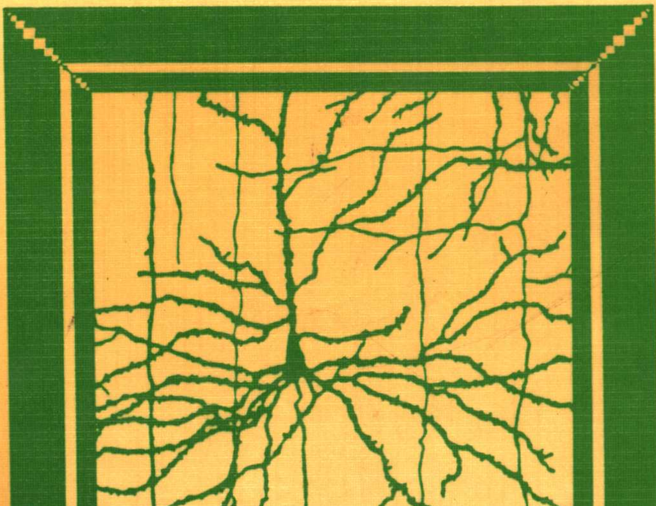
Physics of Neural Networks

B. Müller J. Reinhardt

Neural Networks

An Introduction



Springer-Verlag

B. Müller J. Reinhardt

Neural Networks

An Introduction

With 83 Figures

Springer-Verlag
Berlin Heidelberg New York London
Paris Tokyo Hong Kong Barcelona

Professor Dr. Berndt Müller
Department of Physics, Duke University,
Durham, NC 27706, USA

Dr. Joachim Reinhardt
Institut für Theoretische Physik, J.-W.-Goethe-Universität, Postfach 11 19 32,
D-6000 Frankfurt 1, Fed. Rep. of Germany

Series Editors:

Professor Dr. Jan Leonard van Hemmen
Physik-Department,
Technische Universität München,
D-8046 Garching, Fed. Rep. of Germany

Professor Eytan Domany
Department of Electronics,
Weizmann Institute,
76100 Rehovot, Israel

Professor Klaus Schulten
Department of Physics,
University of Illinois,
Urbana, IL 61801, USA

ISBN 3-540-52380-4 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-52380-4 Springer-Verlag New York Berlin Heidelberg

2155/3150-543210 - Printed on acid-free paper

Physics of
Neural Networks

In memory of

Jörg Briechle

Preface

The mysteries of the human mind have fascinated scientists and philosophers alike for centuries. Descartes identified our ability to think as the foundation stone of ontological philosophy. Others have taken the human mind as evidence of the existence of supernatural powers, or even of God. Serious scientific investigation, which began about half a century ago, has partially answered some of the simpler questions (such as how the brain processes visual information), but has barely touched upon the deeper ones concerned with the nature of consciousness and the possible existence of mental features transcending the biological substance of the brain, often encapsulated in the concept "soul".

Besides the physiological and philosophical approaches to these questions, so impressively presented and contrasted in the recent book by Popper and Eccles [Po77], studies of formal networks composed of binary-valued information-processing units, highly abstracted versions of biological neurons, either by mathematical analysis or by computer simulation, have emerged as a third route towards a better understanding of the brain, and possibly of the human mind. Long remaining – with the exception of a brief period in the early 1960s – a rather obscure research interest of a small group of dedicated scientists scattered around the world, neural-network research has recently sprung into the limelight as a "fashionable" research field. Much of this surge of attention results, not from interest in neural networks as models of the brain, but rather from their promise to provide solutions to technical problems of "artificial intelligence" that the traditional, logic-based approach did not yield.

The quick rise to celebrity (and the accompanying struggle for funding and fame) has also led to the emergence of a considerable amount of exaggeration of the virtues of present neural-network models. Even in the most successful areas of application of neural networks, i.e. content-addressable (associative) memory and pattern recognition, relatively little has been learned which would look new to experts in the various fields. The really hard problem, viz. position- and distortion-invariant recognition of patterns, has not yet been solved by neural networks in a satisfactory way, although we all know from experience that our brains can do it. In fact, it is difficult to pinpoint any technical problem where neural networks have been shown to yield solutions that are superior to those previously known. The standard argument, that neural networks can do anything a traditional computer can, but do not need to be programmed, does not weigh too strongly. A great deal of thought must go into the design of the network architecture and training strategy appropriate for a specific problem, but little experience and few rules are there to help.

