

# Physik neuronaler Netzwerke

## Mathematische Modelle liefern neue Erkenntnisse über das Zusammenwirken von Nervenzellen

Von Wolfgang Kinzel

**In den letzten drei Jahren beschäftigen sich weltweit immer mehr theoretische Physiker mit den kooperativen Eigenschaften neuronaler Netzwerke. Mit den Modellen und Methoden der Festkörperphysik ist es gelungen, das Zusammenwirken der Neuronen mathematisch zu berechnen. Obwohl nur wesentliche einfache Mechanismen der einzelnen Zelle berücksichtigt werden, ist das Ganze zu komplexer Informationsverarbeitung fähig. Ein solches Netzwerk kann an Hand von Beispielen selbsttätig lernen, Information zu speichern und wieder zu erkennen. Das System arbeitet schnell und extrem fehlertolerant. Sowohl die Neurobiologie wie auch die Informatik profitieren von diesen Erkenntnissen.**

Es ist immer wieder beeindruckend, wie leistungsfähig unsere heutigen Computer sind. Ein heutiger Superrechner kann eine Million großer Zahlen in einer Sekunde addieren, multiplizieren oder vergleichen. Auf einem Quadratcentimeter kann man mehr als eine Million Transistoren unterbringen, von denen jeder in einer milliardstel Sekunde schalten kann. Sowohl in der Hardware wie auch der Software ist der Fortschritt nicht aufzuhalten.

Um so überraschender ist es, daß ein kleines Kind Fähigkeiten besitzt, an die ein heuti-

ger Superrechner nicht herankommt. Für uns ist es kein Problem, in einer Einkaufsstraße voller Menschen im Vorbeihasten das Gesicht einer bekannten Person zu erkennen. Im Bruchteil einer Sekunde haben wir die Person erkannt, unabhängig vom Menschengewühl, von der Bewegung, von Größe und Orientierung, von Licht und Schatten – und vom neuen Hut unseres Bekannten. Gleichzeitig stehen Erinnerungen und viele andere Informationen über die Person in uns bereit. Das Gehirn macht dies scheinbar mühelos, ein Kind kann dies schon. Ein Computer kann zwar Bilder

elektronisch aufzeichnen, aber er schafft es nur in ganz einfachen Fällen, mit stundenlangen Rechnungen, den Inhalt von Bildern zu verstehen. Woran liegt das?

### Wie arbeitet das Gehirn?

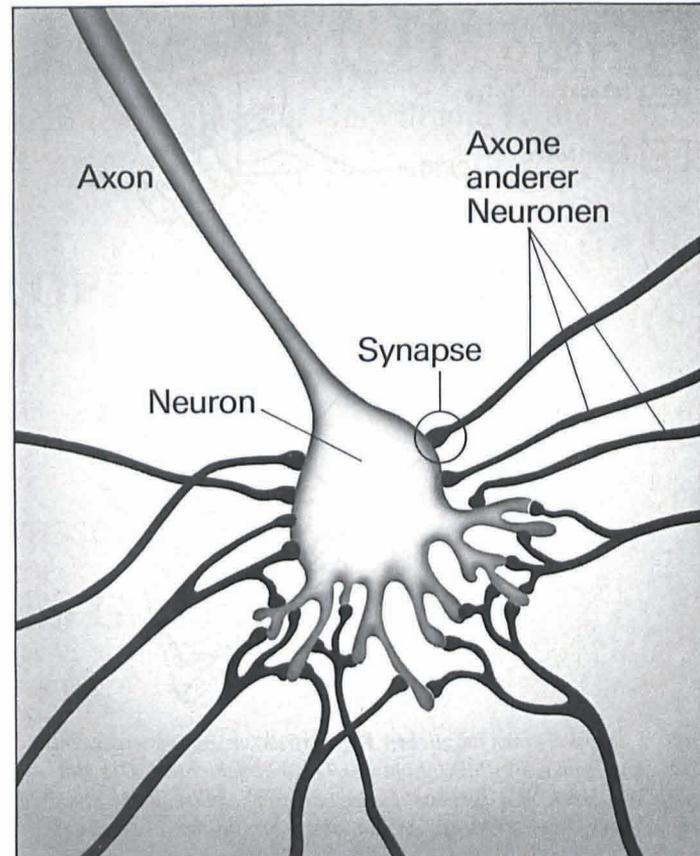
Die materielle Basis solcher Erkennungsleistung sind die Nervenzellen. Im Gehirn haben wir etwa hundert Milliarden solcher Neuronen; die Gehirns substanz besteht aus einem dichten, recht homogenen und scheinbar ungeordneten Geflecht von stark verästelten Zellen. Ebenso wie die Transistoren im Computer, kann man die Neuronen als winzige elektronische Schalter auffassen; jede Zelle sendet elektrische Impulse aus und steuert damit andere Zellen.

