

INTRODUCTION TO THE THEORY OF NEURAL COMPUTATION

*John Hertz
Anders Krogh
Richard G. Palmer*



A LECTURE NOTES VOLUME IN THE

SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY

INTRODUCTION TO THE THEORY OF NEURAL COMPUTATION



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

INTRODUCTION TO THE THEORY OF NEURAL COMPUTATION

John Hertz

NORDITA

Anders Krogh

Niels Bohr Institute

Richard G. Palmer

Duke University and the Santa Fe Institute

Lecture Notes Volume I

**SANTA FE INSTITUTE
STUDIES IN THE SCIENCES OF COMPLEXITY**



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

First published 1991 by Westview Press

Published 2018 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of the Taylor & Francis Group, an Informa business

Copyright © 1991 Taylor & Francis Group LLC

No claim to original U.S. Government works

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Hertz, John A.
Introduction to the theory of neural computation / John A. Hertz,
Richard G. Palmer, Anders S. Krogh.
p. cm.—(Santa Fe Institute studies in the sciences of complexity.
Lecture notes: v. 1) (Computation and neural systems series)
Includes index.
1. Neural computers. 2. Neural circuitry. I. Palmer, Richard G. II.
Krogh, Anders S. III. Title. IV. Series. V. Series: Computation and neural
systems series.
QA76.5.H475 1991 006.3—dc20 91-701
—ISBN 0-201-51560-1 (pbk.)

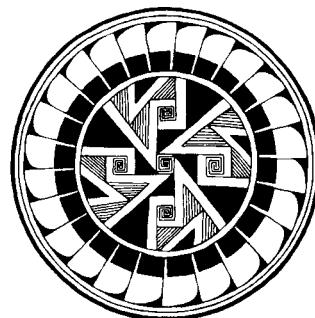
ISBN 13: 978-0-201-51560-2 (pbk)

This volume was typeset using TeXtures on a Macintosh II computer.

About the Santa Fe Institute

The *Santa Fe Institute* (SFI) is a private, independent organization dedicated to multidisciplinary scientific research and graduate education in the natural, computational, and social sciences. The driving force behind its creation in 1984 was the need to understand those complex systems that shape human life and much of our immediate world—evolution, the learning process, the immune system, the world economy. The intent is to make the new tools now being developed at the frontiers of the computational sciences and in the mathematics of nonlinear dynamics more readily available for research in the applied physical, biological, and social sciences.

All titles from the *Santa Fe Institute Studies in the Sciences of Complexity* series will carry this imprint, the SFI logo, which is based on a Mimbres pottery design (circa A.D. 950–1150), drawn by Betsy Jones.



Santa Fe Institute Editorial Board
September 1990

Dr. L. M. Simmons, Jr., *Chair*
Executive Vice President, Santa Fe Institute

Professor Kenneth J. Arrow
Department of Economics, Stanford University

Professor W. Brian Arthur
Dean & Virginia Morrison Professor of Population Studies and Economics, Food Research Institute, Stanford University

Dr. David K. Campbell
Director, Center for Nonlinear Studies, Los Alamos National Laboratory

Dr. George A. Cowan
President, Santa Fe Institute and Senior Fellow Emeritus, Los Alamos National Laboratory

Professor Marcus W. Feldman
Director, Institute for Population & Resource Studies, Stanford University

Professor Murray Gell-Mann
Division of Physics & Astronomy, California Institute of Technology

Professor John H. Holland
Division of Computer Science & Engineering, University of Michigan

Dr. Bela Julesz
Head, Visual Perception Research, AT& T Bell Laboratories

Professor Stuart Kauffman
School of Medicine, University of Pennsylvania

Dr. Alan S. Perelson
Theoretical Division, Los Alamos National Laboratory

Professor David Pines
Department of Physics, University of Illinois

Santa Fe Institute Studies in the Sciences of Complexity

PROCEEDINGS VOLUMES

Volume	Editor	Title
I	David Pines	Emerging Syntheses in Science, 1987
II	Alan S. Perelson	Theoretical Immunology, Part One, 1988
III	Alan S. Perelson	Theoretical Immunology, Part Two, 1988
IV	Gary D. Doolen et al.	Lattice Gas Methods of Partial Differential Equations, 1989
V	Philip W. Anderson et al.	The Economy as an Evolving Complex System, 1988
VI	Christopher G. Langton	Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, 1988
VII	George I. Bell & Thomas G. Marr	Computers and DNA, 1989
VIII	Wojciech H. Zurek	Complexity, Entropy, and the Physics of Information, 1990
IX	Alan S. Perelson & Stuart A. Kauffman	Molecular Evolution on Rugged Landscapes: Proteins, RNA and the Immune System, 1990

LECTURES VOLUMES

Volume	Editor	Title
I	Daniel L. Stein	Lectures in the Sciences of Complexity, 1989
II	Erica Jen	1989 Lectures in Complex Systems, 1990

LECTURE NOTES VOLUMES

Volume	Author	Title
I	John Hertz, Anders Krogh, & Richard Palmer	Introduction to the Theory of Neural Computation, 1991
II	Gérard Weisbuch	Complex Systems Dynamics, 1991



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

	Series Foreword by L. M. Simmons, Jr.	xiii
	Foreword by Jack Cowan	xv
	Foreword by Christof Koch	xvii
	Preface	xix
ONE	Introduction	1
1.1	Inspiration from Neuroscience	2
1.2	History	6
1.3	The Issues	8
TWO	The Hopfield Model	11
2.1	The Associative Memory Problem	11
2.2	The Model	13
2.3	Statistical Mechanics of Magnetic Systems	25
2.4	Stochastic Networks	32
2.5	Capacity of the Stochastic Network	35
THREE	Extensions of the Hopfield Model	43
3.1	Variations on the Hopfield Model	43
3.2	Correlated Patterns	49
3.3	Continuous-Valued Units	53
3.4	Hardware Implementations	58
3.5	Temporal Sequences of Patterns	63
FOUR	Optimization Problems	71
4.1	The Weighted Matching Problem	72
4.2	The Travelling Salesman Problem	76
4.3	Graph Bipartitioning	79
4.4	Optimization Problems in Image Processing	81

FIVE	Simple Perceptrons	89
5.1	Feed-Forward Networks	90
5.2	Threshold Units	92
5.3	Proof of Convergence of the Perceptron Learning Rule	100
5.4	Linear Units	102
5.5	Nonlinear Units	107
5.6	Stochastic Units	110
5.7	Capacity of the Simple Perceptron	111
SIX	Multi-Layer Networks	115
6.1	Back-Propagation	115
6.2	Variations on Back-Propagation	120
6.3	Examples and Applications	130
6.4	Performance of Multi-Layer Feed-Forward Networks	141
6.5	A Theoretical Framework for Generalization	147
6.6	Optimal Network Architectures	156
SEVEN	Recurrent Networks	163
7.1	Boltzmann Machines	163
7.2	Recurrent Back-Propagation	172
7.3	Learning Time Sequences	176
7.4	Reinforcement Learning	188
EIGHT	Unsupervised Hebbian Learning	197
8.1	Unsupervised Learning	197
8.2	One Linear Unit	199
8.3	Principal Component Analysis	204
8.4	Self-Organizing Feature Extraction	210
NINE	Unsupervised Competitive Learning	217
9.1	Simple Competitive Learning	218
9.2	Examples and Applications of Competitive Learning	223
9.3	Adaptive Resonance Theory	228
9.4	Feature Mapping	232
9.5	Theory of Feature Mapping	240
9.6	The Travelling Salesman Problem	244
9.7	Hybrid Learning Schemes	246
TEN	Formal Statistical Mechanics of Neural Networks	251
10.1	The Hopfield Model	251
10.2	Gardner Theory of the Connections	265

APPENDIX	Statistical Mechanics	275
A.1	The Boltzmann-Gibbs Distribution	275
A.2	Free Energy and Entropy	277
A.3	Stochastic Dynamics	279
	Bibliography	281
	Subject Index	307
	Author Index	321



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Series Foreword

We are witnessing the creation of new sciences of complexity, sciences that may well occupy the center of intellectual life in the twenty-first century. The Santa Fe Institute was founded to assist at the birth of these new sciences. Those involved in this activity are proceeding under the conviction that there is a common set of principles shared by the disparate complex systems under study, that the time is ripe to understand those principles, and that it is essential to develop them and the associated tools for dealing in a systematic way with complex systems.

Complex systems typically do not fit within the confines of one of the traditional disciplines, but require for their successful study knowledge and techniques from several disciplines. Thus one task of the Institute has been to find new ways to encourage cooperative research among scholars from different fields. The Studies in the Sciences of Complexity is one means that the Institute has adopted for accelerating the development of the sciences of complexity. These volumes make available to the scholarly community the results of conferences and workshops sponsored by the Institute, lectures presented in the Complex Systems Summer School, other lecture notes, and monographs by active researchers.

The sciences of complexity are emerging in part as a synthesis of some of the traditional sciences, including biology, computer science, physics, and mathematics. In part they are emerging as a result of new ideas, new questions, and new techniques only recently developed. Among these latter are the emergence of heretofore undreamed of computer power on the scientist's desktop and the not unrelated progress in nonlinear dynamics, computer graphics, and adaptive programs. These newly emerging tools and techniques also offer the prospect of new collaboration between the traditional sciences and the social sciences, a collaboration that will

extend modeling techniques to incorporate realistic detailed models of human behavior. Thus, this Series is intended to range broadly across many fields of intellectual endeavor incorporating work in all the areas listed above. The apparently disparate topics, however, share common themes that relate them to the emergent sciences of complexity.

The Santa Fe Institute, and hence this Series, would not exist without the support of farsighted individuals in government funding agencies and private foundations who have recognized the promise of the new approaches to complex systems research being fostered here. It is a pleasure to acknowledge the broad research grants received by the Institute from the Department of Energy, the John D. and Catherine T. MacArthur Foundation, and the National Science Foundation that, together with numerous other grants, have made possible the work of the Institute.

L. M. Simmons, Jr.

*Santa Fe, New Mexico
October 1, 1990*

Foreword

The past decade has seen an explosive growth in studies of neural networks. In part this was the result of technological advances in personal and main-frame computing, enabling neural network investigators to simulate and test ideas in ways not readily available before 1980. Another major impulse was provided by Hopfield's work on neural networks with symmetric connections. Such networks had previously been dismissed as not brain-like and therefore not worth studying. I myself fell into this trap some twenty-five years ago when I formulated what are now termed the standard equations for studying neural networks, those using the so-called squashing or logistic function. It was to Hopfield's credit that he "stepped back from biological reality" as Toulouse has put it, and uncovered an interesting set of properties and uses for symmetric networks. What followed is an interesting episode in the sociology of science. Hopfield's papers triggered an explosion, particularly in the statistical physics community, leading to a whole series of dramatic advances in the understanding of symmetric networks and their properties, especially in respect of their utility as distributed memory stores, and as solvers of constrained optimization problems, e.g., small versions of the famous Traveling Salesman Problem.

At more-or-less the same time, other developments in neural networks, possibly even more important, were taking place, culminating in the publication by Rumelhart, Hinton, and Williams of the now well-known "Back-Propagation Algorithm" for solving the fundamental problem of training neural networks to compute desired functions, a problem first formulated by Rosenblatt in the late 1950's in his now classical work on Perceptrons. Again this paper triggered a massive explosion of work on trainable neural networks which continues to this day.

The authors of this book, Palmer, Krogh, and Hertz, are statistical physicists who have experienced these developments. They have sought to provide an introduction to the theory behind all the hoopla, and to summarize the current state

of the subject. They have, wisely, eschewed neurobiology from their coverage, and have concentrated on what they know best, statistical mechanics, and how it is applied to neural networks. In my opinion they have succeeded admirably in providing a clear and readable account of the statistical mechanical ideas underlying neural networks, including some account of the analogy between neural networks and spinglasses, and the famous Replica Method for analyzing such materials. They have also done justice to Back-Propagation in providing an up-to-date treatment of Recurrent Back-Propagation in its various manifestations. Readers who take the trouble to follow the mathematics outlined in this book will be rewarded with valuable insights into how neural networks really work. One cannot ask for much more in any scientific publication.

Jack Cowan

*Mathematics Department
The University of Chicago
September 24, 1990*

Foreword

It is quite clear, as convincingly illustrated in this textbook, that the theory underlying learning and computing in networks of linear threshold units has developed into a mature subfield existing somewhere between physics, computer science, and neurobiology. We have not only a growing number of examples where learning techniques are successfully applied to practical problems such as recognizing handwritten postal mail codes or protein structures or cases where theories of unsupervised Hebbian learning mimicking certain aspects of neuronal development but now possess a solid understanding of why these algorithms perform so well on certain types of processing or why they fail, why certain features—such as hidden units—are necessary and how these approaches to learning relate to more traditional methods used in statistics to estimate a poorly sampled or unknown function in the presence of noise. Thus, it appears that neural networks are here to stay after three consecutive cycles of enthusiasm and skepticism, first peaking in the 1940's with McCulloch and Pitt's seminar work, then again in the 1960's with Rosenblatt's perceptron convergence theorem and its denouement by Minsky and Papert, and finally for a third time in the 1980's with Hopfield's energy approach and the modern era of multilayered networks ushered in by the backpropagation learning technique. The influence of the neural network learning paradigm on Artificial Intelligence will be profound, so much so that we will need to modify our basic notion of the Turing test as an operational definition of intelligence to encompass at least some rudimentary learning abilities. At this point in time, it is still too early to describe the long-term effect of neural networks on neurobiology and experimental neuroscience. While neural network analysis has been one of the key impulses behind "computational neuroscience," we have yet to develop specific instances where such an analysis has been used to successfully analyze and understand some real neurobiological circuits.

This monograph succinctly captures these trends and summarizes the current state of the art by way of highlighting the analogies to statistical mechanics and electric circuit theory as well as by discussing various practical applications. It is done without overdue emphasis on a formal mathematical treatment, appealing rather to the intuition of the reader. Throughout the book, the emphasis is on those features of neural networks relevant to information processing, storage and recall, that is to computation and function, linking physics to computing machines.

The Computation and Neural Systems Series—Over the past 600 million years, biology has solved the problem of processing massive amounts of noisy and highly redundant information in a constantly changing environment by evolving networks of billions of highly interconnected nerve cells. It is the task of scientists—be they mathematicians, physicists, biologists, psychologists, or computer scientists—to understand the principles underlying information processing in these complex structures. At the same time, researchers in machine vision, pattern recognition, speech understanding, robotics, and other areas of artificial intelligence can profit from understanding features of existing nervous systems. Thus, a new field is emerging: the study of how computations can be carried out in extensive networks of heavily interconnected processing elements, whether these networks are carbon- or silicon-based. Addison-Wesley's new “Computation and Neural Systems” series will reflect the diversity of this field with textbooks, course materials, and monographs on topics ranging from the biophysical modeling of dendrites and neurons, to computational theories of vision and motor control, to the implementation of neural networks using VLSI or optics technology, to the study of highly parallel computational architectures.

