpha\sum_{j=1}^{n} A_{ij}x_j(t)) \tag{25}$$

using symmetric coefficients

$$A_{ij} = A_{ji} \tag{26}$$

and a special type of nonlinear signal function $S(w)$

that characterizes the model. The signal function is a symmetric ramp function:

$$S(w) = \begin{cases} F & \text{if } w \geq F \\ w & \text{if } -F < w < F. \\ -F & \text{if } w \leq -F \end{cases} \tag{27}$$

Thus each STM trace $x_i$ obeys a *linear* equation until its argument reaches the *hard saturation* limit $F$.

The BSB model has been used to discuss categorical perception in terms of its formal contrast enhancement property that each $x_i$ tends to approach a limiting value $\pm F$, and thus that the vector $(x_1, x_2, \ldots, x_n)$ tends to approach a corner of the box $(\pm F, \pm F, \ldots, \pm F)$ as time goes on. An alternative explanation of contrast enhancement by a nonlinear feedback network was provided in Grossberg (1973) using a sigmoid signal function, rather than a function linear near zero, coupled to the soft saturation dynamics of a shunting network, rather than the hard saturation of a symmetric ramp (see Section 15). This is still a topic undergoing theoretical discussion (Anderson, Silverstein, Ritz, & Jones, 1977; Grossberg, 1978b, 1987d).

The BSB model can be rewritten as an additive model with no input and a special signal function that satisfies (17). Hence it is a special case of model (13). To see this, rewrite (25) in the form

$$x_i(t+1) = S\Big(\sum_{j=1}^{n} B_{ij}x_j(t)\Big) \tag{28}$$

using the coefficient

$$B_{ij} = \delta_{ij} + \alpha A_{ij} \tag{29}$$

where $\delta_{ij} =$ if $i = j$ and 0 if $i \neq j$. By (26), it follows that

$$B_{ij} = B_{ji}. \tag{30}$$

Although (28) is written in discrete time for computational convenience, it needs to be expressed in continuous time in order to represent a physical model, as in

$$\frac{d}{dt}x_i = -x_i + S\Big(\sum_{j=1}^{n} B_{ij}x_j\Big). \tag{31}$$

Define the new variables $y_i$ by

$$y_i = \sum_{j=1}^{n} B_{ij}x_j. \tag{32}$$

Then

$$\frac{d}{dt}y_i = -y_i + \sum_{j=1}^{n} B_{ij}S(y_j). \tag{33}$$

Comparison of (33) with (19) shows that the BSB model is an additive model such that each $I_i = 0$. Because this simple change of coordinates is so important in neural modeling, I give it a name: *$S\,\Sigma$ Exchange.*

The observation that, via $S\,\Sigma$ Exchange, a nonlinear signal of a sum, as in (31), can be rewritten as a sum of nonlinear signals, as in (33), shows that a number of models which have been treated as distinct are, in reality, mathematically identical. In contrast, this type of transformation cannot be carried out on shunting models such as (9) and (12).

The Liapunov function for (33) is found by directly substituting into model (13) expressed in terms of the variables $y_i$:

$$\frac{d}{dt}y_i = a_i(y_i)\Big[b_i(y_i) - \sum_{j=1}^{n} c_{ij}d_j(y_j)\Big]. \tag{34}$$

Since $a_i(y_i) = 1$, $b_i(y_i) = -y_i$, $c_{ij} = -B_{ij}$, and $d_j(y_j) = S(y_j)$, substitution into (14) yields

$$V = \sum_{i=1}^{n} \int^{y_i} \xi_i S'(\xi_i)d\xi_i - \tfrac{1}{2} \sum_{k=1}^{n} B_{jk}S(y_j)S(y_k). \tag{35}$$

Using the definitions in (27), (29), and (32), Equation (35) can be rewritten in terms of the original variables $x_i$ as follows:

$$V = -\frac{\alpha}{2} \sum_{j,k=1}^{n} A_{jk}x_jx_k. \tag{36}$$

Golden (1986) has derived (36) from a direct analysis of the BSB model.

### C. The McCulloch-Pitts Model

This classical model takes the form

$$x_i(t+1) = \operatorname{sgn}\Big(\sum_{j=1}^{n} A_{ij}x_j(t) - B_i\Big). \tag{1}$$

Letting

$$M(w) = \operatorname{sgn}(w - B_i), \tag{37}$$

(1) can be rewritten as

$$x_i(t+1) = M\Big(\sum_{j=1}^{n} A_{ij}x_j(t)\Big). \tag{38}$$

As in the analysis of (31), (38) can be rewritten in continuous time in terms of the variables $y_i$ via $S\,\Sigma$ Exchange:

$$\frac{d}{dt}y_i = -y_i + \sum_{j=1}^{n} A_{ij}M(y_j) \tag{39}$$

and is thus also a symmetric additive model with zero inputs. In addition, its signal function $M(y_j)$ has a zero derivative ($M'(y_j) = 0$) except at $y_j = 0$. Substitution of this additional property into (35) shows that the Liapunov function for the continuous time McCulloch-Pitts model is

$$V = -\tfrac{1}{2} \sum_{j,k=1}^{n} A_{jk}M(y_j)M(y_k), \tag{40}$$

which is the continuous time version of the discrete time Liapunov function described by Hopfield (1982).

### D. The Boltzmann Machine

The STM equation of the Boltzmann machine (Ackley, Hinton, & Sejnowski, 1985) has the same form as (31) and (38), and is thus also an additive equation with symmetric coefficients. Its signal function is the sigmoid logistic function

$$f(w) = \frac{1}{1 + e^{-w}}, \tag{41}$$

which satisfies (17) and is thus a special case of model (13). Thus the Boltzmann machine is a specialized additive model regulated by simulated annealing as developed by Geman (1983, 1984), Geman and Geman (1984), and Kirkpatrick, Gelatt, and Vecchi (1982, 1983).

### E. The Hartline-Ratliff-Miller Model

The $S\,\Sigma$ Exchange is not the only change of variables whereby CAM models can be transformed into an additive model format. For example, the STM equation (6) of the classical Hartline-Ratliff-Miller model is transformed into an additive model under an exponential change of variables

$$x_i(t) = \int_0^t e^{-(t-s)/\tau}\, r_j(s)ds. \tag{42}$$

Then (6) becomes

$$\frac{d}{dt} x_i = -\frac{1}{\tau} x_i - \sum_{j=1}^{n} \left[\frac{1}{\tau} x_j - r_{ij}\right]^{+} k_{ij} + e_i. \tag{43}$$

### F. Shunting Cooperative-Competitive Feedback Network

All additive models lead to constant amplification functions $a_i(x_i)$ and linear self-feedback functions $b_i(x_i)$. The need for the more general model (13) becomes apparent when the shunting STM equation (9) is analyzed. Consider, for example, a class of shunting models in which each node can receive excitatory and inhibitory inputs $I_i$ and $J_i$, respectively, and each node can excite itself and can inhibit other nodes via nonlinear feedback. Such networks model on-center off-surround interactions among cells which obey membrane equations (Grossberg, 1973; Hodgkin, 1964; Kandel & Schwartz, 1981; Katz, 1966; Plonsey & Fleming, 1969). In particular, let

$$\frac{d}{dt} x_i = -A_i x_i + (B_i - x_i)[I_i + f_i(x_i)] - (x_i + C_i)[J_i + \sum_{j=1}^{n} D_{ij} g_j(x_j)]. \tag{44}$$

In (44), each $x_i$ can fluctuate within the finite interval $[-C_i, B_i]$ in response to the constant inputs $I_i$ and $J_i$, the state-dependent positive feedback signal $f_i(x_i)$, and the negative feedback signals $D_{ij} g_j(x_j)$. It is assumed that

$$D_{ij} = D_{ji} \geq 0 \tag{45}$$

and that

$$g_j'(x_j) \geq 0. \tag{46}$$

In order to write (44) in Cohen-Grossberg form, it is convenient to introduce the variables

$$y_i = x_i + C_i. \tag{47}$$

In applications, $C_i$ is typically nonnegative. Since $x_i$ can vary within the interval $[-C_i, B_i]$, $y_i$ can vary within the interval $[0, B_i + C_i]$ of nonnegative numbers. In terms of these variables, (44) can be written in the form

$$\frac{d}{dt} y_i = a_i(y_i)[b_i(y_i) - \sum_{j=1}^{n} C_{ij} d_j(y_j)] \tag{34}$$

where

$$a_i(y_i) = y_i \quad \text{(nonconstant!)}, \tag{48}$$

$$b_i(y_i) = \frac{1}{x_i}[A_i C_i - (A_i + J_i)x_i + (B_i + C_i - x_i) \times (I_i + f_i(x_i - C_i))] \quad \text{(nonlinear!)}, \tag{49}$$

$$C_{ij} = D_{ij}, \tag{50}$$

and

$$d_j(y_j) = g_j(y_j - C_j) \quad \text{(noninvertible!)}. \tag{51}$$

Unlike the additive model, the amplification function $a_i(y_i)$ in (48) is not a constant. In addition, the self-signal function $b_i(y_i)$ in (49) is not necessarily linear, notably because the feedback signal $f_i(x_i - C_i)$ is often nonlinear in applications of the shunting model; in particular it is often a sigmoid or multiple sigmoid signal function (Ellias & Grossberg, 1975; Grossberg, 1973, 1977, 1978c; Grossberg & Levine, 1975; Sperling, 1981). Sigmoid signal functions, and approximations thereto, also appear in applications of the additive model and its variants (Ackney, Hinton, & Sejnowski, 1985; Amari & Arbib, 1982; Freeman, 1975, 1979; Grossberg, 1969a, 1982; Grossberg & Kuperstein, 1986; Hinton & Anderson, 1981; Hopfield, 1984; Rumelhart & McClelland, 1986). Such applications do not require the full generality of the Liapunov function (13) because the nonlinear signal function can then be absorbed into the terms $d_j(x_j)$.

Property (16) follows from the fact that $a_i(y_i) = y_i \geq 0$. Property (17) follows from the assumption that the negative feedback signal function $g_j$ is monotone non-decreasing. Cohen and Grossberg (1983) proved that $g_j$ need not be invertible. A signal threshold may exist below which $g_j = 0$ and above which $g_j$ may grow in a nonlinear way. The inclusion of nonlinear signals with thresholds better enables the model to deal with fluctuations due to subthreshold noise. On the other hand, thresholds are not the only mechanisms which can suppress noise in a cooperative-competitive feedback network (see Section 15D).

### G. Masking Field Model

In many applications of the shunting and additive models, the coefficients $c_{ij}$ in (13) may be asymmetric, thereby rendering the Liapunov function (14) inapplicable. Asymmetric coefficients typically occur in problems relating to the learning and recognition of temporal order in behavior. Consequently, a number of mathematical methods were developed from the earliest days of the continuous-nonlinear approach to analyse models with asymmetric interaction coefficients.

On the other hand, certain network models may have asymmetric interaction coefficients, yet be reduceable to the form (13) with symmetric interaction coefficients through a suitable change of variables. The masking field model is a shunting network of this type. The masking field model was introduced in Grossberg (1978a; reprinted in Grossberg, 1982) to explain data about speech learning, word recognition, and the learning of adaptive sensory-motor plans. It has been further developed through computer simulations in Cohen and Grossberg (1986, 1987). A masking field is a multiple-scale, self-similar, automatically gain controlled, cooperative-competitive nonlinear feedback network (Figure 3) which can generate a compressed but distributed STM representation of an input pattern as a whole, of its most salient parts, and of predictive codes which represent larger input patterns of which it forms a part. The masking field model is thus a specialized type of vector quantization scheme (Gray, 1984). Its multiple-scale self-similar properties imply its asymmetric interaction coefficients.

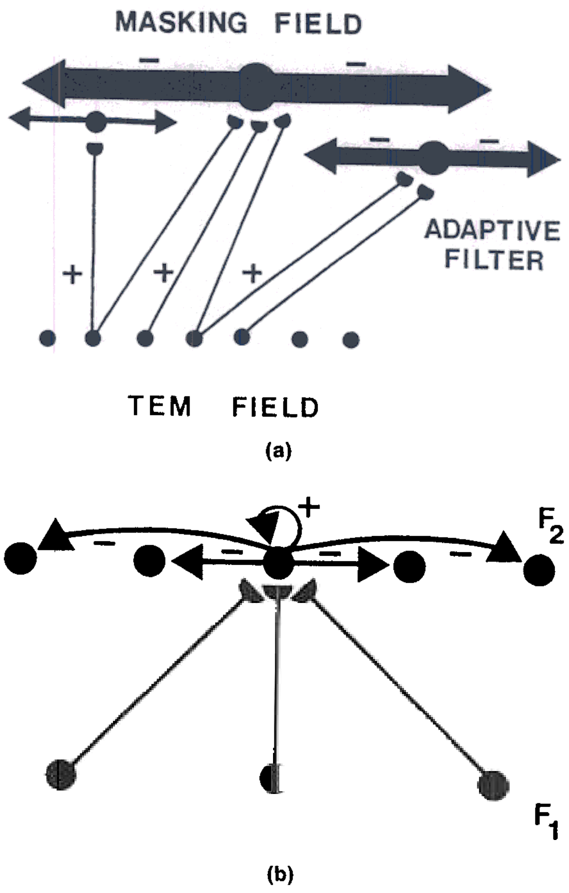


FIGURE 3. Masking field interactions: (a) Cells from an item field $F_1$ grow randomly to a masking field $F_2$ along positionally sensitive gradients. The nodes in the masking field grow so that larger item groupings, up to some optimal size, can activate nodes with broader and stronger inhibitory interactions. Thus the $F_1 \rightarrow F_2$ connections and the $F_2 \leftrightarrow F_2$ interactions exhibit properties of self-similarity; (b) The interactions within a masking field $F_2$ include positive feedback from a node to itself and negative feedback from a node to its neighbors. Long term memory (LTM) traces at the ends of $F_1 \rightarrow F_2$ pathways (designated by hemidisks) adaptively tune the filter defined by these pathways to amplify the $F_2$ reaction to item groupings which have previously succeeded in activating their target $F_2$ nodes. (Reprinted with permission from Cohen & Grossberg, 1987, p. 1868.)

The STM equation of a typical masking field is defined by

$$\frac{d}{dt} x_i^{(J)} = -Ax_i^{(J)} + (B - x_i^{(J)})[\sum_{j\in J} E_j p_{ji}^{(J)} + D|J|f(x_i^{(J)})]$$
$$- (x_i^{(J)} + C) \frac{\Sigma_{m,K} g(x_m^{(K)})|K|(1 + |K \cap J|)}{\Sigma_{m,K}|K|(1 + |K \cap J|)}. \quad (52)$$

In (52), $x_i^{(J)}$ is the STM trace of the $i$th masking field node that receives excitatory input $\Sigma_{j\in J} E_j p_{ji}^{(J)}$ from the unordered set $J$ of input items. Notation $|J|$ counts the number of items in set $J$ and thereby keeps track of the number of spatial scales that go into each version of the model.

The inhibitory interaction coefficient

$$\frac{|K|(1 + |K \cap J|)}{\Sigma_{m,K}|K|(1 + |K \cap J|)} \quad (53)$$

in (52) is an asymmetric function of $J$ and $K$. Despite this fact, (52) can be written in Cohen-Grossberg form as

$$\frac{d}{dt} y_i^{(J)} = a_i^{(J)}(y_i^{(J)})[b_i^{(J)}(y_i^{(J)}) - \sum_{m,K} c_{JK} d^{(K)}(y_m^{(K)})] \quad (54)$$

with symmetric coefficients

$$c_{JK} = c_{KJ} = 1 + |K \cap J| \quad (55)$$

in terms of the variables

$$y_i^{(J)} = F_{|J|}^{-1}(x_i^{(J)} + C) \quad (56)$$

where

$$F_{|J|} = \sum_{m,K} |K|(1 + |K \cap J|). \quad (57)$$

This is seen as follows. Since $F_{|J|}$ is the denominator of (53), it can be used to divide term $x_i^{(J)} + C$ in (52). Then the asymmetric term $|K|$ in the numerator of (53) can be absorbed into the definition of $g$ in (54). Then by redefining and rearranging terms as in (47)–(51), equation (54) holds with

$$a_i^{(J)}(y_i^{(J)}) = F_{|J|}^{-1} y_i^{(J)} \quad (58)$$

$$b_i^{(J)}(y_i^{(J)}) = \frac{1}{y_i^{(J)}}\left[AC - AF_{|J|} y_i^{(J)} + \left(\frac{B + C}{F_{|J|}} - y_i^{(J)}\right)(I_i^{(J)}\right.$$
$$\left. + D|J|F_{|J|} f(F_{|J|} y_i^{(J)} - C))\right] \quad (59)$$

where

$$I_i^{(J)} = F_{|J|} \sum_{j \in J} E_j p_{ji}^{(J)}, \tag{60}$$

and

$$d^{(K)}(y_m^{(K)}) = |K| g(F_{|K|} y_m^{(K)} - C). \tag{61}$$

Thus the masking field model is a specialized Cohen-Grossberg model.

### H. Bidirectional Associative Memories: Symmetrizing an Asymmetric Interaction Matrix

Other procedures have also been devised for dealing with systems having asymmetric coefficients. For example, given an arbitrary $n \times m$ coefficient matrix $Z = \|z_{ij}\|$ from a network level $F_1$ to a network level $F_2$ with STM traces $x_i$ and $y_j$, respectively. Kosko and Guest (in press) and Kosko (1987) have shown that (13) and (14) can be used to construct feedback pathways from $F_2$ to $F_1$ so that the two-level feedback network $F_1 \leftrightarrow F_2$ has convergent trajectories.

For example, if the bottom-up interaction $F_1 \rightarrow F_2$ obeys an additive equation

$$\frac{d}{dt} y_j = -A_j y_j + \sum_k f_k(x_k) z_{kj} + I_j, \tag{62}$$

then the top-down interaction $F_2 \rightarrow F_1$ is defined to obey an additive equation

$$\frac{d}{dt} x_i = -B_i x_i + \sum_l g_l(y_l) z_{il} + J_i, \tag{63}$$

where $I_j$ and $J_i$ are input terms. This definition creates a symmetric interaction matrix by closing the top-down feedback loop, since if $f_i(x_i)$ influences $y_j$ with coefficient $z_{ij}$ in (62), then $g_j(y_j)$ influences $x_i$ with the same coefficient $z_{ij}$. Thus, by defining an augmented vector $(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ of STM activities, system (62)–(63) as a whole define an additive model (19) with an $(n + m) \times (n + m)$ symmetric coefficient matrix.

The same procedure can be used to symmetrize many other neural network models. Kosko and Guest (in press) have described optical implementations for this procedure, and Kosko (1987) has used the symmetrized additive model to discuss minimization of fuzzy entropy.

### I. Volterra-Lotka, Gilpin-Ayala, and Eigen-Schuster Models

The Cohen-Grossberg model was designed to also include models which arose in other areas of biology than neural network theory. For example, it includes the classical

*Volterra-Lotka Model*

$$\frac{d}{dt} x_i = A_i x_i (1 - \sum_{j=1}^{n} B_{ij} x_j) \tag{64}$$

of population biology (Lotka, 1956), the

*Gilpin-Ayala Model*

$$\frac{d}{dt} x_i = A_i x_i \left[ 1 - \left( \frac{x_i}{B_i} \right)^{\theta_i} - \sum_{j=1}^{n} C_{ij} \left( \frac{x_j}{B_j} \right) \right], \tag{65}$$

also from population biology (Gilpin & Ayala, 1973), and the

*Eigen-Schuster Model*

$$\frac{d}{dt} x_i = x_i (A_i x_i^{p-1} - q \sum_{j=1}^{n} A_j x_j^p) \tag{66}$$

from the theory of macromolecular evolution (Eigen & Schuster, 1978). In all of these models, either the amplification function $a_i(x_i)$ is non-constant, or the self-signal function $b_i(x_i)$ is nonlinear, or both.

The specialized models summarized in Sections 9A–9I illustrate that model (13) and Liapunov function (14) embody a general principle for designing CAM devices from cooperative-competitive feedback models. These models are said to be *absolutely stable* because the CAM property is not destroyed by changing the parameters, inputs, or initial values of the model. The persistence of the CAM property under arbitrary parameter changes enables learning to change system parameters in response to unpredictable input environments without destroying CAM. The STM transformation executed by a network with adaptively altered parameters can differ significantly from its original STM transformation. A finer analysis is needed to choose models, as in Sections 9A–9I, which are optimally designed to carry out specialized processing tasks.

The Cohen-Grossberg analysis emphasizes the critical role of mathematical analysis in classifying and understanding very large systems of nonlinear neural networks (VLSN). Without such an integrative approach, it is difficult to tell whether or not a model is really new computationally, or whether it is a special case of a known model in slightly different coordinates or notation. For example, many scientists have not realized that models (31) and (33) are mathematically equivalent. Table 2 describes the relationships between models disclosed by such an analysis. Thus the BSB model enjoys a CAM property for the same reason that

TABLE 2
CAM Models in Decreasing Generality

| | | |
|---|---|---|
| CG (1983) | ADDITIVE (1967) | MP (1943)<br>BSB (1977)<br>BM (1985)<br>BAM (1987) |
| | SHUNTING (1973) | MF (1978, 1986) |

Organization in terms of decreasing generality of the models described in Section 9. Abbreviations: CG = Cohen-Grossberg; MP = McCulloch-Pitts; BSB = Brain-State-in-a-Box; BM = Boltzmann Machine; BAM = Bidirectional Associative Memory; MF = Masking Field.

any additive or CG model does. On the other hand, the BSB model may have special properties that may make it ideal for certain tasks, or it may be too specialized to accomplish certain tasks which are better dealt with using a shunting model.

## 10. OTHER LIAPUNOV METHODS

A considerable amount of work was done on finding Liapunov functions for special cases of (13) before the appearance of Cohen and Grossberg (1983). A global Liapunov method was also developed which is in some respects more general than that of Cohen and Grossberg (1983).