Es ist sicherlich nicht die Anzahl der Neuronen, die unserem Gehirn so überlegene Fähigkeiten vermittelt, obwohl immer noch hundert Kubikmeter hochintegrierter Elektronik nötig sind, um so viele Transistoren wie Neuronen zu erhalten. Es liegt auch nicht an der Geschwindigkeit der elektrischen Schalter, denn ein Neuron reagiert in einer tausendstel Sekunde, während ein Transistor einmillionenmal schneller schaltet.

Ein wesentlicher Unterschied zwischen Rechner und Gehirn ist die Arbeitsweise der Schaltelemente. Im Rechner werden Schritt für Schritt die Befehle ausgeführt,



Abb. 1: Neuronen sind die Schaltzellen des Gehirns – im Mikrofoto drei sogenannte Purkinje-Zellen aus dem Kleinhirn. Die Grafik zeigt ihren Aufbau schematisch: Axone anderer Neuronen „schalten“ ihre Signale über Synapsen auf das Neuron, das daraus eine Art Summensignal bildet und über sein Axon an andere Neuronen fortleitet.



die der Mensch vorher in einem Programm festgelegt hat. Dabei sind nur wenige Transistoren beteiligt. Die Neuronen dagegen funktionieren als Netzwerk. Das heißt, sehr viele Nervenzellen sind gleichzeitig an der Informationsverarbeitung beteiligt. Man glaubt heute, daß der grundlegende Unterschied die Netzwerkstruktur ist, und man beginnt an einfachen Modellen zu verstehen, warum Netzwerke ganz andere Eigenschaften als herkömmliche Computer haben.

Ein Nervennetzwerk besteht aus hundert Milliarden von Neuronen. Jedes Neuron ist mit etwa zehntausend anderen direkt verbunden. Die Kontaktstellen zwischen den Nervenzellen – die sogenannten Synapsen (siehe Abb. 1) – sind für Lernen und Erinnern verantwortlich. Die Neuronen senden elektrische Impulse aus, die über die Synapsen an andere Neuronen weitergegeben werden. Komplizierte biochemische Prozesse in den Synapsen bestimmen, wie stark die ankommenden Impulse auf die anliegende Nervenzelle wirken. Jede Nervenzelle sammelt nun die ankommende Information, und je nach ihrer Stärke sendet sie wiederum Impulse aus. Das Gehirn ist also ein dicht vernetztes System von Zellen, die andauernd gegenseitig Informationen austauschen und auf ankommende Informationen reagieren. Kann man unser Denken, das Erkennen von Personen oder das Gedächtnis aus der Funktion von Nervenzellen verstehen?

Die Neurobiologie – die Wissenschaft, die sich mit der Funktion der Nervenzellen befaßt – hat in den letzten Jahrzehnten wichtige Erkenntnisse gewonnen. So ist die einzelne Zelle recht gut verstanden. Auch die komplizierten Vorgänge in den Kontaktstellen – den Synapsen – werden immer mehr erforscht. Für die Aufdeckung der Leitung der elektrischen Ströme entlang der Neuronen und für die Untersuchung der Nervenbahnen vom Auge zur Großhirnrinde gab es Nobelpreise. Dagegen sind die höheren Funktionen des Gehirns heute noch ein Rätsel. Wie wirken die Neuronen zusammen beim Erkennen einer Person? Gibt es eine Nervenzelle, die nur dann Impulse sendet, wenn Sie Ihre Großmutter sehen? Wie ist Ihre Großmutter im Gehirn gespeichert; in einer Synapse oder verteilt im ganzen Gehirn? Wie ruft das Gehirn alle Assoziationen ab, die Ihnen einfallen, wenn Sie

an eine bekannte Person denken? Alle diese Dinge sind im wesentlichen noch unverstanden; es gibt sogar Wissenschaftler, die bezweifeln, daß sich das Gehirn selbst verstehen kann.