Christof Koch

Pasadena, California
September 21, 1990

Preface

We generally like our titles shorter than an *Introduction to the Theory of Neural Computation*, but all those words are important in understanding our purpose:

Neural Computation

Our subject matter is computation by artificial neural networks. The adjective “neural” is used because much of the inspiration for such networks comes from neuroscience, *not* because we are concerned with networks of real neurons. Brain modelling is a different field and, though we sometimes describe biological analogies, our prime concern is with what the *artificial* networks can do, and why. It is arguable that “neural” should be purged from the vocabulary of this field—perhaps *Network Computation* would have been more accurate in our title—but at present it is firmly ensconced. We do however avoid most other biological terms in non-biological contexts, including “neuron” (*unit*) and “synapse” (*connection*).

Theory

We emphasize the *theoretical* aspects of neural computation. Thus we provide little or no coverage of applications in engineering or computer science; implementations in hardware or software; or implications for cognitive science or artificial intelligence. There are recent books on all these topics and we prefer to complement rather than to compete. On the other hand, we feel that even those whose interest in the subject is completely practical may benefit from a broad theoretical perspective. We are no doubt biased by the fact that we are theorists by trade, but in our own experience we found this background to be essential in using neural networks for practical applications (not described in this book).

Introduction

Our book is intended as an introduction. This has implications at both extremes: where we start from, and how far we go. We try to start from the beginning, and assume little of the reader beyond some mathematical training. We do not assume any prior knowledge of neural networks, or of physics, engineering, or computer science. There are local exceptions to this ideal, but nothing that is central.

On the other hand we do *not* go to the end. The theory built up around neural networks is huge, and we cannot hope to cover it all. We do discuss most of the major architectures and theoretical concepts, but at varying depth. We stop short of very intricate or technical analysis, and of most directions that we consider dead ends. We aim not at mathematical rigor, but at conveying understanding through mathematics. Understanding should, we feel, consist not only of “knowing what,” but also of “knowing how”; especially *knowing how to go on* [Wittgenstein, 1958]. With that in mind we are usually not satisfied with simply stating or deducing a given result, but instead try to show the reader how to think about it, how to handle and hold it.

Bibliography and Coverage

At the same time we try to provide access to the research literature for further reading. We selected the bibliography with the primary aim of assisting the reader, and only with the secondary aim of attributing credit. When there was a choice we picked the more readable source, especially in references to the associated areas of mathematics, computer science, and physics.

It may be tempting to consider this book as a comprehensive survey of what has been done in the theory of neural computation. It does have that character in some sections, particularly those near the forefront, where we try to describe just *who* has done *what* recently. But that is not our overall aim and we have *not* tried to be complete. Omissions may be for reasons of ignorance, complexity, irrelevance, obscurity, pedagogy, space, or many other reasons. We apologize only for our ignorance.

Approach

Our selection and treatment of material reflects our background as physicists. This background has helped us to understand how these complex systems function, often in terms of physical analogies. Others might find easier paths into the subject area from computer science, statistics, or psychology, and there could be written equally good or better books along these lines. We tell the story the way that we are best able to understand it, and hope that our readers find the perspective enlightening.

We often view the analysis of artificial neural networks as a statistical mechanics problem. Like many systems in condensed matter physics, these networks are large collections of interacting entities with emergent properties, and it should not be

surprising that related mathematical tools are useful. Nevertheless, this is a new feature in the engineering world, where one normally expects every minute aspect of the operation of machinery (or software) to follow an explicit plan or program. But systems are becoming so complex that this kind of detailed engineering is neither possible nor desirable. Some of the networks described in this book illustrate how systems can design themselves with relatively little external guidance. The user will never know all the internal details, but will need methods like those we describe to analyze how the whole thing works.

In our experience some people, even among those working on neural networks, appear to be frightened by the prospect of having to learn statistical mechanics. They also question its necessity. Often, however, their equations turn out to be exactly the same ones that we would have written. Anyone trying to analyze the typical behavior of a many-component system is doing statistical mechanics whether it is called that or not. We hope that the doubters will not be put off by a few partition functions here and there, but will benefit by seeing many problems put in a more unified perspective.

This is not to say that we employ *only* a statistical mechanics or “physics” approach, or that one has to be a physicist to read this book. Far from it. Explicit statistical mechanics is used only rarely, and is explained where it arises (and in the Appendix). It often underlies and motivates our approach, but is not often visible at the surface. And one certainly does not need to be a physicist; we have tried hard not to assume anything (besides mathematics) of the reader, and to avoid or explain words and ideas specific to physics.

Prerequisites

There are no prerequisites besides mathematics. The mathematical level varies somewhat, with more required for the starred (*) sections and for Chapter 10 than for the rest. These sections may be omitted however without loss of continuity. On the whole we expect our readers to know something about multi-variate calculus, ordinary differential equations, basic probability and statistics, and linear algebra. But in most of these areas a general familiarity should be enough. The exception for some readers may be linear algebra, where we use vectors and matrices, inner products, matrix inversion, eigenvalues and eigenvectors, and diagonalization of quadratic forms. However, eigenvalues and eigenvectors are not used in any essential way until Chapter 8, and diagonalization appears only in starred sections. Kohonen [1989] provides an introduction to the linear algebra needed for neural network theory.

Acknowledgments

This book began as lecture notes for a half-semester course given at Duke University (and broadcast on the North Carolina teleclassroom network) in the spring of 1988. The audience for these lectures was very broad, including *néophytes* from neurosciences

and cognitive sciences as well as computer science, engineering, and physics. Later versions were used as the basis of summer school lectures at Santa Fe in June 1988 [Palmer, 1989], and for a one-semester course for physics and computer science students at the University of Copenhagen in the fall of 1989. We thank all of the students in all of these courses for constructive feedback that led to successive improvements.

We owe a debt of gratitude to many of our colleagues in Durham, Copenhagen, and Santa Fe who encouraged, supported and helped us in this work. These include Ajay, Alan, Benny, Corinna, Dave, Ingi, David, Frank, Gevne, Jack, John, Jun, Kurt, Lars, Marjorie, Mike, Per, Ronda, Søren, Stu, Tamas, and Xiang. Two of us (JH and AK) also thank the Physics Department at Duke for their hospitality in the spring of 1988, when this whole enterprise got started. AK thanks the Carlsberg Foundation for generous financial support.

Finally, we reserve our deepest appreciation for our wives and families. It is a hackneyed theme to thank loved ones for patience and understanding while a book was being written; but now we know why, and do give heartfelt thanks.

Richard Palmer
Anders Krogh
John Hertz

Durham and Copenhagen, August 1990

Electronic mail addresses for the authors: palmer@phy.duke.edu
askrogh@nbivax.nbi.dk
hertz@nordita.dk

Introduction

Anyone can see that the human brain is superior to a digital computer at many tasks. A good example is the processing of visual information: a one-year-old baby is much better and faster at recognizing objects, faces, and so on than even the most advanced AI system running on the fastest supercomputer.

The brain has many other features that would be desirable in artificial systems.

- It is robust and fault tolerant. Nerve cells in the brain die every day without affecting its performance significantly.
- It is flexible. It can easily adjust to a new environment by “learning”—it does not have to be programmed in Pascal, Fortran or C.
- It can deal with information that is fuzzy, probabilistic, noisy, or inconsistent.
- It is highly parallel.
- It is small, compact, and dissipates very little power.

Only in tasks based primarily on simple arithmetic does the computer outperform the brain!

This is the real motivation for studying neural computation. It is an alternative computational paradigm to the usual one (based on a programmed instruction sequence), which was introduced by von Neumann and has been used as the basis of almost all machine computation to date. It is inspired by knowledge from neuroscience, though it does not try to be biologically realistic in detail. It draws its methods in large degree from statistical physics, and that is why the lectures on which this book is based originally formed part of a physics course. Its potential applications lie of course mainly in computer science and engineering. In addition it may be of value as a modelling paradigm in neuroscience and in sensory and cognitive psychology.

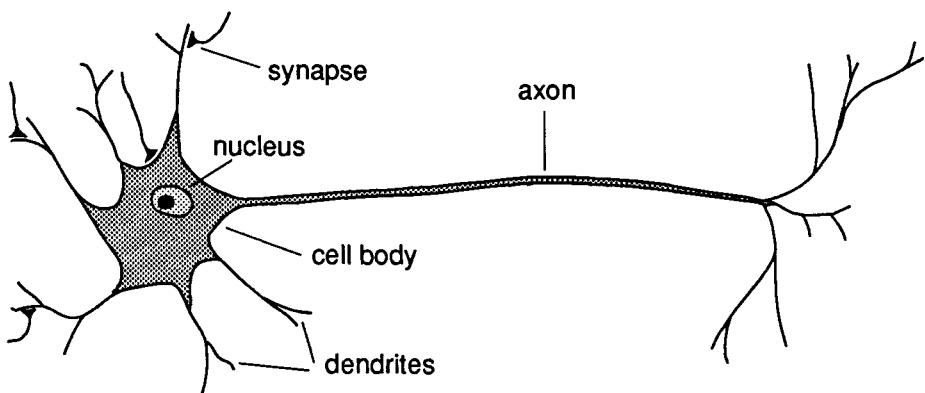


FIGURE 1.1 Schematic drawing of a typical neuron.

The field is also known as neural networks, neurocomputation, associative networks, collective computation, connectionism, and probably many other things. We will use all these terms freely.

1.1 Inspiration from Neuroscience

Today's research in neural computation is largely motivated by the possibility of making artificial computing networks. Yet, as the term "neural network" implies, it was originally aimed more towards modelling networks of real neurons in the brain. The models are extremely simplified when seen from a neurophysiological point of view, though we believe that they are still valuable for gaining insight into the principles of biological "computation." Just as most of the details of the separate parts of a large ship are unimportant in understanding the behavior of the ship (e.g., that it floats, or transports cargo), so many details of single nerve cells may be unimportant in understanding the *collective* behavior of a network of cells.

Neurons

The brain is composed of about 10^{11} neurons (nerve cells) of many different types. Figure 1.1 is a schematic drawing of a single neuron. Tree-like networks of nerve fiber called **dendrites** are connected to the **cell body** or soma, where the cell nucleus is located. Extending from the cell body is a single long fiber called the **axon**, which eventually branches or **arborizes** into strands and substrands. At the ends of these are the transmitting ends of the **synaptic junctions**, or **synapses**, to other neurons. The receiving ends of these junctions on other cells can be found

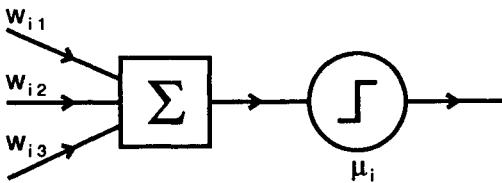


FIGURE 1.2 Schematic diagram of a McCulloch-Pitts neuron. The unit fires if the weighted sum $\sum_j w_{ij} n_j$ of the inputs reaches or exceeds the threshold μ_i .

both on the dendrites and on the cell bodies themselves. The axon of a typical neuron makes a few thousand synapses with other neurons.

The transmission of a signal from one cell to another at a synapse is a complex chemical process in which specific transmitter substances are released from the sending side of the junction. The effect is to raise or lower the electrical potential inside the body of the receiving cell. If this potential reaches a threshold, a pulse or **action potential** of fixed strength and duration is sent down the axon. We then say that the cell has “fired”. The pulse branches out through the axonal arborization to synaptic junctions to other cells. After firing, the cell has to wait for a time called the **refractory period** before it can fire again.

McCulloch and Pitts [1943] proposed a simple model of a neuron as a binary threshold unit. Specifically, the model neuron computes a weighted sum of its inputs from other units, and outputs a one or a zero according to whether this sum is above or below a certain threshold:

$$n_i(t+1) = \Theta\left(\sum_j w_{ij} n_j(t) - \mu_i\right). \quad (1.1)$$

See Fig. 1.2. Here n_i is either 1 or 0, and represents the state of neuron i as **firing** or **not firing** respectively. Time t is taken as discrete, with one time unit elapsing per processing step. $\Theta(x)$ is the **unit step function**, or Heaviside function:

$$\Theta(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{otherwise.} \end{cases} \quad (1.2)$$

The weight w_{ij} represents the strength of the synapse connecting neuron j to neuron i . It can be positive or negative corresponding to an **excitatory** or **inhibitory** synapse respectively. It is zero if there is no synapse between i and j . The cell-specific parameter μ_i is the threshold value for unit i ; the weighted sum of inputs must reach or exceed the threshold for the neuron to fire.

Though simple, a McCulloch-Pitts neuron is computationally a powerful device. McCulloch and Pitts proved that a synchronous assembly of such neurons is capable in principle of **universal computation** for suitably chosen weights w_{ij} . This means that it can perform any computation that an ordinary digital computer can, though not necessarily so rapidly or conveniently.

Real neurons involve many complications omitted from this simple description. The most significant ones include:

- Real neurons are often not even approximately threshold devices as described above. Instead they respond to their input in a continuous way. This is sometimes referred to as a **graded response**. But the nonlinear relationship between the input and the output of a cell is a universal feature. Our working hypothesis is that it is the nonlinearity that is essential, not its specific form. In any case, continuous-valued units can be modelled too, and are sometimes more convenient to deal with than threshold units.
- Many real cells also perform a *nonlinear* summation of their inputs, which takes us a bit further from the McCulloch-Pitts picture. There can even be significant logical processing (e.g., AND, OR, NOT) within the dendritic tree. This can in principle be taken care of by using several formal McCulloch-Pitts neurons to represent a single real one, though there has been little work along these lines so far. We will generally ignore this complication, since the simple McCulloch-Pitts picture is already very rich and interesting to study.
- A real neuron produces a sequence of pulses, not a simple output level. Representing the firing rate by a single number like n_i , even if continuous, ignores much information—such as pulse phase—that might be carried by such a pulse sequence. The majority of experts do not think that phase plays a significant role in most neuronal circuits, but agreement is incomplete.
- Neurons do not all have the same fixed delay ($t \rightarrow t + 1$). Nor are they updated synchronously by a central clock. We will in fact use asynchronous updating in much of this book.
- The amount of transmitter substance released at a synapse may vary unpredictably. This sort of effect can be modelled, at least crudely, by a stochastic generalization of the McCulloch-Pitts dynamics.