In the former category, MacArthur (1970) described a quadratic Liapunov function for proving local asymptotic stability of isolated equilibrium points of Volterra-Lotka systems with symmetric coefficients. Goh and Agnew (1977) described a global Liapunov function for Volterra-Lotka and Gilpin-Ayala systems in cases where only one equilibrium point exists. Liapunov functions were also described for Volterra-Lotka systems whose off-diagonal terms are relatively small (Kilmer, 1972; Takeuchi, Adachi, & Tokumaru, 1978). Such constraints are, however, too limiting for the design of CAM systems aimed at transforming and storing a large variety of patterns.

## 11. TESTING THE GLOBAL CONSISTENCY OF DECISIONS IN COMPETITIVE SYSTEMS

An alternative approach began with the global analysis in Grossberg (1973) of the nonlinear dynamics of shunting cooperative-competitive feedback networks. The goal of this analysis was to design CAM networks capable of transforming and stably storing in STM large numbers of patterns (see Section 15). The first analyses carried out direct proofs of the STM transformation and storage properties for small classes of shunting networks which arose in specialized applications. Later articles (Ellias & Grossberg, 1975; Grossberg & Levine, 1975; Levine, 1979; Levine & Grossberg, 1976) classified the global CAM behavior of increasingly large sets of networks.

These results led to the progressive development in Grossberg (1977, 1978c, 1978d, 1980a) of a global Liapunov method for classifying the dynamical behaviors of a wider variety of competitive dynamical systems. A competitive dynamical system is, for present purposes, defined by a system of differential equations such that

$$\frac{d}{dt} x_i = f_i(x_1, x_2, \ldots, x_n) \tag{67}$$

where

$$\frac{\partial f_i}{\partial x_j} \le 0, \quad i \ne j, \tag{68}$$

and the $f_i$ are chosen to generate bounded trajectories. By (68), increasing the activity $x_j$ of a given population can only decrease the growth rates $d/dt\, x_i$ of other populations, $i \ne j$, or may not influence them at all. No constraint is placed upon the sign of $\partial f_i/\partial x_i$. Typically, cooperative behavior occurs within a population and competitive behavior occurs between populations, as in the on-center off-surround networks (44). Since this Liapunov method led to results which are still of current interest and which seem amenable to further development, some of its most salient points will be summarized here.

The method makes mathematically precise the simple intuitive idea that a competitive system can be understood by keeping track of who is winning the competition. To do this, write (67) in the form

$$\frac{d}{dt} x_i = a_i(x_i) M_i(x) \quad x = (x_1, x_2, \ldots, x_n), \tag{69}$$

which factors out the amplification function $a_i(x_i) \ge 0$. Then define

$$M^+(x) = \max\{M_i(x): i = 1, 2, \ldots, n\} \tag{70}$$

and

$$M^-(x) = \min\{M_i(x): i = 1, 2, \ldots, n\}. \tag{71}$$

These variables track the largest and smallest rates of change, and are used to keep track of who is winning. Using these functions, it is easy to see that there exists a property of *ignition*: Once a trajectory enters the *positive ignition region*

$$R^+ = \{x: M^+(x) \ge 0\} \tag{72}$$

or the *negative ignition region*

$$R^- = \{x: M^-(x) \le 0\}, \tag{73}$$

it can never leave it. If $x(t)$ never enters the set

$$R^* = R^+ \cap R^-, \tag{74}$$

then each variable $x_i(t)$ converges monotonically to a limit. The interesting behavior in a competitive system occurs in $R^*$. In particular, if $x(t)$ never enters $R^+$, each $x_i(t)$ decreases to a limit; then the competition never gets started. The set

$$S^+ = \{x: M^+(x) = 0\} \tag{75}$$

acts like a competition threshold, which is called the *positive ignition hypersurface.*

We therefore consider a trajectory after it has entered $R^*$. For simplicity, redefine the time scale so that the trajectory is in $R^*$ at time $t = 0$. The Liapunov functional for any competitive system is then defined as

$$L(x_t) = \int_0^t M^+(x(v))dv. \tag{76}$$

The Liapunov property is a direct consequence of positive ignition:

$$\frac{d}{dt} L(x_t) = M^+(x(t)) \ge 0.$$

This functional provides the "energy" that forces trajectories through a series of competitive decisions, which are also called *jumps*. Jumps keep track of the state which is undergoing the *maximal* rate of change at any time ("who's winning"). If $M^+(x(t)) = M_i(x(t))$ for times $S \leq t < T$ but $M^+(x(t)) = M_j(x(t))$ for times $T \leq t < U$, then we say that the system *jumps* from node $v_i$ to node $v_j$ at time $t = T$. A jump from $v_i$ to $v_j$ can only occur on the *jump set*

$$J_{ij} = \{x \in R^*: M^+(x) = M_i(x) = M_j(x)\}. \tag{78}$$

The Liapunov functional $L(x_t)$ moves the system through these decision hypersurfaces through time. The geometry of $S^+$, $S^-$, and the jump sets $J_{ij}$, together with the energy defined by $L(x_t)$, can be used to globally analyze the dynamics of the system. In particular, due to the positive ignition property (77), the limit

$$\lim_{t\to\infty} L(x_t) = \int_0^\infty M^+(x(v))dv \tag{79}$$

always exists, and is possibly infinite.

The following results illustrate the use of these concepts (Grossberg, 1978d):

*Theorem* 1: Given any initial data $x(0)$, suppose that

$$\int_0^\infty M^+(x(v))dv < \infty. \tag{80}$$

Then the limit $x(\infty) = \lim_{t\to\infty} x(t)$ exists.

*Corollary* 1: If in response to initial data $x(0)$, all jumps cease after some time $T < \infty$, then $x(\infty)$ exists.

Speaking intuitively, this result means that after all local decisions, or jumps, have been made in response to an initial state $x(0)$, then the system can settle down to a global decision, or CAM $x(\infty)$. In particular, if $x(0)$ leads to only finitely many jumps because there exists a jump tree, or partial ordering of decisions, then $x(\infty)$ exists. This fact led to the analysis of circumstances under which no jump cycle, or repetitive series of jumps, occurs in response to $x(0)$, and hence that jump trees exist.

Further information follows readily from (80). Since $M^+(x(t)) \geq 0$ for all $t \geq 0$, it also follows that $\lim_{t\to\infty} M^+(x(t)) = 0$. This tells us to look for the equilibrium points $x(\infty)$ on the positive ignition hypersurface $S^+$ in (75):

*Corollary* 2: If $\int_0^\infty M^+(x(t))dt < \infty$, then $x(\infty) \in S^+$.

Thus the positive ignition surface is the place where the competition both ignites and is stored if no jump cycle exists. Using this result, an analysis was made of conditions under which no jump cycle exists in response to any initial vector $x(0)$, and hence all trajectories approach an equilibrium or CAM state.

The same method was also used to prove that a competitive system can generate sustained oscillations if it contains globally inconsistent decisions. These results are important for understanding the role of symmetric coefficients in the design of CAM systems. They identified circumstances under which, in response to initial data $x(0)$,

$$\int_0^\infty M^+(x(v))dv = \infty, \tag{81}$$

thus that infinitely many jumps occur, hence a jump cycle occurs, and finally that the trajectory undergoes undamped oscillations.

This method was used to provide a global analysis of the oscillations taking place in the May-Leonard (1975) model of the voting paradox. In this specialized Volterra-Lotka model,

$$\begin{aligned}\frac{d}{dt}x_1 &= x_1(1 - x_1 - \alpha x_2 - \beta x_3)\\ \frac{d}{dt}x_2 &= x_2(1 - \beta x_1 - x_2 - \alpha x_3)\\ \frac{d}{dt}x_3 &= x_3(1 - \alpha x_1 - \beta x_2 - x_3)\end{aligned} \tag{82}$$

and the parameters are chosen to satisfy $\beta > 1 > \alpha$ and $\alpha + \beta > 2$. System (82) represents the following intuitive situation. Three "candidates" are run against each other in pairwise elections. If $v_1$ wins over $v_2$, $v_2$ wins over $v_3$, and $v_3$ wins over $v_1$, what happens when all three candidates run against each other? If the winning relationship were transitive, then $v_1$ could win over himself! Thus the voting paradox illustrates how a globally inconsistent decision scheme can arise.

In (82) the relationship "$v_i$ wins over $v_j$" is represented by "$v_i$ inhibits $v_j$ more than $v_j$ inhibits $v_i$." In particular, $v_1 > v_2 > v_3 > v_1$. May and Leonard (1975) did computer simulations which showed that the trajectories of (82) oscillate. Grossberg (1978d) proved that the trajectories oscillate because system (82) generates a globally inconsistent decision scheme, characterized by a jump cycle $v_1 \to v_2 \to v_3 \to v_1$ with $L(x_\infty) = \infty$, for almost all trajectories.

The interaction matrix

$$\begin{pmatrix} 1 & \alpha & \beta \\ \beta & 1 & \alpha \\ \alpha & \beta & 1 \end{pmatrix} \tag{83}$$

of system (82) can be chosen arbitrarily close to a symmetric matrix by letting $\alpha$ and $\beta$ approach 1 without violating the constraint $\beta > 1 > \alpha$ and $\alpha + \beta > 2$. Thus there exist competitive systems whose matrices are arbitrarily close to symmetric matrices almost all of whose trajectories oscillate, albeit slowly. There also exist competitive systems without jump cycles whose coefficients are not symmetric, yet approach equilibrium points, because they satisfy Theorem 1. Although symmetry may be sufficient to generate CAM, as in model (13), the concepts of jump cycle and jump tree illustrate that one needs to analyze more global geometrical concepts to understand the relationship between a system's symmetry and its emergent CAM properties.

The Liapunov functional method led to the Cohen and Grossberg (1983) analysis in the following way. The Liapunov functional method was used to prove a theorem about the global CAM behavior of the competitive *adaptation level systems*

$$\frac{d}{dt} x_i = a_i(x)[b_i(x_i) - c(x)] \tag{84}$$

which were identified through an analysis of many specialized networks. In system (84), each state-dependent amplification function $a_i(x)$ and self-signal function $b_i(x_i)$ can be chosen with great generality without destroying the system's ability to reach equilibrium because there exists a state-dependent *adaptation level* $c(x)$ against which each $b_i(x_i)$ is compared. Such an adaptation level $c(x)$ defines a strong type of long-range symmetry within the system.

The examples which motivated the analysis of (84) were additive networks

$$\frac{d}{dt} x_i = -A_i x_i + \sum_k f_k(x_k) B_{ki} + I_i \tag{85}$$

and shunting networks

$$\frac{d}{dt} x_i = -A_i x_i + (B_i - x_i)[I_i + \sum_k f_k(x_k) C_{ki}] - (x_i + D_i)[J_i + \sum_k g_k(x_k) E_{ki}] \tag{86}$$

in which the symmetric coefficients $B_{ki}$, $C_{ki}$, and $E_{ki}$ took on different values when $k = i$ and when $k \neq i$. Examples in which the symmetric coefficients varied with $|k - i|$ in a graded fashion were also studied through computer simulations (Ellias & Grossberg, 1975; Levine & Grossberg, 1976), but an adequate global mathematical convergence proof was not available before Cohen and Grossberg (1983).

In the proof of the global convergence theorem (Grossberg, 1978c, 1980a) for systems of the form (84), it was shown that each $x_i(t)$ gets trapped within a sequence of decision boundaries that get laid down through time at the abscissa values of the highest peaks in the graphs of the functions $b_i$. The size and location of these peaks reflect the statistical rules, which can be chosen extremely complex, that give rise to the output signals from the totality of cooperating subpopulations within each node $v_i$. In particular, a $b_i$ with multiple peaks can be generated when a population's positive feedback signal function is a multiple-sigmoid function which adds up output signals from multiple randomly defined subpopulations within $v_i$.

After all the decision boundaries get laid down, each $x_i$ is trapped within a single valley of its $b_i$ graph. This valley acts, in some respects, like a classical potential. Correspondingly, it was proved that after all the $x_i$ get trapped in such valleys, the function

$$B[x(t)] = \max\{b_i(x(t)): i = 1, 2, \ldots, n\} \tag{87}$$

is a Liapunov function. This Liapunov property was used to complete the proof of the theorem.

The adaptation level model (84) is in some ways more general and in some ways less general than model (13). Cohen and I began our study of (13) with the hope that we could use the symmetric coefficients in (13) to prove that no jump cycles exist, and thus that all trajectories approach equilibrium as a consequence of the general Theorem 1. Such a proof is greatly to be desired because it would be part of a more general theory and, by using geometrical concepts such as jump set and ignition surface, it would clarify how to perturb off the symmetric case without generating oscillations such as the voting paradox (82). As it is, the Liapunov function (14) does not necessarily require that the system (13) be competitive because, by (18), $(d/dt)\ V_i \leq 0$ whether or not the coefficients $c_{ij}$ are all nonnegative.

Hirsch (1982, 1985) has proved powerful global theorems about the class of cooperative systems

$$\frac{d}{dt} x_i = f_i(x_1, x_2, \ldots, x_n) \tag{88}$$

where

$$\frac{\partial f_i}{\partial x_j} \geq 0, \quad i \neq j. \tag{89}$$

One of the outstanding mathematical problems in neural network theory is to find more general methods than the Cohen and Grossberg, Grossberg, and Hirsch results for designing mixed cooperative-competitive feedback systems with desired global behavior.

## 12. STABLE PRODUCTION STRATEGIES FOR A COMPETITIVE MARKET

The properties of adaptation level systems may prove useful in areas far removed from neural networks. To illustrate this possible range, consider the problem of how to design a competitive market such that every competing firm can choose one of infinitely many production strategies, each choice is unknown to the other competitors, yet the market generates a stable price and each firm balances its books.

Let $x_i$ denote the amount produced by firm $i$ of the commodity; $P(x)$ denote the market price per item of the commodity, where $x = (x_1, x_2, \ldots, x_n)$; $C_i(x_i)$ denote the cost per item of firm $i$; and $A_i(x)$ ($\geq 0$) denote a multiplier chosen by firm $i$. Let the firms agree to govern their individual production plans according to the adaptation level system

$$\frac{d}{dt} x_i = x_i A_i(x)[P(x) - C_i(x_i)].$$

The competitive property of the market is expressed by the conditions

$$\frac{\partial P}{\partial x_i} < 0, \quad i = 1, 2, \ldots, n. \tag{91}$$

In order to play this market, each firm compares its private cost function with the publicly known market price. If $A_i(x)$ depends only on $x_i$, then this is all that the firm needs to know to determine its production rate $dx_i/dt$. If $A_i(x)$ depends on amount $x_j$ produced by other firms, then each firm needs also to know how much the other firms are producing. In either case, no firm knows the internal strategies $A_i(x)$ and $C_i(x_i)$ of the other firms, which can be very complex. Nor does any firm need to know the function $P(x)$, which can also be very complex. All it needs to know are the values of $P(x(t))$ through time, which it can read in a trade newspaper.

By the adaptation level convergence theorem, limits $\lim_{t\to\infty} P(x(t))$ and $\lim_{t\to\infty} C_i(x_i(t))$ exist and are equal. Thus the market price is stable and every firm breaks even. If the definition of $C_i(x_i)$ also includes a savings factor, then the savings functions of all the firms would also be satisfied.

In this generality, the theorem does not say what firms will get rich. It only says that if firms are willing to play the game, then they can attain some much-valued properties of market stability and predictability. Just as the existence of stable CAM in neural networks must be supplemented by a mathematical classification theory which determines who, if anyone, will win a specially designed competition, the existence of a stable market must be supplemented by an analysis of how firms should choose their strategies to maximize their gains despite ignorance of their competitors' strategies.

Before turning to a discussion of some recent specialized architectures, I shall further discuss two issues that naturally arise from the preceding text:

1. Why bother studying shunting interactions? Why aren't the simpler additive interactions always sufficient?
2. Are symmetric coefficients necessary to achieve stable learning and memory storage? In Section 11, it was noted that the answer is "no" for CAM systems whose storage is in short term memory (STM). The answer is also well-known to be "no" for associative learning systems whose storage is in long term memory (LTM). This is true for networks designed to accomplish associative pattern learning as well as for networks designed for spatiotemporal pattern recognition and planned sensory-motor performance. Some asymmetric associative networks which arise in adaptive pattern recognition and adaptive sensory-motor control are discussed in Sections 16–20.

## 13. SENSITIVE VARIABLE-LOAD PARALLEL PROCESSING BY SHUNTING COOPERATIVE-COMPETITIVE NETWORKS: AUTOMATIC GAIN CONTROL AND TOTAL ACTIVITY NORMALIZATION

The value of shunting networks is clarified by their ability to help overcome one of the problems which has confronted recent investigators who have been using additive networks. Amit, Gutfreund, and Sompolinsky (1987, p. 2294) found spurious memory states in the additive model that they studied. Their analysis led them to conclude that "there must be some global control on the dynamics of the network, which prevents too high or too low activity." In other words, it is important to carefully regulate the network's total activation through time. The importance of this property, called *total activity normalization*, has been recognized in the neural network literature for the past two decades, and is one of the basic properties of shunting cooperative-competitive networks (Grossberg, 1970b, 1972a, 1973, 1982).

More generally, shunting networks provide a design for sensitive variable-load parallel processors. Suppose that the STM traces or activations $x_1, x_2, \ldots, x_n$ at a network level fluctuate within fixed finite limits at their respective network nodes. Setting a bounded operating range for each $x_i$ has the advantage that fixed decision criteria, such as output thresholds, can also be defined. On the other hand, if a large number of intermittent input sources converge on the nodes through time, then a serious design problem arises, due to the fact that the total input converging on each node can vary wildly through time. I have called this problem the *noise-saturation dilemma*: If the $x_i$ are sensitive to large inputs, then why do not small inputs get lost in internal system noise? If the $x_i$ are sensitive to small inputs, then why do they not all saturate at their maximum values in response to large inputs?

Shunting cooperative-competitive networks possess automatic gain control properties capable of generating an infinite dynamic range within which input patterns can be effectively processed, thereby solving the noise-saturation dilemma. Specialized shunting networks have been classified in terms of their specific pattern processing and memory storage properties, thereby providing a storehouse of networks which serves as a resource for solving particular computational problems. Since the design and properties both of feedforward and feedback shunting networks have been reviewed in a number of places (Grossberg, 1981, 1982, 1987d), the present summary considers briefly only the simplest feedforward and feedback networks to convey some of the main ideas. First the simplest feedforward network will be described to illustrate how it solves the sensitivity problem raised by the noise-saturation dilemma.

Let a spatial pattern $I_i = \theta_i I$ of inputs be processed by the cells $v_i$, $i = 1, 2, \ldots, n$. Each $\theta_i$ is the constant relative size, or reflectance, of its input $I_i$ and $I$ is the variable total input size. In other words, $I = \sum_{k=1}^{n} I_k$, so that $\sum_{k=1}^{n} \theta_k = 1$. How can each cell $v_i$ maintain its sensitivity to $\theta_i$ when $I$ is parametrically increased? How is saturation avoided?

To compute $\theta_i = I_i(\sum_{k=1}^{n} I_k)^{-1}$, each cell $v_i$ must have information about all the inputs $I_k$, $k = 1, 2, \ldots, n$. Moreover, since $\theta_i = I_i(I_i + \sum_{k \neq i} I_k)^{-1}$, increasing $I_i$ increases $\theta_i$ whereas increasing any $I_k$, $k \neq i$, decreases $\theta_i$. When this observation is translated into an anatomy for delivering feedforward inputs to the cells $v_i$, it suggests that $I_i$ excites $v_i$ and that all $I_k$, $k \neq i$, inhibit $v_i$. This rule represents the simplest feedforward on-center off-surround anatomy (Figure 4a).

How does the on-center off-surround anatomy activate and inhibit the cells $v_i$ via mass action? Let each $v_i$ possess $B$ excitable sites of which $x_i(t)$ are excited and $B - x_i(t)$ are unexcited at each time $t$. Then at $v_i$, $I_i$ excites $B - x_i$ unexcited sites by mass action, and the total inhibitory input $\sum_{k \neq i} I_k$ inhibits $x_i$ excited sites by mass action. Moreover, excitation $x_i$ can spontaneously decay at a fixed rate $A$, so that the cell can return to an equilibrium point (arbitrarily set equal to 0) after all inputs cease. These rules say that

$$\frac{d}{dt} x_i = -Ax_i + (B - x_i)I_i - x_i \sum_{k \neq i} I_k. \tag{92}$$

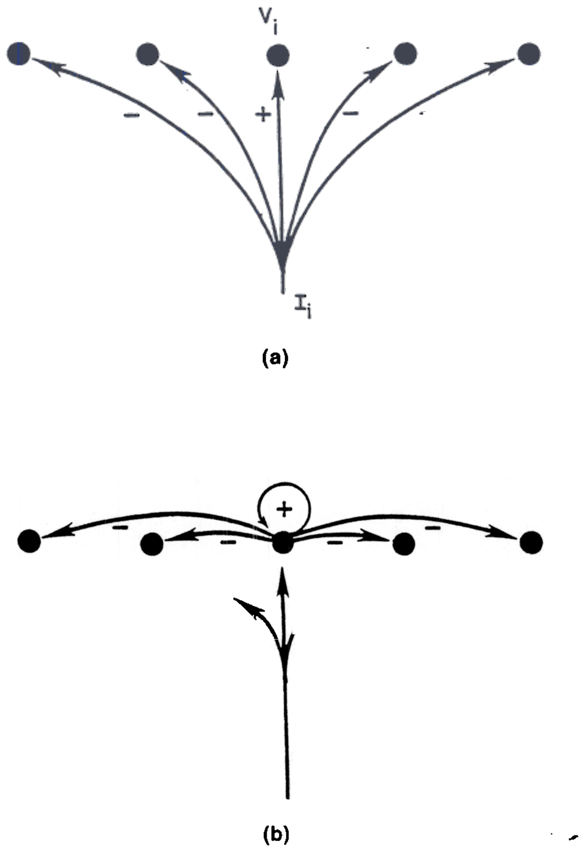


FIGURE 4. Two types of on-center off-surround networks: (a) A feedforward network in which the input pathways define the on-center off-surround interactions; (b) A feedback network in which interneurons define the on-center off-surround interactions.