## Mathematisches Modell

In den letzten Jahren hat man allerdings dazu einige neue Erkenntnisse gewinnen können. Und zwar nicht mit Experimenten an Ratten oder Affen, sondern mit mathematischen Modellen. Und es sind nicht Biologen, sondern Physiker, die die Modelle gelöst haben. Natürlich ist das Gehirn viel zu kompliziert, als daß man es in Mathematik fassen könnte. Aber man kann sich auf einige wesentliche Mechanismen beschränken und versuchen, das Zusammenwirken der Neuronen damit zu beschreiben. Erste Ansätze dazu waren schon sehr erfolgreich und auch am Institut für Theoretische Physik III der Universität Gießen wurden in den letzten zwei Jahren wichtige Ergebnisse erzielt. Man beginnt zu verstehen, wie durch das Zusammenwirken vieler einfacher Einzelteile komplexe neue Fähigkeiten entstehen. Viele Wissenschaftler und Ingenieure sind sogar überzeugt, daß man diese Erkenntnisse zur Konstruktion neuartiger Computer benutzen kann.

Das Modell beschränkt sich auf folgende Mechanismen: Jedes Neuron hat nur zwei Zustände; entweder es feuert (sendet Impulse aus) oder es ruht. Die Impulse laufen über die Synapsen zu den anliegenden Neuronen und werden dort verstärkt oder abgeschwächt, je nach Art der Kontaktstelle (erregende und hemmende Synapsen). Jede Synapse wird durch eine Zahl  $W$  beschrieben, die die Stärke des elektrischen Potentials in der Empfängerzelle angibt, das durch die einlaufenden Impulse der Sendezelle aufgebaut wird. Ein Neuron sammelt sämtliche ankommenden Signale und addiert sie entsprechend ihrer Stärken auf. Wenn die Summe der Signale größer als ein Schwellenwert ist, feuert das Neuron, sonst ruht es. Die momentane Information ist das Aktivitätsmuster der gesamten Nervenzellen; sie ist also – wie im Computer – binär geschrieben, als Folge von 1 (= Neuron feuert) und 0 (= Neuron ruht).

Damit wird das Nervennetz in eine mathematische Form gebracht. Der Signalaus-

tausch zwischen den Nervenzellen über die Synapsen wird so zu mathematischen Gleichungen mit vielen Unbekannten, nichtlinear und mit Rückkopplung. Solche Gleichungen sind im allgemeinen nur schwer oder gar nicht zu lösen, aber in diesem Fall ist den Physikern mit den Methoden der theoretischen Festkörperphysik die Lösung gelungen. Man kann die Eigenschaften des Netzwerkes in eine mathematische Form bringen, mit der man vieles berechnen und verstehen kann. Warum kann die Festkörperphysik etwas zur Neurobiologie beitragen?

Die statistische Mechanik hat seit Anfang dieses Jahrhunderts Methoden entwickelt, um das Zusammenwirken von unendlich vielen Atomen zu beschreiben. Man versteht, warum Wasser bei  $0^\circ\text{C}$  plötzlich zu Eis wird, warum manche Metalle bei sehr tiefen Temperaturen ihren elektrischen Widerstand verlieren, oder warum Eisen magnetisch wird. In den letzten 13 Jahren hat man sich mit ungeordneten Materialien – den Spingläsern – beschäftigt, deren magnetische Eigenschaften immer noch nicht ganz verstanden sind. Zum Beispiel gehört eine Gold-Eisen-Legierung dazu. In diesen Substanzen gibt es einen Wettbewerb konkurrierender Kräfte zwischen den atomaren „Magnetnadeln“, den man mit einem einfachen mathematischen Modell beschreiben kann. Die Richtung der magnetischen Momente (oben/unten) kann auch als 1 oder 0 geschrieben werden, und die konkurrierenden Kräfte können als Wettbewerb zwischen erregenden und hemmenden Synapsen aufgefaßt werden; somit kann das Modell auch als neuronales Netzwerk aufgefaßt werden. Allerdings sind die Eigenschaften des ganzen Systems nur mit schwieriger Mathematik zu verstehen; dies ist aber den theoretischen Physikern für diese Art von Modellen gelungen.

## Wiedererkennen

Die mathematische Lösung zeigt, daß ein neuronales Netzwerk als assoziativer Speicher arbeiten kann, d. h. eine Menge von Daten wird durch geeignete Wahl der Synapsenstärken gespeichert. Das System kann nun automatisch Teilinformation vervollständigen, also Daten assoziieren (siehe

Abb. 2). Ein verrauschtes Bild wird wieder vollständig restauriert. Oder zu einem Namen, der auch noch fehlerhaft geschrieben sein darf, findet das System von selbst alle persönlichen Daten. Das funktioniert mit sehr vielen Bildern gleichzeitig. Jedes Bild ist ein Aktivierungsmuster der Neuronen, in das das ganze Netzwerk stabil einrasten kann. Abweichungen davon laufen von selbst wieder in die gespeicherten Muster zurück.

Das Bemerkenswerte daran ist, daß das Netzwerk ganz anders arbeitet als ein Computer. Ein Computer speichert Daten wie ein Stapellager. Jedes Fach hat eine Nummer, und wenn Sie Daten suchen, müssen Sie die Nummer kennen oder alle Fächer absuchen. Bei unvollständigen Daten müssen Sie alle Fächer mit Ihrer Information vergleichen. Das Ganze wird Schritt für Schritt von einem Programm gesteuert und benötigt viele Rechenschritte. Das Netzwerk arbeitet ganz anders: Es benötigt kein Programm und hat keine numerierten Schubladen. Alle Daten sind überall in allen Synapsen ein wenig gespeichert. Der Speicher ist also über das ganze Netz verteilt. Das Abrufen von unvollständigen oder fehlerhaften Daten geschieht nach Inhalt; das Netz vervollständigt und verbessert von selbst die Information. Dabei arbeiten alle Neuronen gleichzeitig nach den einfachen Regeln des Modells.

Das führt dazu, daß das Netzwerk extrem fehlertolerant ist. Ein Computer arbeitet mit Ja/Nein-Logik streng nach Vorschrift. Ein kleiner Fehler im Programm oder in der Hardware zerstört meistens den gesamten Ablauf. Im Netzwerk dagegen geschieht alles nur mehr oder weniger genau. Selbst wenn ein großer Teil der Neuronen und Synapsen zerstört wird, funktioniert das Erkennen von fehlerhafter Teilinformation noch. Das Neuronennetz ist äußerst flexibel; wie im biologischen Vorbild können andere Teile Aufgaben übernehmen. Im

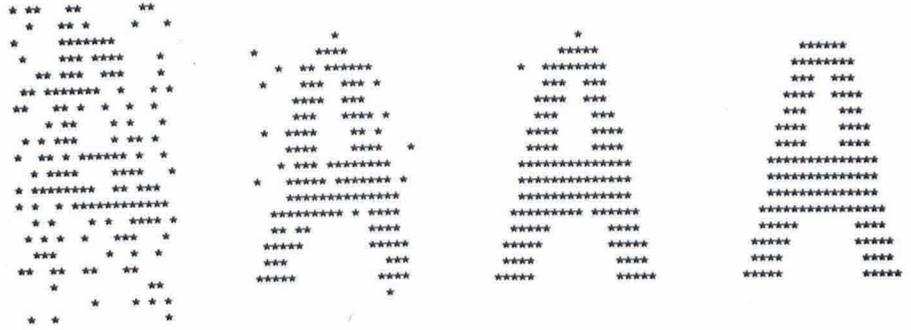


Abb. 2: Dieses Computermodell arbeitet mit  $20 \times 20$  in einem Rechteck angeordneten „Neuronen“. Die Sterne kennzeichnen die aktiven Neuronen. Nicht gezeigt sind die inaktiven Neuronen sowie das Netz, das die Neuronen koppelt. Es hat 30 Buchstabenmuster „gelernt“. Hier erkennt es in nur drei Schritten das aufgeprägte unscharfe Muster als „A“.

Gehirn sterben täglich etwa tausend Neuronen ab, ohne daß wir dies merken, und Verletzungen werden oft von anderen Teilen des Gehirns ausgeglichen.

Die bisherigen Forschungen haben gezeigt, daß neue Eigenschaften durch das Zusammenwirken vieler Einzelteile entstehen können. Im Modell des Nervennetzes sind die Einzelteile nur einfache Schalter (Neuronen) und Verbindungen (Synapsen). Einfache Regeln bestimmen das Zusammenwirken. Dennoch ist das Netzwerk als Ganzes zu komplexer Informationsverarbeitung fähig: Fehlerhafte Daten werden erkannt und schnell mit den dazugehörigen Informationen assoziiert, verschiedene Aufgaben werden selbsttätig an Hand von Beispielen gelernt.

### Lernen

Das Modell kann aber noch mehr, und wiederum liefert uns die mathematische Lösung die Eigenschaften. Es kann nämlich auch selbsttätig gewisse Aufgaben lernen, ohne daß ein Programmierer dem System sagen muß, was genau geschehen soll. Das

Lernen geschieht – wie im biologischen Nervensystem – durch Änderung der synaptischen Stärken: Eine Synapse wird durchlässiger, wenn sie oft benutzt wird. So kann das System durch anpassungsfähige Synapsen Aufgaben lernen. Eine mögliche Aufgabe wäre wieder, eine Menge von Daten assoziativ zu speichern. Wiederum ermöglicht der Wettbewerb von hemmenden und erregenden Kontaktstellen die besonderen Eigenschaften des Netzwerkes.

Lernalgorithmen für solche Netzwerke wurden schon Anfang der sechziger Jahre im Rahmen der Kybernetik vorgeschlagen. Einer davon war das sogenannte „Perzeptron“, für das ein mathematischer Konvergenzbeweis existiert. Aber erst jetzt ist es Dr. Manfred Opper in unserem Institut gelungen, allgemeine Gesetzmäßigkeiten über den Lernprozeß analytisch zu berechnen – mit den Methoden der theoretischen Festkörperphysik.

Man kann nun versuchen, Anwendungen für diese Fähigkeiten zu finden, oder man kann das Modell benutzen, um biologische Vorgänge besser zu verstehen. Zunächst ein Beispiel zur Biologie: In den ersten Lebensjahren sterben die Hälfte der Synapsen ab,

obwohl wir in dieser Zeit wichtige elementare Dinge lernen. In der Erbinformation ist also nicht die detaillierte Verschaltung unseres Gehirns enthalten, sondern es wird zu-

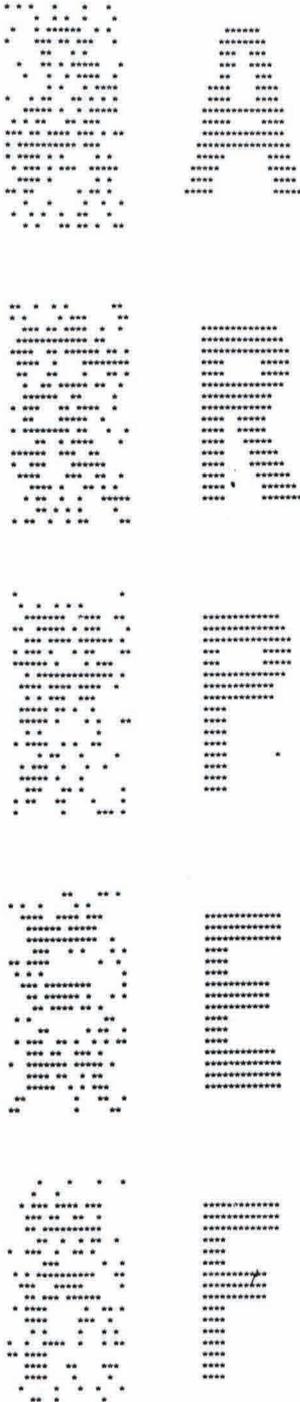


Abb. 3: Die  $20 \times 20$  Neuronen dieses Modells waren zunächst durch 80 000 zufällig gewählte Kopplungen total miteinander vernetzt. Damit es die fünf gezeigten Buchstabenmuster „lernte“, wurden 60 000 der Kopplungen abgebaut, nämlich diejenigen, die bei den Mustern „frustriert“ waren. Wie die Beispiele zeigen, führt auch diese nicht gerade besonders ökonomische Lernmethode zu hoher Erkennungsfähigkeit.

nächst Überfluß produziert, der durch den Lernprozeß wieder abgebaut wird. Die Frage an das Modell ist nun: Kann ein Netzwerk allein dadurch lernen, indem Synapsen absterben? Die mathematischen Gleichungen zeigen, daß dies funktioniert (siehe Abb. 3).

Aber auch erste Anwendungen wurden schon demonstriert, bis jetzt nur als Simulation des Netzwerkes auf einem herkömmlichen Computer. Aber schon werden auch elektronische Chips dazu gebaut, so daß bald die Netzwerke auch in Hardware vorhanden sind. So hat ein Netzwerk gelernt, laut zu lesen; d. h. ein geschriebener englischer Text wird vom Modell laut vorgelesen. Zwar gibt es auch herkömmliche Programme, die dies können; aber das Netzwerk hat dies selbsttätig gelernt, ohne Programm, mit einfachen Mechanismen, schnell und ohne Expertenwissen, nur an Hand von Beispielen.

Wieder andere Netzwerke haben gelernt, medizinische Diagnosen in Notfällen zu stellen, Backgammon zu spielen oder gesprochene Sprache zu erkennen. Dies sind alles erste Versuche und Demonstrationen. Zur Zeit wird weltweit auf diesem neuen Gebiet geforscht, und neue Erkenntnisse und Anwendungen werden mit Sicherheit kommen.

Dieses Beispiel zeigt, daß sich in der Grundlagenforschung oft überraschende interdis-

ziplinäre Zusammenhänge ergeben, zum Teil mit interessanten praktischen Anwendungen. