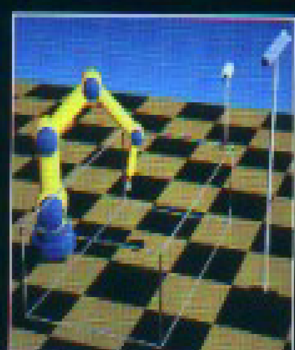
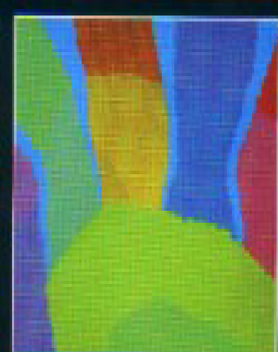
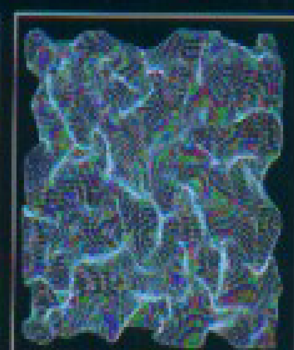
Neural Computation and Self-Organizing Maps

AN INTRODUCTION

Helge Ritter
Thomas Martinetz
Klaus Schulten

Neural Computation and Self-Organizing Maps - An Introduction

by Helge Ritter, Thomas Martinetz, and Klaus Schulten Addison-Wesley, New York, 1992

About the Book:

This book is a comprehensive introduction to neural networks and neural information processing. It describes the most important models of neural networks and how they contribute to our understanding of information and organization processes in the brain. One of the few generally recognized organizational principles of the nervous system, the development of cortical feature maps (brain maps), is described in detail, and the reader is introduced to the biological background and the mathematical properties of self-organizing maps as important functional building blocks of the brain. Examples show how neural networks can solve important information processing tasks, including the development of sensory maps, the traveling salesman problem, and visuomotor control of robots.

About the Cover:

The images portray (from left to right): (I) a robot arm, learning movement control by using self-organizing maps: (ii) a model of mapping from the hand onto the somatosensory cortex of the brain: (iii) a theoretical study of phase transitions in topology-conserving maps.

FOREWORD

The wave of interest in the artificial neural networks (ANNs) that started in the mid-1980s was inspired by new prospects not visible ten years earlier. First of all it should be realized that ANNs have been intended for a new *component technology*. There are many computation-intensive tasks such as preprocessing of natural signals, pattern classification and recognition, coordination of movements in complex mechanisms, decision making on the basis of extensive but uncertain data, and high-definition animated graphics that can no longer be handled by digital computers. Even supercomputers are soon unable to cope with the growing dimensionality of such problems. It has become more and more obvious that one has to resort to special analog computing methods; with the aid of modern VLSI technology and optics it will be possible to produce analog picowatt circuits by the billions, and so the cost of massively parallel computation can be cut to a fraction. The breakthroughs in the analog semiconductor and active optical component technology around 1980 were thus crucial for the acceptance of the ANN computing principles.

Before digital technology can be replaced or at least augmented by the "neuromorphic" technology in practice, one must fully understand *what* and *how* to compute. As even the most fundamental operations are different from those of digital computing, and the innumerable system parameters of the ANNs are time variable, designers are faced with new phenomena, and they have to learn how to deal with them. This revolution in the paradigms and standards will not be easy; however, if the ANNs prove cost-effective in practice, this change will be inevitable. Therefore, we must welcome every teaching effort in this new field. Books, especially monographs, of which the present one is an excellent example, are invaluable aids in education of these new technologies.

The excitement about the ANNs has also been accompanied by certain beliefs that we finally understand how the brain works. The collective computations thereby performed, and the automatic adaptive changes of the system parameters and structures have often been identified with mental processes and learning ability. However, it may already have become clear even to the most enthusiastic supporter of these ideas that mere increase in the parallel computing capacity is not sufficient for the duplication or even imitation of the brain functions. Every biological cell makes use of tens of informations processing principles of which only two or three have been utilized in the ANNs and the immensely complex structures of the biological nervous networks have been formed in innumerable cycles of evolution, under continuous bombardment of complex signals from the environment and other sources of natural information. There exists yet nothing similar in the ANNs, which are usually only dedicated to some restricted tasks.

While it is obvious that the ANNs cannot accurately imitate even the simplest biological circuits, it is also necessary to realize that the *functions* and *processes* at work in the nervous systems are not at all that mysterious; since they are based on physical and chemical phenomena, it is possible to approximate their behavior, at least on some level of abstraction. For their *understanding* it will then be sufficient to set up a model that takes into account a number of the basic operations and relationships of the elementary functions in the spatial and temporal domain. If certain essential modeling assumptions are made, one cannot avoid starting to see phenomena that very much resemble those observed in the biological systems. This is an irrefutable fact, and can certainly be interpreted as partial explanation of these phenomena.

When working with the ANNs, it is therefore necessary to realize that while the principles and components thereby applied have been inspired by brain-theoretic considerations, the artificial implementations need not necessarily do exactly the same as their biological counterparts. It may not be possible to achieve the complexity, flexible learning ability, and capability of high-level abstraction of experiences characteristic of biological organisms. On the other hand, the stability and accuracy of the artificial components can be orders of magnitude higher than those of the biological ones. In some tasks it can be a significant advantage that the ANNs do not exhibit fatigue, and are not panicked in alarming situations. It is plausible that in the future the computing capacity of the ANNs can be increased much beyond that of the biological systems. All this gives us promises of development that we may not yet fully foresee.

Teuvo Kohonen

Professor, Helsinki University of Technology
Research Professor at the Academy of Finland

PREFACE

The understanding of biological brains—with their capacity for learning as well as for the processing of sensory impressions and the control of movements—is one of the most fascinating research challenges of our time. In the still-young discipline of *Neural Computation*, scientists from such distinct fields as biology, information theory, physics, mathematics, psychology, and medicine have joined forces to pursue this challenge. *Neural Computation* seeks to simulate "biological intelligence" in artificial *"neural networks"* the structure and dynamics of which attempt to imitate the function of biological neural systems.

In the past few years, a number of promising successes have been achieved in this endeavor, triggering lively research activities in diverse research groups. The present book took shape during this period. Its aim is to provide an introduction to the field of neural computation and it is equally intended for those working in the fields of computer science, physics, biology, mathematics, engineering, psychology, and medicine, as well as for all those readers with an interest in computer models of neural networks and of the brain.

The first part of the book gives a general overview of the most important current models of neural nets together with a short sketch of the relevant biological background. The second part of the book is devoted to the central question of how functional neural circuitry in the brain can arise by means of a *self-organizing process*. It is shown how, by means of a few simple mechanisms, neural layers can learn representations or "maps" of important stimulus features under the influence of nothing more than a random sequence of sensory stimuli. A series of examples demonstrates the simulation of observed organization processes in the brain. However, these examples also show how solutions of abstract tasks from traditional information science can be obtained by the same mechanisms. The third part of this book is concerned with the question of what extensions of these mechanisms are required in order to enable the learning of simple motor skills, such as the balancing of a pole or the control of eye movements. Building on this foundation, the fourth part of the book describes several studies concerning problems of robot control. It is shown how a neural network can implement the coordination of robot arm movements under visual feedback control. Finally, the last part of the book treats important theoretical questions connected with the learning process, in particular the question of convergence and the influence of the element of chance during the learning phase.

At several points, a basic mathematical knowledge of elementary analysis and linear algebra may be useful to the reader, but they are not required in large parts of the book. Only in the last part, which is concerned with a more thorough mathematical analysis, some familiarity with vector analysis will be helpful.

Here, we would like to thank all those who have contributed to the creation of this book. We are grateful to the friendship and advice we received from Hans-Ulrich Bauer, Joachim Buhmann, Anita Govindjee, Leo van Hemmen, Karl Hess, Teuvo Kohonen, Christoph von der Malsburg, Sabine Martinetz, Jeanette Rubner, Zan Schulten, Werner von Seelen, Larry Smarr, Paul Tavan, and Udo Weigelt. We want to mention particularly our colleague Klaus Obermayer, whose work on self-organizing maps enriched our own views in many important ways, and who provided two of the color pictures on the front cover. Daniel Barsky, Ron Kates, and Markus Tesch have helped us tremendously with the translation from the original German text. Allan Wylde and Pam Suwinsky of Addison Wesley have been patient supporters. The book would have been impossible without grants which we received over the years from the National Science Foundation, the National Institute of Health, the State of Illinois, as well as from the German Ministry of Research and Development. Thomas Martinetz received a fellowship from the Volkswagen Foundation. Computer time and much good advice had been available to us from the National Center for Supercomputing supported by the National Science Foundation.

We are especially grateful to the Beckman Institute of the University of Illinois where we had the privilege to work and in whose stimulating atmosphere the book could be completed.

This book was written with the TeX typesetting system and the program Textures. Typesetting and graphics were produced on a Macintosh II computer.

Urbana, Illinois, and Bielefeld, Germany
October 1991
Thomas Martinetz, Helge Ritter, Klaus Schulten

CONTENTS

0. INTRODUCTION AND OVERVIEW

The emergence of the electronic computer and its incredibly rapid development have revived humankind's age-old curiosity about the working of the brain and about the nature of the human mind. The availability of the computer as a research tool has raised hopes of at least partial answers to these questions. There are three reasons for regarding this hope as justified.

First, the accomplishments of computers have forced us to define, in a precise manner, our concepts of the phenomenon "mind" — in this context generally under the heading of "intelligence." The rapid evolution of computers also demands a redefinition of the previously clear and unproblematic concept of "machine." In particular, the high flexibility made possible through programming has led us to regard the capabilities of computers as being separate from their material substrate, the *hardware*, but rather as residing in their program, the *software*. This "hardware-software duality" has enriched our conceptual framework on the relationship between mind and matter.

Secondly, as a tool, the computer has tremendously accelerated scientific progress, including progress in areas that are important for a better understanding of the brain. For example, computers made it possible to carry out and evaluate many neurophysiological, psychophysical, and cognitive experiments. Other relevant branches of science, in particular computer science and its subfield " artificial intelligence" (" AI") came into being as computers became available.

Thirdly, with the ability to manufacture computer hardware of high enough performance, discoveries concerning the functioning of the brain, in addition to their former purely intellectual benefit, have also become valuable for their technical applicability. This circumstance has opened up important resources for theoretical studies of the brain, and will probably continue to do so.

However, the demand for practical applications of artificial intelligence made evident the limitations of previous concepts of hardware and software. Characteristically, today's computers solve problems that are difficult for humans, but fail miserably at everyday tasks that humans master without a great deal of effort. This circumstance, so far, has limited the use of computers to narrow problem areas and indicates a fundamental difference between AI methods and the operation of biological nervous systems.

Most computers, until recently, were based on the so-called *von Neumann architecture*. They derive their performance from one or only a few central processors which carry out long sequential programs at extremely high speed. Therefore, signal-propagation times within the computer have already begun to emerge as limiting factors to further gains in speed. At the same time, efforts to master multifaceted problem situations, *e.g.*, those encountered in driving a car, by means of conventional programming techniques lead to programs of a complexity that can no longer be managed reliably.

A way out of this dilemma requires an abandonment of the von Neumann architecture used up to now, and instead to apply a large number of computational processors *working in parallel*. For the programming of such computers, new kinds of algorithms are required that must allow a distribution of computational tasks over a great number of processors. In order to keep the necessary task of integrating such algorithms into complex software systems manageable, the algorithms must be error tolerant and capable of learning. These features seem to be realized in biological brains with nerve cells as processors, which, by technical standards, are slow computational elements and of only limited reliability, but which on the other hand are present in huge numbers, processing sensory data and motor tasks concurrently.

In order to make this "biological know-how" available, the interdisciplinary research area of *Neural Computation* has developed in the last few years. While its main aim is an understanding of the principles of information processing employed by biological nervous systems, this discipline also seeks to apply the insights gained to the construction of new kinds of computers with more flexible capabilities. In the pursuit of this goal, Neural Computation combines the efforts of computer scientists, neurobiologists, physicists, engineers, mathematicians, psychologists, and physicians.

Although we are still far from a true understanding of how the brain works, a great deal of progress has been made, especially in the last few years. On the experimental side, modern techniques are opening new "windows into the brain". Today, optical dyes allow one to stain living brain tissue such that the optical properties of the dyes provide a measure of the electrical activation of the nerve cells. In this way, optical recording of neural activity patterns has become possible. Other staining methods allow precise reconstructions of the three-dimensional shape of single nerve cells. By means of modern computer tomographic methods (PET, NMR) the momentary metabolic activity level of brain tissue can be recorded up to spatial resolutions in the range of millimeters. Sensitive magnetic field detectors based on superconducting devices (SQUID's) can measure the spatial distribution of brain currents with a similar resolution from outside the head. This has made it possible to monitor patterns of brain activity with noninvasive methods — and, therefore, also in humans — and to investigate its dependence on experimentally preselected mental tasks.

Nevertheless, the task of integrating the multitude of experimental data collected up to now into predictive theories of information processing in the brain is anything but simple. The first efforts go back to 1943, when McCulloch and Pitts originally postulated that nerve cells play the role of " logical elements," *i.e.*, evaluate Boolean (logical) functions. With the advent of digital computers, a strong additional motivation for the further development of these ideas arose, since quantitatively formulated models were suddenly no longer dependent on mathematical analysis alone, often both very difficult and feasible only to a limited degree, but could now be investigated in computer simulations. At this time the " perceptron" was being developed by Rosenblatt (1958). Rosenblatt derived a network model capable of learning to classify patterns making use only of simple principles for the change of connection strengths between neurons. These resembled the principles previously suggested by the psychologist Hebb (1949) on theoretical grounds to explain memory performance. Thus, the "perceptron" represents one of the first "brain models" that could successfully demonstrate the ability to "learn."

At about the same time the availability of computers led to the advent of a competing research direction, which regarded orientation toward the structure of biological nerve systems as of little aid in the investigation and simulation of intelligence. Instead, this direction attempted a more direct approach: by introducing sufficiently elaborate programming based on "problem solution heuristics," it was hoped that ultimately the goal of intelligent machines would be reached. Due to rapid initial successes, this research direction, now forming most of traditional AI, managed to push the investigation of neural networks for a number of years into obscurity. Even so, a series of important insights were gained in the theory of neural networks during this period: examples are the discovery of models for associative memory (Taylor 1956; Steinbuch 1961), models for self-organization of feature detectors (von der Malsburg 1973) and of ordered neural connections (Willshaw and von der Malsburg 1976), as well as pioneering studies concerning mathematical properties of important classes of network models by Amari, Grossberg, Kohonen, and numerous other researchers.

A highly significant stimulus for the further development of the subject was contributed by Hopfield (1982). Exploiting the formal equivalance between network models with "Boolean" neural units and physical systems consisting of interacting "elementary magnets" or "spins" (Cragg and Temperley 1954, 1955; Caianiello 1961; Little 1974; Little and Shaw 1975), he showed that the dynamic of such networks can be described by an *energy function* and that patterns stored in these networks can be regarded as attractors in a high-dimensional phase space. As a consequence, a whole arsenal of mathematical methods of statistical physics became available for the analysis of such network models. Many questions previously approachable only by computer simulations found an elegant mathematical solution (see, *e.g.*, Amit et al. 1985ab, Derrida et al. 1987, Gardner 1988, Buhmann et al. 1989). At the same time, new kinds of network models were found, two of which deserve special mention because of their promise: The *backpropagation model* (rediscovered several times, most recently by Rumelhart et al. 1986) constituted a significant improvement of the earlier perceptron models. In spite of a few aspects that are implausible from a biological point of view, its extremely broad applicability triggered considerable new research activity. Kohonen's model of *self-organizing neural maps* (Kohonen 1982a) represented an important abstraction of earlier models of von der Malsburg and Willshaw; the model combines biological plausibility with proven applicability in a broad range of difficult data processing and optimization problems.

All of these models provide us with a much more refined picture of the function of the brain than could have been anticipated a few decades ago. Nevertheless, most of the work has yet to be done. Compared to the capabilities of biological systems, the performance of our present " neurocomputers" is quite rudimentary. We still are unable to relate more than a relatively small number of experimental observations to properties of our models. There are still enormous gaps between the complexity of the brain, our theoretical models, and the capacity of today's computers. However, the modest amount of "biological know-how" which has been accumulated in order to bridge these gaps is already promising and suggests that further research will be rewarding. In particular, a new generation of computers with thousands of processors has put us in a position to simulate at least small areas of the brain in much greater detail than previously possible, and to use for the first time realistic numbers of neurons and synapses for such simulations (Obermayer et al. 1990a-c, 1991).

The first part of the book furnishes an overview of the major concepts on which much of the current work in Neural Computation is based. In Chapter 1, our present view of the brain as a " neurocomputer" is briefly outlined. The second chapter contains a sketch of the biological background, emphasizing its significance in understanding the various brain models. The third chapter introduces the most prominent model approaches of neural networks, including the perceptron model, the Hopfield model, and the backpropagation algorithm. A particular type of network, Kohonen's model of self-organizing maps, is the focus of Chapter 4. This network model is capable both of reproducing important aspects of the structure of biological neural nets and of a wide range of practical applications. It will serve as a basis for much of the discussion in the later chapters.

The later parts of the book take the reader through a series of typical issues in neural computation. We devote each chapter to an information-processing task that is characteristic of those confronting a biological organism in its environment. It is not our intention in the later chapters to present a complete survey of the by now large field of neural computation. Rather than a broad overview of the many different approaches, we present a highly focused and detailed description of network models based on self-organizing maps.

As an introductory example, it is shown in Chapter 5 how an adaptive " neural frequency map" can be formed in the cortex of a bat, which enables the bat to perform an extremely precise analysis of sonar ultrasound signals. Chapter 6 is concerned with the relationship of this example to the solution of a task appearing completely different at first glance, namely the determination of a route that is as short as possible in the " traveling salesman problem." A further example (Chapter 7) considers the creation of an ordered connectivity between touch receptors of the hand surface and the cortical area responsible for the sense of touch in the brain. Here, as in the case of the bat, we are concerned with the processing of *sensory information*. However, in nature this is never an end in itself. The processing of sensory information always has as its eventual goal the triggering and control of *motor functions*, probably the oldest task of biological nerve systems. This points to the need for the investigation of strategies for *neural control* and for *learning to execute movements*; it also brings us to the theme of the third part of the book.

The task of balancing a pole already contains a number of important features of motor control problems, and it is therefore discussed thoroughly in Chapter 8. We show how a neural network can learn the task of balancing a pole, first in a version with the help of a "teacher," then in an improved version by "independent trial and error." The main point of this task is to learn how to maintain an unstable equilibrium. An equally important aspect of motor function is the support of our sensory perception. In vision, for example, this purpose is served by unconscious, sudden eye movements. The precise " calibration" of these movements is provided by a permanently operating adaptation process, and Chapter 9 describes a simple neural network model demonstrating such capability in a computer simulation.

It is clear that, for the control of their movements, biological organisms and intelligent robots are confronted with tasks that are in many respects similar. Hence, in Part IV of the book, we turn our attention to issues of robotics (Chapter 10). In Chapter 11, it is shown how a robot arm observed by two cameras can learn in the course of a training phase to position its "hand" within the field of view of the cameras by means of visual feedback. Here, by trial and error, the network gradually learns to take properly into account the geometry of the arm and the visual world "seen" by the cameras.

The capability of proper positioning forms the basis for the more complex motor behavior of object gripping. Chapter 12 demonstrates that this ability can also be acquired by a network through learning. However, in view of the higher complexity of the procedure, a network with a hierarchical construction is required. Chapter 12 offers an interesting example of the implementation and training of nets structured in this way.

For the control problems of Chapters 11 and 12, consideration of purely geometrical relationships, *i.e.*, the so-called *kinematics* of the robot, is sufficient. However, for sudden movements, *arm inertia* also plays a role. Chapter 13 shows how the network can take such *dynamic aspects* into account. Here, the network learns the control of “ ballistic arm movements” in a training phase by triggering short torques about the joints of the robot arm.

The preceding examples attempt to illustrate the multitude of tasks that biological brains have learned to master in the course of evolution. At best, we can solve a few isolated tasks today, and in many cases we must develop new solution heuristics which are often ad hoc and without substantial theoretical foundation. This may be compared to the situation prevailing in chemistry during the middle ages, when many chemical reactions were indeed known empirically, but it was not yet appreciated that the huge number of distinct chemical substances could be attributed to barely one-hundred chemical elements. The number of different “ neural modules” in the brain appears to be of the same order of magnitude as the number of chemical elements. This suggests that in the area of information processing, a reduction of the great variety of phenomena to a manageable number of “elements” might also exist.

Our present level of understanding provides us with little more than a vague idea of which principles might be fundamental in this reduction. However, we have available some network models that are encouragingly versatile. The present book illustrates this by demonstrating that the solution of the tasks discussed above can succeed using only a few variants of a single network model, Kohonen’s “self-organizing neural map” (Kohonen 1982a). The biological basis for this model is the organization encountered in many regions of the brain in the form of two-dimensional neuron layers. These layers receive their input signals from nerve fibers emerging either from other neural layers or from peripheral sensory receptors. As a rule, the activities in the individual nerve fibers encode different features of the input stimulus. The nerve fibers coming into contact with a neuron thus determine which input features are particularly effective in exciting this neuron. As experiments show, the connections between neurons and incoming nerve fibers are frequently structured in such a way that adjacent neurons respond to similar input features. This corresponds to a mapping of the (usually higher-dimensional) space of stimulus features, which are coded in the nerve fibre activities, to the two-dimensional neuron layer. Important similarity relationships of abstract stimulus features can be translated into spatial relations of excited neurons of a two-dimensional layer in the manner of a “ topographic map.” Kohonen’s model explains the creation of appropriate connection patterns and the resulting “ maps” of stimulus features as a consequence of a few simple assumptions. The connection pattern forms step by step during a learning process requiring as its only information a sufficiently long sequence of input stimuli. By means of appropriate variants of the basic model, this procedure can be exploited for a broad spectrum of interesting information-processing tasks.

Our book is not limited to the discussion of a series of examples. Rather, each example serves to introduce a mathematical analysis of some particular aspect of the model and, in the course of the discussion, serves as an illustration of the application of a number of important mathematical methods to concrete questions of Neural Computation. The mathematical aspect takes center stage in Part V of the book. First, in Chapter 14, the relationship of the model to procedures for data compression and to factor analysis for the determination of "hidden variables" is presented. This is followed by a discussion of those aspects of the model whose investigation requires a higher degree of mathematical sophistication. The learning process is treated as a stochastic process and described by means of a partial differential equation. Statements concerning convergence properties and statistical fluctuations of the learning process can then be made. The capacity for automatic selection of the most important feature dimensions is discussed mathematically in greater depth, and the relationship to the periodic structure of certain sensory maps in the brain is pointed out. Finally, Chapter 15 discusses the use of local linear transformations as output (needed to solve control tasks), and provides a mathematical analysis of the improvement of the learning process as a consequence of " neighborhood cooperation" between processing units.