If a fixed spatial pattern, $I_i = \theta_i I$ is presented and the background input $I$ is held constant for awhile, each $x_i$ approaches an equilibrium value. This value is easily found by setting $dx_i/dt = 0$ in (92). It is

$$x_i = \theta_i \frac{BI}{A + I} \tag{93}$$

Note that the relative activity $X_i = x_i(\sum_{k=1}^{n} x_k)^{-1}$ equals $\theta_i$ no matter how large $I$ is chosen; there is no saturation. This is due to automatic gain control by the inhibitory inputs. In other words, $\sum_{k \neq i} I_k$ multiplies $x_i$ in (92). The total gain in (92) is found by writing

$$\frac{d}{dt} x_i = -(A + I)x_i + BI_i. \tag{94}$$

The gain is the coefficient of $x_i$, namely $-(A + I)$, since if $x_i(0) = 0$,

$$x_i(t) = \theta_i \frac{BI}{A + I}(1 - e^{-(A+I)t}). \tag{95}$$

Both the steady state and the gain of $x_i$ depend on the input strengths. This is characteristic of mass action, or shunting networks but not of additive networks. Many alternative models cannot retune themselves in response to parametric shifts in background intensity.

The simple law (93) combines two types of information: information about pattern $\theta_i$, or "reflectances," and information about background activity, or "luminance." In visual psychophysics, the tendency towards reflectance processing helps to explain brightness constancy, and the rule $I(A + I)^{-1}$ helps to explain the Weber-Fechner law (Cornsweet, 1970).

Another property of (93) is that the total activity

$$x = \sum_{k=1}^{n} x_k = \frac{BI}{A + I} \tag{96}$$

is independent of the number of active cells. This *normalization* rule is a conservation law which says, for example, that a network that receives a fixed total luminance, making one part of the field brighter tends to make another part of the field darker. This property helps to explain brightness contrast (Cornsweet, 1970; Grossberg & Todorović, in press). Brightness constancy and contrast are two sides of a coin: on one side is Weber-law modulated reflectance processing, as in (93), and on the other side is a normalization rule, as in (96).

Equation (93) can be written in another form that expresses a different physical intuition. If we plot the intensity of an on-center input in logarithmic coordinates $K_i$, then $K_i = \ln(I_i)$ and $I_i = \exp(K_i)$. Also write the total off-surround input as $J_i = \sum_{k \neq i} I_k$. Then (93) can be written in logarithmic coordinates as

$$x_i(K_i, J_i) = \frac{Be^{K_i}}{A + e^{K_i} + J_i}. \quad (97)$$

How does the response $x_i$ at $v_i$ change if we parametrically change the off-surround input $J_i$? The answer is that $x_i$'s entire response curve to $K_i$ is shifted, and thus its dynamic range is not compressed. Such a shift occurs, for example, in bipolar cells of the *Necturus* retina (Werblin, 1971) and in a modified form in the psychoacoustic data of Iverson and Pavel (1981). The shift property says that

$$x_i(K_i + S, J_i^{(1)}) = x_i(K_i, J_i^{(2)}) \quad (98)$$

for all $K_i \geq 0$, where the amount of shift $S$ caused by changing the total off-surround input from $J_i^{(1)}$ to $J_i^{(2)}$ is predicted to be

$$S = \ln\left(\frac{A + J_i^{(1)}}{A + J_i^{(2)}}\right) \quad (99)$$

## 14. PHYSIOLOGICAL INTERPRETATION OF SHUNTING MECHANISMS AS A MEMBRANE EQUATION

Equation (92) is a special case of a law that occurs in vivo; namely, the membrane equation on which cellular neurophysiology is based. The membrane equation is the voltage equation that appears in the Hodgkin-Huxley equations mentioned in Section 6C. This equation embodies the classical electrical circuit interpretation (Hodgkin, 1964; Katz, 1966; Plonsey & Fleming, 1969) which is used to physically interpret the additive and shunting neural networks.

The membrane equation describes the voltage $V(t)$ of a cell by the law

$$C\frac{\partial V}{\partial t} = (V^+ - V)g^+ + (V^- - V)g^- + (V^p - V)g^p. \quad (100)$$

In (100), $C$ is a capacitance; $V^+$, $V^-$, and $V^p$ are constant excitatory, inhibitory, and passive saturation points, respectively; and $g^+$, $g^-$, and $g^p$ are excitatory, inhibitory, and passive conductances, respectively. We will scale $V^+$ and $V^-$ so that $V^+ > V^-$. Then in vivo $V^+ \geq V(t) \geq V^-$ and $V^+ > V^p \geq V^-$. Often $V^+$ represents the saturation point of a $\text{Na}^+$ channel and $V^-$ represents the saturation point of a $\text{K}^+$ channel. There is also symmetry-breaking in (100) because $V^+ - V^p$ is usually much larger than $V^p - V^-$. This symmetry-breaking operation, which is usually mentioned in the experimental literature without comment, achieves an important noise suppression property when it is coupled to an on-center off-surround anatomy.

To see why (92) is a special case of (100), suppose that (100) holds at each cell $v_i$. Then at $v_i$, $V = x_i$. Set $C = 1$ (rescale time), $V^+ = B$, $V^- = V^p = 0$, $g^+ = I_i$, $g^- = \sum_{k \neq i} I_k$, and $g^p = A$.

The reflectance processing and Weber law properties (93), the total activity normalization property (96), and the shift property (98) set the stage for the design and classification of more complex feedforward and feedback on-center off-surround shunting networks. Some results classifying feedforward on-center off-surround networks are reviewed in Grossberg (1981, 1987d).

## 15. SIGMOID FEEDBACK, CONTRAST ENHANCEMENT, AND SHORT TERM MEMORY STORAGE BY SHUNTING FEEDBACK NETWORKS

Feedback additive and shunting networks possess useful CAM properties that eventually led to the Cohen-Grossberg model reviewed in Section 9. During the last few years, many investigators have realized the importance of sigmoid feedback signals for generating effective pattern processing and CAM properties; (e.g., Ackley, Hinton, & Sejnowski, 1985; Hopfield, 1984; McClelland & Rumelhart, 1986). The first complete global analysis which rigorously demonstrated these properties was provided in Grossberg (1973). There the importance of sigmoid feedback was clarified by classifying the manner in which different types of feedback signal functions—linear, slower-than-linear, faster-than-linear, and sigmoid—transform input patterns and store the transformed patterns in STM. The simplest shunting on-center off-surround feedback network was chosen for this demonstration because it possessed the key properties of: (a) solving the noise-saturation dilemma by using the interaction between automatic gain control and on-center off-surround interactions, (b) normalizing or conserving its total activity, and (c) being capable of absolutely stable STM.

This simplest such network is defined by the equations

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)[I_i + f(x_i)] - x_i[J_i + \sum_{k \neq i} f(x_k)], \quad (101)$$

$i = 1, 2, \ldots, n$ (Figure 3b). Suppose that the inputs $I_i$ and $J_i$ acting before $t = 0$ establish an arbitrary initial activity pattern $(x_1(0), x_2(0), \ldots, x_n(0))$ before being shut off at $t = 0$. How does the choice of the feedback signal function $f(w)$ control the transformation of this pattern at $t \to \infty$? The answer is schematized in Table 3.

Table 3 displays choices of the feedback signal function $f(w)$ and the corresponding function $g(w) = w^{-1}f(w)$ which measures how much $f(w)$ deviates from linearity at prescribed activity levels $w$. The network's responses to these choices are summarized using the functions $X_i = x_i(\sum_{k=1}^{n} x_k)^{-1}$ and $x = \sum_{k=1}^{n} x_k$. The relative activity $X_i$ of the $i$th node computes how the network transforms the input pattern through time.

The functions $X_i$ play the role for feedback networks that the reflectances $\theta_i$ in (93) play for feedforward networks. The total activity $x$ measures how well the network normalizes the total network activity and whether the pattern is stored ($x(\infty) = \lim_{t \to \infty} x(t) > 0$) or not ($x(\infty) = 0$). Variable $x$ plays the role of the total input $I$ in (93).

Using these functions, (101) can be rewritten as the system

$$\frac{d}{dt} X_i = BX_i \sum_{k=1}^{n} X_k[g(X_i x) - g(X_k x)] \tag{102}$$

and

$$\frac{d}{dt} x = x[-A + (B - x) \sum_{k=1}^{n} X_k g(X_k x)]. \tag{103}$$

Using system (102)–(103), the following types of results were proved.

### A. Linear Signal: Perfect Pattern Storage and Noise Amplification

If $f(w)$ is chosen linear, as in $f(w) = Cw$, then $g(w) = C$ = constant. Hence by (102), all $d/dt\ X_i = 0$, so that $X_i(t)$ = constant in response to an arbitrary initial pattern $x_i(0)$. This system thus possesses a continuum of nondistorting CAM states. Why, therefore, is not a linear feedback signal a perfect choice for sensory pattern processing?

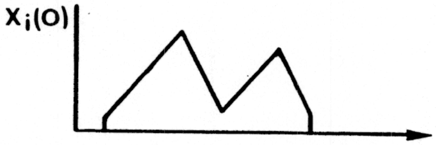
The answer becomes clear through consideration of the total activity variable $x(t)$. In the linear case, (103) reduces to

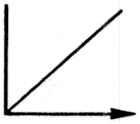
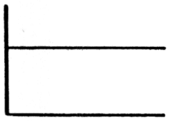
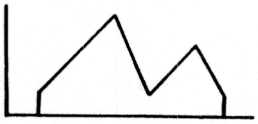
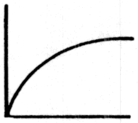
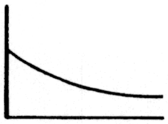
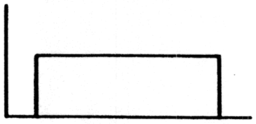
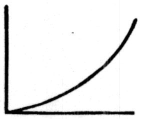
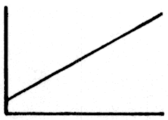
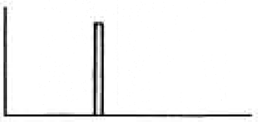
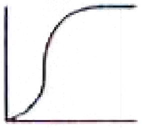
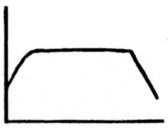
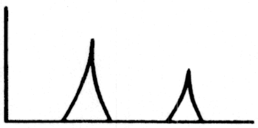
$$\frac{d}{dt} x = x[-A + (B - x)C]. \tag{104}$$

Hence such a system either cannot store any pattern ($x(\infty) = 0$ if $B - AC^{-1} < 0$), or it amplifies noise as vigorously as it amplifies signals if it is capable of CAM ($x(\infty) = B - AC^{-1}$ if $B - AC^{-1} > 0$, no matter how small $x(0) > 0$ is chosen). This amplification property generalizes to other models, and challenges those models, such as the BSB model (Section 9B), whose feedback signals are linear at small activity values.

Equation (102) suggests how to define slower-than-linear, faster-than-linear, and sigmoid feedback signals for purposes of pattern processing. Just as a linear $f(w) = Cw$ generates a constant $g(w) = C$, a slower-than-linear $f(w)$, such as $Cw(D + w)^{-1}$, generates a decreasing $g(w)$, such as $C(D + w)^{-1}$; a faster-than-linear $f(w)$, such as $Cw^n$ with $n > 1$, generates an increasing $g(w)$, such as $Cw^{n-1}$; and a sigmoid $f(w)$, such as $Cw^n(D^n + w^n)^{-1}$ with $n > 1$, generates a hill-shaped $g(w)$, such as $Cw^{n-1}(D^n + w^n)^{-1}$.

TABLE 3



| f(w) | g(w) | $X_i(\infty) = \frac{X_i(\infty)}{\sum_k X_k(\infty)}$ | $x(\infty) = \sum_k X_k(\infty)$ |
|---|---|---|---|
|  |  |  | AMPLIFIES NOISE |
|  |  |  | AMPLIFIES NOISE |
|  |  |  | QUENCHES NOISE |
|  |  |  | QUENCHES NOISE |

### B. Slower-than-Linear Signal: Pattern Compression and Noise Amplification

Slower-than-linear signal functions $f(w)$ also amplify noise. Analysis of (102) and (103) shows that a slower-than-linear feedback signal exerts a compressive effect on the reverberating activity pattern which obliterates all differences between initially active nodes ($X_i(\infty) = \frac{1}{n}$) if the network is capable of CAM ($x(\infty) > 0$).

### C. Faster-than-Linear Signal: Winner-Take-All, Noise Suppression, and Total Activity Quantization in an Emergent Finite State Machine

A faster-than-linear signal function can tell the difference between small and large initial values by amplifying and storing only sufficiently large activities. Analysis of (102) and (103) shows that a faster-than-linear signal function amplifies the largest initial activities so much more than smaller initial activities that it makes a *choice*: Only the node with the largest initial activity gets stored in STM. This is a remarkable property from several perspectives.

It shows how a very large network of nodes can quickly choose a winner in a single processing step without any search, simply by letting its nodes compete. Choice networks were originally designed for use in the many applications wherein the computational task is to choose a winner from noisy data. Feldman and Ballard (1982) have called this choice property winner-take-all. Platt and Hopfield (1986) have, for example, mentioned this property in their discussion of error-correcting codes.

A faster-than-linear signal function also generates remarkable normalization and quantization properties in the total activity domain. Combined with the winner-take-all property, the quantization property shows that a faster-than-linear signal function generates emergent properties of a finite-state machine, even though system (101) is defined by continuous laws. In particular, at large times, (103) is approximated by the equation

$$\frac{d}{dt}x \sim x[-A + (B - x)g(x)]. \tag{105}$$

Thus the stored total activity $x(\infty)$ is a root of the equation

$$g(w) = \frac{A}{B - w} \tag{106}$$

where both $g(w)$ and $A(B - w)^{-1}$ are increasing functions of $w$. The stored total activity is normalized because the roots of (106) are independent of the number $n$ of competing nodes.

Total activity quantization and noise suppression supplement the normalization property if the following hypotheses are satisfied. Suppose that $A > Bg(0)$ and that there are $m$ roots $E_1 < E_2 < \cdots < E_m < B$ of equation (106). Then the roots $E_1, E_3, \cdots$ are unstable equilibrium points of $x(t)$, whereas the roots $E_2 < E_4 < \cdots$ are stable equilibrium points of $x(t)$. Root $E_1$ defines the level below which $x(t)$ is treated as noise and suppressed. Roots $E_2, E_4, \cdots$ are stable, quantized, normalized limit values of $x(\infty)$. Function $g(w)$ can also be chosen to equal $A(B - w)$ along an interval of values, thereby leading to a continuum of stable equilibrium values. Thus, one can use system (101) with a faster-than-linear feedback signal function to design infinitely many finite-state machines or continuous energy spectrum machines capable of rapidly making choices in noise and possessing as many normalized asymptotic activity levels as one pleases.

### D. Sigmoid Signal: Tunable Filter, Quenching Threshold, Noise Suppression, and Normalization

Although a faster-than-linear signal function suppresses noise, it does so with such vigor that only the node with maximal initial activity survives in CAM. In many applications, one needs a spatially distributed CAM code, albeit one that contrast-enhances and thereby compresses an input pattern before the transformed pattern is stored in CAM. A sigmoid signal function generates these properties. Indeed, the classification of signal function properties in (A)–(C) shows that a signal function $f(w)$ which suppresses noise must be faster-than-linear at small activity values $w$. In addition, every physical signal function is bounded at large activity values $w$, thereby suggesting the use of a hybrid signal function which is faster-than-linear at small activities, slower-than-linear at large activities, and thus, by continuity, approximately linear in between; viz., a sigmoid signal function (Table 3).

Grossberg (1973) proved that a sigmoid signal function generates a *quenching threshold* (QT): Activities less than the QT are suppressed, whereas the pattern of activities that exceeds the QT is contrast-enhanced before being stored in STM. Speaking heuristically, the QT property is a consequence of pattern processing properties of faster-than-linear and linear signal functions combined with normalization properties in the total energy domain: The faster-than-linear property at small activity levels begins to contrast-enhance the input pattern as the total activity shifts due to normalization. As the partially contrast-enhanced activity pattern is normalized, it is influenced by the (approximately) linear range of the sigmoid signal function, which stores whatever pattern it detects (Table 3), including the partially contrast-enhanced pattern. Thus a sigmoid signal function can be used to design a noise-suppressing network with infinitely many stable equilibrium points, representing partially contrast-enhanced, or compressed, input patterns.

Any network that possesses a QT can be adaptively tuned. By increasing or decreasing the QT, the criterion of which activities represent functional signals—and hence should be processed and stored in STM—and of which activities represent functional noise—and hence should be suppressed—can be flexibly modified through time. An increase in the QT can cause all but the largest activities to be quenched. Then the network behaves like a choice machine. A sudden decrease in the QT can cause all recently presented input signals to be stored. If a novel or unexpected event suddenly decreases the QT, then all recently presented data can be stored in CAM until the cause of the unexpected event can be determined and learned. This property is important in actively regulating the focus of attention of a neural network sensory processor (Grossberg, 1982).

It cannot be overemphasized that the existence of the QT and its desirable tuning and CAM properties follow from the use of a nonlinear sigmoid signal function. When these properties were first proved in the early 1970s, the popularity of linear control models and of digital serial models in applications to intelligent systems prevented their acceptance, or often even their toleration. The recent popularity of connectionist models and of Liapunov methods have turned the obscure into the obvious, which is a sure sign of major progress.

The QT has been explicitly computed in a special case (Grossberg, 1973, pp. 355–359). In system (101) with $I_i = J_i = 0$, let

$$f(w) = Cwg(w) \tag{107}$$

where $C \geq 0$, $g(w)$ is increasing for $0 \leq w \leq x^{(1)}$, and $g(w) = 1$ for $x^{(1)} \leq w \leq B$. Then

$$QT = \frac{x^{(1)}}{B - AC^{-1}}. \tag{108}$$

Thus all the parameters of the network influence the QT. An important open problem is to compute the QT of more general cooperative-competitive networks that arise in computational applications.

In summary, several factors work together to generate desirable pattern transformation and STM-CAM properties: the dynamics of mass action, the geometry of competition, and the statistics of competitive feedback signals work together to define a unified network module whose several parts are designed in a coordinated fashion through development. How such a network module is self-organized in vivo is a profound open problem whose solution is worthy of a major scientific effort. A great deal more mathematical work will also be needed to fully understand even the properties of those shunting and additive networks which have already arisen in applications. For example, mixed cooperative-competitive nonlinear feedback networks have been designed to analyze and predict properties of emergent visual segmentation (Grossberg, 1987a, 1987b; Grossberg & Mingolla, 1985a, 1985b, 1987). Although the computer simulations of these networks were guided by previous theorems about competitive and cooperative nonlinear feedback networks, no global theorems have yet been proved about these mixed cooperative-competitive feedback networks. The additional insights that such theorems are bound to bring are much to be desired if only because of the great practical importance of emergent visual segmentation in a number of applications.

## 16. COMPETITIVE LEARNING MODELS

Another application of a choice network occurs in competitive learning models. In the simplest competitive learning model, normalized input patterns pass through an adaptive filter before the maximal filter output is chosen by a winner-take-all network. The winning population then triggers associative pattern learning within the vector of LTM traces which sent its inputs through the adaptive filter. Such a competitive learning model is a particular type of adaptive vector quantization scheme (Gray, 1984) which possesses Bayesian processing properties (Duda & Hart, 1973). In cognitive psychology, competitive learning properties are used to model categorical perception (Anderson, Silverstein, Ritz, & Jones, 1977; Elman, Diehl, & Buchwald, 1977; Hary & Massaro, 1982; Miller & Liberman, 1979; Pastore, 1981; Sawusch & Nusbaum, 1979; Sawusch, Nusbaum, & Schwab, 1980; Schwab, Sawusch, & Nusbaum, 1981; Studdert-Kennedy, 1980). During categorical perception, input patterns are classified into mutually exclusive recognition categories which are separated by sharp categorical boundaries. A sudden switch in pattern classification can occur if an input pattern is deformed so much that it crosses one of these boundaries.

The development of competitive learning models was achieved through an interaction between results of Grossberg (1970b, 1972b, 1973) and of Malsburg (1973), leading in Grossberg (1976a, 1976b) to the model in several forms which have subsequently been further analyzed and applied by a number of authors (e.g., Amari & Takeuchi, 1978; Bienenstock, Cooper, & Munro, 1982; Carpenter & Grossberg, 1985, 1987a; Grossberg, 1982; Grossberg & Kuperstein, 1986; Rumelhart & Zipser, 1985). Kohonen (1984) has made particularly strong use of competitive learning models in his work on self-organizing maps. His important theorem on the statistical distribution of such maps extends to the stochastic case the Grossberg (1976a) theorem on the stability of map learning in response to sparsely distributed input patterns. Kohonen's analysis of the distribution properties of a self-organizing map in the stochastic case reflects the property in the deterministic case that LTM vectors in the adaptive filter are a time average of their learned input vectors, and thus track the distribution of these input vectors within their convex hull under suitable conditions. Such deterministic analyses include stochastic analyses in the sense that they predict the time course of learning in response to arbitrary sequences of input patterns, including stochastically controlled input patterns. A historical discussion of the development of competitive learning models is given in Grossberg (1987e).

Although competitive learning models are useful in many situations, their learning becomes unstable in response to a variety of input environments, as was first shown in Grossberg (1976a). An effort to understand how to design an adaptive pattern recognition and map learning system that could self-stabilize its learning in response to arbitrary input environments led to the introduction of adaptive resonance theory in Grossberg (1976b). In this theory, competitive learning mechanisms are embedded in a larger network which includes learned top-down expectations and other modulatory mechanisms that were identified through an analysis of data concerning perception, cognition, conditioning, attention, event-related potentials, and neurophysiology.

In addition to suggesting mechanistic explanations of many interdisciplinary data from these subjects, the theory also made a number of predictions which have since been partially supported by experiments. For example, Grossberg (1976b) predicted that both norepinephrine (NE) mechanisms and attentional mechanisms modulate the adaptive development of thalamocortical visual feature detectors. Kasamatsu and Pettigrew (1976) and Pettigrew and Kasamatsu (1978) described NE modulation of feature detector development and Singer (1982) reported attentional modulation. Grossberg (1978a) predicted a word length effect in word recognition paradigms. Samuel, van Santen, and Johnston (1982, 1983) reported a word length effect in word superiority experiments. Grossberg (1978a, 1980b) predicted a hippocampal generator of the P300 event-related potential. Halgren et al. (1980) reported the existence of a hippocampal P300 generator in humans. The existence and correlations between other event-related potentials, such as processing negativity (PN), early positive wave (P120), and N200 were also predicted in these theoretical articles (see Banquet & Grossberg, in press, for further discussion).

The next section describes a number of the key properties which emerged from such data analyses and subsequent mathematical developments and computational analyses of Carpenter and Grossberg (1985, 1987a).

## 17. STABLE SELF-ORGANIZATION OF PATTERN RECOGNITION CODES

A number of basic computational distinctions can be used to differentiate neural network architectures and to clarify the applications for which they are best suited. Since a given computational property may be advantageous for one application and disadvantageous for a different application, such a classification is well worth keeping in mind.

In this section, I compare and contrast properties of the adaptive resonance theory (ART) for the stable self-organization of pattern recognition codes with properties of alternative models for the learning of pattern recognition codes. Since ART was introduced in Grossberg (1976b), it has undergone substantial development and analysis. A number of articles which contributed to the theory are brought together in several books (Grossberg, 1982, 1987c, 1987d, 1988). The discussion herein will be based upon properties of an ART architecture, called ART 1, whose key properties were developed, proved mathematically, and illustrated through extensive computer simulations in Carpenter and Grossberg (1987a). The architecture and processing cycle of ART 1 are summarized in Figures 5 and 6. The main computational distinctions to be discussed are outlined under separate headings:

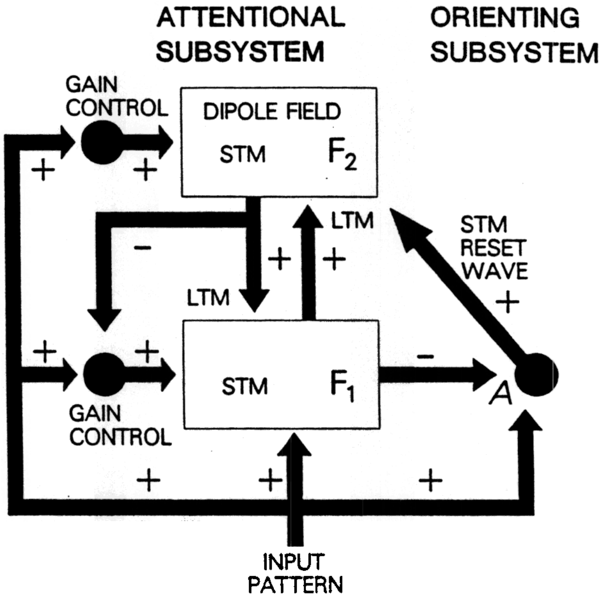


FIGURE 5. Anatomy of the ART 1 architecture: Two successive stages, $F_1$ and $F_2$, of the attentional subsystem encode patterns of activation in short term memory (STM). Bottom-up and top-down pathways between $F_1$ and $F_2$ contain adaptive long term memory (LTM) traces which multiply the signals in these pathways. The remainder of the circuit modulates these STM and LTM processes. Modulation by gain control enables $F_1$ to distinguish between bottom-up input patterns and top-down priming, or template, patterns, as well as to match these bottom-up and top-down patterns. Gain control signals also enable $F_2$ to react supraliminally to signals from $F_1$ while an input pattern is on. The orienting subsystem generates a reset wave to $F_2$ when mismatches between bottom-up and top-down patterns occur at $F_1$. This reset wave selectively and enduringly inhibits active $F_2$ cells until the input is shut off. (Reprinted with permission from Carpenter & Grossberg, 1987a, p. 56.)

### A. Real-Time (On-Line) Learning Versus Lab-Time (Off-Line) Learning

An ART architecture is designed to run in real-time, or on-line, when it is implemented in hardware. Various other architectures can only be run in lab-time, or off-line. This is perhaps the most important distinction that separates neural network architectures. Although some problems, such as the traveling salesman problem (Hopfield & Tank, 1985, 1986), can be run off-line, other problems, such as learning to recognize novel objects in a rapidly changing environment, must be solved on-line. Many of the computational properties which set apart ART architectures from other currently available learning algorithms are imposed to enable them to learn and recognize well in real-time.

### B. Nonstationary Unexpected World Versus Stationary Controlled World

Real-time environments are often nonstationary; their statistical properties can change unexpectedly through time. In addition, the world does not stop in a real-time environment. A ceaseless flow of input patterns of variable complexity can occur. ART architectures are designed to cope with nonstationary worlds of unlimited complexity. In contrast, many alternative learning and recognition schemes are off-line models that work well only in a stationary world whose inputs are carefully controlled both in number and statistical properties. The following discussion sharpens this basic difference.

### C. Self-Organization Versus Teacher as a Source of Expected Output

An ART architecture self-organizes its recognition code, without a teacher, through a direct interaction with its input environment. Self-organizing networks contrast sharply with learning networks which require an external teacher who presents an explicit correct answer, in the code of the network, for comparison with every output generated by the network, as in back propagation. Back propagation uses the Adeline learning rule of Widrow (1962). It is a steepest descent algorithm which was discovered by Werbos (1974), rediscovered and further developed by Parker (1982, 1985, 1986) under the name learning-logic, and popularized and applied to cognitive science applications by Rumelhart, Hinton, and Williams (1986) under the name back propagation.

The off-line nature of back-propagation, at least as it is used in many applications, is illustrated by the fact that its teaching signals often have no analog with analogous learning experiences in vivo. For example, the popular NETtalk back-propagation simulation of Sejnowski and Rosenberg (1986) uses a phoneme-by-phoneme matching scheme that has no analog during human real-time learning to read. The use of such a pre-coded teacher also has major implications for the structure of the learned code—notably the invariants of this code (see items F and G below)—and the way in which matching occurs within the network (see item H below).

### D. Self-Stabilization Versus Capacity Catastrophe

An ART architecture can self-stabilize its learning in response to arbitrarily many inputs. New inputs may either refine the criteria for accessing already established recognition codes, or may initiate the learning of a new recognition code, until the full capacity of the architecture is utilized. Input patterns which cannot refine prior knowledge, or which are first experienced after full capacity is utilized, are rejected by the architecture's self-stabilizing mechanisms.

In architectures which cannot self-stabilize their learning, later input patterns can wash away the learning of prior input patterns, leading to an unstable cycle of learning and forgetting. These architectures include essentially all the classical versions of autoassociators, competitive learning mechanisms, and steepest descent algorithms such as back propagation.

Effective use of a non-self-stabilizing architecture depends upon results which estimate the architecture's capacity. The capacity estimates the maximum number of input patterns which the architecture can learn, recognize, or remember. Typically, an autoassociator's capacity is $\sim .15n$, where the autoassociator's memory is defined by an $n \times n$ matrix (Anderson, 1983; Hopfield, 1984; Kohonen, 1984; McEliece, Posner, Rodemich, & Venkatesh, 1980; Psaltis & Park, 1986; Venkatesh, 1986). Thus an autoassociator cannot effectively use its full capacity. Moreover, if the number of input patterns exceeds this capacity, a capacity catastrophe occurs which renders the architecture's output unreliable. Such non-self-stabilizing architectures are thus inherently off-line machines whose lab-time world of inputs is under strict control. In an ART architecture, estimates of capacity play a different role than in an autoassociator or steepest-descent algorithm, since no catastrophe occurs when the input world contains more patterns than the architecture can encode.

### E. Maintain Plasticity in an Unexpected World Versus Externally Shut Off Plasticity

An ART architecture retains its plasticity, or ability to learn, for all time; that is, the parameters which enable its adaptive weights, or long-term memory (LTM) traces, to learn are not switched off as time goes on. The self-stabilization property is due to a dynamic buffering scheme which protects these LTM traces from changing except during appropriate circumstances.

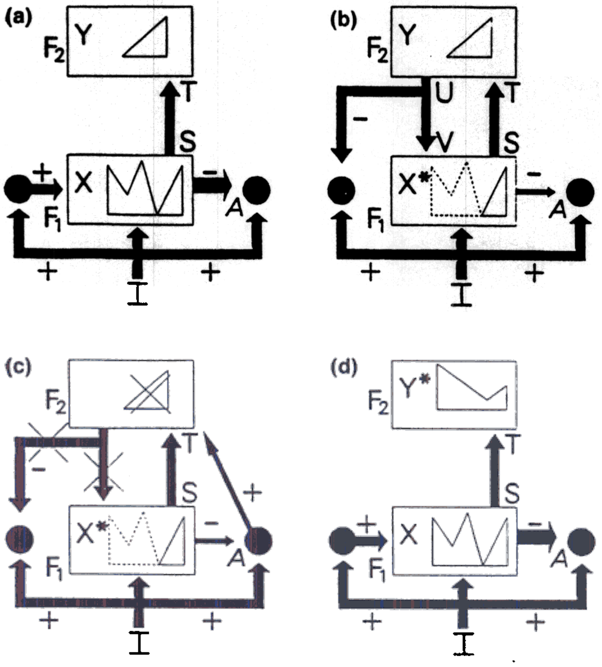


FIGURE 6. Search for a correct $F_2$ code: (a) The input pattern I generates the specific STM activity pattern X at $F_1$ as it nonspecifically activates $A$. Pattern X both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM pattern Y across $F_2$; (b) Pattern Y generates the top-down signal pattern U which is transformed into the template pattern V. If V mismatches I at $F_1$, then a new STM activity pattern X* is generated at $F_1$. The reduction in total STM activity which occurs when X is transformed into X* causes a decrease in the total inhibition from $F_1$ to $A$; (c) Then the input-driven activation of A can release a nonspecific arousal wave to $F_2$, which resets the STM pattern Y at $F_2$; (d) After Y is inhibited, its top-down template is eliminated, and X can be reinstated at $F_1$. Now X once again generates input pattern T to $F_2$, but since Y remains inhibited T can activate a different STM pattern Y* at $F_2$. If the top-down template due to Y* also mismatches I at $F_1$, then the rapid search for an appropriate $F_2$ code continues. (Reprinted with permission from Carpenter & Grossberg, 1987a, p. 61.)

Thus, if an unfamiliar event occurs at any time in the future, an ART architecture can learn about it, without destabilizing its prior knowledge, just so long as it has not already committed its full capacity to prior learning.

In contrast, if a non-self-stabilizing architecture is exposed to a never-ending time-series of input patterns, as in vivo, then it will experience a fatal capacity catastrophe unless its plasticity is shut off; that is, unless the parameters of its individual LTM traces are switched to a no-learning mode by an external or internalized teacher.

Such a switching-off of plasticity does not work well in nonstationary environments whose properties are not predictable in advance. If learning is switched off too soon, then important later events cannot be learned. If learning is switched off too late, then important earlier learning may be washed away due to a capacity catastrophe. An omniscient teacher is needed to switch off learning at just the right time in response to an arbitrary input environment. If an effective model of an omniscient teacher is available, however, then a potentially unstable learning device will not be needed.

### F. Self-Scaling Computational Units

An ART architecture can learn to distinguish arbitrary pairs of input patterns. In contrast, a number of alternative recognition architectures depend upon orthogonality, linear predictability, or other statistical constraints on input patterns in order to function well. In order to achieve this property, an ART architecture self-scales the processing of its input patterns. Individual input features are automatically given less weight when they occur within more complex input patterns. A side benefit of this type of feature normalization is that sufficiently small noisy changes in a complex input pattern may not force recoding of the pattern.

As a result of the self-scaling property, an input feature may be encoded in LTM when it occurs within a simple input pattern, yet be rejected from LTM when it occurs within a complex input pattern. In the former case, the input feature becomes a critical feature by being learned by the LTM code that helps to recognize the pattern. In the latter case, the input feature becomes a noise element by being deleted from the LTM code that helps to recognize the pattern. This decision depends upon the entire history of learning that has preceded the presentation of an input pattern which contains the feature. Thus, the concept of critical feature is an emergent property of the network rather than a property which can be defined solely by choice of an input filter.

### G. Learn Internal Expectations Versus Impose External Costs

An ART architecture learns its own top-down expectations as a function of the unique input environment to which it is exposed. These expectations are emergent internal representations which capture invariant statistical properties of the entire input environment. Such learned top-down expectations are a key ingredient of ART architectures. They function as prototypes for an entire recognition category, and abstractly encode the similarity properties which are shared by all exemplars of the category. On the other hand, the manner in which these expectations are learned and manipulated sets them apart from classical prototype ideas. They may also be interpreted as "costs" which the architecture learns for itself, such that different costs are learned in response to different input environments (Figure 7). In order to avoid misinterpretations of these expectations due to such analogies with previous ideas, Carpenter and Grossberg (1987a) have called them *critical feature patterns*.

In contrast, architectures which use an external teacher often represent this teacher as a set of target patterns, or externally imposed costs. In NETtalk, for example, the mismatch between a target phoneme output code and the actual phoneme output code is used to drive learning at all the model's stages, from the visual representation of input letters to the auditory representation of output phonemes. Thus, an auditory mismatch is used to determine the learned properties of a visual code. In vivo, by contrast, many objects can be visually recognized based upon visual invariants, even if the objects have no names; for example, the familiar face of a check-out person at the supermarket. The verbal phrase "check-out person" does not determine how we visually recognize such a person's face, any more than we would visually recognize the same face differently if the same job were renamed "cashier attendant."

In ART, self-organization of invariant critical feature patterns can occur within each modality before intermodality transformations between these invariants, such as between a visual representation of a face and an auditory representation of a name, are learned via associative mechanisms (Grossberg, 1978a, 1982).

### H. Active Attentional Focusing and Priming Versus Passive Weight Change

A top-down learned expectation in ART is actively matched against bottom-up information (Figure 6).

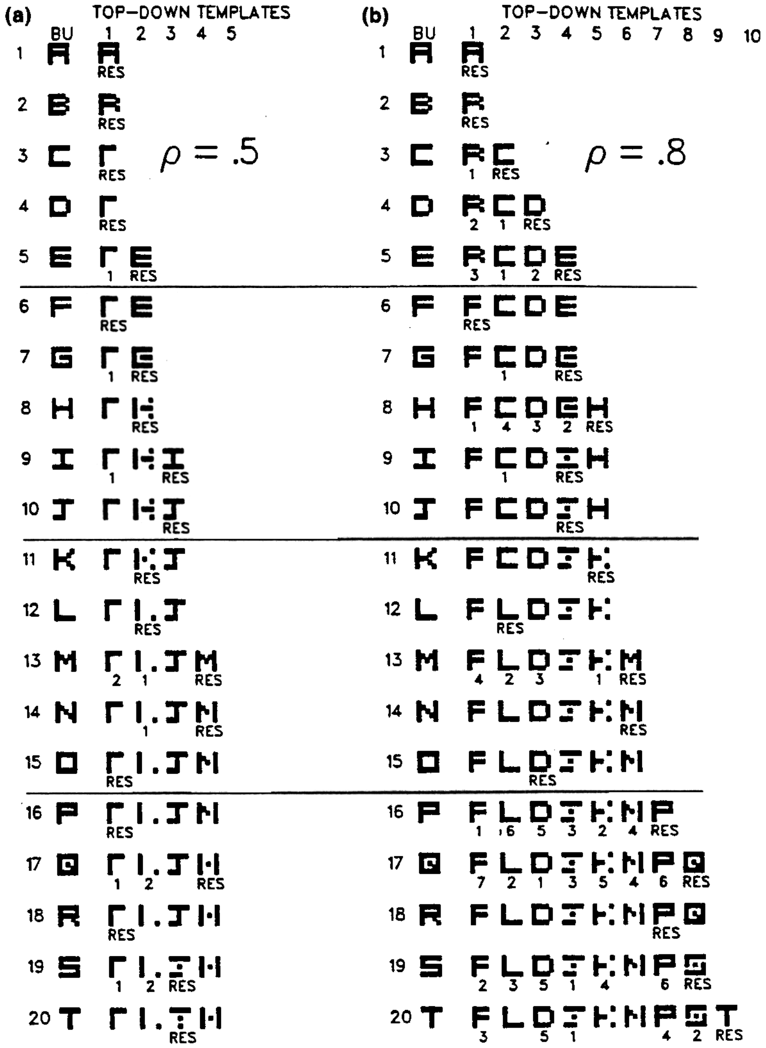


FIGURE 7. Alphabet learning: Code learning in response to the first presentation of the first 20 letters of the alphabet is shown. Two different vigilance levels were used, (a) $\rho = .5$ and (b) $\rho = .8$. Each row represents the total code that is learned after the letter at the left-hand column of the row is presented at $F_1$. Each column represents the critical feature pattern that is learned through time by the $F_2$ node listed at the top of the column. The critical feature patterns do not, in general, equal the pattern exemplars which change them through learning. Instead, each critical feature pattern acts like a prototype for the entire set of these exemplars, as well as for unfamiliar exemplars which share invariant properties with familiar exemplars. The simulation illustrates the "fast learning" case, in which the altered LTM traces reach a new equilibrium in response to each new stimulus. Slow learning is more gradual than this. (Reprinted with permission from Carpenter & Grossberg, 1987a, p. 72.)

This matching process takes place within a processing stage, or level, and can rapidly reorganize the activations, or short-term memory (STM) traces, that are computed at the nodes of this stage. In particular, a top-down critical feature pattern can prime a lower level to get ready for an exemplar from an expected class of input patterns. It can amplify and thereby speed up processing of an input pattern that belongs to an expected class, while it actively reorganizes the information processing of this input. A top-down expectation can also attenuate processing of unexpected inputs through a mismatch between the bottom-up input and the top-down expectation.

In this way, a top-down expectation enables an ART architecture to function like an intentional machine that generates an active focus of attention. In contrast, the top-down mechanisms of a teacher-driven architecture such as back propagation does not directly prime the system, or focus its attention, or reorganize its fast information processing. Instead back propagation merely causes slow changes in the bottom-up adaptive weights of the network (Figure 8).

### I. Closing Versus Opening the Fast-Slow Feedback Loop

The bottom-up and top-down interactions in an ART architecture close the feedback loop between fast activations at network levels and slower learned changes in the pathways between levels (Figure 6). This fundamental property is what enables the ART architecture to learn stably in real-time. In contrast, back propagation opens this feedback loop (Figure 8), which makes this network easier to understand but computationally less robust.

### J. Expectant Priming Versus Grinding All Memory Cycles

Due to the closed feedback cycle in ART, a top-down prime that is locked into STM by a large gain control signal can prevent any input that is not in the expected recognition category from activating higher levels of the architecture, much as the verbal command to look for oranges enables one to avoid being too distracted by other objects.

In contrast, an architecture without a top-down expectation capable of actively suppressing mismatched exemplars will become fully engaged by any input patterns that may happen to occur. Its memory cycles may thus be so fully engaged by irrelevant inputs when a crucial event occurs that it cannot respond in time.

### K. Learning in the Approximate Match Phase Versus in the Mismatch Phase: Hypothesis Testing Avoids the Noise Catastrophe

Due to its possession of learned top-down expectations which actively suppress mismatched input patterns while prototypically deforming and amplifying matched input patterns, an ART architecture learns in the *approximate match* phase. In other words, if an input pattern causes the read-out of a learned top-down expectation which matches it well enough to prevent a reset event, then the matching process selects the information that is consistent between the input and the expectation. If these matched data include novel elements, they are used to refine the learned code of that recognition category. In other words, if the compressed recognition code, or hypothesis, generated by the input pattern does not cause reset due to read-out of its top-down expectation, then the architecture deems this input pattern to be consistent enough with that hypothesis to refine or update the hypothesis based upon any additional information that the pattern may contain. If reset does occur and leads to selection of an uncommitted code, then this code learns a complete representation of the input pattern, which is intuitively reasonable because the pattern could not be assimilated into any previously learned code.

Because an ART architecture learns in the approximate match stage, it is not degraded by noise fluctuating at its input level. If a noise pattern is very different from its learned codes, the architecture dynam-

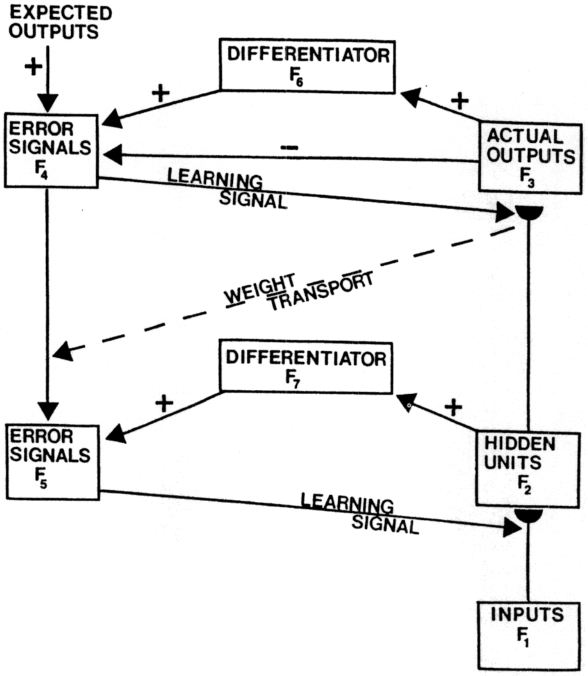


FIGURE 8. Circuit diagram of the back propagation model: In addition to the processing levels $F_1$, $F_2$, $F_3$, there are also levels $F_4$, $F_5$, $F_6$, and $F_7$ to carry out the computations which control the learning process. The transport of learned weights from the $F_2 \rightarrow F_3$ pathways to the $F_4 \rightarrow F_5$ pathways show that this algorithm cannot represent a learning process in the brain. All feedback within the levels $F_1 \rightarrow F_2 \rightarrow F_3$ is expressed through learning signals which slowly change bottom-up adaptive weights but do not quickly reorganize the network's fast information processing.

ically buffers these codes against relearning and selects an uncommitted code to learn the noise. If the noise does not repeat itself often enough and learning proceeds slowly, the noise will not cause significant learning within the uncommitted code. It will not, in any case, interfere with the codes of sufficiently different patterns even if learning proceeds quickly.

Architectures without a fast-slow feedback loop, such as back propagation, are designed so that learning occurs in the mismatch mode. Thus if noise can activate the teacher, or expected output stage, then all the LTM traces of the architecture can eventually be recoded. This noise catastrophe can be prevented only if the noise is infrequent and learning is slow. In ART, even frequent noise is simply treated as a new source of input that does not endanger the corpus of previously learned information.

Although learning in the mismatch phase can cause a noise catastrophe in an architecture which is designed to learn a recognition code in real-time, it is a useful type of learning in a number of alternative situations, notably during adaptive sensory-motor control (see Sections 18–21).

### L. Fast or Slow Learning: The Oscillation Catastrophe

Another consequence of learning in the approximate match phase is that ART can learn at either a fast or slow rate. Fast learning occurs when each LTM trace can reach a new equilibrium value on a single learning trial. Successive learning trials may or may not cause the LTM trace to assume different equilibrium values. Slow learning may require many trials before the LTM traces reach any equilibrium point of the system.

When learning occurs in the mismatch phase, as in back propagation, fast learning can cause wild oscillations to occur in the network's LTM traces, since each mismatch can drag the LTM traces to a totally different region of phase space. Thus such architectures must learn slowly, which emphasizes their off-line character. Contrast the vivid recollection of an exciting movie after one-trial on-line learning in vivo.

### M. Self-Adjusting Parallel Memory Search and Global Energy Landscape Upheaval Versus Search Trees and Local Minima

Learning in the approximate match phase could be disastrous were it not for the existence in ART of a self-adjusting parallel memory search that maintains its efficiency as the learned code becomes more complex. When an input pattern causes read-out of a top-down expectation with which it cannot form an approximate match, the attenuation of activity due to the mismatch event resets the compressed code, or hypothesis, that caused the mismatch (Figure 6c). A parallel memory search is hereby triggered that rapidly tests a series of such hypotheses. This hypothesis testing scheme actively and globally reorganizes the energy landscape of the architecture. In this way, the architecture circumvents the problem of local minima that has plagued alternative architectures such as autoassociators, simulated annealing, and the Boltzmann machine.

The need to escape local minima has influenced the design of all content-addressable memory and learning architectures. Architectures which do not possess self-adjusting hypothesis testing schemes typically use two alternative approaches. In autoassociators, simulated annealing, and the Boltzmann machine, noise is used to help the classification to jump out of a local minimum. Such a scheme does not globally reorganize the energy landscape based upon the outcome of active information processing. Instead the nonspecific action of noise tries to exploit the fact that the local minima of a fixed energy landscape may be easier to escape than a global minimum, and that a nonspecific external temperature parameter may be used to control the rate of approach to equilibrium.

Steepest-descent methods, such as back propagation, use a mismatch to drag the system's LTM traces out of local minima, and do so slowly enough to reduce the potentially destabilizing effects of the oscillation catastrophe.

Traditional AI architectures often include search trees to escape erroneous classifications. Although a search tree may be efficient at one stage of learning, it cannot remain efficient at an arbitrary stage of learning unless it has a self-adjusting capability, as in ART.

### N. Rapid Direct Access Versus Increase of Recognition Time with Code Complexity

In ART, as learning self-stabilizes in response to a set of input patterns, the search mechanism is automatically disengaged. Thereafter rapid direct access to the recognition code occurs in response to familiar input exemplars as well as to novel exemplars that share invariant properties with familiar input exemplars. Direct access occurs because read-out of the top-down expectation of a familiar input pattern always leads to an approximate match (Carpenter & Grossberg, 1987a). Then the initial bottom-up event that led to top-down readout (Figure 6) can generate a resonant recognition event without causing reset. Thus an ART architecture reconciles the ostensibly conflicting demands of direct access and search. It uses reset and self-regulating search to build globally self-consistent recognition codes which avoid local minima in response to arbitrarily complex time-series of input patterns. It uses direct access to recognize familiar events with a speed as fast as one's hardware can run.

In contrast, an AI architecture with a search tree takes longer and longer to recognize an event as it needs to search more and more codes. If this were the case in vivo, it could take orders of magnitude more time to recognize our parents when we were 30 years old than it did when we were 5 years old. Fortunately, this is false.

### O. Asynchronous Versus Synchronous Learning

In ART 1, each LTM trace oscillates at most once through time in response to an arbitrary time-series of binary input patterns (Carpenter & Grossberg, 1987a). Were it not for the architecture's active reorganization of the energy landscape through hypothesis testing, this remarkable monotonicity property could easily cause the architecture to learn a local minimum. As it is, learning is allowed to occur only when the energy landscape reaches a configuration that leads to an approximate match. Due to these properties, an ART 1 architecture can learn asynchronously or synchronously: If input patterns come in too quickly to generate an approximate match or to terminate a search, then no learning occurs. If the input pattern stays on long enough for some learning to occur, then it does not matter too much how long it stays on because, due to the monotonicity property, the LTM traces will tend to change in the same direction no matter how long it stays on.

In contrast, if mismatch drives learning in such a way that each LTM trace can oscillate persistently as it approaches equilibrium, then variable durations of the input patterns may destabilize the learning process. The limiting case of this property is the oscillation catastrophe, which occurs if sufficiently many input patterns stay on long enough for LTM traces to reach equilibrium on each such learning trial.

### P. Discriminative Tuning via Attentional Vigilance

Although an ART architecture self-organizes its learning, it can be tuned by environmental feedback to learn coarser or finer discriminations; that is, it can learn to categorize the same set of input patterns into larger or smaller groupings depending upon how strictly the performance demands of the external environment are imposed.

Such environmental feedback acts to change a single parameter of the network that is called the vigilance parameter (Carpenter & Grossberg, 1987a). This parameter determines how fine the mismatch between bottom-up input and top-down expectation must be in order to reset the code which reads-out the expectation. A large vigilance parameter demands a high degree of match to prevent reset, hence finely categorizes the input patterns. A small vigilance parameter tolerates larger mismatches before forcing reset, hence more coarsely categorizes the input patterns.

This tuning parameter depends for its existence upon the fact that bottom-up and top-down matches occur as part of the real-time feedback cycle that determines the classification of input patterns. In particular, it can change dramatically the global reorganization of the energy landscape that regulates self-regulating search and learning by modifying the overall attentiveness, or sensitivity, of the circuit.

### Q. Towards a General-Purpose Machine for Cognitive Hypothesis Testing, Data Search, and Classification

In ART, adaptive pattern recognition is a form of hypothesis discovery, testing, learning, and recognition in response to a nonstationary input environment. This property would seem to be essential for any cognitive theory which hopes to understand how ever-more-complex knowledge invariants are discovered, tested, and recognized, at any level of abstraction.

In particular, future ART machines may be designed to learn, search, and classify complex hierarchically organized data files. The properties of automatic self-stabilization and self-scaling will be essential for such an architecture to work reliably and quickly on enormous data bases. The advantages of such a self-organizing system become greater as the data sets are chosen larger, because the decision task for hand-sorting so much data or the discovery of rules for automatically sorting all the data, especially when new data are added later on, rapidly become unmanageably difficult as a function of scale.

The postulates of ART thus define a class of models for a broad range of cognitive science applications, which include examples that may or may not even possess an orienting subsystem and search mechanism (Figure 5), as discussed by Carpenter and Grossberg (in press). McClelland (1987) has proposed that models of this class be renamed the "interactive activation framework" after the interactive activation model that was introduced by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982). The postulates of the interactive activation model are, however, inconsistent with those of an ART model. The postulates of the interactive activation model have been abandoned in favor of ART postulates because they are inconsistent with key cognitive data and possess some undesirable computational properties, as noted in Grossberg (1984, 1986, 1987e). Since ART was an established cognitive theory (Grossberg, 1978a, 1980b) before the interactive activation model was published, it seems historically and scientifically justified to retain the name Adaptive Resonance Theory for such models.

## 18. INTERNALLY REGULATED LEARNING AND PERFORMANCE IN NEURAL MODELS OF SENSORY-MOTOR CONTROL: ADAPTIVE VECTOR ENCODERS AND COORDINATE TRANSFORMATIONS

In many real-time nonlinear neural network architectures other than ART architectures, learning is reg-

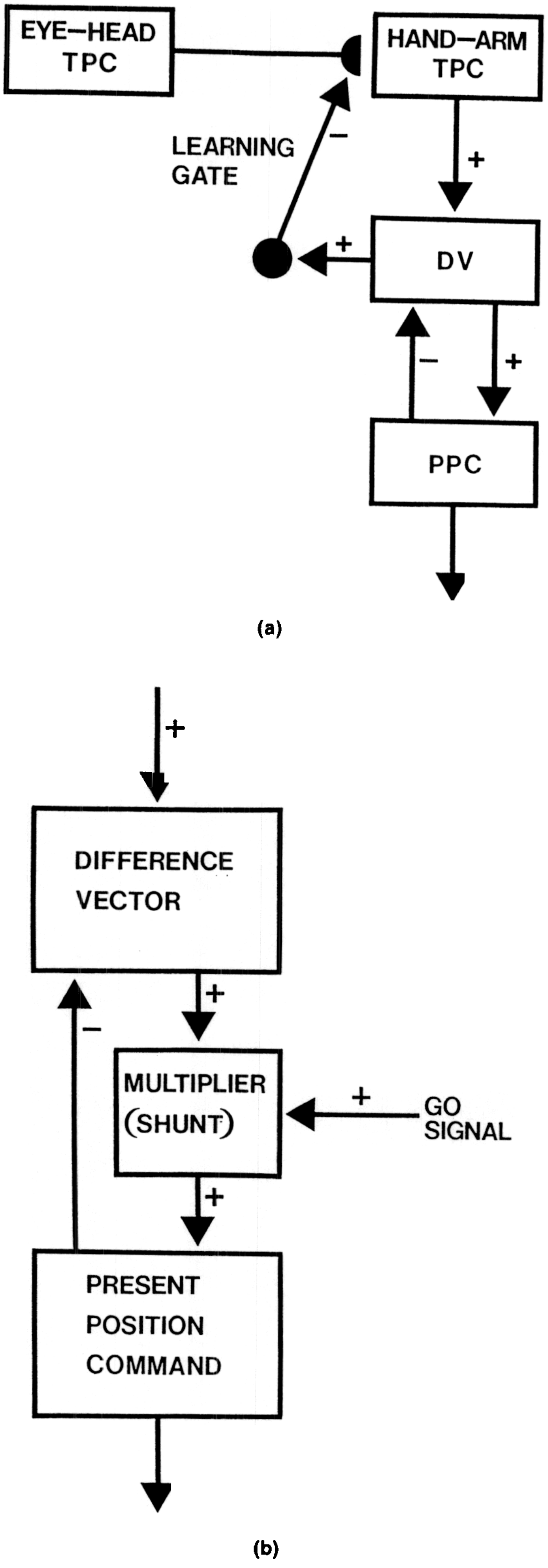


FIGURE 9. (a) Learning in sensory-motor pathways is gated by a difference vector (DV) process which matches target position command (TPC) with present position command (PPC) to prevent incorrect associations from forming between eye-head TPCs and hand-arm TPCs; (b) A GO signal gates execution of a primed movement vector and regulates the rate at which the movement vector updates the present position command. (Reprinted with permission from Bullock & Grossberg, 1987a).

ulated by pattern matches and mismatches. For example, learning of intermodality associative maps between the target position commands of different sensory-motor systems (Figure 9a) is gated on and off by intramodality pattern matches and mismatches, respectively. These same matching processes transform automatically such a target position command into a synchronous multijoint trajectory which automatically compensates for variable initial positions in a manner that quantitatively explains a large body of data about human and monkey arm movements (Bullock & Grossberg, in press-a). Thus just as in an ART model, model neural architectures which have been identified for the learning and performance of arm movements use internal matching processes to regulate both the fast information processing and the slower learning of the system.

The circuit depicted in Figure 9 schematizes key processing stages in a Vector Integration to Endpoint, or VITE, circuit. In its simplest form, the VITE circuit obeys the equations:

*Difference Vector*

$$\frac{d}{dt}V_i = \alpha(-V_i + T_i - P_i) \tag{109}$$

and

*Present Position Command*

$$\frac{d}{dt}P_i = G[V_i]^+, \tag{110}$$

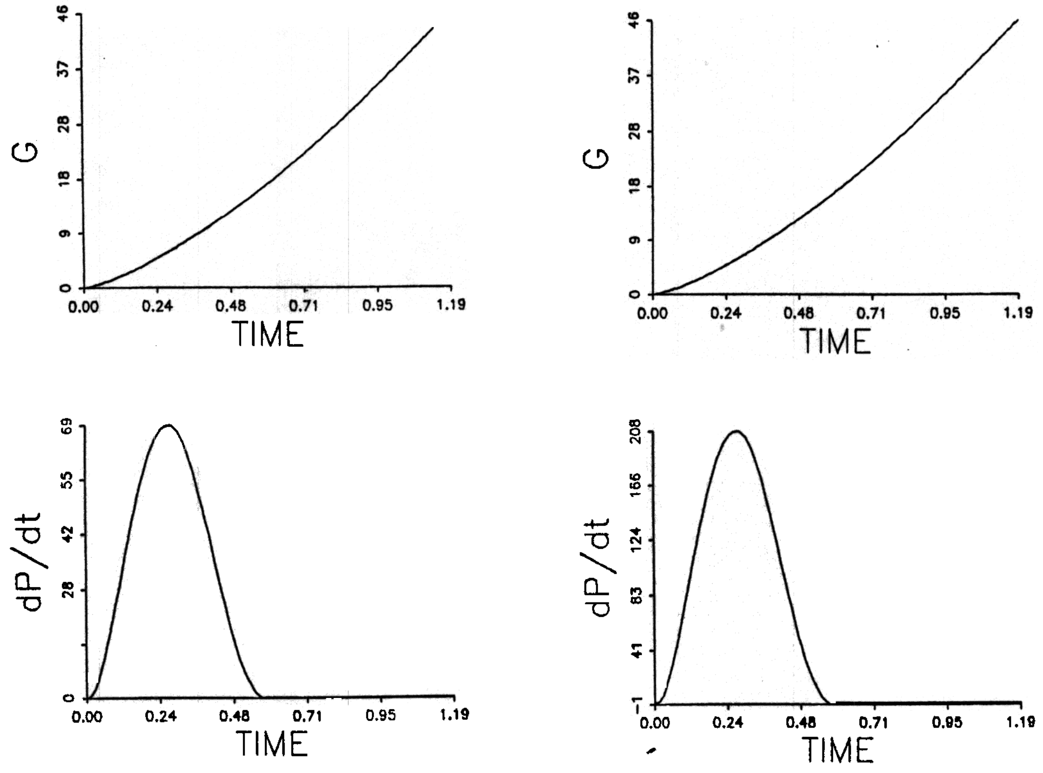
where $[V_i]^+ = \max(V_i, 0)$. Equations (109) and (110) describe a generic component of a target position command $(T_1, T_2, \ldots, T_n)$, a difference vector $(V_1, V_2, \ldots, V_n)$, and a present position command $(P_1, P_2, \ldots, P_n)$ in response to a time-varying velocity command, or GO signal $G(t)$; see Figure 9b. The difference vector computes a mismatch between target position and present position, and is used to update present position at a variable rate determined by $G(t)$ until the present position matches the target position.

Such a scheme permits multiple muscles, or other motor effectors, to contract synchronously even though the total amount of contraction, scaled by $T_i(0) - P_i(0)$, may be different for each effector (Figure 10). Unlike many alternative schemes for motor control, present position in (110) is not computed using inflow signals from the muscles. Rather, it is determined by nonlinear integration of vectors computed by matching an outflowing target position command with feedback from outflowing present position signals.

The VITE circuit is not sufficient in itself to accomplish all the tasks required of a variable-speed variable-load arm movement system. In concert with several parallel circuits, however, it can generate flexible and adaptive trajectories without suffering from the combinatorial explosions and rigid performance of control systems which preplan an entire trajectory. Herein I outline some of the adaptive control issues which arise in the design of such neural architectures for movement control and mention where internal matching processes regulate their processes of learning and performance. Quantitative neural network solutions to such problems are suggested by Bullock and Grossberg (in press-a, in press-b) and Grossberg and Kuperstein (1986).

The computation of a hand or arm's present position illustrates the complexity of the problem. As mentioned above, two general types of present position signals have been identified in discussions of motor control: *outflow* signals and *inflow* signals. Figure 11 schematizes the difference between these signal sources. An outflow signal carries a movement command from the brain to a muscle (Figure 11a). Signals that branch off from the efferent brain-to-muscle pathway in order to register present position signals are called *corollary discharges* (Helmholtz, 1962; von Holst & Mittelstaedt, 1950). An *inflow* signal carries present position information from a muscle to the brain (Figure 11b). A primary difference between outflow and inflow is that a change in outflow signals is triggered only when an observer's brain generates a new movement command. A new inflow signal can, in contrast, be generated by passive movements of the limb. Both outflow and inflow signals are used in multiple ways to provide different types of information about present position. The following summary itemizes some of the ways in which these signals are used in our theory.

Although one role of an outflow signal is to move a limb by contracting its target muscles, or motor plant, the operating characteristics of the motor plant are not known a priori to the outflow source. It is not known a priori either how much the muscle will actually contract in response to an outflow signal of prescribed size, or how much the limb will move in response to a prescribed muscle contraction. In addition, even if the outflow system somehow possessed this information at one time, it might turn out to be the wrong information at a later time, because muscle plant characteristics can change through time due to development, aging, exercise, changes in blood supply, or minor tears. (State-dependent and history-dependent plant changes may occur on the factory assembly line or in a freely-moving robot, no less than in a living muscle.) Thus the relationship between the size of an outflow movement



(a) (b)

FIGURE 10. With equal GO signals, movements of different size have equal durations and perfectly superimposable velocity profiles after velocity axis rescaling. (a, b): GO signals and velocity profiles for 20 and 60 unit movements lasting 560 ms. (Reprinted with permission from Bullock & Grossberg, 1987a).

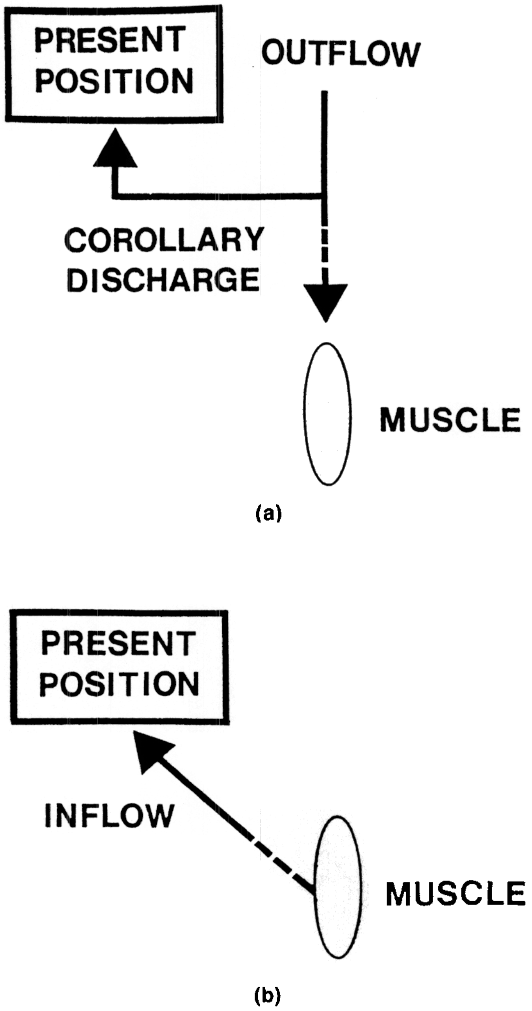


FIGURE 11. Both outflow (a) and inflow (b) signals contribute to the brain's estimate of the limb's present position, but in different ways.

command and the amount of muscle contraction is, in principle, undeterminable without additional information which characterizes the muscle plant's actual response to outflow signals.

To establish a satisfactory correspondence between outflow movement signals and actual muscle contractions, the motor system needs to compute reliable present position signals which represent where the outflow command tells the muscle to move, as well as reliable present position signals which represent the state of contraction of the muscle. Corollary discharges and inflow signals can provide these different types of information. Grossberg and Kuperstein (1986) have shown how a match between corollary discharges and inflow signals can be used to modify, through an automatic learning process, the total outflow movement signal to the muscle in a way that effectively compensates for changes in the muscle plant (Figure 12). Mismatches act as error signals which change the gain of the total outflow movement signal. This automatic gain control process generates adaptively a linear correspondence between an outflow movement command and the amount of muscle contraction even if the muscle plant is nonlinear. The process which matches outflow and inflow signals to linearize the muscle plant response through learning is called *adaptive linearization* of the muscle plant.

The cerebellum is implicated by both the theoretically derived circuit and experimental evidence as the site of learning. Early cerebellar learning models were proposed by Albus (1971), Brindley (1964), Grossberg (1964, 1969d, 1972b), and Marr (1969). Later models and experimental support were provided by Fujita (1982a, 1982b), Ito (1974, 1982, 1984), McCormick and Thompson (1984), Optican and Robinson (1980), Ron and Robinson (1973), Vilis and Hore (1986), and Vilis, Snow, and Hore (1983). The present model introduces new features which are critical to its success in correcting behaviorally well-characterized types of movements errors.

For example, an adaptive gain (AG) stage in our theory—which is interpreted as a model cerebellum—is used by multiple circuits that contribute to both eye movement and arm movement accuracy and postural stability. Each of these circuits involves different—and specific—input, output, and error signal pathways, but

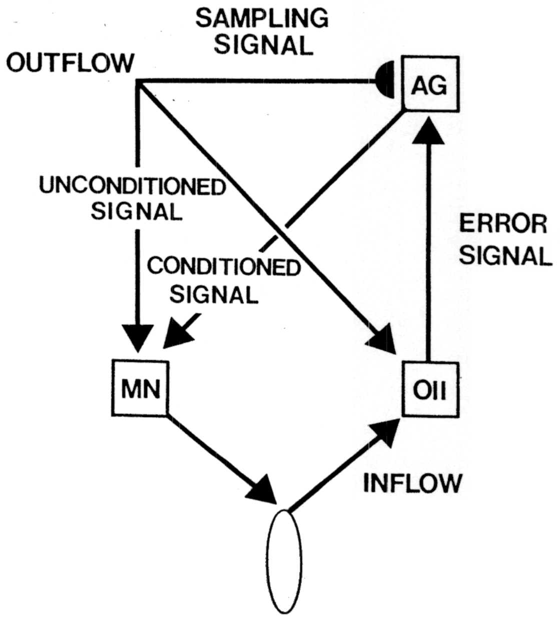


FIGURE 12. Some main features of the muscle linearization network, or MLN: The outflow-inflow interface (OII) registers matches and mismatches between outflow signals and inflow signals. Mismatches generate error signals to the adaptive gain (AG) stage. These error signals change the gain of the conditioned movement signal to the motoneurons (MN). Such an MLN adaptively linearizes the responses of a nonlinear muscle plant to outflow signals. The outflow signals can therefore also be used as a source of accurate corollary discharges of present eye position. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 136.)

all of these pathways are mediated by the same internal AG stage architecture. These AG stage results significantly extend recent data and models concerning the cerebellum's role in the conditioning of movements (Fujita, 1982a, 1982b; Ito, 1984; McCormick & Thompson, 1984; Optican & Robinson, 1980). For example, push-pull opponent processing of the error signals which govern adaptive gain changes at the AG stage is a novel, and computationally critical, part of this model. Such an arrangement enables the AG stage to correct undershoot, overshoot, or directionally skewed eye movement errors, as well as errors due to sustained wearing of curvature-distorting contact lens.

Although adaptive linearization is, like back-propagation, a form of error-driven learning in the mismatch mode, it is not susceptible to the instabilities of back-propagation because it merely changes the gains of internal command pathways via internal mismatches. It does not put at risk the spatial encoding of the commands that activate the pathways and is not subjected to noise fluctuations from the external environment.

Given that corollary discharges are matched with inflow signals to linearize the relationship between muscle plant contraction and outflow signal size, outflow signals can also be used in other ways to provide important information about present position. As Figure 9 illustrates, outflow present position signals can then be matched with target position commands to generate a trajectory with synchronous properties. Thus outflow signals are used in at least three ways, and all of these ways are automatically registered: They send movement signals to target muscles; they generate corollary discharges which are matched with inflow signals to guarantee linear muscle contractions even if the muscle plant is nonlinear; and they generate corollary discharges which are matched with target position signals to generate synchronous trajectories.

Inflow signals are also used in several ways. One way has already been itemized. A second use of inflow signals is suggested by the following gedanken example. When you are sitting in an armchair, let your hands drop passively towards your sides. Depending upon a multitude of accidental factors, your hands and arms can end up in any of infinitely many final positions. If you are then called upon to make a precise movement with your arm-hand system, this can be done with great accuracy. Thus the fact that your hands and arms start out this movement from an initial position which was not reached under active control by an outflow signal does not impair the accuracy of the movement.

Much evidence suggests, however, that comparison between target position and present position information is used to move the arms and that, as in Figure 9, this present position information is computed from outflow signals. In contrast, during the passive fall of an arm under the influence of gravity, changes in outflow signal commands are not responsible for the changes in position of the limb. Since the final position of a passively falling limb cannot be predicted in advance, it is clear that inflow signals must be used to update present position when an arm is moved passively by an external force, even though outflow signals are used to update present position when the arm moves actively under neural control.

This conclusion calls attention to a closely related issue that must be dealt with to understand the neural bases of skilled movement: How does the motor system know that the arm is being moved passively due to an external force, and not actively due to a changing outflow command? Such a distinction is needed to prevent inflow information from contaminating outflow commands when the arm is being actively moved. The motor system uses internally generated signals to make the distinction between active movement and passive movement, or postural, conditions. Computational gates are opened and shut based upon whether these internally generated signals are on or off.

Bullock and Grossberg (in press-a) have suggested that the GO signal schematized in Figure 9b helps to computationally define the postural state. Offset of the GO signal is hypothesized to open a learning gate which enables inflow signals to be adaptively recalibrated until they are computed in the same measurement scale as outflow signals (Figure 13). This type of learning occurs

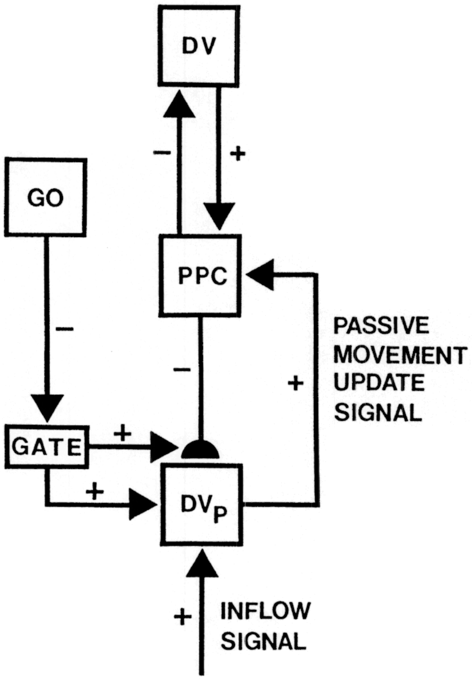


FIGURE 13. A passive update of position (PUP) circuit. An adaptive path *PPC* → $DV_P$ calibrates PPC-outflow signals in the same scale as inflow signals during intervals of posture. During passive movements, output from DV equals zero. Hence the passive difference vector $DV_P$ updates the PPC until it equals the new position caused by any passive movements that may occur due to the application of external forces.

in the mismatch mode, rather than the approximate match mode, using the mismatch between the coordinate systems as an error signal to drive the learning process. Offset of the GO signal is also hypothesized to open a gate which enables an outflow-inflow mismatch due to a passive movement to update the outflow present position command (Figure 13).

The circuit which accomplishes both of these learning and update functions is called the Passive Update of Position, or PUP, circuit. The equations of a typical PUP circuit are:

*Present Position Command*

$$\frac{d}{dt} P_i = G[V_i]^+ + G_p[M_i]^+ \tag{111}$$

*Outflow-Inflow Match*

$$\frac{d}{dt} M_i = -\beta M_i + \gamma I_i - z_i P_i \tag{112}$$

*Adaptive Gain Control*

$$\frac{d}{dt} z_i = \delta G_p(-\epsilon z_i + [M_i]^+). \tag{113}$$

Equation (111) supplements equation (110) with an update signal $G_p[M_i]^+$ that is turned on only when the passive gating function, or "pauser" signal, $G_p$ becomes positive in the passive, or postural, state. Function $z_i$ in (113) is an LTM trace, or associative weight, which adaptively recalibrates the gain of outflow signals $P_i$ until they are in the same scale as outflow signals $\gamma I_i$ in (112).

In summary, offset of the GO signal within the VITE circuit enables a pauser signal within the PUP circuit to drive its learning and reset functions. Such pauser-modulated learning of mismatches seems to occur in several adaptive sensory-motor control circuits. For example, Grossberg and Kuperstein (1986) have suggested that a pauser signal which defines the postural state of the ballistic eye movement system enables a mismatch signal analogous to $M_i$ in (112) to adaptively recode the representation of a light-activated target position command into motor coordinates. This adaptive recoding scheme, which is called the Head-Muscle Interface, or HMI, permits the recoded target position to be matched against the present position of the eye, which is also coded in motor coordinates. Such a matching process generates a movement command in the form of a difference vector which measures the mismatch between target position and present position, much as in the VITE circuit of Figure 9.

Typical equations for an HMI circuit are:

*Head-Muscle Match*

$$\frac{d}{dt} x_i = -Ax_i + G\left(\sum_{j=1}^{n} S_j\right)\left(-\sum_{j=1}^{n} S_j z_{ji} + I_i\right), \tag{114}$$

*Adaptive Gain Control*

$$\frac{d}{dt} z_{ji} = P\{-Bz_{ji} + S_j[x_i]^+\},$$

and

*Vector*

$$V = ([x_1]^+, [x_2]^+, \ldots, [x_n]^+)$$

where

*Match Gate*

$$G\left(\sum_{j=1}^{n} S_j\right) = \begin{cases} 1 & \text{if } \sum_{j=1}^{n} S_j > 0 \\ 0 & \text{if } \sum_{j=1}^{n} S_j = 0 \end{cases} \tag{117}$$

and $[x_j]^+ = \max(x_j, 0)$. In (114), $I_i$ is the corollary discharge signal that represents the position of the $i$th muscle, $S_j$ is the light-activated representation of the $j$th target position, $z_{ji}$ is the LTM trace that adaptively adjusts the gain between $S_j$ and $I_i$, $P$ is a gating signal that switches on in the postural mode, and $V$ is the output vector. Variable $x_i$ in (114) plays the role of variable $M_i$ in (112).

Due to the general importance of schemes such as the PUP circuit of (111)–(113) and the HMI circuit of (114)–(117) for adaptively recalibrating coordinate systems using vector computations, I have called all schemes of this type Adaptive Vector Encoders.

A third role for inflow signals is needed due to the fact that arms can move at variable velocities while carrying variable loads. Because an arm is a mechanical system embedded in a Newtonian world, an arm can generate unexpected amounts of inertia and acceleration when it tries to move novel loads at novel velocities. During such a novel motion, the commanded outflow position of the arm and its actual position may significantly diverge. Inflow signals are needed to compute mismatches leading to partial compensation for this uncontrolled component of the movement.

Such novel movements are quite different from our movements when we pick up a familiar fountain pen or briefcase. When the object is familiar, we can predictively adjust the gain of the movement to compensate for the expected mass of the object. This type of automatic gain control can, moreover, be flexibly switched on and off using signal pathways that can be activated by visual recognition of a familiar object. Inflow signals are used in the learning process which enables such automatic gain control signals to be activated in an anticipatory fashion in response to familiar objects (Bullock & Grossberg, in press-b).

## 19. EXTERNAL ERROR SIGNALS FOR LEARNING ADAPTIVE MOVEMENT GAINS: PUSH-PULL OPPONENT PROCESSING

The previous discussion outlined several of the types of learning whereby *internal* mismatches generate error

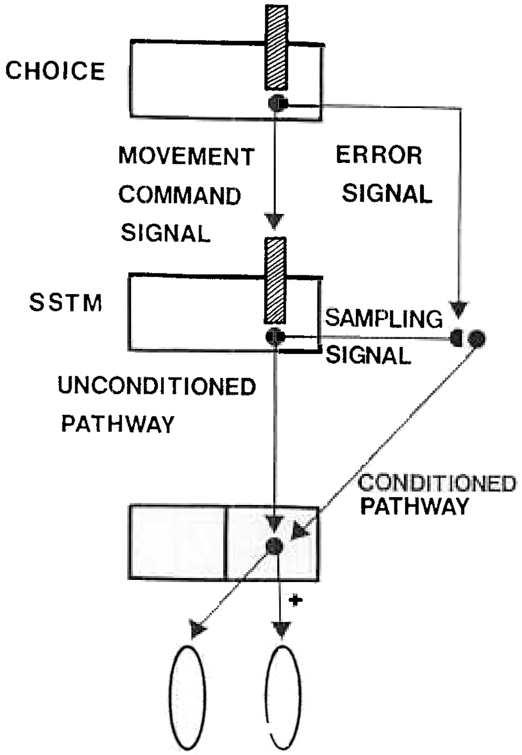


FIGURE 14. The representation of the chosen first light gives rise to an unconditioned movement signal and a conditioned movement signal. The unconditioned signal causes movements that are corrected by the conditioned movement signal via learning. The conditioned pathway carries sampling signals whose strength can be altered by second-light mediated error signals. These sampling signals give rise to the conditioned movement signal. The representation of the first light must be stored until after the end of the saccade, so that the second-light mediated error signal can act. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 38.)

signals capable of adaptively recalibrating a coordinate system or adaptively changing the gain of a movement command. *External* error signals are also used to alter the gain of a movement command. Such error signals function like an external "teacher." Unlike the hypothetical teachers employed in many examples of back-propagation, these error signals correspond to events which actually occur in the external environment during real-time learning.

Grossberg and Kuperstein (1986) have, for example, demonstrated how the accuracy of eye movements which do not successfully foveate a light can be improved by using the position of the light on the retina after the movement terminates as an error signal (Figure 14). Such an error signal adaptively changes the gain of the movement command. Mathematical and computer analyses demonstrate how to design such a system so that external error signals due to nonfoveated lights cooperate with internal error signals due to outflow-inflow mismatches to generate movements capable of adaptively maintaining their accuracy without the intervention of a human teacher (Figure 15).

The success of this learning model depends critically upon the hypothesis that errors are corrected in a push-pull, or opponent, fashion (Figure 16). The eye muscles, or other effectors, are assumed to be organized in reciprocal push-pull pairs, such as a pair for pulling to the right (R) and the left (L). Typical learning rules for such pairs are (expressed for convenience in discrete time ($n$):

*Hemifield Gradient Learning Rule*

$$z_{Ri}(n+1) = \delta z_{Ri}(n) + [L(E_n)]^+ \tag{118}$$

$$z_{Li}(n+1) = \delta z_{Li}(n) + [-L(E_n)]^+ \tag{119}$$

and the

*Fractured Somatotopy Learning Rule*

$$z_{Ri}(n+1) = [\delta z_{Ri}(n) + L(E_n)]^+ \tag{120}$$

$$z_{Li}(n+1) = [\delta z_{Li}(n) - L(E_n)]^+ \tag{121}$$

where in both cases $E_n$ represents the position of the light error on the $n$th trial, and $L(E_n)$ is the error signal by which it drives the learning of adaptive gains. Function $L(w)$ is assumed to be an increasing function of $w \geq 0$, and to be an odd function of $w$; viz., $L(w) = -L(-w)$. Variables $z_{Ri}$ and $z_{Li}$ are the LTM traces controlled by the $i$th command source to the right and left motor effectors.

Due to the push-pull organization of the learning process, the output signal $O_R(n)$ to the right muscle depends upon the differences $z_{Ri}(n) - z_{Li}(n)$ of these LTM traces, whereas the output signal $O_L(n)$ to the left muscle depends upon the differences $z_{Li}(n) - z_{Ri}(n)$. Such push-pull terms suggest a physical way to instantiate the types of formal comparisons between increments and baseline terms that have been hypothesized in a number of learning models (Rescorla & Wagner, 1972; Sutton & Barto, 1981). Opponent processing has also been assumed to regulate learning in real-time neural network models of classical and instrumental conditioning (Grossberg, 1972a; Grossberg & Schmajuk, 1987).

## 20. MATCH-INVARIANTS: INTERNALLY REGULATED LEARNING OF AN INVARIANT SELF-REGULATING TARGET POSITION MAP

Among the most important types of problems in neural network theory are those which concern the adaptive emergence of recognition invariants. Many different types of invariants can be identified; for example, the emergent invariants encoded by the critical feature patterns in an ART architecture enable the architecture to group all exemplars that share certain similarity properties into a single recognition category. As in ART, a number of other types of invariants are learned through a match-regulated process which gates

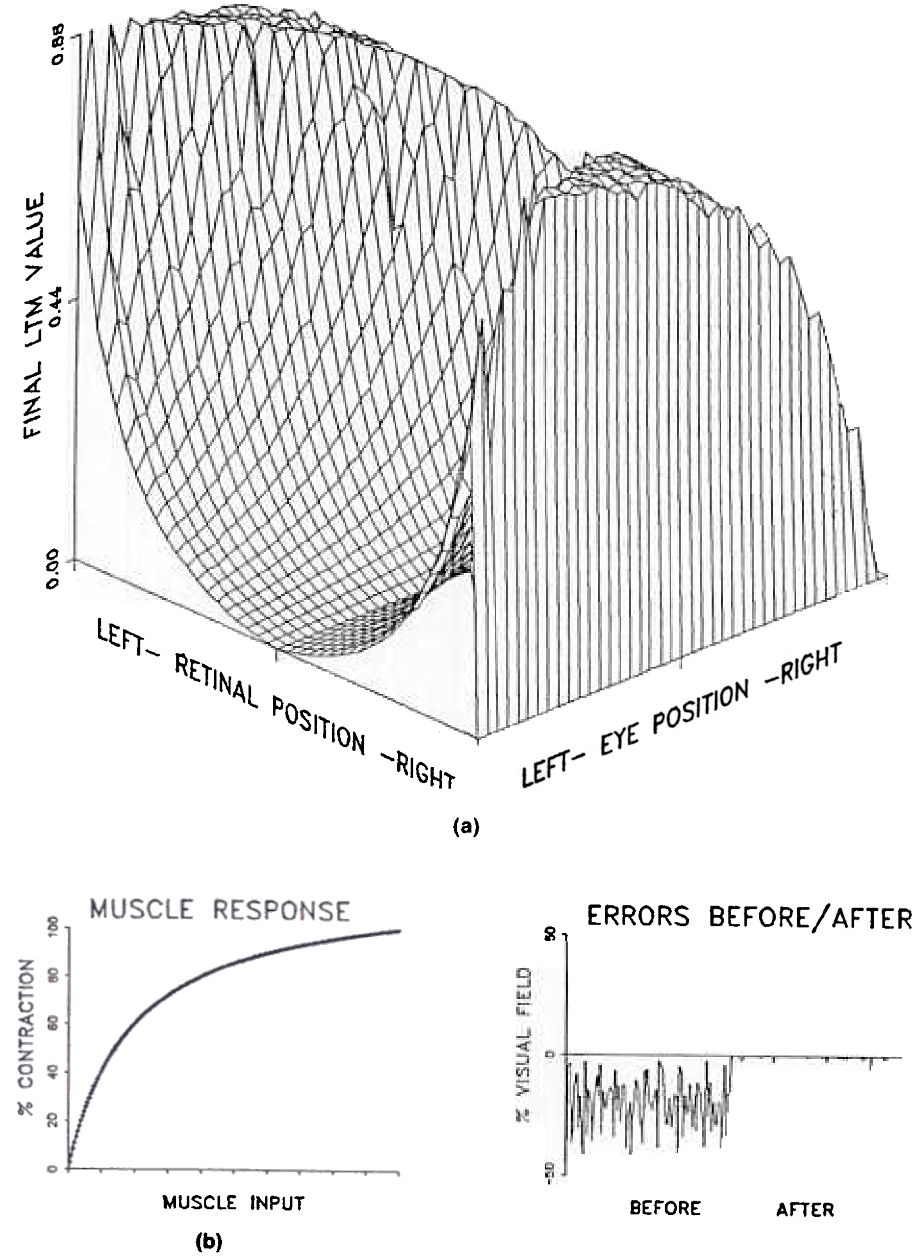


FIGURE 15. Computer simulation of saccadic error correction model with sampling from a non-invariant target position map using a slower-than-linear muscle function and a linear learning function. (a) Topographic distribution of LTM trace values after learning; (b) Muscle response function used in the simulation; (c) Errors in 100 trials before learning begins and 100 trials after learning ends. Negative values correspond to undershoots and positive values correspond to overshoots. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 89.)

on and off the learning mechanisms that enable individual events to be grouped together adaptively. I call such invariants *match invariants* to distinguish them from invariants that arise (say) due to a passive filtering process, such as Fourier-Mellin filtering.

Another model of a match invariant process was developed by Grossberg and Kuperstein (1986). This model shows how a matching process which defines the postural state can regulate the learning of an invariant self-regulating target position map in egocentric, or head-centered, coordinates. This problem arises when one considers how a visual signal to a moveable eye, or camera system, can be efficiently converted into an eye-tracking movement command.

To solve this problem, the positions of lights registered on the retina of an eye need to be converted into a head-coordinate frame so they can be compared with present eye positions which are also computed in a head-coordinate frame. In order to convert the position of a light on the retina into a target position in head coordinates, one needs to join together information about the light's retinal position with information about

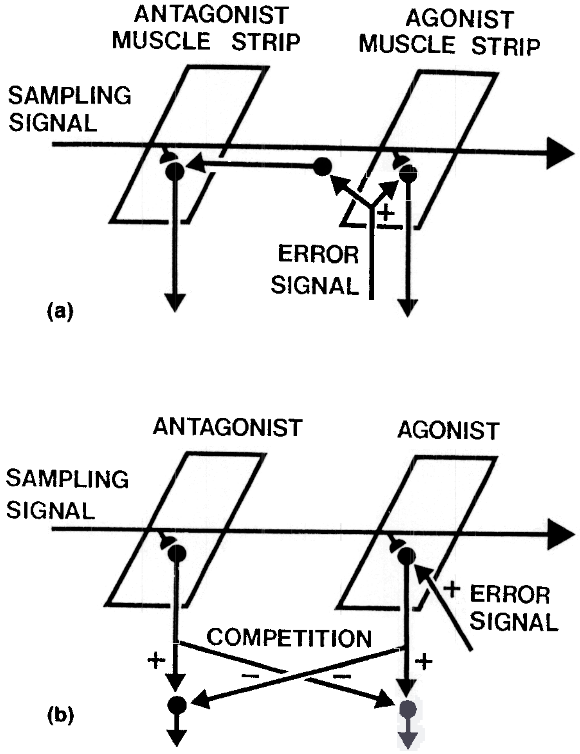


FIGURE 16. Two ways to achieve opponent conditioning of agonist-antagonist muscles: (a) An error signal increases the conditioned gain at the agonist muscle strip and decreases the conditioned gain at the antagonist muscle strip; (b) An error signal increases the conditioned gain at the agonist muscle strip. Competition between agonist and antagonist muscle strip outputs causes the decrease in the net antagonist output. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 70.)

present position of the eye in the head (Figure 17). Kuperstein and I suggested that this type of transformation is learned. Otherwise, the retinal system and the eye position system—which are widely separated in the brain and designed according to different internal constraints—would have to be pre-wired with perfectly chosen parameters for their mutual interaction. We have shown how such a transformation can be learned even if parameters are coarsely chosen initially and if significant portions of either system are damaged or even destroyed. This type of learning exhibits properties which are of general interest in other biological movement systems, in cognitive psychology, and in the design of freely moving robots. I will therefore describe its major elements here.

The most important properties of this transformation are that it is many-to-one, invariant, and self-regulating. As Figure 17 illustrates, many combinations of retinal position and eye position correspond to a single target position with respect to the head. When a single target position representation is activated by all of these possible combinations, the transformation is said to be invariant (Figure 18a). The key difficulty in understanding how such an invariant transformation is learned arises from its many-to-one property. The many-to-one property implies that each retinal position and each eye position can activate many target positions in head coordinates (Figure 18b). Even after learning takes place, each pair of retinal and eye positions can activate many target positions in head coordinates, but only the correct target position should receive the maximal total activation.

What prevents learning due to one pair of retinal and eye positions from contradicting learning due to a different pair of positions? In particular, if pairing retinal position $R_1$ with eye position $E_1$ strengthens the pathways from these positions to target position $T_1$, then why does not future pairing of $R_1$ with a different eye position $E_2$ continue to maximally excite $T_1$ instead

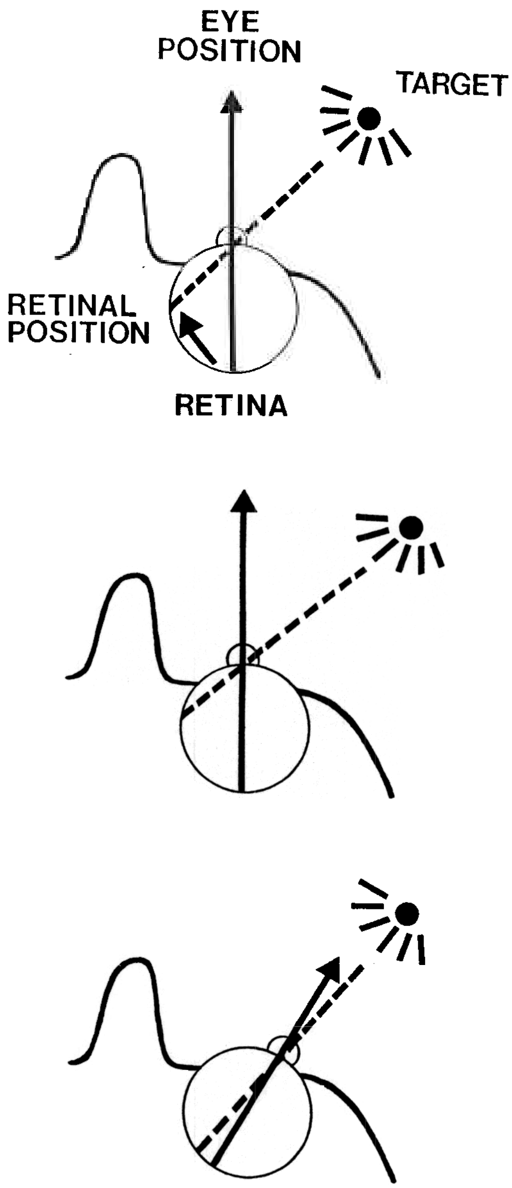


FIGURE 17. Many combinations of retinal position and eye position can encode the same target position.

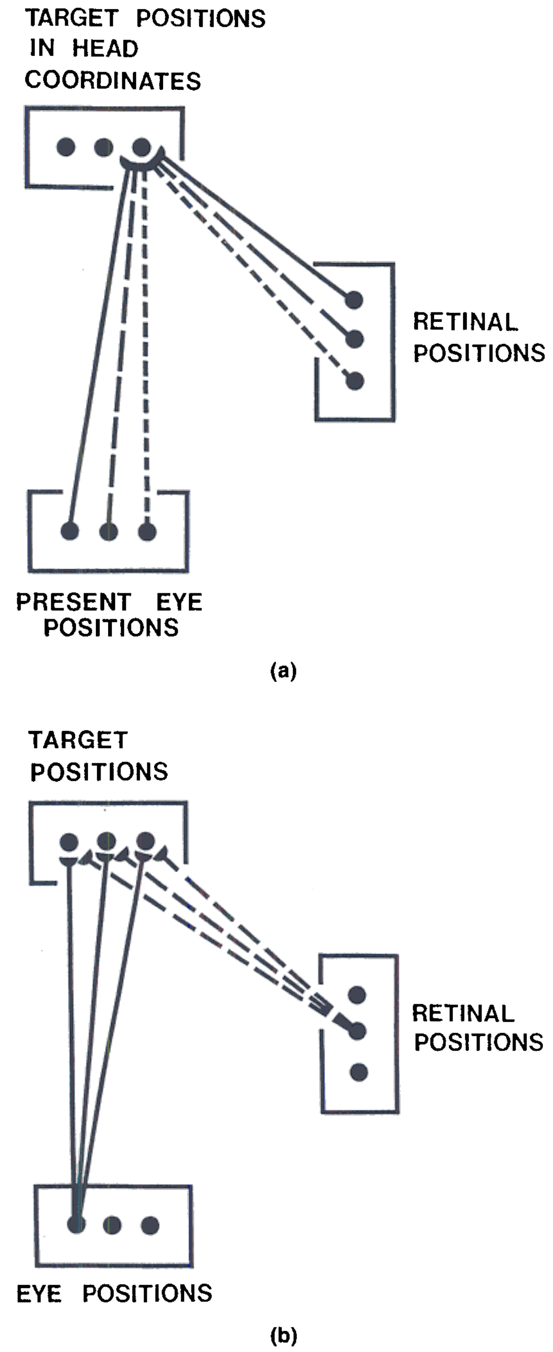


FIGURE 18. (a) When the many combinations of retinal position and eye position that correspond to each fixed target position all activate the same internal representation of that target position in head coordinates, the ensemble of such head coordinate representations is said to form an invariant map; (b) Every eye position and retinal position can send signals to many target position representations.

of the correct target position corresponding to $R_1$ and $E_2$? How is a globally consistent rule learned by a network, despite the fact that all computations in the network are local? How can a target position map be *implicitly* defined, such that each eye position and retinal position, taken separately, activates a large number of target positions, yet in combination always maximally activate the correct target position?

Finally, the property of self-regulation means that the map can correct itself even if a large fraction of the retinal positions and/or eye positions are destroyed, or if their parameters are otherwise altered through time. Destruction of a single retinal position eliminates all the combinations which that position made with all eye positions to activate target positions. In a similar fashion, destroying a single eye position can disrupt all target positions with which it was linked. A self-regulating map must thus be able to reorganize all of its learned changes to maintain its global self-consistency after removal of any of its components.

The self-regulation property is illustrated by the computer simulation from Grossberg and Kuperstein (1986) that is summarized in Figure 19. Each row in Figure 19 depicts learning of target positions corresponding to a different number of retinal and eye positions. More combinations of positions are represented in each successive row. The first column in each row depicts an intermediate learning stage, and the second column depicts a late learning stage. The abscissa plots topographic positions across the retinal and eye positions maps, whereas the ordinate plots the sizes of the adaptive path strengths, or learned long term memory (LTM) traces, in the pathways from these maps to the target position map. Such a scheme clarifies how the eye-head target position map that is schematized in Figure 9 may self-organize.

The LTM traces in Figure 19 were randomly chosen before learning began. A comparison of panels (b), (d), and (f) shows that the LTM traces can reorganize themselves when more combinations of positions are associated in such a way as to (approximately) preserve that part of the map which was learned when fewer combinations of positions were associated. This self-regulation property also holds when more combinations are replaced by fewer combinations, or if the initial LTM traces are not randomly chosen (Figure 20).

## 21. PRESYNAPTIC COMPETITION FOR LONG TERM MEMORY: SELF-REGULATING COMPETITIVE LEARNING

A complete model of how an invariant self-regulating target position map can be learned, as well as variants of this model, are described in Grossberg and Kuperstein (1986). Herein I emphasize one key point about the model.

The invariance and self-regulation properties of the TPM are due to the fact that all the LTM traces whose pathways project to a single TPM cell readjust themselves in a compensatory fashion when any one of these LTM traces changes due to learning (Figure 18). We suggested that the synaptic endings in which these LTM traces are computed contain autoreceptors (Cubeddu,

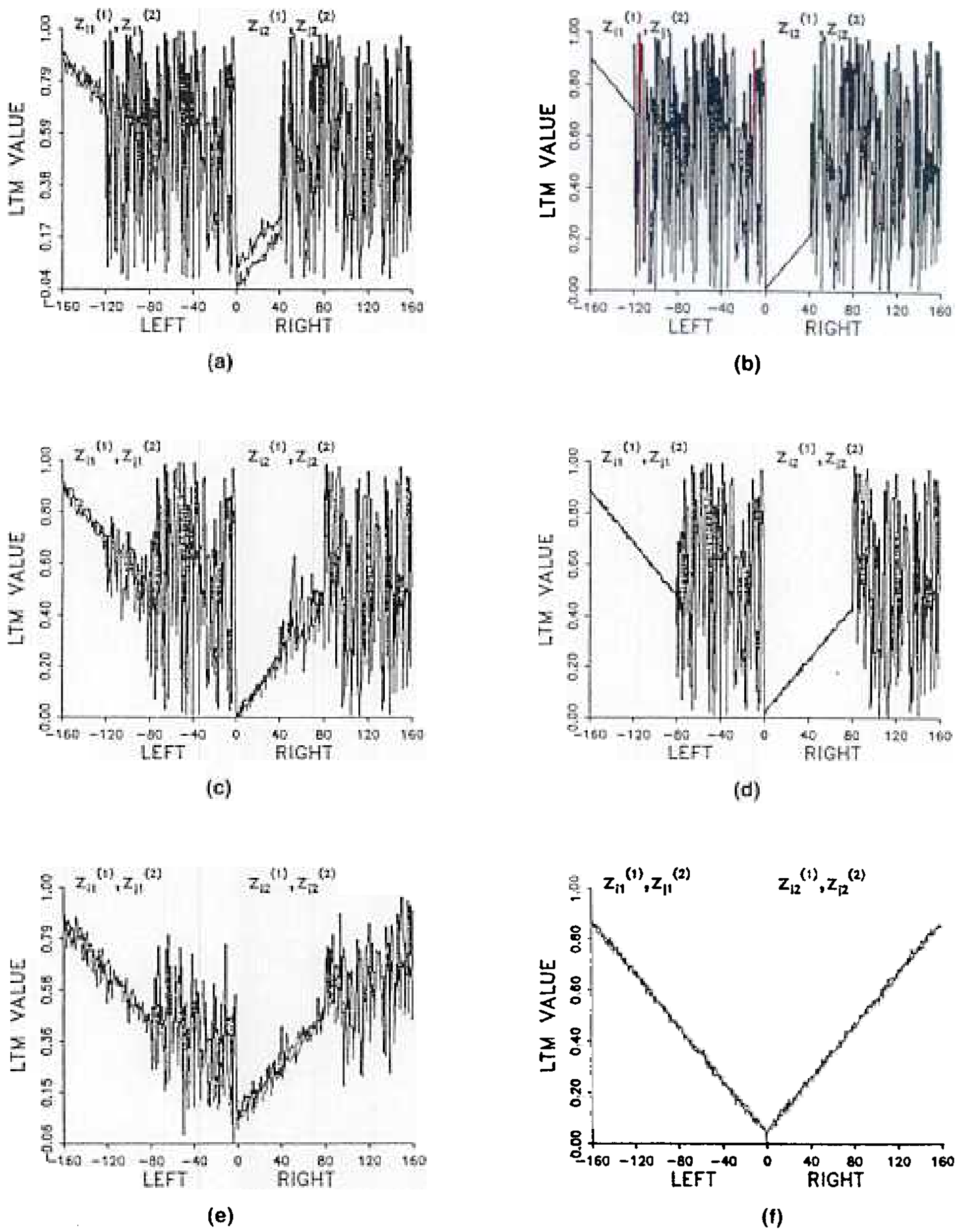


**FIGURE 19. Expansion of LTM spatial maps due to increase of the number of light positions and eye positions being correlated. The learned LTM values in corresponding positions agree, thereby illustrating map self-regulation. Initial values of LTM traces were randomly chosen. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 246.)**

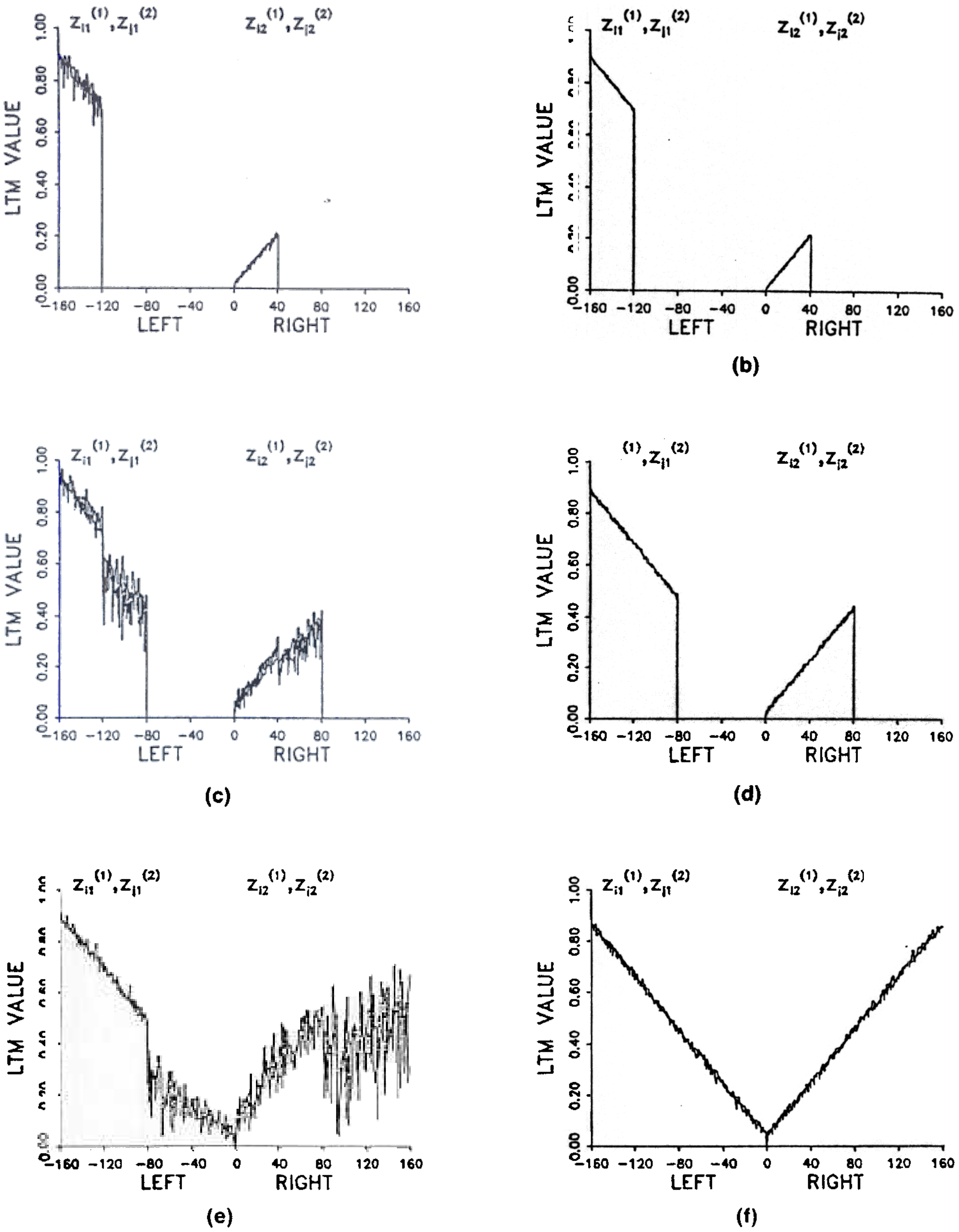
Hoffmann, & James, 1983; Dubocovich & Weiner, 1982; Groves & Tepper, 1983; Groves, Fenster, Tepper, Nakamura, & Young, 1981; Niedzwiecki, Mailman, & Cubeddu, 1984; Siever & Sulser, 1984; Tepper, Young, & Groves, 1984). In a network whose cells contain autoreceptive synapses, when a transmitter is released by one synaptic ending, a portion of it can undergo reuptake via the autoreceptors of other active and nearby synaptic endings. Reuptake has an inhibitory effect on the LTM trace of each active synaptic ending. Thus autoreceptors realize a type of presynaptic competition among all the LTM traces whose pathways converge upon the same cell within the TPM. Autoreceptors hereby mediate a novel type of self-regulating competitive learning.

Such an LTM trace obeys an equation of the form:

*Autoreceptive Associator*

$$\frac{d}{dt} z_{ij} = \epsilon S_i\left[-Fz_{ij} + Gx_j - H \sum_{k=1}^{n} S_k z_{kj}\right]. \qquad (122)$$

In (122), $z_{ij}$ is the LTM trace in the pathway from the $i$th cell in the retinotopic map or eye position map to the $j$th cell in the TPM; $S_i$ is the signal emitted by the $i$th cell into this pathway; and $x_j$ is the activity of the $j$th TPM cell. The terms $\epsilon$, $F$, $G$, and $H$ are constants.



(b)

(c)

(d)

(e)

(f)

**FIGURE 20. Same as in Figure 19, except initial values of LTM traces were chosen equal to zero. Note that learned spatial maps in Figures 19 and 20 agree, thereby illustrating the ability of map learning to overcome noise. (Reprinted with permission from Grossberg & Kuperstein, 1986, p. 239.)**

Equation (122) says that reuptake via autoreceptors of a fraction of released transmitter, as in term $-H \times \sum_{k=1}^{n} S_k z_{kj}$, inhibits the growth of the corresponding LTM trace.

Match-regulated learning processes also occur in applications to such spatio-temporal learning problems as speech and language learning, and planned sensory-motor control (Grossberg, 1982, 1987b; Grossberg & Kuperstein, 1986). The totality of such known examples illustrate how combinations of internal matching processes and external error signals derived from natural real-time environments can regulate stable self-organized learning, recognition, and action in response to noisy and unpredictable environments.

## REFERENCES

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**, 147–169.

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, **10**, 25–61.

Amari, S., & Arbib, M. A. (Eds.). (1982). *Competition and cooperation in neural networks*. New York: Springer-Verlag.

Amari, S., & Takeuchi, A. (1978). Mathematical theory on formation of category detecting nerve cells. *Biological Cybernetics,* **29,** 127–136.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1987). Information storage in neural networks with low levels of activity. *Physical Review A,* **35,** 2293–2303.

Anderson, J. A. (1968). A memory model utilizing spatial correlation functions. *Kybernetik,* **5,** 113–119.

Anderson, J. A. (1983). Cognitive and psychological computation with neural models. *IEEE Transactions* **SMC-13, 5,** 799–815.

Anderson, J. A., Silverstein, J. W., Ritz, S. R., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review,* **84,** 413–451.

Arend, L. E., Buehler, J. N., & Lockhead, G. R. (1971). Difference information in brightness perception. *Perception and Psychophysics,* **9,** 367–370.

Banquet, J.-P., & Grossberg, S. (in press). Probing cognitive processes through the structure of event-related potentials during learning: An experimental and theoretical analysis. *Applied Optics.*

Beck, J. (1972). *Surface color perception.* Ithaca, NY: Cornell University Press.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience,* **2,** 32–48.

Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Brindley, G. S. (1964). The use made by the cerebellum of the information that it receives from sense organs. *International Brain Research Organization Bulletin,* **3,** 80.

Bullock, D., & Grossberg, S. (in press-a). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review.*

Bullock, D., & Grossberg, S. (in press-b). Neuromuscular realization of planned trajectories: Adaptive and automatic mechanisms.

Caianiello, E. R. (1961). Outline of a theory of thought and thinking machines. *Journal of Theoretical Biology,* **1,** 204–235.

Campbell, L., & Garnett, W. (1882). *The life of James Clerk Maxwell.* London: Macmillan.

Carpenter, G. A. (1977a). A geometric approach to singular perturbation problems with applications to nerve impulse equations. *Journal of Differential Equations,* **23,** 335–367.

Carpenter, G. A. (1977b). Periodic solutions of nerve impulse equations. *Journal of Mathematical Analysis and Applications,* **58,** 152–173.

Carpenter, G. A. (1979). Bursting phenomena in excitable membranes. *SIAM Journal on Applied Mathematics,* **36,** 334–372.

Carpenter, G. A. (1981). Normal and abnormal signal patterns in nerve cells. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (pp. 49–90). Providence, RI: American Mathematical Society.

Carpenter, G. A., & Grossberg, S. (1985). Category learning and adaptive pattern recognition: A neural network model. *Proceedings of the Third Army Conference on Applied Mathematics and Computing,* 37–56.

Carpenter, G. A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing,* **37,** 54–115.

Carpenter, G. A., & Grossberg, S. (1987b). Associative learning, adaptive pattern recognition, and cooperative-competitive decision making by neural networks. In H. Szu (Ed.), *Optical and hybrid computing* (218–247). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.

Carpenter, G. A., & Grossberg, S. (in press). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics.*

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions* **SMC-13,** 815–826.

Cohen, M. A., & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics,* **36,** 428–456.

Cohen, M., & Grossberg, S. (1986). Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology,* **5,** 1–22.

Cohen, M., & Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics,* **26,** 1866–1891.

Cornsweet, T. N. (1970). *Visual perception.* New York: Academic Press.

Cubeddu, L. X., Hoffmann, I. S., & James, M. K. (1983). Frequency-dependent effects of neuronal uptake inhibitors on the autoreceptor-mediated modulation of dopamine and acetylcholine release from the rabbit striatum. *Journal of Pharmacology and Experimental Therapeutics,* **226,** 88–94.

Denker, J. S. (Ed.). (1986). *Neural networks for computing.* New York: American Institute of Physics.

Dubocovich, M. L., & Weiner, N. (1982). Modulation of the stimulation-evoked release of 3H-dopamine through activation of dopamine autoreceptors of the D-2 subtype in the isolated rabbit retina. In M. Kohsaka *et al.* (Eds.), *Advances in the biosciences, Volume 37: Advances in dopamine research.* New York: Pergamon Press.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Eigen, M., & Schuster, P. (1978). The hypercycle: A principle of natural self-organization, B: The abstract hypercycle. *Naturwissenschaften,* **65,** 7–41.

Ellias, S. A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics,* **20,** 69–98.

Elman, J. H., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America,* **62,** 971–974.

Ermentrout, G. B., & Cowan, J. D. (1979). Temporal oscillations in neuronal nets. *Journal of Mathematical Biology,* **7,** 265–286.

Ermentrout, G. B., & Cowan, J. D. (1980). Large scale spatially organized activity in neural nets. *SIAM Journal on Applied Mathematics,* **38,** 1–21.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science,* **6,** 205–254.

Freeman, W. J. (1975). *Mass action in the nervous system.* New York: Academic Press.

Freeman, W. J. (1979). EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb. *Biological Cybernetics,* **35,** 221–234.

Fujita, M. (1982a). Simulation of adaptive modification of the vestibulo-ocular reflex with an adaptive filter model of the cerebellum. *Biological Cybernetics,* **45,** 207–214.

Fujita, M. (1982b). Adaptive filter model of the cerebellum. *Biological Cybernetics,* **45,** 195–206.

Geman, S. (1983, April). Stochastic relaxation methods for image restoration and expert systems. Proceedings of ARO workshop on unsupervised image analysis. Brown University, Providence, RI.

Geman, S. (1984). Stochastic relaxation methods for image restoration and expert systems. In D. B. Cooper, R. L. Launer, & D. E. McClure (Eds.), *Automated image analysis: Theory and experiments.* New York: Academic Press.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6,** 721–741.

Gilpin, M. E., & Ayala, F. J. (1973). Global models of growth and competition. *Proceedings of the National Academy of Sciences,* **70,** 3590–3593.

Glazebrook, R. T. (1905). *James Clerk Maxwell and modern physics.* New York: Macmillan.

Goh, B. S., & Agnew, T. T. (1977). Stability in Gilpin and Ayala's models of competition. *Journal of Mathematical Biology,* **4**, 275–279.

Golden, R. M. (1986). The "Brain-state-in-a-box" neural model is a gradient descent algorithm. *Journal of Mathematical Psychology,* **30**, 73–80.

Gray, R. M. (1984, April). Vector quantization. *IEEE ASSP Magazine,* 4–29.

Grossberg, S. (1964). *The theory of embedding fields with applications to psychology and neurophysiology.* New York: Rockefeller Institute for Medical Research.

Grossberg, S. (1967). Nonlinear difference-differential equations in prediction and learning theory. *Proceedings of the National Academy of Sciences,* **58**, 1329–1334.

Grossberg, S. (1968a). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences,* **60**, 758–765.

Grossberg, S. (1968b). Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proceedings of the National Academy of Sciences,* **59**, 368–372.

Grossberg, S. (1969a). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics,* **1**, 319–350.

Grossberg, S. (1969b). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, I. *Journal of Mathematics and Mechanics,* **19**, 53–91.

Grossberg, S. (1969c). On the serial learning of lists. *Mathematical Biosciences,* **4**, 201–253.

Grossberg, S. (1969d). On learning of spatiotemporal patterns by networks with ordered sensory and motor components, I: Excitatory components of the cerebellum. *Studies in Applied Mathematics,* **48**, 105–132.

Grossberg, S. (1970a). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II. *Studies in Applied Mathematics,* **49**, 135–166.

Grossberg, S. (1970b). Neural pattern discrimination. *Journal of Theoretical Biology,* **27**, 291–337.

Grossberg, S. (1971a). On the dynamics of operant conditioning. *Journal of Theoretical Biology,* **33**, 225–255.

Grossberg, S. (1971b). Pavlovian pattern learning by nonlinear neural networks. *Proceedings of the National Academy of Sciences,* **68**, 828–831.

Grossberg, S. (1972a). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences,* **15**, 253–285.

Grossberg, S. (1972b). Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik,* **10**, 49–57.

Grossberg, S. (1972c). Pattern learning by functional-differential neural networks with arbitrary path weights. In K. Schmitt (Ed.), *Delay and functional-differential equations and their applications* (pp. 121–160). New York: Academic Press.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics,* **52**, 217–257.

Grossberg, S. (1974). Classical and instrumental learning by neural networks. In R. Rosen and F. Snell (Eds.), *Progress in theoretical biology.* New York: Academic Press.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics,* **23**, 121–134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics,* **23**, 187–202.

Grossberg, S. (1977). Pattern formation by the global limits of a nonlinear competitive interaction in *n* dimensions. *Journal of Mathematical Biology,* **4**, 237–256.

Grossberg, S. (1978a). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.

Grossberg, S. (1978b). Do all neural models really look alike? *Psychological Review,* **85**, 592–596.

Grossberg, S. (1978c). Competition, decision, and consensus. *Journal of Mathematical Analysis and Applications,* **66**, 470–493.

Grossberg, S. (1978d). Decisions, patterns, and oscillations in nonlinear competitive systems with applications to Volterra-Lotka systems. *Journal of Theoretical Biology,* **73**, 101–130.

Grossberg, S. (1980a). Biological competition: Decision rules, pattern formation, and oscillations. *Proceedings of the National Academy of Sciences,* **77**, 2338–2342.

Grossberg, S. (1980b). How does a brain build a cognitive code? *Psychological Review,* **87**, 1–51.

Grossberg, S. (Ed.). (1981). Adaptive resonance in development, perception, and cognition. In *Mathematical psychology and psychophysiology* (pp. 107–156). Providence, RI: American Mathematical Society.

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston: Reidel Press.

Grossberg, S. (1984). Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory,* **7**, 263–283.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines, Vol. 1: Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color, and brightness perception, I: Monocular theory. *Perception and Psychophysics,* **41**, 87–116.

Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory. *Perception and Psychophysics,* **41**, 117–158.

Grossberg, S. (Ed.). (1987c). *The adaptive brain, I: Cognition, learning, reinforcement, and rhythm.* Amsterdam: Elsevier/North-Holland.

Grossberg, S. (Ed.). (1987d). *The adaptive brain, II: Vision, speech, language, and motor control.* Amsterdam: Elsevier/North-Holland.

Grossberg, S. (1987e). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science,* **11**, 23–63.

Grossberg, S. (Ed.). (1988). *Neural networks and natural intelligence.* Cambridge, MA: MIT Press.

Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Amsterdam: Elsevier/North-Holland.

Grossberg, S., & Levine, D. (1975). Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology,* **53**, 341–380.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review,* **92**, 173–211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics,* **38**, 141–171.

Grossberg, S., & Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading. *Computer Vision, Graphics, and Image Processing,* **37**, 116–165.

Grossberg, S., & Pepe, J. (1971). Spiking threshold and overarousal effects in serial learning. *Journal of Statistical Physics,* **3**, 95–125.

Grossberg, S., & Schmajuk, N. A. (1987). Neural dynamics of Pavlovian conditioning: Conditioned reinforcement, habituation, and opponent processing. *Psychobiology,* **15**, 195–240.

Grossberg, S., & Todorović, D. (in press). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics.*

Groves, P. M., Fenster, G. A., Tepper, J. M., Nakamura, S., & Young, S. J. (1981). Changes in dopaminergic terminal excitability induced by amphetamine and hatoperidol. *Brain Research,* **221,** 425–431.

Groves, P. M., & Tepper, J. M. (1983). Neuronal mechanisms of action of amphetamine. In I. Creese (Ed.), *Stimulants: Neurochemical, behavioral and clinical perspectives* (pp. 81–129). New York: Raven Press.

Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., & Crandall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science,* **210,** 803–805.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception and Psychophysics,* **32,** 409–418.

Hastings, S. P. (1976). The existence of periodic solutions to Nagumo's equation. *Quarterly Journal of Mathematics,* **27,** 123–134.

Hastings, S. P. (1982). Single and multiple pulse waves for the FitzHugh-Nagumo equations. *SIAM Journal of Applied Mathematics,* **42,** 247–260.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hecht-Nielsen, R. (1986). Performance limits of optical, electro-optical, and electronic neurocomputers. In H. Szu (Ed.), *Hybrid and optical computing* (pp. 277–306). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.

Helmholtz, H. von (1962). *Treatise on physiological optics* (J. P. C. Southall, Trans.). New York: Dover. (Original work published in 1866)

Hestenes, D. (1987). How the brain works: The next great scientific revolution. In C. R. Smith (Ed.), *Maximum entropy in Bayesian spectral analysis and estimation problems.* Boston: Reidel Press.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hirsch, M. W. (1982). Systems of differential equations that are competitive or cooperative, I: Limit sets. *SIAM Journal of Mathematical Analysis,* **13,** 167–179.

Hirsch, M. W. (1985). Systems of differential equations that are competitive or cooperative, II: Convergence almost everywhere. *SIAM Journal of Mathematical Analysis,* **16,** 423–439.

Hodgkin, A. L. (1964). *The conduction of the nervous impulse.* Liverpool: Liverpool University.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology,* **117,** 500–544.

Hodgson, J. E. P. (Ed.). (1983). *Oscillations in mathematical biology.* New York: Springer-Verlag.

Holst, E. von, & Mittelstaedt, H. (1950). The reafference principle: Interaction between the central nervous system and the periphery. *Naturwissenschaften,* **37,** 464–476.

Hopfield, J. J. (1982). Neuronal networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences,* **79,** 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences,* **81,** 3058–3092.

Hopfield, J. E., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics,* **52,** 141–152.

Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science,* **233,** 625–633.

Hurvich, L. M. (1981). *Colour vision.* Sunderland, MA: Sinauer Associates.

Ito, M. (1974). The control mechanisms of cerebellar motor systems. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences third study program* (pp. 293–303). Cambridge, MA: MIT Press.

Ito, M. (1982). Cerebellar control of the vestibulo-ocular reflex—Around the flocculus hypothesis. *Annual Review of Neuroscience,* **5,** 275–296.

Ito, M. (1984). *The cerebellum and neural control.* New York: Raven Press.

Iverson, G. J., & Pavel, M. (1981). Invariant properties of masking phenomena in psychoacoustics and their theoretical consequences. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology.* Providence, RI: American Mathematical Society.

Kaczmarek, L. K., & Babloyantz, A. (1977). Spatiotemporal patterns in epileptic seizures. *Biological Cybernetics,* **26,** 199–208.

Kandel, E. R., & Schwartz, J. H. (1981). *Principles of neural science.* New York: Elsevier/North-Holland.

Kasamatsu, T., & Pettigrew, J. D. (1976). Depletion of brain catecholamines: Failure of ocular dominance shift after monocular occlusion in kittens. *Science,* **194,** 206–209.

Katz, B. (1966). *Nerve, muscle, and synapse.* New York: McGraw-Hill.

Kilmer, W. L. (1972). On some realistic constraints in prey-predator mathematics. *Journal of Theoretical Biology,* **36,** 9–22.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1982). *Optimization by simulated annealing.* IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

Kirkpatrick, S., Gelatt, D. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science,* **220,** 671–680.

Koenigsberger, L. (1906). *Hermann von Helmholtz* (F. A. Welby, Trans.). Oxford: Clarendon Press.

Kohonen, T. (1971). A class of randomly organized associative memories. *Acta Polytechnica Scandinavica,* El. 25.

Kohonen, T. (1977). *Associative memory—A system-theoretical approach.* New York: Springer-Verlag.

Kohonen, T. (1984). *Self-organization and associative memory.* New York: Springer-Verlag.

Kosko, B. (1987). *Bidirectional associative memories.* Manuscript submitted for publication.

Kosko, B., & Guest, C. (in press). Optical bidirectional associative memories. *Society for Photo-optical and Instrumentation Engineering (SPIE) Proceedings: Image Understanding,* **758.**

Land, E. H. (1977). The retinex theory of color vision. *Scientific American,* **237,** 108–128.

Levine, D. S. (1979). Existence of a limiting pattern for a system of nonlinear equations describing inter-population competition. *Bulletin of Mathematical Biology,* **41,** 617–628.

Levine, D. S. (1983). Neural population modeling and psychology: A review. *Mathematical Biosciences,* **66,** 1–86.

Levine, D. S., & Grossberg, S. (1976). Visual illusions in neural networks: Line neutralization, tilt aftereffects, and angle expansion. *Journal of Theoretical Biology,* **61,** 477–504.

Levy, W. B. (1985). Associative changes at the synapse: LTP in the hippocampus. In W. B. Levy, J. Anderson, & S. Lehmkuhle (Eds.), *Synaptic modification, neuron selectivity, and nervous system organization* (pp. 5–33). Hillsdale, NJ: Erlbaum.

Levy, W. B., Brassel, S. E., & Moore, S. D. (1983). Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience,* **8,** 799–808.

Levy, W. B., & Desmond, N. L. (1985). The rules of elemental synaptic plasticity. In W. B. Levy, J. Anderson, & S. Lehmkuhle (Eds.), *Synaptic modification, neuron selectivity, and nervous system organization* (pp. 105–121). Hillsdale, NJ: Erlbaum.

Lotka, A. J. (1956). *Elements of mathematical biology.* New York: Dover.

MacArthur, R. H. (1970). Species packing and competitive equilibrium for many species. *Theoretical Population Biology,* **1,** 1–11.

Mach, E. (1914). *The analysis of sensation and the relation of the physical to the psychical* (C. M. Williams, Trans., revised by S. Waterlow). London: Open Court Publishing.

Malsburg, C. von der (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik,* **14,** 85–100.

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology (London),* **202,** 437–470.

May, R. M., & Leonard, W. J. (1975). Nonlinear aspects of competition between three species. *SIAM Journal on Applied Mathematics,* **29,** 243–253.

McClelland, J. L. (1987). *The case for interactionism in language processing.* Preprint.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review,* **88,** 375–407.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing* (Vol. II). Cambridge, MA: MIT Press.

McCormick, D. A., & Thompson, R. F. (1984). Cerebellum: Essential involvement in the classically conditioned eyelid response. *Science,* **223,** 296–299.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics,* **5,** 115–133.

McEliece, R. J., Posner, E. C. Rodemich, E. R., & Venkatesh, S. S. (1986). The capacity of the Hopfield associative memory. *IEEE Transactions in Information Theory.*

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics,* **25,** 457–465.

Mollon, J. D., & Sharpe, L. T. (Eds.). (1983). *Colour vision.* New York: Academic Press.

Mueller, P., Martin, T., & Putzrath, F. (1962). General principles of operations in neuron nets with application to acoustical pattern recognition. In E. E. Bernard & M. R. Kare (Eds.), *Biological prototypes and synthetic systems* (Vol. 1, pp. 192–212). New York: Plenum Press.

Nathan, O., & Norden, H. (Eds.). (1960). *Einstein on peace.* New York: Schocken Books.

Niedzwiecki, D. M., Mailman, R. B., & Cubeddu, L. X. (1984). Greater potency of mesoridazine and sulforidazine compared with the parent compound, thioridazine, on striatal dopamine autoreceptors. *Journal of Pharmacology and Experimental Therapeutics,* **228,** 636–639.

Optican, L. M., & Robinson, D. A. (1980). Cerebellar-dependent adaptive control of primate saccadic system. *Journal of Neurophysiology,* **44,** 1058–1076.

Parker, D. B. (1982, October). *Learning-logic* (Invention Report 581-64, File 1). Office of Technology Licensing, Stanford University.

Parker, D. B. (1985, April). *Learning-Logic, TR-47.* Center for Computational Research in Economics and Management Science, MIT.

Parker, D. B. (1986). A comparison of algorithms for neuron-like cells. In J. S. Denker (Ed.), *Neural networks for computing* (pp. 327–332). New York: American Institute of Physics.

Pastore, R. E. (1981). Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech* (pp. 165–205). Hillsdale, NJ: Erlbaum.

Pettigrew, J. D., & Kasamatsu, T. (1978). Local perfusion of noradrenaline maintains visual cortical plasticity. *Nature,* **271,** 761–763.

Platt, J. C., & Hopfield, J. J. (1986). Analog decoding using neural networks. In J. S. Denker (Ed.), *Neural networks for computing* (pp. 364–369). New York: American Institute of Physics.

Plonsey, R., & Fleming, D. G. (1969). *Bioelectric phenomena.* New York: McGraw-Hill.

Poggio, T., & Koch, C. (1987). Synapses that compute motion. *Scientific American,* **256,** 46–52.

Psaltis, D., & Park, C. H. (1986). Nonlinear discriminative functions and associative memories. In J. S. Denker (Ed.), *Neural networks for computing.* New York: American Institute of Physics.

Ratliff, F. (1965). *Mach bands: Quantitative studies on neural networks in the retina.* New York: Holden-Day.

Ratliff, F., Hartline, H. K., & Miller, W. H. (1963). Spatial and temporal aspects of retinal inhibitory interactions. *Journal of the Optical Society of America,* **53,** 110–120.

Rauschecker, J. P., & Singer, W. (1979). Changes in the circuitry of the kitten's visual cortex are gated by postsynaptic activity. *Nature,* **280,** 58–60.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning, II: Current research and theory.* New York: Appleton-Century-Crofts.

Ricciardi, L., & Scott, A. (Eds.). (1982). *Biomathematics in 1980.* Amsterdam: North-Holland.

Ron, S., & Robinson, D. A. (1973). Eye movements evoked by cerebellar stimulation in the alert monkey. *Journal of Neurophysiology,* **36,** 1004–1021.

Rosenblatt, F. (1962). *Principles of neurodynamics.* Washington, DC: Spartan Books.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception, Part 2: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review,* **89,** 60–94.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing* (Vol. I). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science,* **9,** 75–112.

Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1982). Length effects in word perception: We is better than I but worse than you or them. *Journal of Experimental Psychology: Human Perception and Performance,* **8,** 91–105.

Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1983). Reply to Matthei: We really is worse than you or them, and so are ma and pa. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 321–322.

Sawusch, J. R., & Nusbaum, H. C. (1979). Contextual effects in vowel perception, I: Anchor-induced contrast effects. *Perception and Psychophysics,* **25,** 292–302.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. (1980). Contextual effects in vowel perception, II: Evidence for two processing mechanisms. *Perception and Psychophysics,* **27,** 421–434.

Schwab, E. C., Sawusch, J. R., & Nusbaum, H. C. (1981). The role of second formant transitions in the stop-semivowel distinction. *Perception and Psychophysics,* **29,** 121–128.

Scott, A. C. (1977). *Neurophysics.* New York: Wiley-Interscience.

Sejnowski, T. J., & Rosenberg, C. R. (1986, January). *NETtalk: A parallel network that learns to read aloud.* Johns Hopkins University, MD.

Siever, L., & Sulser, F. (1984). Regulations of amine neurotransmitter systems: Implication for the major psychiatric syndromes and their treatment. *Psychopharmacology Bulletin,* **20,** 500–504.

Singer, W. (1982). The role of attention in developmental plasticity. *Human Neurobiology,* **1,** 41–43.

Singer, W. (1983). Neuronal activity as a shaping factor in the self-organization of neuron assemblies. In E. Basar, H. Flohr, H. Haken, & A. J. Mandell (Eds.), *Synergetics of the brain* (pp. 89–101). New York: Springer-Verlag.

Sperling, G. (1981). Mathematical models of binocular vision. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (pp. 281–300). Providence, RI: American Mathematical Society.

Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America,* **58,** 1133–1145.

Studdert-Kennedy, M. (1980). Speech perception. *Language and Speech,* **23,** 45–66.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review,* **88,** 135–170.

Szu, H. (1986). Three layers of vector outer product neural networks for optical pattern recognition. In H. Szu (Ed.), *Optical and hybrid computing* (pp. 312–330). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.

Takeuchi, Y., Adachi, N., & Tokumaru, H. (1978). The stability of generalized Volterra equations. *Journal of Mathematical Analysis and Applications,* **62,** 453–473.

Tepper, J. M., Young, S. J., & Groves, P. M. (1984). Autoreceptor-mediated changes in dopaminergic terminal excitability: Effects of increases in impulse flow. *Brain Research,* **309,** 309–316.

Venkatesh, S. S. (1986). Epsilon capacity of neural networks. In J. S. Denker (Ed.), *Neural networks for computing.* New York: American Institute of Physics.

Vilis, T., & Hore, J. (1986). A comparison of disorders in saccades and in fast and accurate elbow flexions during cerebellar dysfunction. In H. J. Freund, U. Büttner, B. Cohen, & J. Noth (Eds.), *The oculomotor and skeletal motor systems: Differences and similarities.* New York: Elsevier.

Vilis, T., Snow, R., & Hore, J. (1983). Cerebellar saccadic dysmetria is not equal in the two eyes. *Experimental Brain Research,* **51,** 343–350.

Werblin, F. S. (1971). Adaptation in a vertebrate retina: Intracellular recordings in *Necturus. Journal of Neurophysiology,* **34,** 228–241.

Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences.* Unpublished doctoral thesis, Harvard University, Cambridge, MA.

Werbos, P. (1982). Applications of advances in nonlinear sensitivity analysis. In A. V. Balakrishnan, M. Thoma, R. F. Drenick, & F. Kozin (Eds.), *Lecture notes in control and information sciences, Vol. 38: System modeling and optimization.* Proceedings of the 10th IFIP Conference. New York: Springer-Verlag.

Widrow, B. (1962). Generalization and information storage in networks of Adaline neurons. In M. C. Yovits, G. T. Jacobi, & G. D. Goldstein (Eds.), *Self-organizing systems.* Washington, DC: Spartan Books.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal,* **12,** 1–24.