Then why, the reader may ask, have we as nonexperts taken the trouble to write this book and the computer programs on the disk? One motivation, quite honestly, was our own curiosity. We wanted to see for ourselves what artificial neural networks can do, what their merits are and what their failures. Not having done active research in the field, we have no claim to fame. Whether our lack of prejudice outweighs our lack of experience, and maybe expertise, the reader must judge for herself (or himself).

The other, deeper reason is our firm belief that neural networks are, and will continue to be, an indispensable tool in the quest for understanding the human brain and mind. When the reader feels that this aspect has not received its due attention in our book, we would not hesitate to agree. However, we felt that we should focus more on the presentation of physical concepts and mathematical techniques that have been found to be useful in neural-network studies. Knowledge of proven tools and methods is basic to progress in a field that still has more questions to discover than it has learned to ask, let alone answer.

To those who disagree (we know some), and to the experts who know everything much better, we apologize. The remaining readers, if there are any, are invited to play with our computer programs, hopefully capturing some of the joy we had while devising them. We hope that some of them may find this book interesting and stimulating, and we would feel satisfied if someone is inspired by our presentation to think more deeply about the important problems concerning the mind and brain.

This book developed out of a course on neural-network models with computer demonstrations that we taught to physics students at the J.W. Goethe University in the winter semester 1988/89. The interest in the lecture notes accompanying the course, and the kind encouragement of Dr. H.-U. Daniel of Springer-Verlag, have provided the motivation to put it into a form suitable for publication. In line with its origin, the present monograph addresses an audience mainly of physicists, but we have attempted to limit the "hard-core" physics sections to those contained in Part II, which readers without an education in theoretical physics may wish to skip. We have also attempted to make the explanations of the computer programs contained on the enclosed disk self-contained, so that readers mainly interested in "playing" with neural networks can proceed directly to Part III.

PC

References to the demonstration programs are indicated in the main text by this "PC logo". We encourage all readers to do the exercises and play with these programs.

Durham and Frankfurt
July 1990

Berndt Müller
Joachim Reinhardt

Contents

I Models of Neural Networks

III Computer Codes

Part I

Models of Neural Networks

1. The Structure of the Central Nervous System

1.1 The Neuron

Although the human central nervous system has been studied by medical doctors ever since the late Middle Ages, its detailed structure began to be unraveled only a century ago. In the second half of the nineteenth century two schools contended for scientific prevalence: the *reticularists* claimed that the nervous system formed a continuous, uninterrupted network of nerve fibres, whereas the *neuronists* asserted that this neural network is composed of a vast number of single, interconnected cellular units, the *neurons*. As often in the course of science, the struggle between these two doctrines was decided by the advent of a new technique, invented by Camillo Golgi around 1880, for the staining of nerve fibres by means of a bichromate silver reaction. This technique was ingeniously applied by the Spanish doctor Santiago Ramon y Cajal in 1888 to disprove the doctrine of reticularism by exhibiting the tiny gaps between individual neurons. The modern science of the human central nervous system thus has just celebrated its first centennial![1]

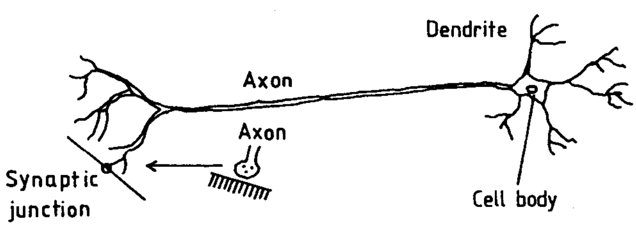


Fig. 1.1. Structure of a typical neuron (schematic).

The detailed investigation of the internal structure of neural cells, especially after the invention of the electron microscope some 50 years ago, has revealed that all neurons are constructed from the same basic parts, independent of their size and shape (see Fig. 1.1): The bulbous central part is called the cell body or *soma* ; from it project several root-like extensions, the *dendrites*, as well as a single tubular fibre, the *axon*, which ramifies at its end into a number of small branches. The size of the soma of a typical neuron is about 10–80 μm, while dendrites and axons have a diameter of a few μm. While the dendrites

[1] Golgi and Ramon y Cajal shared the 1906 Nobel prize in medicine for their discoveries.

serve as receptors for signals from adjacent neurons, the axon's purpose is the transmission of the generated neural activity to other nerve cells or to muscle fibres. In the first case the term *interneuron* is often used, whereas the neuron is called a *motor neuron* in the latter case. A third type of neuron, which receives information from muscles or sensory organs, such as the eye or ear, is called a *receptor neuron.*

The joint between the end of an axonic branch, which assumes a plate-like shape, and another neuron or muscle is called a *synapse*. At the synapse the two cells are separated by a tiny gap only about 200 nm wide (the *synaptic gap* or *cleft*), barely visible to Ramon y Cajal, but easily revealed by modern techniques. Structures are spoken of in relation to the synapse as *presynaptic* and *postsynaptic*, e.g. postsynaptic neuron. The synapses may be located either directly at the cell body, or at the dendrites, of the subsequent neuron, their strength of influence generally diminishing with increasing distance from the cell body. The total length of neurons shows great variations: from 0.01 mm for interneurons in the human brain up to 1 m for neurons in the limbs.

Nervous signals are transmitted either electrically or chemically. Electrical transmission prevails in the interior of a neuron, whereas chemical mechanisms operate between different neurons, i.e. at the synapses. Electrical transmission[2] is based on an electrical discharge which starts at the cell body and then travels down the axon to the various synaptic connections. In the state of inactivity, the interior of the neuron, the *protoplasm*, is negatively charged against the surrounding neural liquid. This resting potential of about −70 mV is supported by the action of the cell membrane, which is impenetrable for Na^+ ions, causing a deficiency of positive ions in the protoplasm (see Fig. 1.2).

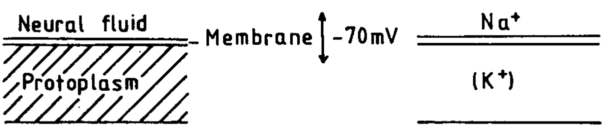


Fig. 1.2. Structure of an axon.

Signals arriving from the synaptic connections result in a transient weakening, or *depolarization*, of the resting potential. When this is reduced below −60 mV, the membrane suddenly loses its impermeability against Na^+ ions, which enter into the protoplasm and neutralize the potential difference, as illustrated in the left part of Fig. 1.3. This discharge may be so violent that the interior of the neuron even acquires a slightly positive potential against its surroundings. The membrane then gradually recovers its original properties and regenerates the resting potential over a period of several milliseconds. During this recovery period the neuron remains incapable of further excitation. When the recovery is completed, the neuron is in its resting state and can "fire" again.

The discharge, which initially occurs in the cell body, then propagates along the axon to the synapses (see Fig. 1.3, right part). Because the depolarized

[2] Detailed studies of the mechanisms underlying electrical signal transmission in the nervous system were pioneered by Sir John Eccles, Alan Lloyd Hodgkin, and Andrew Huxley, who were jointly awarded the 1963 Nobel prize in medicine.

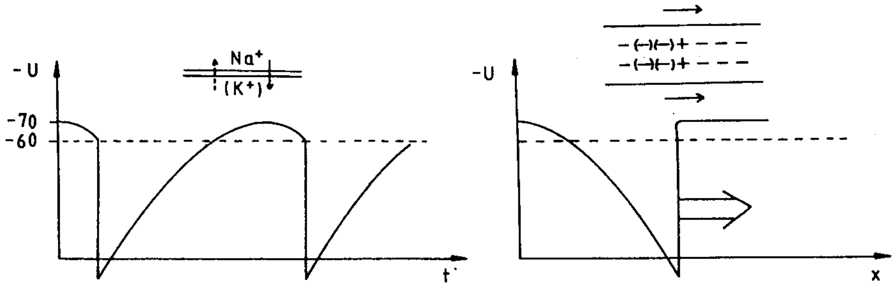


Fig. 1.3. Temporal sequence of activity spikes of a neuron (*left*), which travel along the axon as depolarization waves (*right*).

parts of the neuron are in a state of recovery and cannot immediately become active again, the pulse of electrical activity always propagates in one direction: away from the cell body. Since the discharge of each new segment of the axon is always complete, the intensity of the transmitted signal does not decay as it propagates along the nerve fibre. One might be tempted to conclude that signal transmission in the nervous system is of a digital nature: a neuron is either fully active, or it is inactive. However, this conclusion would be wrong, because the intensity of a nervous signal is coded in the frequency of succession of the invariant pulses of activity, which can range from about 1 to 100 per second (see Fig. 1.4). The interval between two electrical spikes can take any value (longer than the regeneration period), and the combination of analog and digital signal processing is utilized to obtain optimal quality, security, and simplicity of data transmission.