A simple generalization of the McCulloch-Pitts equation (1.1) which includes some of these features is

$$n_i := g\left(\sum_j w_{ij} n_j - \mu_i\right). \quad (1.3)$$

The number n_i is now continuous-valued and is called the **state** or **activation** of unit i . The threshold function $\Theta(x)$ of (1.1) has been replaced by a more general nonlinear function $g(x)$ called the **activation function**, **gain function**, **transfer function**, or **squashing function**. Rather than writing the time t or $t+1$ explicitly as we did in (1.1), we now simply give a rule for updating n_i whenever that occurs.¹ Units are often updated *asynchronously*: in random order at random times.

Nowhere in this book do we attempt a detailed description of networks of real neurons, or of other neurobiological structures or phenomena. Kandel and Schwartz [1985] give an excellent introduction. We do sometimes appeal to biological realism, and do describe a few models of cortical organization in Chapters 8 and 9, but the emphasis is generally on the computational abilities of network models, not on their

¹Note that we use the symbol “ $:=$ ” to emphasize that the right-hand side is assigned to the left-hand side upon update; the equality is not continuously true.

direct applicability to brain modelling. Nevertheless, despite the intimidating detail and complexity of real brains, we do believe that the kind of theory discussed in this book is relevant to neuroscience. But the connection is not so much at the level of detailed modelling as at the level of *algorithms and representation* [Marr, 1982]. That is, this kind of approach can help in formulating and testing what sort of computational algorithms the brain is using in different tasks. While the biological and artificial implementations of the algorithms are very different, there can be many features in common at the algorithmic level.

When discussing artificial neural networks it remains commonplace to talk of “neurons” and “synapses”, even though the network components are far simpler than their biological counterparts. We prefer to use the terms “units” and “connections” (or “weights”) except when discussing networks that are intended as direct models of brain structures. Other terms in use for the units include “processing elements” and “neurodes”.

Parallel Processing

In computer science terms, we can describe the brain as a parallel system of about 10^{11} processors. Using the simplified model (1.3) above, each processor has a very simple program: it computes a weighted sum of the input data from other processors and then outputs a single number, a nonlinear function of this weighted sum. This output is then sent to other processors, which are continually doing the same kind of calculation. They are using different weights and possibly different gain functions; the coefficients w_{ij} are in general different for different i , and we could also make $g(x)$ be site-dependent. These weights and gain functions can be thought of as local data stored by the processors.

The high connectivity of the network (i.e., the fact that there are many terms in the sum in (1.3)), means that errors in a few terms will probably be inconsequential. This tells us that such a system can be expected to be robust and its performance will degrade gracefully in the presence of noise or errors. In the brain itself cells die all the time without affecting the function, and this robustness of the biological neural networks has probably been essential to the evolution of intelligence.

The contrast between this kind of processing and the conventional von Neumann kind could not be stronger. Here we have very many processors, each executing a very simple program, instead of the conventional situation where one or at most a few processors execute very complicated programs. And in contrast to the robustness of a neural network, an ordinary sequential computation may easily be ruined by a single bit error.

It is worth remarking that the typical cycle time of neurons is a few milliseconds, which is about a million times slower than their silicon counterparts, semiconductor gates. Nevertheless, the brain can do very fast processing for tasks like vision, motor control, and decisions on the basis of incomplete and noisy data, tasks that are far beyond the capacity of a Cray supercomputer. This is obviously possible only because billions of neurons operate simultaneously. Imagine the capabilities of a

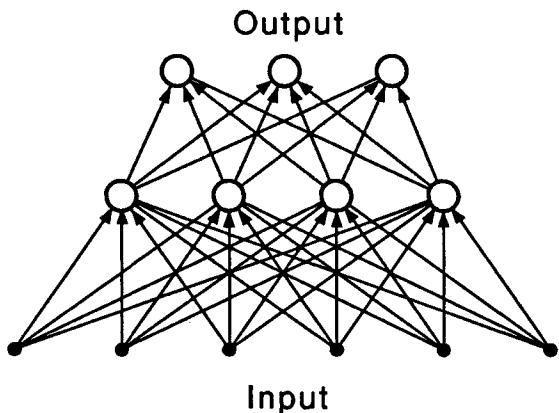


FIGURE 1.3 A two-layer perceptron.

system which could operate in parallel like this but with switching times of current semiconductor devices!

1.2 History

The history of these sorts of ideas in psychology originates with Aristotle. Yet as a basis for computational or neural modelling we can trace them to the paper of McCulloch and Pitts [1943], which introduced the model described above.

During the next fifteen years there was considerable work on the detailed logic of threshold networks. They were realized to be capable of universal computation and were analyzed as finite-state machines; see Minsky [1967]. The problem of making a reliable network with unreliable parts was solved by the use of redundancy [von Neumann, 1956], leading later to *distributed* redundant representations [Winograd and Cowan, 1963].

At the opposite extreme to detailed logic, continuum theories were also developed. Known as **neurodynamics** or **neural field theory**, this approach used differential equations to describe activity patterns in bulk neural matter [Rashevsky, 1938; Wiener, 1948; Beurle, 1956; Wilson and Cowan, 1973; Amari, 1977].

Around 1960 there was a wave of activity centered around the group of Frank Rosenblatt, focusing on the problem of how to find appropriate weights w_{ij} for particular computational tasks. They concentrated on networks called **perceptrons**, in which the units were organized into layers with feed-forward connections between one layer and the next. An example is shown in Fig. 1.3. Very similar networks called **adalines** were invented around the same time by Widrow and Hoff [1960; Widrow, 1962].

For the simplest class of perceptrons without any intermediate layers, Rosenblatt [1962] was able to prove the convergence of a **learning algorithm**, a way to change the weights iteratively so that a desired computation was performed. Many

people expressed a great deal of enthusiasm and hope that such machines could be a basis for artificial intelligence.

There was however a catch to the learning theorem, forcefully pointed out by Minsky and Papert [1969] in their book *Perceptrons*: the theorem obviously applies only to those problems which the structure is *capable of computing*. Minsky and Papert showed that some rather elementary computations could *not* be done by Rosenblatt's one-layer² perceptron. The simplest example is the **exclusive or** (XOR) problem: a single output unit is required to turn on ($n = +1$) if one or the other of two input lines is on, but not when neither or both inputs are on.

Rosenblatt had also studied structures with more layers of units and believed that they could overcome the limitations of the simple perceptrons. However, there was no learning algorithm known which could determine the weights necessary to implement a given calculation. Minsky and Papert doubted that one could be found and thought it more profitable to explore other approaches to artificial intelligence. With this most of the computer science community left the neural network paradigm for almost 20 years.

Still, there were a number of people who continued to develop neural network theory in the 1970's. A major theme was **associative content-addressable memory**, in which different input patterns become associated with one another (i.e., trigger the same response) if sufficiently similar. These had actually been proposed much earlier [Taylor, 1956; Steinbuch, 1961], and were later revived or rediscovered by Anderson [1968, 1970; Anderson and Mozer, 1981], Willshaw et al. [1969], Marr [1969, 1971] and Kohonen [1974–1989]. Grossberg [1967–1987] made a comprehensive reformulation of the general problem of learning in networks. Marr [1969, 1970, 1971] developed network theories of the cerebellum, cerebral neocortex, and hippocampus, assigning specific functions to each type of neuron. A number of people, including Marr [1982], von der Malsburg [1973], and Cooper [1973; Nass and Cooper, 1975], studied the development and functioning of the visual system.

Another thread of development can be traced to Cragg and Temperley [1954, 1955]. They reformulated the McCulloch-Pitts network as a spin (magnetic) system of the sort familiar in physics. Memory was believed to reside in the hysteresis of the domain patterns expected for such a system. Caianiello [1961] then constructed a **statistical** theory, using ideas from statistical mechanics, and incorporated learning in a way which drew on the ideas of Hebb [1949] about learning in the brain. The same theme was taken up in the 1970's by Little [1974; Little and Shaw, 1975, 1978] and again in 1981 by Hopfield [1982]. Hopfield was able to add some helpful physical insight by introducing an **energy function**, and by emphasizing the notion of memories as dynamically stable attractors. Hinton and Sejnowski [1983, 1986] and Peretto [1984] constructed formulations using **stochastic units** which follow the dynamics (1.1) or (1.3) only approximately, making "mistakes" with a certain probability analogous to temperature in statistical mechanics. The real power of

²We never count input lines as units in numbering layers. Figure 1.3 is thus a two-layer network. Until recently it would often have been called a three-layer network, but the convention is changing.

statistical mechanics was then brought to bear on the stochastic network problem by Amit et al. [1985a, b; Amit, 1989], using methods developed in the theory of random magnetic systems called **spin glasses**.

Perhaps the most influential development in this decade, however, takes up the old thread of Rosenblatt's perceptrons where it was cut 20 years ago. Various people have developed an algorithm which works quite well for adjusting the weights connecting units in successive layers of multi-layer perceptrons. Known as **back-propagation**, it appears to have been found first by Werbos [1974] in the mid-70's, and then independently rediscovered around 1985 by Rumelhart, Hinton, and Williams [1986a, b], and by Parker [1985]. Le Cun [1985] also proposed a related algorithm. Though not yet the holy grail of a completely general algorithm able to teach an arbitrary computational task to a network, it can solve many problems (such as XOR) which the simple one-layer perceptrons could not. Much current activity is centered on back-propagation and its extensions.

Many of the important early papers have been collected in Anderson and Rosenfeld [1988], including many of those mentioned here. This is an excellent collection for those interested in the history of neural networks. We also recommend the review article by Cowan and Sharp [1988a, b], which we drew on for this section.

1.3 The Issues

Massive parallelism in computational networks is extremely attractive in principle. But in practice there are many issues to be decided before a successful implementation can be achieved for a given problem:

- What is the best architecture? Should the units be divided into layers, or not? How many connections should be made between units, and how should they be organized? What sort of activation functions $g(x)$ should be used? What type of updating should be used: synchronous or asynchronous, deterministic or stochastic? How many units are needed for a given task?
- How can a network be programmed? Can it learn a task or must it be pre-designed? If it can learn a task, how many examples are needed for good performance? How many times must it go through the examples? Does it need the right answers during training, or can it learn from correct/incorrect reinforcement? Can it learn in real-time while functioning, or must the training phase be separated from the performance phase?
- What can the various types of network do? How many different tasks can they learn? How well? How fast? How robust are they to missing information, incorrect data, and unit removal or malfunction? Can they *generalize* from known tasks or examples to unknown ones? What classes of input-to-output functions can they represent?

- How can a network be built in hardware? What are the advantages and disadvantages of different hardware implementations, and how do they compare to simulation in software?

These questions are obviously coupled and cannot be answered independently. The architecture, for instance, strongly influences what the network can do, and what hardware options are available.

Much of this book will be concerned with refining and answering the above questions. However we will generally approach them from a theoretical point of view, rather than from a design one. That is, we will attempt to understand the behavior of networks as a function of their architecture, and only rarely raise the question of designing networks to fulfill particular goals. But of course the two viewpoints are not independent, and a strong understanding of principles is invaluable for good design.

Three of the issues raised above deserve a little more comment here, as general background before we become involved in details.

Hardware

Almost everything in the field of neural computation has been done by simulating the networks on serial computers, or by theoretical analysis. Neural network VLSI chips are far behind the models, as is natural at this point. The main problem with making neural network chips is that one needs a *lot* of connections, often some fraction of the square of the number of units. The space taken up by the connections is usually the limiting factor for the size of a network. The neural chips made so far contain of the order of 100 units, which is too few for most practical applications.

Potential alternatives to integrated circuit chips include optical computers. The field is very young, but electro-optical and optical associative memories have already been proposed or built.

Efficient hardware is crucially important in the long term if we are going to take full advantage of the capabilities of neural networks, and there is growing activity in this area. However, it is largely beyond the scope of this book; we return to hardware issues only briefly in Section 3.4.

Generalization

The reason for much of the excitement about neural networks is their ability to generalize to new situations. After being trained on a number of examples of a relationship, they can often induce a complete relationship that interpolates and extrapolates from the examples in a sensible way. But what is meant by *sensible* generalization is often not clear. In many problems there are almost infinitely many possible generalizations. How does a neural network—or a human for that matter—choose the “right” one? As an example one could train a neural network on three of the four XOR relations mentioned earlier, and it would be very unlikely that any of

the known types of networks would actually generalize to the full XOR. Nevertheless neural networks commonly make very useful generalizations that would be judged sensible in human terms.

Programming

Like most of the work done in neural networks, much of this book is concerned with the problem of programming or learning: how do we choose the connection weights so the network can do a specific task?

We will encounter some examples where we can choose the weights *a priori* if we are a little clever. This embeds some information into the network by design. But such problems are the exception rather than the rule. In other cases we can often “teach” the network to perform the desired computation by iterative adjustments of the w_{ij} strengths. This may be done in two main ways:

- **Supervised learning.** Here the learning is done on the basis of direct comparison of the output of the network with known correct answers. This is sometimes called **learning with a teacher**. It includes the special case of **reinforcement learning**, where the only feedback is whether each output is correct or incorrect, not what the correct answer is.
- **Unsupervised learning.** Sometimes the learning goal is not defined at all in terms of specific correct examples. The only available information is in the correlations of the input data or signals. The network is expected to create categories from these correlations, and to produce output signals corresponding to the input category.

There are many exciting implications of the possibility of *training* a network to do a computation. Instead of having to specify every detail of a calculation, we simply have to compile a training set of representative examples. This means that we can hope to treat problems where appropriate rules are very hard to know in advance, as in expert systems and robotics. It may also spare us a lot of tedious (and expensive) software design and programming even when we do have explicit rules. John Denker has remarked that “neural networks are the second best way of doing just about anything.” The *best* way is to find and use the right rules or the optimum algorithm for each particular problem, but this can be inordinately expensive and time consuming. There is plenty of scope for a second best approach based on learning by example.

References

- Abu-Mostafa, Y.S. (1989). The Vapnik-Chervonenkis Dimension: Information versus Complexity in Learning. *Neural Computation* 1, 312317.
- Abu-Mostafa, Y.S. and D. Psaltis (1987). Optical Neural Computers. *Scientific American* 256, March 1987, 8895.
- Ackley, D.H. , G.E. Hinton , and T.J. Sejnowski (1985). A Learning Algorithm for Boltzmann Machines. *Cognitive Science* 9, 147169. Reprinted in Anderson and Rosenfeld [1988].
- Ackley, D.H. and M.S. Littman (1990). Generalization and Scaling in Reinforcement Learning. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 550557. San Mateo: Morgan Kaufmann.
- Ahalt, S.C. , A.K. Krishnamurthy , P. Chen , and D.E. Melton (1990). Competitive Learning Algorithms for Vector Quantization. *Neural Networks* 3, 277290.
- Almeida, L.B. (1987). A Learning Rule for Asynchronous Perceptrons with Feedback in a Combinatorial Environment. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 609618. New York: IEEE.
- Almeida, L.B. (1988). Backpropagation in Perceptrons with Feedback. In *Neural Computers* (Neuss 1987), eds. R. Eckmiller and Ch. von der Malsburg , 199208. Berlin: Springer-Verlag.
- Alspector, J. and R.B. Allen (1987). A Neuromorphic VLSI Learning System. In *Advanced Research in VLSI: Proceedings of the 1987 Stanford Conference*, ed. P. Losleben , 313349. Cambridge: MIT Press.
- Alspector, J. , R.B. Allen , V. Hu , and S. Satyanarayana (1988). Stochastic Learning Networks and Their Electronic Implementation. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 921. New York: American Institute of Physics.
- 282 Amari, S.A. (1977). Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics* 27 , 7787.
- Amari, S.A. (1980). Topographic Organization of Nerve Fields. *Bulletin of Mathematical Biology* 42, 339364.
- Amit, D. (1988). Neural Networks for Counting Chimes. *Proceedings of the National Academy of Sciences, USA* 85, 21412145.
- Amit, D. (1989). *Modelling Brain Function*. Cambridge: Cambridge University Press.
- Amit, D. , H. Gutfreund , and H. Sompolinsky (1985a). Spin-Glass Models of Neural Networks. *Physical Review A* 32, 10071018.
- Amit, D. , H. Gutfreund , and H. Sompolinsky (1985b). Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks. *Physical Review Letters* 55, 15301533.
- Amit, D. , H. Gutfreund , and H. Sompolinsky (1987 a). Statistical Mechanics of Neural Networks Near Saturation. *Annals of Physics* 173, 3067.
- Amit, D. , H. Gutfreund , and H. Sompolinsky (1987 b). Information Storage in Neural Networks with Low Levels of Activity. *Physical Review A* 35, 22932303.
- Anderson, C.H. and D.C. Van Essen (1987). Shifter Circuits: A Computational Strategy for Dynamic Aspects of Visual Processing. *Proceedings of the National Academy of Sciences, USA* 84, 62976301.
- Anderson, D.Z. (1986). Coherent Optical Eigenstate Memory. *Optics Letters* 11, 5658.
- Anderson, J.A. (1968). A Memory Model Using Spatial Correlation Functions. *Kybernetik* 5, 113119.
- Anderson, J.A. (1970). Two Models for Memory Organization. *Mathematical Biosciences* 8, 137160.
- Anderson, J.A. and M.C. Mozer (1981). Categorization and Selective Neurons. In *Parallel Models of Associative Memory*, eds. G.E. Hinton and J.A. Anderson , 213236. Hillsdale: Erlbaum.
- Anderson, J.A. and E. Rosenfeld , eds. (1988). *Neurocomputing: Foundations of Research*. Cambridge: MIT Press.
- Anderson, S. , J.W.L. Merrill , and R. Port (1989). Dynamic Speech Categorization with Recurrent Networks. In *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 398406. San Mateo: Morgan Kaufmann.
- Angniol, B. , G. de La Croix Vaubois , and J.-Y. Le Texier (1988). Self-Organizing Feature Maps and the Travelling Salesman Problem. *Neural Networks* 1, 289293.
- Arbib, M.A. (1987). *Brains, Machines, and Mathematics*. Berlin: Springer-Verlag.
- Bachmann, C.M. , L.N. Cooper , A. Dembo , and O. Zeitouni (1987). A Relaxation Model for Memory with High Storage Density. *Proceedings of the National Academy of Sciences, USA* 84, 75297531.

- 283 Bachrach, J. (1988). Learning to Represent State. Masters Thesis, University of Massachusetts, Amherst.
- Baldi, P. and K. Hornik (1989). Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima. *Neural Networks* 2, 5358.
- Barhen, J. , S. Gulati , and M. Zak (1989). Neural Learning of Constrained Nonlinear Transformations. *Computer*, June 1989, 6776.
- Barlow, H.B. (1989). Unsupervised Learning. *Neural Computation* 1, 295311.
- Barto, A.G. and P. Anandan (1985). Pattern Recognizing Stochastic Learning Automata. *IEEE Transactions on Systems, Man, and Cybernetics* 15, 360375.
- Barto, A.G. and M.I. Jordan (1987). Gradient Following Without Back-Propagation in Layered Networks. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 629636. New York: IEEE.
- Barto, A.G. , R.S. Sutton , and C.W. Anderson (1983). Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems. *IEEE Transactions on Systems, Man, and Cybernetics* 13, 834846. Reprinted in Anderson and Rosenfeld [1988].
- Barto, A.G. , R.S. Sutton , and C.J.C.H. Watkins (1991). Learning and Sequential Decision Making. In *Learning and Computational Neuroscience*, eds. M. Gabriel and J.W. Moore . Cambridge: MIT Press.
- Baum, E.B. (1986). Towards Practical Neural Computation for Combinatorial Optimization Problems. In *Neural Networks for Computing* (Snowbird 1986), ed. J.S. Denker , 5358. New York: American Institute of Physics.
- Baum, E.B. and D. Haussler (1989). What Size Net Gives Valid Generalization? *Neural Computation* 1, 151160.
- Baum, E.B. and F. Wilczek (1988). Supervised Learning of Probability Distributions by Neural Networks. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 5261. New York: American Institute of Physics.
- Becker, S. and Y. Le Cun (1989). Improving the Convergence of Back-Propagation Learning with Second Order Methods. In *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 2937. San Mateo: Morgan Kaufmann.
- Bengio, Y. , R. Cardin , and R. De Mori (1990). Speaker Independent Speech Recognition with Neural Networks and Speech Knowledge. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 218225. San Mateo: Morgan Kaufmann.
- Beurle, R.L. (1956). Properties of a Mass of Cells Capable of Regenerating Pulses. *Philosophical Transactions of the Royal Society of London B* 240, 5594.
- Bienenstock, E.L. , L.N. Cooper , and P.W. Munro (1982). Theory for the Development of Neuron Selectivity: Orientation Specificity and Binocular Interaction in Visual Cortex. *Journal of Neuroscience* 2, 3248. Reprinted in Anderson and Rosenfeld [1988].
- 284 Bilbro, G. , R. Mann , T.K. Miller , W.E. Snyder , D.E. Van den Bout , and M. White (1989). Optimization by Mean Field Annealing. In *Advances in Neural Information Processing Systems I* (Denver 1988), ed. D.S. Touretzky , 9198. San Mateo: Morgan Kaufmann.
- Bilbro, G.L. and W. Snyder (1989). Range Image Restoration Using Mean Field Annealing. In *Advances in Neural Information Processing Systems I* (Denver 1988), ed. D.S. Touretzky , 594601. San Mateo: Morgan Kaufmann.
- Binder, K. and D.W. Heerman (1988). Monte Carlo Simulation in Statistical Mechanics. Berlin: Springer-Verlag.
- Binder, K. and A.P. Young (1986). Spin Glasses: Experimental Facts, Theoretical Concepts, and Open Questions. *Reviews of Modern Physics* 58, 801976.
- Blakemore, C. and G.F. Cooper (1970). Development of the Brain Depends on the Visual Environment. *Nature* 228, 477478.
- Block, H.D. (1962). The Perceptron: A Model for Brain Functioning. *Reviews of Modern Physics* 34, 123135. Reprinted in Anderson and Rosenfeld [1988].
- Blumer, A. , A. Ehrenfeucht , D. Haussler , and M. Warmuth (1986). Classifying Learnable Geometric Concepts with the Vapnik-Chervonenkis Dimension. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing* (Berkeley 1986), 273282. Salem: ACM.
- Bohr, H. , J. Bohr , S. Brunak , R.M.J. Cotterill , B. Lautrup , L. Nrskov , O.H. Olsen , and S.B. Petersen (1988). Protein Secondary Structure and Homology by Neural Networks: The α -Helices in Rhodopsin. *FEBS Letters* 241, 223228.
- Bourland, H. and Y. Kamp (1988). Auto-Association by Multilayer Perceptrons and Singular Value Decomposition. *Biological Cybernetics* 59, 291294.

- Brandt, R.D. , Y. Wang , A.J. Laub , and S.K. Mitra (1988). Alternative Networks for Solving the Travelling Salesman Problem and the List-Matching Problem. In IEEE International Conference on Neural Networks (San Diego 1988), vol. II, 333340. New York: IEEE.
- Brunak, S. and B. Lautrup (1989). Liniedeling med et Neuralt Netvirk. Skrifter for Anvendt Matematik og Lingvistik 14, 5574.
- Brunak, S. and B. Lautrup (1990). Neural Networks: Computers with Intuition. Singapore: World Scientific.
- Bryson, A.E. and Y.-C. Ho (1969). Applied Optimal Control. New York: Blais-dell.
- Buhmann, J. and K. Schulten (1987). Noise-Driven Temporal Association in Neural Networks. Europhysics Letters 4, 12051209.
- Buhmann, J. and K. Schulten (1988). Storing Sequences of Biased Patterns in Neural Networks with Stochastic Dynamics. In Neural Computers (Neuss 1987), eds. R. Eckmiller and Ch. von der Malsburg , 231242. Berlin: Springer-Verlag.
- Burr, D.J. (1988). An Improved Elastic Net Method for the Travelling Salesman Problem. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 6976. New York: IEEE.
- Caianiello, E.R. (1961). Outline of a Theory of Thought and Thinking Machines. Journal of Theoretical Biology 1, 204235.
- 285 Canning, A. and E. Gardner (1988). Partially Connected Models of Neural Networks. Journal of Physics A 21, 32753284.
- Carpenter, G.A. and S. Grossberg (1987a). A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine. Computer Vision, Graphics, and Image Processing 37, 54115.
- Carpenter, G.A. and S. Grossberg (1987b). ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns. Applied Optics 26, 49194930.
- Carpenter, G.A. and S. Grossberg (1988). The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network. Computer, March 1988, 7788.
- Casdagli, M. (1989). Nonlinear Prediction of Chaotic Time Series. Physica 35D, 335356.
- Cater, J.P. (1987). Successfully Using Peak Learning Rates of 10 (and Greater) in Back-Propagation Networks with the Heuristic Learning Algorithm. In IEEE First International Conference on Neural Networks (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 645651. New York: IEEE.
- Caudill, M. (1989). Neural Networks Primer. San Francisco: Miller Freeman.
- Chauvin, Y. (1989). A Back-Propagation Algorithm with Optimal Use of Hidden Units. In Advances in Neural Information Processing Systems I (Denver 1988), ed. D.S. Touretzky , 519526. San Mateo: Morgan Kaufmann.
- Cleeremans, A. , D. Servan-Schreiber , and J.L. McClelland (1989). Finite State Automata and Simple Recurrent Networks. Neural Computation 1, 372381.
- Cohen, M.A. and S. Grossberg (1983). Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks. IEEE Transactions on Systems, Man, and Cybernetics 13, 815826.
- Cohen, M.S. (1986). Design of a New Medium for Volume Holographic Information Processing. Applied Optics 25, 22282294.
- Coolen, A.C.C. and C.C.A.M. Gielen (1988). Delays in Neural Networks. Europhysics Letters 7, 281285.
- Cooper, L.N. (1973). A Possible Organization of Animal Memory and Learning. In Collective Properties of Physical Systems (24th Nobel Symposium), eds. B. Lundqvist and S. Lundqvist , 252264. New York: Academic Press. Reprinted in Anderson and Rosenfeld [1988].
- Cortes, C. and J.A. Hertz (1989). A Network System for Image Segmentation. In International Joint Conference on Neural Networks (Washington 1989), vol. I, 121127. New York: IEEE.
- Cortes, C. , A. Krogh , and J.A. Hertz (1987). Hierarchical Associative Networks. Journal of Physics A 20, 44494455.
- Cottrell, G.W. , P. Munro , and D. Zipser (1987). Learning Internal Representations from Gray-Scale Images: An Example of Extensional Programming. In Ninth Annual Conference of the Cognitive Science Society (Seattle 1987), 462473. Hillsdale: Erlbaum.
- Cottrell, M. and J.C. Fort (1986). A Stochastic Model of Retinotopy: A Self Organizing Process. Biological Cybernetics 53, 405411.
- 286 Cover, T.M. (1965). Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. IEEE Transactions on Electronic Computers 14, 326334.
- Cowan, J.D. and D.H. Sharp (1988a). Neural Nets and Artificial Intelligence. Ddalus, Proceedings of the American Academy of Arts and Sciences 117, 85121.

- Cowan, J.D. and D.H. Sharp (1988b). Neural Nets. *Quarterly Reviews of Biophysics* 21, 365427.
- Cragg, B.G. and H.N.V. Temperley (1954). The Organization of Neurones: A Cooperative Analogy. *EEG and Clinical Neurophysiology* 6, 8592.
- Cragg, B.G. and H.N.V. Temperley (1955). Memory: The Analogy with Ferromagnetic Hysteresis. *Brain* 78, 11, 304316.
- Crick, F. (1989). The Recent Excitement About Neural Networks. *Nature* 337, 129132.
- Crisanti, A. and H. Sompolinsky (1987). Dynamics of Spin Systems with Randomly Asymmetric Bonds: Langevin Dynamics and a Spherical Model. *Physical Review A* 36, 49224939.
- Crisanti, A. , D.J. Amit , and H. Gutfreund (1986). Saturation Level of the Hopfield Model for Neural Network. *Europhysics Letters* 2, 337341.
- Cybenko, G. (1988). Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient. Technical Report, Department of Computer Science, Tufts University, Medford, MA.
- Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems* 2, 303314.
- Dehaene, S. , J.-P. Changeux , and J.-P. Nadal (1987). Neural Networks That Learn Temporal Sequences by Selection. *Proceedings of the National Academy of Sciences, USA* 84, 27272731.
- Denker, J. (1986). Neural Network Refinements and Extensions. In *Neural Networks for Computing* (Snowbird 1986), ed. J.S. Denker , 121128. New York: American Institute of Physics.
- Denker, J. , D. Schwartz , B. Wittner , S. Solla , R. Howard , L. Jackel , and J. Hopfield (1987). Large Automatic Learning, Rule Extraction, and Generalization. *Complex Systems* 1, 877922.
- Derrida, B. , E. Gardner , and A. Zippelius (1987). An Exactly Soluble Asymmetric Neural Network Model. *Europhysics Letters* 4, 167173.
- Derthick, M. (1984). Variations on the Boltzmann Machine. Technical Report CMU-CS-84-120, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- DeSieno, D. (1988). Adding a Conscience to Competitive Learning. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. I, 117124. New York: IEEE.
- Devijver, P.A. and J. Kittler (1982). *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall.
- Diederich, S. and M. Opper (1987). Learning of Correlated Patterns in Spin-Glass Networks by Local Learning Rules. *Physical Review Letters* 58, 949952.
- 287 Duda, R.O. and P.E. Hart (1973). *Pattern Classification and Scene Analysis*. New York: Wiley.
- Durbin, R. and D. Willshaw (1987). An Analogue Approach to the Travelling Salesman Problem Using an Elastic Net Method. *Nature* 326, 689691.
- Elman, J.L. (1990). Finding Structure in Time. *Cognitive Science* 14, 179211.
- Elman, J.L. and D. Zipser (1988). Learning the Hidden Structure of Speech. *Journal of the Acoustical Society of America* 83, 16151626.
- Fahlman, S.E. (1989). Fast-Learning Variations on Back-Propagation: An Empirical Study. In *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 3851. San Mateo: Morgan Kaufmann.
- Fahlman, S.E. and C. Lebiere (1990). The Cascade-Correlation Learning Architecture. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 524532. San Mateo: Morgan Kaufmann.
- Farhat, N.H. (1987). Optoelectronic Analogs of Self-Programming Neural Nets: Architectures and Methods for Implementing Fast Stochastic Learning by Simulated Annealing. *Applied Optics* 26, 50935103.
- Farhat, N.H. , D. Psaltis , A. Prata , and E. Paek (1985). Optical Implementation of the Hopfield Model. *Applied Optics* 24, 14691475. Reprinted in Anderson and Rosenfeld [1988].
- Farmer, D. and J. Sidorowich (1987). Predicting Chaotic Time Series. *Physical Review Letters* 59, 845848.
- Farmer, D. and J. Sidorowich (1988). Exploiting Chaos to Predict the Future and Reduce Noise. In *Evolution, Learning, and Cognition*, ed. W.C. Lee , 277330. Singapore: World Scientific.
- Feldman, J.A. and D.H. Ballard (1982). Connectionist Models and Their Properties. *Cognitive Science* 6, 205254. Reprinted in Anderson and Rosenfeld [1988].

- Feller, W. (1968). An Introduction to Probability Theory and Its Applications, vol. I. New York: Wiley.
- Fischer, K.H. and J.A. Hertz (1990). Spin Glasses. Cambridge: Cambridge University Press.
- Flidak, P. (1989). Adaptive Network for Optimal Linear Feature Extraction. In International Joint Conference on Neural Networks (Washington 1989), vol. I, 401405. New York: IEEE.
- Franzini, M.A. (1987). Speech Recognition with Back Propagation. In Proceedings of the Ninth Annual Conference of the IEEE Engineering in Medicine and Biology Society (Boston 1987), 17021703. New York: IEEE.
- Frean, M (1990). The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks. *Neural Computation* 2, 198209.
- Fu, Y. and P.W. Anderson (1986). Application of Statistical Mechanics to NP-Complete Problems in Combinatorial Optimization. *Journal of Physics A* 19, 16051620.
- Fukushima, K. (1975). Cognitron: A Self-Organizing Multilayered Neural Network. *Biological Cybernetics* 20, 121136.
- 288 Fukushima, K. (1980). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics* 36, 193202.
- Fukushima, K. , S. Miyake , and T. Ito (1983). Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 13, 826834. Reprinted in Anderson and Rosenfeld [1988].
- Gallant, S.I. (1986). Optimal Linear Discriminants. In Eighth International Conference on Pattern Recognition (Paris 1986), 849852. New York: IEEE.
- Gardner, E. (1987). Maximum Storage Capacity in Neural Networks. *Europhysics Letters* 4, 481485.
- Gardner, E. (1988). The Space of Interactions in Neural Network Models. *Journal of Physics A* 21, 257270.
- Gardner, E. and B. Derrida (1988). Optimal Storage Properties of Neural Network Models. *Journal of Physics A* 21, 271284.
- Gardner, E. and B. Derrida (1989). Three Unfinished Works on the Optimal Storage Capacity of Networks. *Journal of Physics A* 22, 19831994.
- Gardner, E. , H. Gutfreund , and I. Yekutieli (1989). The Phase Space of Interactions in Neural Networks with Definite Symmetry. *Journal of Physics A* 22, 19952008.
- Garey, M.R. and D.S. Johnson (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness. New York: Freeman.
- Geman, S and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721741. Reprinted in Anderson and Rosenfeld [1988].
- Geszti, T. (1990). Physical Models of Neural Networks. Singapore: World Scientific.
- Geszti, T. and F. Pzmndi (1987). Learning Within Bounds and Dream Sleep. *Journal of Physics A* 20, L1299L1303.
- Glauber, R.J. (1963). Time-Dependent Statistics of the Ising Model. *Journal of Mathematical Physics* 4, 294307.
- Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Reading: Addison-Wesley.
- Golomb, D. , N. Rubin , and H. Sompolinsky (1990). Willshaw Model: Associative Memory with Sparse Coding and Low Firing Rates. *Physical Review A* 41, 18431854.
- Gorman, R.P. and T.J. Sejnowski (1988a). Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets. *Neural Networks* 1, 7589.
- Gorman, R.P. and T.J. Sejnowski (1988b). Learned Classification of Sonar Targets Using a Massively-Parallel Network. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36, 11351140.
- 289 Graf, H.P. , L.D. Jackel , R.E. Howard , B. Straughn , J.S. Denker , W. Hubbard , D.M. Tennant , and D. Schwartz (1986). VLSI Implementation of a Neural Network Memory with Several Hundreds of Neurons. In Neural Networks for Computing (Snowbird 1986), ed. J.S. Denker , 182187. New York: American Institute of Physics.
- Graf, D.H. and W.R. LaLonde (1988). A Neural Controller for Collision-Free Movement of General Robot Manipulators. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 7784. New York: IEEE.
- Gray, R.M. (1984). Vector Quantization. *IEEE ASSP Magazine*, April 1984, 429.
- Grossberg, S. (1967). Nonlinear Difference-Differential Equations in Prediction and Learning Theory. *Proceedings of the National Academy of Sciences, USA* 58, 13291334.

- Grossberg, S. (1968a). Some Nonlinear Networks Capable of Learning a Spatial Pattern of Arbitrary Complexity. *Proceedings of the National Academy of Sciences, USA* 59, 368372.
- Grossberg, S. (1968b). Some Physiological and Biochemical Consequences of Psychological Postulates. *Proceedings of the National Academy of Sciences, USA* 60, 758765.
- Grossberg, S. (1969). Embedding Fields: A Theory of Learning with Physiological Implications. *Journal of Mathematical Psychology* 6, 209239.
- Grossberg, S. (1972). Neural Expectation: Cerebellar and Retinal Analogs of Cells Fired by Learnable or Unlearned Pattern Classes. *Kybernetik* 10, 4957.
- Grossberg, S. (1976a). Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors. *Biological Cybernetics* 23, 121134. Reprinted in Anderson and Rosenfeld [1988].
- Grossberg, S. (1976b). Adaptive Pattern Classification and Universal Recoding: II. Feedback, Expectation, Olfaction, Illusions. *Biological Cybernetics* 23, 187202.
- Grossberg, S. (1980). How Does the Brain Build a Cognitive Code? *Psychological Review* 87, 151. Reprinted in Anderson and Rosenfeld [1988].
- Grossberg, S. (1987a). *The Adaptive Brain*, 2 vols. Amsterdam: Elsevier.
- Grossberg, S. (1987b). Competitive Learning: From Interactive Activation to Adaptive Resonance. *Cognitive Science* 11, 2363.
- Grossman, T. (1990). The CHIR Algorithm for Feed Forward Networks with Binary Weights. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 516523. San Mateo: Morgan Kaufmann.
- Grossman, T. , R. Meir , and E. Domany (1989). Learning by Choice of Internal Representations. *Complex Systems* 2, 555575.
- Gutfreund, H. (1988). Neural Networks with Hierarchically Correlated Patterns. *Physical Review A* 37, 570577.
- Gutfreund, H. and M. Mezard (1988). Processing of Temporal Sequences in Neural Networks. *Physical Review Letters* 61, 235238.
- 290 Gutzmann, K. (1987). Combinatorial Optimization Using a Continuous State Boltzmann Machine. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. III, 721734. New York: IEEE.
- Gyrgyi, G. (1990). Inference of a Rule by a Neural Network with Thermal Noise. *Physical Review Letters* 64, 29572960.
- Gyrgyi, G. and N. Tishby (1990). Statistical Theory of Learning a Rule. In *Neural Networks and Spin Glasses*, eds. W.K. Theumann and R. Koeberle . Singapore: World Scientific.
- Hanson, S.J. and L. Pratt (1989). A Comparison of Different Biases for Minimal Network Construction with Back-Propagation. In *Advances in Neural Information Processing Systems I* (Denver 1988), ed. D.S. Touretzky , 177185. San Mateo: Morgan Kaufmann.
- Harp, S.A. , T. Samad , and A. Guha (1990). Designing Application-Specific Neural Networks Using the Genetic Algorithm. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 447454. San Mateo: Morgan Kaufmann.
- Hartman, E.J. , J.D. Keeler , and J.M. Kowalski (1990). Layered Neural Networks with Gaussian Hidden Units As Universal Approximations. *Neural Computation* 2, 210215.
- Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley. Partially reprinted in Anderson and Rosenfeld [1988].
- Hecht-Nielsen, R. (1987a). Combinatorial Hypercompression. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 455461. New York: IEEE.
- Hecht-Nielsen, R. (1987b). Counterpropagation Networks. *Applied Optics* 26, 49794984.
- Hecht-Nielsen, R. (1988). Applications of Counterpropagation Networks. *Neural Networks* 1, 131139.
- Hecht-Nielsen, R. (1989). Theory of the Backpropagation Neural Network. In *International Joint Conference on Neural Networks* (Washington 1989), vol. I, 593605. New York: IEEE.
- Hegde, S.U. , J.L. Sweet , and W.B. Levy (1988). Determination of Parameters in a Hopfield/Tank Computational Network. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. II, 291298. New York: IEEE.
- van Hemmen, J.L. , L.B. Ioffe , R. Kihnn , and M. Vaas (1990). Increasing the Efficiency of a Neural Network through Unlearning. *Physica* 163A, 386392.
- van Hemmen, J.L. and R. Khn (1986). Nonlinear Neural Networks. *Physical Review Letters* 57, 913916.
- van Hemmen, J.L. and R.G. Palmer (1979). The Replica Method and a Solvable Spin Glass Model. *Journal of Physics A* 12, 563580.

- Hertz, J.A. (1989). A Gauge Theory in Computational Vision: A Model for Outline Extraction. *Physica Scripta* 39, 161167.
- Hertz, J.A. (1990). Statistical Dynamics of Learning. In *Characterizing Complex Systems*, ed. H. Bohr , 3439. Singapore: World Scientific.
- 291 Hertz, J.A. , G. Grinstein , and S. Solla (1986). Memory Networks with Asymmetric Bonds. In *Neural Networks for Computing (Snowbird 1986)*, ed. J.S. Denker , 212218. New York: American Institute of Physics.
- Hertz, J.A. , G. Grinstein , and S. Solla (1987). Irreversible Spin Glasses and Neural Networks. In *Heidelberg Colloquium on Glassy Dynamics (Heidelberg 1986)*, eds. J.L. van Hemmen and I. Morgenstern , 538546. Berlin: Springer-Verlag.
- Hertz, J.A. , A. Krogh , and G.I. Thorbergsson (1989). Phase Transitions in Simple Learning. *Journal of Physics A* 22, 21332150.
- Herz, A. , B. Sulzer , R. Khn , and J.L. van Hemmen (1989). Hebbian Learning Reconsidered: Representation of Static and Dynamic Objects in Associative Neural Nets. *Biological Cybernetics* 60, 457467.
- Hinton, G.E. (1986). Learning Distributed Representations of Concepts. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society (Amherst 1986)*, 112. Hillsdale: Erlbaum.
- Hinton, G.E. (1989). Deterministic Boltzmann Learning Performs Steepest Descent in Weight Space. *Neural Computation* 1, 143150.
- Hinton, G.E. and T.J. Sejnowski (1983). Optimal Perceptual Inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Washington 1983)*, 448453. New York: IEEE.
- Hinton, G.E. and T.J. Sejnowski (1986). Learning and Relearning in Boltzmann Machines. In *Parallel Distributed Processing*, vol. 1, chap. 7. See Rumelhart, McClelland, et al. [1986].
- Hopfield, J.J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences, USA* 79, 25542558. Reprinted in Anderson and Rosenfeld [1988].
- Hopfield, J.J. (1984). Neurons with Graded Responses Have Collective Computational Properties Like Those of Two-State Neurons. *Proceedings of the National Academy of Sciences, USA* 81, 30883092. Reprinted in Anderson and Rosenfeld [1988].
- Hopfield, J.J. (1987). Learning Algorithms and Probability Distributions in Feed-Forward and Feed-Back Networks. *Proceedings of the National Academy of Sciences, USA* 84, 84298433.
- Hopfield, J.J. , D.I. Feinstein , and R.G. Palmer (1983). Unlearning Has a Stabilizing Effect in Collective Memories. *Nature* 304, 158159.
- Hopfield, J.J. and D.W. Tank (1985). Neural Computation of Decisions in Optimization Problems. *Biological Cybernetics* 52, 141152.
- Hopfield, J.J. and D.W. Tank (1986). Computing with Neural Circuits: A Model. *Science* 233, 625633.
- Hopfield, J.J. and D.W. Tank (1989). Neural Architecture and Biophysics for Sequence Recognition. In *Neural Models of Plasticity*, eds. J.H. Byrne and W.O. Berry , 363377. San Diego: Academic Press.
- Hornik, K. , M. Stinchcombe , and H. White (1989). Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2, 359366.
- 292 Hsu, K , D. Brady , and D. Psaltis (1988). Experimental Demonstration of Optical Neural Computers. In *Neural Information Processing Systems (Denver 1987)*, ed. D.Z. Anderson , 377386. New York: American Institute of Physics.
- Huang, K. (1987). *Statistical Mechanics*. New York: Wiley.
- Huang, W.Y. and R.P. Lippmann (1988). Neural Net and Traditional Classifiers. In *Neural Information Processing Systems (Denver 1987)*, ed. D.Z. Anderson , 387396. New York: American Institute of Physics.
- Hubel, D.H. and T.N. Wiesel (1962). Receptive Fields, Binocular Interaction, and Functional Architecture in the Cats Visual Cortex. *Journal of Physiology (London)* 160, 106154.
- Hueter, G.J. (1988). Solution of the Travelling Salesman Problem with an Adaptive Ring. In *IEEE International Conference on Neural Networks (San Diego 1988)*, vol. I, 8592. New York: IEEE.
- Hush, D.R. and J.M. Salas (1988). Improving the Learning Rate of Back-Propagation with the Gradient Reuse Algorithm. In *IEEE International Conference on Neural Networks (San Diego 1988)*, vol. I, 441447. New York: IEEE.
- Jacobs, R.A. (1988). Increased Rates of Convergence Through Learning Rate Adaptation. *Neural Networks* 1, 295307.

- Johnson, D.S. , C.R. Aragon , L.A. McGeoch , and C. Schevon (1989). Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning. *Operations Research* 37, 865891. Parts II and III are expected to appear in 1990.
- Jolliffe, I.T. (1986). Principal Component Analysis. New York: Springer-Verlag.
- Jordan, M.I. (1986). Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. In Proceedings of the Eighth Annual Conference of the Cognitive Science Society (Amherst 1986), 531546. Hillsdale: Erlbaum.
- Jordan, M.I. (1989). Serial Order: A Parallel, Distributed Processing Approach. In Advances in Connectionist Theory: Speech, eds. J.L. Elman and D.E. Rumelhart . Hillsdale: Erlbaum.
- Kahng, A.B. (1989). Travelling Salesman Heuristics and Embedding Dimension in the Hopfield Model. In International Joint Conference on Neural Networks (Washington 1989), vol. I, 513520. New York: IEEE.
- Kammen, D.M. and A.L. Yuille (1988). Spontaneous Symmetry-Breaking Energy Functions and the Emergence of Orientation Selective Cortical Cells. *Biological Cybernetics* 59, 2331.
- Kandel, E.R. and J.H. Schwartz (1985). Principles of Neural Science (2nd ed.). New York: Elsevier.
- Kanter, I. and H. Sompolinsky (1987). Associative Recall of Memory Without Errors. *Physical Review A* 35, 380392.
- Kerszberg, M. and A. Zippelius (1990). Synchronization in Neural Assemblies. *Physica Scripta* T33, 5464.
- Kinzel, W. and M. Opper (1990). Dynamics of Learning. In Physics of Neural Networks, eds. E. Domany , J.L. van Hemmen , and K. Schulten , vol. I. Berlin: Springer-Verlag.
- Kirkpatrick, S. , C.D. Gelatt Jr. , and M.P. Vecchi (1983). Optimization by Simulated Annealing. *Science* 220, 671680. Reprinted in Anderson and Rosenfeld [1988].
- 293 Kirkpatrick, S. and G. Toulouse (1985). Configuration Space Analysis of Travelling Salesman Problems. *Journal de Physique (Paris)* 46, 12771292.
- Kleinfeld, D. (1986). Sequential State Generation by Model Neural Networks. *Proceedings of the National Academy of Sciences, USA* 83, 94699473.
- Kleinfeld, D. and H. Sompolinsky (1989). Associative Network Models for Central Pattern Generators. In Methods in Neuronal Modeling: From Synapses to Networks, eds. C. Koch and I. Segev , 195246. Cambridge: MIT Press.
- Klopff, A.H. (1982). The Hedonistic Neuron: A Theory of Memory, Learning, and Intelligence. Washington: Hemisphere.
- Koch, C. , J. Luo , C. Mead , and J. Hutchinson (1988). Computing Motion Using Resistive Networks. In Neural Information Processing Systems (Denver 1987), ed. D.Z. Anderson , 422431. New York: American Institute of Physics.
- Koch, C. , J. Marroquin , and A. Yuille (1986). Analog Neuronal Networks in Early Vision. *Proceedings of the National Academy of Sciences, USA* 83, 42634267.
- Kohonen, T. (1974). An Adaptive Associative Memory Principle. *IEEE Transactions on Computers* C-23, 444445.
- Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics* 43, 5969. Reprinted in Anderson and Rosenfeld [1988].
- Kohonen, T. (1989). Self-Organization and Associative Memory (3rd ed.). Berlin: Springer-Verlag.
- Kohonen, T. , G. Barna , and R. Chrisley (1988). Statistical Pattern Recognition with Neural Networks: Benchmarking Studies. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 6168. New York: IEEE.
- Kohonen, T. , K. Mksara , and T. Saramki (1984). Phonotopic Maps Insightful Representation of Phonological Features for Speech Recognition. In Proceedings of the Seventh International Conference on Pattern Recognition (Montreal 1984), 182185. New York: IEEE.
- Kramer, A.H. and A. Sangiovanni-Vincentelli (1989). Efficient Parallel Learning Algorithms for Neural Networks. In Advances in Neural Information Processing Systems I (Denver 1988), ed. D.S. Touretzky , 4048. San Mateo: Morgan Kaufmann.
- Krauth, W. and M. Mezard (1989). The Cavity Method and the Travelling-Salesman Problem. *Europhysics Letters* 8, 213218.
- Krishnaiah, P.R. and L.N. Kanal , eds. (1982). Classification, Pattern Recognition, and Reduction of Dimensionality. *Handbook of Statistics*, vol. 2. Amsterdam: North Holland.
- Krogh, A. and J.A. Hertz (1988). Mean Field Analysis of Hierarchical Associative Networks with Magnetization. *Journal of Physics A* 21, 22112224.
- 294 Krogh, A. and J.A. Hertz (1990). Hebbian Learning of Principal Components. In Parallel Processing in Neural Systems and Computers (Dsseldorf 1990), eds. R. Eckmiller , G.

- Hartmann , and G. Hauske , 183186. Amsterdam: Elsevier.
- Krogh, A. , G.I. Thorbergsson , and J.A. Hertz (1990). A Cost Function for Internal Representations. In Advances in Neural Information Processing Systems II (Denver 1989), ed. D.S. Touretzky , 733740. San Mateo: Morgan Kaufmann.
- Kullback, S. (1959). Information Theory and Statistics. New York: Wiley.
- Kushner, H.J. and D.S. Clark (1978). Stochastic Approximation Methods for Constrained and Unconstrained Systems. New York: Springer-Verlag.
- Lapedes, A. and R. Farber (1986a). A Self-Optimizing, Nonsymmetrical Neural Net for Content Addressable Memory and Pattern Recognition. *Physica* 22D, 247259.
- Lapedes, A. and R. Farber (1986b). Programming a Massively Parallel, Computation Universal System: Static Behavior. In Neural Networks for Computing (Snowbird 1986), ed. J.S. Denker , 283298. New York: American Institute of Physics.
- Lapedes, A. and R. Farber (1987). Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.
- Lapedes, A. and R. Farber (1988). How Neural Nets Work. In Neural Information Processing Systems (Denver 1987), ed. D.Z. Anderson , 442456. New York: American Institute of Physics.
- Lawler, E.L. (1976). Combinatorial Optimization: Networks and Matroids. New York: Holt-Rinehart-Winston.
- Lawler, E.L. , J.K. Lenstra , A.H.G. Rinnooy Kan , and D.B. Shmoys , eds. (1985). The Travelling Salesman Problem. Chichester: Wiley.
- Le Cun, Y. (1985). Une Procedure d'Apprentissage pour Rseau Seuil Assymtrique. In Cognitiva 85: A la Frontire de lIntelligence Artificielle des Sciences de la Connaissance des Neurosciences (Paris 1985), 599604. Paris: CESTA.
- Le Cun, Y. , B. Boser , J.S. Denker , D. Henderson , R.E. Howard , W. Hubbard , and L.D. Jackel (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1, 541551.
- Le Cun, Y. , B. Boser , J.S. Denker , D. Henderson , R.E. Howard , W. Hubbard , and L.D. Jackel (1990). Handwritten Digit Recognition with a Back-Propagation Network. In Advances in Neural Information Processing Systems II (Denver 1989), ed. D.S. Touretzky , 396404. San Mateo: Morgan Kaufmann.
- Le Cun, Y. , J.S. Denker , and S.A. Solla (1990). Optimal Brain Damage. In Advances in Neural Information Processing Systems II (Denver 1989), ed. D.S. Touretzky , 598605. San Mateo: Morgan Kaufmann.
- von Lehman, A. , E.G. Paek , P.F. Liao , A. Marrakchi , and J.S. Patel (1988). Factors Influencing Learning by Back-Propagation. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 335341. New York: IEEE.
- 295 Liang, F.M. (1983). Ph.D. Thesis. Stanford University. See Appendix H of The TEXBook, D.E. Knuth (Reading: Addison Wesley 1986).
- Lin, S. and B.W. Kernighan (1973). An Effective Heuristic Algorithm for the Travelling Salesman Problem. *Operations Research* 21, 498516.
- Linsker, R. (1986). From Basic Network Principles to Neural Architecture. *Proceedings of the National Academy of Sciences, USA* 83, 75087512, 83908394, 87798783.
- Linsker, R. (1988). Self-Organization in a Perceptual Network. *Computer*, March 1988, 105117.
- Lippmann, R.P. (1987). An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, April 1987, 422.
- Lippmann, R.P. (1989). Review of Neural Networks for Speech Recognition. *Neural Computation* 1, 138.
- Lister, R. (1990). Segment Reversal and the TSP. In Intl. Joint Conference on Neural Networks (Washington 1990), vol. 1, 424427. Hillsdale: Erlbaum.
- Little, W.A. (1974). The Existence of Persistent States in the Brain. *Mathematical Biosciences* 19, 101120.
- Little, W.A. and G.L. Shaw (1975). A Statistical Theory of Short and Long Term Memory. *Behavioral Biology* 14, 115133. Reprinted in Anderson and Rosenfeld [1988].
- Little, W.A. and G.L. Shaw (1978). Analytic Study of the Memory Storage Capacity of a Neural Network. *Mathematical Biosciences* 39, 281290.
- Luenerger, D.G. (1986). Linear and Nonlinear Programming. Reading: Addison-Wesley.
- Ma, S.-K. (1985). Statistical Mechanics. Philadelphia: World Scientific.
- MacKay, D.J.C. and K.D. Miller (1990). Analysis of Linskers Simulation of Hebbian Rules. *Neural Computation* 2, 173187.

- Mackey, M.C. and L. Glass (1977). Oscillation and Chaos in Physiological Control Systems. *Science* 197, 287.
- Makram-Ebeid, S. , J.-A. Sirat , and J.-R. Viala (1989). A Rationalized Back-Propagation Learning Algorithm. In International Joint Conference on Neural Networks (Washington 1989), vol. II, 373380. New York: IEEE.
- von der Malsburg, Ch. (1973). Self-Organization of Orientation Sensitive Cells in the Striate Cortex. *Kybernetik* 14, 85100. Reprinted in Anderson and Rosenfeld [1988].
- von der Malsburg, Ch. and J.D. Cowan (1982). Outline of a Theory for the Ontogenesis of Iso-Orientation Domains in Visual Cortex. *Biological Cybernetics* 45, 4956.
- Mandelbrot, B.B. (1982). *The Fractal Geometry of Nature*. San Francisco: Freeman.
- Marchand, M. , M. Golea , and P. Rujn (1990). A Convergence Theorem for Sequential Learning in Two-Layer Perceptrons. *Europhysics Letters* 11, 487492.
- Marr, D. (1969). A Theory of Cerebellar Cortex. *Journal of Physiology (London)* 202, 437470.
- 296 Marr, D. (1970). A Theory for Cerebral Neocortex. *Proceedings of the Royal Society of London B* 176, 161234.
- Marr, D. (1971). Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London B* 262, 2381.
- Marr, D. (1982). *Vision*. San Francisco: Freeman. Partially reprinted in Anderson and Rosenfeld [1988].
- Marr, D. and T. Poggio (1976). Cooperative Computation of Stereo Disparity. *Science* 194, 283287. Reprinted in Anderson and Rosenfeld [1988].
- Mattis, D. (1976). Solvable Spin Systems with Random Interactions. *Physics Letters* 56A, 421422.
- Mazaika, P.K. (1987). A Mathematical Model of the Boltzmann Machine. In IEEE First International Conference on Neural Networks (San Diego 1987), eds. M. Caudill and C. Butler , vol. III, 157163. New York: IEEE.
- McClelland, J.L. and J.L. Elman (1986). Interactive Processes in Speech Perception: The TRACE Model. In Parallel Distributed Processing, vol. 2, chap. 15. See McClelland, Rumelhart, et al. [1986].
- McClelland, J.L. , D.E. Rumelhart, D.E. , and the PDP Research Group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models. Cambridge: MIT Press
- McClelland, J.L. and D.E. Rumelhart (1988). Explorations in Parallel Distributed Processing. Cambridge: MIT Press.
- McCulloch, W.S. and W. Pitts (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics* 5, 115133. Reprinted in Anderson and Rosenfeld [1988].
- McEliece, R.J. , E.C. Posner , E.R. Rodemich , and S.S. Venkatesh (1987). The Capacity of the Hopfield Associative Memory. *IEEE Transactions on Information Theory* 33, 461482.
- McInerny, J.M. , K.G. Haines , S. Biafore , and R. Hecht-Nielsen (1989). Back Propagation Error Surfaces Can Have Local Minima. In International Joint Conference on Neural Networks (Washington 1989), vol. II, 627. New York: IEEE.
- Mead, C. (1989). *Analog VLSI and Neural Systems*. Reading: Addison Wesley.
- Metropolis, N. , A.W. Rosenbluth , M.N. Rosenbluth , A.H. Teller , and E. Teller (1953). Equation of State Calculations for Fast Computing Machines. *Journal of Chemical Physics* 21, 10871092.
- Mzard, M. and J.-P. Nadal (1989). Learning in Feedforward Layered Networks: The Tiling Algorithm. *Journal of Physics A* 22, 21912204.
- Mzard, M. and G. Parisi (1985). Replicas and Optimization. *Journal de Physique Lettres (Paris)* 46, 771778.
- Mzard, M. and G. Parisi (1986). A Replica Analysis of the Travelling Salesman Problem. *Journal de Physique (Paris)* 47, 12851296.
- Mzard, M. and G. Parisi (1988). The Euclidean Matching Problem. *Journal de Physique (Paris)* 49, 20192025.
- Mzard, M. , G. Parisi , and M.A. Virasoro (1987). *Spin Glass Theory and Beyond*. Singapore: World Scientific.
- 297 Miller, G.F. , P.M. Todd , and S.U. Hegde (1989). Designing Neural Networks Using Genetic Algorithms. In Proceedings of the Third International Conference on Genetic Algorithms (Arlington 1989), ed. J.D. Schaffer , 379384. San Mateo: Morgan Kaufmann.
- Minsky, M.L. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs: Prentice-Hall.

- Minsky, M.L. and S.A. Papert (1969). *Perceptrons*. Cambridge: MIT Press. Partially reprinted in Anderson and Rosenfeld [1988].
- Mitchison, G.J. and R.M. Durbin (1989). Bounds on the Learning Capacity of Some Multi-Layer Networks. *Biological Cybernetics* 60, 345356.
- Montana, D.J. and L. Davis (1989). Training Feedforward Networks Using Genetic Algorithms. In Eleventh International Joint Conference on Artificial Intelligence (Detroit 1989), ed. N.S. Sridharan , 762767. San Mateo: Morgan Kaufmann.
- Moody, J. and C. Darken (1988). Learning with Localized Receptive Fields. In Proceedings of the 1988 Connectionist Models Summer School (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 133143. San Mateo: Morgan Kaufmann.
- Moody, J. and C. Darken (1989). Fast Learning in Networks of Locally-Tuned Processing Units. *Neural Computation* 1, 281294.
- Moore, B. (1988). ART1 and Pattern Clustering. In Proceedings of the 1988 Connectionist Models Summer School (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 174185. San Mateo: Morgan Kaufmann.
- Mozer, M.C. (1989). A Focused Back-Propagation Algorithm for Temporal Pattern Recognition. *Complex Systems* 3, 349381.
- Munro, P. (1987). A Dual Back-Propagation Scheme for Scalar Reward Learning. In The Ninth Annual Conference of the Cognitive Science Society (Seattle 1987), 165176. Hillsdale: Erlbaum.
- Nadal, J.-P. , G. Toulouse , J.-P. Changeux , and S. Dehaene (1986). Networks of Formal Neurons and Memory Palimpsests. *Europhysics Letters* 1, 535542.
- Narendra, K. and M.A.L. Thathachar (1989). *Learning Automata: An Introduction*. Englewood Cliffs: Prentice-Hall.
- Nasrabadi, N.M. and Y. Feng (1988). Vector Quantization of Images Based upon the Kohonen Self-Organizing Feature Maps. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 101108. New York: IEEE.
- Nasrabadi, N.M. and R.A. King (1988). Image Coding Using Vector Quantization: A Review. *IEEE Transactions on Communications* 36, 957971.
- Nass, M.M. and L.N. Cooper (1975). A Theory for the Development of Feature Detecting Cells in Visual Cortex. *Biological Cybernetics* 19, 118.
- Naylor, J. and K.P. Li (1988). Analysis of a Neural Network Algorithm for Vector Quantization of Speech Parameters. *Neural Networks Supplement* 1, 310.
- von Neumann, J. (1956). Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In *Automata Studies*, eds. C.E. Shannon and J. McCarthy , 4398. Princeton: Princeton University Press.
- 298 Niranjan, M. and F. Fallside (1990). Neural Networks and Radial Basis Functions in Classifying Static Speech Patterns. *Computer Speech and Language* 4, 275289.
- Nishimori, H. , T. Nakamura , and M. Shiino (1990). Retrieval of Spatio-Temporal Sequence in Asynchronous Neural Network. *Physical Review A* 41, 33463354.
- Nowlan, S.J. (1988). Gain Variation in Recurrent Error Propagation Networks. *Complex Systems* 2, 305320.
- Oja, E. (1982). A Simplified Neuron Model As a Principal Component Analyzer. *Journal of Mathematical Biology* 15, 267273.
- Oja, E. (1989). Neural Networks, Principal Components, and Subspaces. *International Journal of Neural Systems* 1, 6168.
- Oja, E. and J. Karhunen (1985). On Stochastic Approximation of the Eigenvectors and Eigenvalues of the Expectation of a Random Matrix. *Journal of Mathematical Analysis and Applications* 106, 6984.
- Opper, M. , W. Kinzel , J. Kleinz , and R. Nehl (1990). On the Ability of the Optimal Perceptron to Generalize. *Journal of Physics A* 23, L581L586.
- Orland, H. (1985). Mean-Field Theory for Optimization Problems. *Journal de Physique Lettres (Paris)* 46, 763770.
- Owens, A.J. and D.L. Filkin (1989). Efficient Training of the Back Propagation Network by Solving a System of Stiff Ordinary Differential Equations. In International Joint Conference on Neural Networks (Washington 1989), vol. II, 381386. New York: IEEE.
- Packard, N.H. , J.P. Crutchfield , J.D. Farmer , and R.S. Shaw (1980). Geometry from a Time Series. *Physical Review Letters* 45, 712716.
- Palmer, R.G. (1988). Statistical Mechanics Approaches to Complex Optimization Problems. In *The Economy As an Evolving Complex System*, eds. P.W. Anderson , K.J. Arrow , and D. Pines . SFI Studies in the Sciences of Complexity, proc. vol. 5, 177193. Redwood City: Addison-Wesley.

- Palmer, R.G. (1989). Neural Nets, In *Lectures in the Sciences of Complexity*, ed. D.L. Stein . SFI Studies in the Sciences of Complexity, lect. vol. 1, 439461. Redwood City: Addison-Wesley.
- Papadimitriou, C.H. and K. Steiglitz (1982). *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs: Prentice-Hall.
- Parga, N. and M.A. Virasoro (1986). The Ultrametric Organization of Memories in a Neural Network. *Journal de Physique (Paris)* 47, 18571864.
- Parisi, G. (1986). Asymmetric Neural Networks and the Process of Learning. *Journal of Physics A* 19, L675L680.
- Parker, D.B. (1985). Learning Logic. Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA.
- Parker, D.B. (1987). Optimal Algorithms for Adaptive Networks: Second Order Back Propagation, Second Order Direct Propagation, and Second Order Hebbian Learning. In *IEEE First International Conference on Neural Networks (San Diego 1987)*, eds. M. Caudill and C. Butler , vol. II, 593600. New York: IEEE.
- 299 Parks, M. (1987). Characterization of the Boltzmann Machine Learning Rate. In *IEEE First International Conference on Neural Networks (San Diego 1987)*, eds. M. Caudill and C. Butler , vol. III, 715719. New York: IEEE.
- Pearlmutter, B.A. (1989a). Learning State Space Trajectories in Recurrent Neural Networks. In *International Joint Conference on Neural Networks (Washington 1989)*, vol. II, 365372. New York: IEEE.
- Pearlmutter, B.A. (1989b). Learning State Space Trajectories in Recurrent Neural Networks. *Neural Computation* 1, 263269.
- Pearlmutter, B.A. and G.E. Hinton (1986). G-Maximization: An Unsupervised Learning Procedure for Discovering Regularities. In *Neural Networks for Computing (Snowbird 1986)*, ed. J.S. Denker , 333338. New York: American Institute of Physics.
- Peretto, P. (1984). Collective Properties of Neural Networks: A Statistical Physics Approach. *Biological Cybernetics* 50, 5162.
- Peretto, P. (1988). On Learning Rules and Memory Storage Abilities of Asymmetrical Neural Networks. *Journal de Physique (Paris)* 49, 711726.
- Peretto, P. and J.J. Niez (1986). Collective Properties of Neural Networks. In *Disordered Systems and Biological Organization (Les Houches 1985)*, eds. E. Bienenstock , F. Fogelman-Soulie , and G. Weisbuch , 171185. Berlin: Springer-Verlag.
- Prez, R. , L. Glass , and R. Shlaer (1975). Development of Specificity in the Cat Visual Cortex. *Journal of Mathematical Biology* 1, 275288.
- Personnaz, L. , I. Guyon , and G. Dreyfus (1985). Information Storage and Retrieval in Spin-Glass-Like Neural Networks. *Journal de Physique Lettres (Paris)* 46, 359365.
- Personnaz, L. , I. Guyon , and G. Dreyfus (1986). Collective Computational Properties of Neural Networks: New Learning Mechanisms. *Physical Review A* 34, 42174228.
- Peterson, C. and J.R. Anderson (1987). A Mean Field Theory Learning Algorithm for Neural Networks. *Complex Systems* 1, 9951019.
- Peterson, C. , S. Redfield , J.D. Keeler , and E. Hartman (1990). An Optoelectronic Architecture for Multilayer Learning in a Single Photorefractive Crystal. *Neural Computation* 2, 2534.
- Peterson, C. and B. Sderberg (1989). A New Method for Mapping Optimization Problems onto Neural Networks. *International Journal of Neural Systems* 1, 322.
- Pineda, F.J. (1987). Generalization of Back-Propagation to Recurrent Neural Networks. *Physical Review Letters* 59, 22292232.
- Pineda, F.J. (1988). Dynamics and Architecture for Neural Computation. *Journal of Complexity* 4, 216245.
- Pineda, F.J. (1989). Recurrent Back-Propagation and the Dynamical Approach to Adaptive Neural Computation. *Neural Computation* 1, 161172.
- Plaut, D. , S. Nowlan , and G. Hinton (1986). Experiments on Learning by Back Propagation. Technical Report CMU-CS-86-126, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- 300 Poggio, T. and F. Girosi (1990). Regularization Algorithms for Learning That Are Equivalent to Multilayer Networks. *Science* 247, 978982.
- Poggio, T. , V. Torre , and C. Koch (1985). Computational Vision and Regularization Theory. *Nature* 317, 314319.
- Pomerleau, D.A. (1989). ALVINN: An Autonomous Land Vehicle in a Neural Network. In *Advances in Neural Information Processing Systems I (Denver 1988)*, ed. D.S. Touretzky ,

305313. San Mateo: Morgan Kaufmann.
- Press, W.H. , B.P. Flannery , S.A. Teukolsky , and W.T. Vetterling (1986). Numerical Recipes. Cambridge: Cambridge University Press.
- Qian, N. and T.J. Sejnowski (1988a). Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *Journal of Molecular Biology* 202, 865884.
- Qian, N. and T.J. Sejnowski (1988b). Learning to Solve Random-Dot Stereograms of Dense Transparent Surfaces with Recurrent Back-Propagation. In *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 435443. San Mateo: Morgan Kaufmann.
- Rao, C.R. and S.K. Mitra (1971). Generalized Inverse of Matrices and Its Applications. New York: Wiley.
- Ramanujam, J. and P. Sadayappan (1988). Optimization by Neural Networks. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. II, 325332. New York: IEEE.
- Rashevsky, N. (1938). Mathematical Biophysics. Chicago: University of Chicago Press.
- Rescorla, R.A. and A.R. Wagner (1972). A Theory of Pavlovian Conditioning: The Effectiveness of Reinforcement and Nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, eds. A.H. Black and W.F. Prokasy , 6469. New York: Appleton-Century-Crofts.
- Ricotti, L.P. , S. Ragazzini , and G. Martinelli (1988). Learning of Word Stress in a Sub-Optimal Second Order Back-Propagation Neural Network. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. I, 355361. New York: IEEE.
- Riedel, U. , R. Khn , and J.L. van Hemmen (1988). Temporal Sequences and Chaos in Neural Nets. *Physical Review A* 38, 11051108.
- Ritter, H. and K. Schulten (1986). On the Stationary State of Kohonens Self-Organizing Sensory Mapping. *Biological Cybernetics* 54, 99106.
- Ritter, H. and K. Schulten (1988a). Extending Kohonens Self-Organizing Mapping Algorithm to Learn Ballistic Movements. In *Neural Computers* (Neuss 1987), eds. R. Eckmiller and Ch. von der Malsburg , 393406. Berlin: Springer-Verlag.
- Ritter, H. and K. Schulten (1988b). Convergence Properties of Kohonens Topology Conserving Maps: Fluctuations, Stability, and Dimension Selection. *Biological Cybernetics* 60, 5971.
- Ritter, H. and K. Schulten (1988c). Kohonens Self-Organizing Maps: Exploring Their Computational Capabilities. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. I, 109116. New York: IEEE.
- 301 Robinson, A.J. and F. Fallside (1988). Static and Dynamic Error Propagation Networks with Application to Speech Coding. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 632641. New York: American Institute of Physics.
- Rohwer, R. (1990). The Moving Targets Training Algorithm. In *Advances in Neural Information Processing Systems II* (Denver 1989), ed. D.S. Touretzky , 558565. San Mateo: Morgan Kaufmann.
- Rohwer, R. and B. Forrest (1987). Training Time-Dependence in Neural Networks. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 701708. New York: IEEE.
- Romeo, F.I. (1989). Simulated Annealing: Theory and Applications to Layout Problems. Ph.D. Thesis, University of California at Berkeley Memorandum UCB/ERL-M89/29.
- Rose, D. and V.G. Dobson , eds. (1985). Models of the Visual Cortex. Chichester: Wiley.
- Rosenblatt, F. (1962). Principles of Neurodynamics. New York: Spartan.
- Rubner, J. and K. Schulten (1990). Development of Feature Detectors by Self-Organization. *Biological Cybernetics* 62, 193199.
- Rubner, J. and P. Tavan (1989). A Self-Organizing Network for Principal-Component Analysis. *Europhysics Letters* 10, 693698.
- Rumelhart, D.E. , G.E. Hinton , and R.J. Williams (1986a). Learning Representations by Back-Propagating Errors. *Nature* 323, 533536. Reprinted in Anderson and Rosenfeld [1988].
- Rumelhart, D.E. , G.E. Hinton , and R.J. Williams (1986b). Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing*, vol. 1, chap. 8. See Rumelhart, McClelland , et al. [1986]. Reprinted in Anderson and Rosenfeld [1988].
- Rumelhart, D.E. , J.L. McClelland , and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. Cambridge: MIT Press.
- Rumelhart, D.E. and D. Zipser (1985). Feature Discovery by Competitive Learning. *Cognitive Science* 9, 75112. Also reprinted in *Parallel Distributed Processing*, vol. 1, chap. 5. See

- Rumelhart, McClelland , et al. [1986].
- Saad, D. and E. Marom (1990a). Learning by Choice of Internal Representations An Energy Minimization Approach. *Complex Systems* 4, 107118.
- Saad, D. and E. Marom (1990b). Training Feed Forward Nets with Binary Weights via a Modified CHIR Algorithm. *Complex Systems* 4, 573586.
- Salamon, P. , J.D. Nulton , J. Robinson , J. Petersen , G. Ruppeiner , and L. Liao (1988). Simulated Annealing with Constant Thermodynamic Speed. *Computer Physics Communications* 49, 423428.
- Sanger, T.D. (1989a). Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network. *Neural Networks* 2, 459473.
- 302 Sanger, T.D. (1989b). An Optimality Principle for Unsupervised Learning. In *Advances in Neural Information Processing Systems I* (Denver 1988), ed. D.S. Touretzky , 1119. San Mateo: Morgan Kaufmann.
- Sato, M. (1990). A Real Time Learning Algorithm for Recurrent Analog Neural Networks. *Biological Cybernetics* 62, 237241.
- Scalettar, R. and A. Zee (1988). Emergence of Grandmother Memory in Feed Forward Networks: Learning with Noise and Forgetfulness. In *Connectionist Models and Their Implications: Readings from Cognitive Science*, eds. D. Waltz and J.A. Feldman , 309332. Norwood: Ablex.
- Schwartz, D.B. , V.K. Samalam , S.A. Solla , and J.S. Denker (1990). Exhaustive Learning, *Neural Computation* 2, 371382.
- Scofield, C.L. (1988). Learning Internal Representations in the Coulomb Energy Network. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol, I, 271276. New York: IEEE.
- Sejnowski, T.J. , P.K. Kienker , and G. Hinton (1986). Learning Symmetry Groups with Hidden Units: Beyond the Perceptron. *Physica* 22D, 260275.
- Sejnowski, T.J. and C.R. Rosenberg (1987). Parallel Networks that Learn to Pronounce English Text. *Complex Systems* 1, 145168.
- Sherrington, D. and S. Kirkpatrick (1975). Solvable Model of a Spin Glass. *Physical Review Letters* 35, 17921796.
- Shimohara, K. , T. Uchiyama , and Y. Tokunaga (1988). Back-Propagation Networks for Event-Driven Temporal Sequence Processing. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. I, 665672. New York: IEEE.
- Sietsma, J. and R.J.F. Dow (1988). Neural Net PruningWhy and How. In *IEEE International Conference on Neural Networks* (San Diego 1988), vol. I, 325333. New York: IEEE.
- Silverman, R.H. and A.S. Noetzel (1988). Time-Sequential Self-Organization of Hierarchical Neural Networks. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 709714. New York: American Institute of Physics.
- Simard, P.Y. , M.B. Ottaway , and D.H. Ballard (1989). Analysis of Recurrent Backpropagation. In *Proceedings of the 1988 Connectionist Models Summer School* (Pittsburg 1988), eds. D. Touretzky , G. Hinton , and T. Sejnowski , 103112. San Mateo: Morgan Kaufmann.
- Simic, P.D. (1990). Statistical Mechanics As the Underlying Theory of Elastic and Neural Optimizations. *Network* 1, 89103.
- Sirat, J.-A. and J.-P. Nadal (1990). Neural Trees: A New Tool for Classification. Preprint, Laboratoires d'Electronique Philips, Limeil-Brevannes, France.
- Sivilotti, M.A. , M.A. Mahowald , and C.A. Mead (1987). Real-Time Visual Computations Using Analog CMOS Processing Arrays. In *Advanced Research in VLSI: Proceedings of the 1987 Stanford Conference*, ed. P. Losleben , 295312. Cambridge: MIT Press. Reprinted in Anderson and Rosenfeld [1988].
- Smolensky, P. (1986). Information Processing in Dynamical Systems: Foundations of Harmony Theory. In *Parallel Distributed Processing*, vol. 1, chap. 6. See Rumelhart, McClelland, et al. [1986].
- 303 Softer, B.H. , G.J. Dunning , Y. Owechko , and E. Marom (1986). Associative Holographic Memory with Feedback Using Phase-Conjugate Mirrors. *Optics Letters* 11, 118120.
- Solla, S.A. (1989). Learning and Generalization in Layered Neural Networks: The Contiguity Problem. In *Neural Networks from Models to Applications* (Paris 1988), eds. L. Personnaz and G. Dreyfus , 168177. Paris: I.D.S.E.T.
- Solla, S.A. , E. Levin , and M. Fleisher (1988). Accelerated Learning in Layered Neural Networks. *Complex Systems* 2, 625639.

- Sompolinsky, H. (1987). The Theory of Neural Networks: The Hebb Rules and Beyond. In Heidelberg Colloquium on Glassy Dynamics (Heidelberg 1986), eds. J.L. van Hemmen and I. Morgenstern , 485527. Berlin: Springer-Verlag.
- Sompolinsky, H. , A. Crisanti , and H.J. Sommers (1988). Chaos in Random Neural Networks. *Physical Review Letters* 61, 259262.
- Sompolinsky, H. and I. Kanter (1986). Temporal Association in Asymmetric Neural Networks. *Physical Review Letters* 57, 28612864.
- Soukoulis, C.M. , K. Levin , and G.S. Grest (1983). Irreversibility and Metastability in Spin-Glasses. I. Ising Model. *Physical Review B* 28, 14951509.
- Specht, D.F. (1990). Probabilistic Neural Networks. *Neural Networks* 3, 109118.
- Steinbuch, K. (1961). Die Lernmatrix. *Kybernetik* 1, 3645.
- Stokbro, K. , D.K. Umberger , and J.A. Hertz (1990). Exploiting Neurons with Localized Receptive Fields to Learn Chaos. *Complex Systems* 4, 603622.
- Stornetta, W.S. , T. Hogg , and B.A. Huberman (1988). A Dynamical Approach to Temporal Pattern Processing. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 750759. New York: American Institute of Physics.
- Sutton, R.S. (1984). Temporal Credit Assignment in Reinforcement Learning. Ph. D. Thesis, University of Massachusetts, Amherst.
- Sutton, R.S. (1988). Learning to Predict by the Methods of Temporal Differences. *Machine Learning* 3, 944.
- Sutton, R.S. and A.G. Barto (1991). Time Derivative Models of Pavlovian Reinforcement. In *Learning and Computational Neuroscience*, eds. M. Gabriel and J.W. Moore . Cambridge: MIT Press.
- Szu, H. (1986). Fast Simulated Annealing. In *Neural Networks for Computing* (Snowbird 1986), ed. J.S. Denker , 420425. New York: American Institute of Physics.
- Takens, F. (1981). Detecting Strange Attractors In Turbulence. In *Dynamical Systems and Turbulence. Lecture Notes in Mathematics*, vol. 898 (Warwick 1980), eds. D.A. Rand and L.-S. Young , 366381. Berlin: Springer-Verlag.
- Takeuchi, A. and S. Amari (1979). Formation of Topographic Maps and Columnar Microstructures in Nerve Fields. *Biological Cybernetics* 35, 6372.
- Tank, D.W. and J.J. Hopfield (1986). Simple Neural Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit. *IEEE Transactions on Circuits and Systems* 33, 533541.
- 304 Tank, D.W. and J.J. Hopfield (1987a). Neural Computation by Concentrating Information in Time. *Proceedings of the National Academy of Sciences, USA* 84, 18961900.
- Tank, D.W. and J.J. Hopfield (1987b). Concentrating Information in Time: Analog Neural Networks with Applications to Speech Recognition Problems. In *IEEE First International Conference on Neural Networks* (San Diego 1987), eds. M. Caudill and C. Butler , vol. IV, 455468. New York: IEEE.
- Taylor, W.K. (1956). Electrical Simulation of Some Nervous System Functional Activities. In *Information Theory* (London 1985), ed. C. Cherry , 314328. London: Butterworths.
- Tesauro, G. (1986). Simple Neural Models of Classical Conditioning. *Biological Cybernetics* 55, 187200.
- Tesauro, G. (1990). Neurogammon Wins Computer Olympiad. *Neural Computation* 1, 321323.
- Tesauro, G. and B. Janssens (1988). Scaling Relationships in Back-Propagation Learning. *Complex Systems* 2, 3944.
- Tesauro, G. and T.J. Sejnowski (1988). A Neural Network That Learns to Play Backgammon. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 442456. New York: American Institute of Physics.
- Thakoor, A.P. , A. Moopenn , J. Lambe , and S.K. Khanna (1987). Electronic Hardware Implementations of Neural Networks. *Applied Optics* 26, 50855092.
- Ticknor, A.J. and H. Barrett (1987). Optical Implementations of Boltzmann Machines. *Optical Engineering* 26, 1621.
- Tishby, N. , E. Levin , and S.A. Solla (1989). Consistent Inference of Probabilities in Layered Networks: Predictions and Generalization. In *International Joint Conference on Neural Networks* (Washington 1989), vol. II, 403410. New York: IEEE.
- Toulouse, G. , S. Dehaene , and J.-P. Changeux (1986). Spin Glass Model of Learning by Selection. *Proceedings of the National Academy of Sciences, USA* 83, 16951698.
- Touretzky, D.S. and D.A. Pomerleau (1989). Whats Hidden in the Hidden Layers? *BYTE*, August 1989, 227233.

- Tsodyks, M.V. and M.V. Feigelman (1988). The Enhanced Storage Capacity in Neural Networks with Low Activity Level. *Europhysics Letters* 6, 101105.
- Van den Bout, D.E. and T.K. Miller (1988). A Travelling Salesman Objective Function That Works. In IEEE International Conference on Neural Networks (San Diego 1988), vol. II, 299303. New York: IEEE.
- Van den Bout, D.E. and T.K. Miller (1989). Improving the Performance of the Hopfield-Tank Neural Network Through Normalization and Annealing. *Biological Cybernetics* 62, 129139.
- Vapnik, V.N. (1982). *Estimation of Dependences Based on Empirical Data*. Berlin: Springer-Verlag.
- Vapnik, V.N. and A.Y. Chervonenkis (1971). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability and Its Applications* 16, 264280.
- 305 Vogl, T.P. , J.K. Mangis , A.K. Rigler , W.T. Zink , and D.L. Alkon (1988). Accelerating the Convergence of the Back-Propagation Method. *Biological Cybernetics* 59, 257263.
- Wagner, K. and D. Psaltis (1987). Multilayer Optical Learning Networks. *Applied Optics* 26, 50615076.
- Waibel, A. (1989). Modular Construction of Time-Delay Neural Networks for Speech Recognition. *Neural Computation* 1, 3946.
- Waibel, A. , T. Hanazawa , G. Hinton , K. Shikano , and K. Lang (1989). Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 328339.
- Watrous, R.L. (1987). Learning Algorithms for Connectionist Networks: Applied Gradient Methods of Nonlinear Optimization. In IEEE First International Conference on Neural Networks (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 619627. New York: IEEE.
- Webster (1988). Webster's Ninth New Collegiate Dictionary. Springfield: Merriam-Webster.
- Weisbuch, G. and F. Fogelman-Soulie (1985). Scaling Laws for the Attractors of Hopfield Networks. *Journal de Physique Lettres (Paris)* 46, 623630.
- Werbos, P. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. Thesis, Harvard University.
- Werbos, P.J. (1987). Building and Understanding Adaptive Systems: A Statistical/Numerical Approach to Factory Automation and Brain Research. *IEEE Transactions on Systems, Man, and Cybernetics* 17, 720.
- Werbos, P.J. (1988). Generalization of Backpropagation with Application to a Recurrent Gas Market Model. *Neural Networks* 1, 339356.
- Whitley, D. and T. Hanson (1989). Optimizing Neural Networks Using Faster, More Accurate Genetic Search. In Proceedings of the Third International Conference on Genetic Algorithms (Arlington 1989), ed. J.D. Schaffer , 391396. San Mateo: Morgan Kaufmann.
- Widrow, B. (1962). Generalization and Information Storage in Networks of Adaline Neurons. In *Self-Organizing Systems 1962* (Chicago 1962), eds. M.C. Yovits , G.T. Jacobi , and G.D. Goldstein , 435461. Washington: Spartan.
- Widrow, B. , N.K. Gupta , and S. Maitra (1973). Punish/Reward: Learning with a Critic in Adaptive Threshold Systems. *IEEE Transactions on Systems, Man, and Cybernetics* 3, 455465.
- Widrow, B. and M.E. Hoff (1960). Adaptive Switching Circuits. In 1960 IRE WESCON Convention Record, part 4, 96104. New York: IRE. Reprinted in Anderson and Rosenfeld [1988].
- Wiener, N. (1948). *Cybernetics, or Control and Communication in the Animal and the Machine*. New York: Wiley.
- Williams, R.J. (1987). A Class of Gradient-Estimating Algorithms for Reinforcement Learning in Neural Networks. In IEEE First International Conference on Neural Networks (San Diego 1987), eds. M. Caudill and C. Butler , vol. II, 601608. New York: IEEE.
- 306 Williams, R.J. (1988a). On the Use of Back-Propagation in Associative Reinforcement Learning. In IEEE International Conference on Neural Networks (San Diego 1988), vol. I, 263270. New York: IEEE.
- Williams, R.J. (1988b). Towards a Theory of Reinforcement-Learning Connectionist Systems. Technical Report NU-CCS-88-3, College of Computer Science, Northeastern University, Boston, MA.
- Williams, R.J. and J. Peng (1989). Reinforcement Learning Algorithms As Function Optimizers. In International Joint Conference on Neural Networks (Washington 1989), vol. II, 8995. New York: IEEE.
- Williams, R.J. and D. Zipser (1989a). A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation* 1, 270280.

- Williams, R.J. and D. Zipser (1989b). Experimental Analysis of the Real-Time Recurrent Learning Algorithm. *Connection Science* 1, 87111.
- Willshaw, D.J. , O.P. Buneman , and H.C. Longuet-Higgins (1969). Non-Holographic Associative Memory. *Nature* 222, 960962. Reprinted in Anderson and Rosenfeld [1988].
- Willshaw, D.J. and C. von der Malsburg (1976). How Patterned Neural Connections Can Be Set Up by Self-Organization. *Proceedings of the Royal Society of London B* 194, 431445.
- Wilson, G.V. and G.S. Pawley (1988). On the Stability of the Travelling Salesman Problem Algorithm of Hopfield and Tank. *Biological Cybernetics* 58, 6370.
- Wilson, H.R. and J.D. Cowan (1973). A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue. *Kybernetik* 13, 5580.
- Winograd, S. and J.D. Cowan (1963). Reliable Computation in the Presence of Noise. Cambridge: MIT Press.
- Winters, J.H. and C. Rose (1989). Minimum Distance Automata in Parallel Networks for Optimum Classification. *Neural Networks* 2, 127132.
- Wittgenstein, L. (1958). *Philosophical Investigations*. 2nd ed., translated by G.E.M. Anscombe . Oxford: Blackwell.
- Wittner, B.S. and J.S. Denker (1988). Strategies for Teaching Layered Networks Classification Tasks. In *Neural Information Processing Systems* (Denver 1987), ed. D.Z. Anderson , 850859. New York: American Institute of Physics.
- Yuille, A.L. , D.M. Kammen , and D.S. Cohen (1989). Quadrature and the Development of Orientation Selective Cortical Cells by Hebb Rules. *Biological Cybernetics* 61, 183194.
- Zak, M. (1988). Terminal Attractors for Addressable Memory in Neural Networks. *Physics Letters* 133A, 1822.
- Zak, M. (1989). Terminal Attractors in Neural Networks. *Neural Networks* 2, 259274.