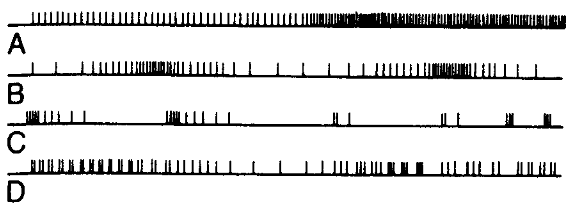


Fig. 1.4. Neuron as pulse-coded analog device: spike trains of some typical neural transmission patterns. The microstructure of the successive intervals is increasingly important from top to bottom (*A*–*D*), indicating messages of growing complexity (from [Bu77]).

The speed of propagation of the discharge signal along the nerve fibre also varies greatly. In the cells of the human brain the signal travels with a velocity of about 0.5–2 m/s. While this allows any two brain cells to communicate within 20–40 ms, which is something like a temporal quantum in the operation of the human central nervous system, it would cause unacceptably long reaction times for peripheral neurons connecting brain and limbs: a person would hit the ground before even knowing that he had stumbled. To increase the speed of propagation, the axons for such neurons are composed of individual segments that are covered by an electrically insulating myelin sheath, which is interrupted from time to time at the so-called Ranvier nodes. The presence of an insulating cover causes the signal to propagate along the axon as in a wave guide from one Ranvier node to the next, triggering almost instantaneous discharge within the whole myelinated segment. This mode of propagation, called *saltatory conduction*, allows for transmission velocities of up to 100 m/s.

The discharge signal traveling along the axon comes to a halt at the synapses, because there exists no conducting bridge to the next neuron or muscle fibre. Transmission of the signal across the synaptic gap is mostly effected by chemical mechanisms. Direct electrical transmission is also known to occur in rare cases, but is of less interest here in view of the much lower degree of adjustability of this type of synapse. In chemical transmission, when the spike signal arrives at the presynaptic nerve terminal, special substances called *neurotransmitters* are liberated in tiny amounts from vesicles contained in the endplate (e.g. about 10^{-17} mol acetylcholin per impulse). The transmitter release appears to be triggered by the influx of Ca^{++} ions into the presynaptic axon during the depolarization caused by the flow of Na^+ ions. The neurotransmitter molecules travel across the synaptic cleft, as shown in Fig. 1.5, reaching the postsynaptic neuron (or muscle fibre) within about 0.5 ms. Upon their arrival at special receptors these substances modify the conductance of the postsynaptic membrane for certain ions (Na^+, K^+, Cl^-, etc.), which then flow in or out of the neuron, causing a polarization or depolarization of the local postsynaptic potential. After their action the transmitter molecules are quickly broken up by enzymes into pieces, which are less potent in changing the ionic conductance of the membrane.

If the induced polarization potential δU is positive, i.e. if the total strength of the resting potential is reduced, the synapse is termed *excitatory*, because the influence of the synapse tends to activate the postsynaptic neuron. If δU is negative, the synapse is called *inhibitory*, since it counteracts excitation of the neuron. Inhibitory synapses often terminate at the presynaptic plates of other axons, inhibiting their ability to send neurotransmitters across the synaptic gap. In this case one speaks of presynaptic inhibition (see Fig. 1.6). There is evidence that *all* the synaptic endings of an axon are either of an excitatory or an inhibitory nature (*Dale's law*),[3] and that there are significant structural differences between those two types of synapses (e.g. the conductance for Na^+ and K^+ changes at excitatory synapses, that for Cl^- at inhibitory synapses).

Under which condition is the postsynaptic neuron stimulated to become active? Although, in principle, a single synapse can inspire a neuron to "fire", this is rarely so, especially if the synapse is located at the outer end of a dendrite. Just as each axon sends synapses to the dendrites and bodies of a number of downstream neurons, so is each neuron connected to many upstream neurons which transmit their signals to it. The body of a neuron acts as a kind of "summing" device which adds the depolarizing effects of its various input

[3] Sir Henry Dale shared the 1936 Nobel prize in medicine with Otto Loewi, who discovered the chemical transmission of nerve signals at the synapse.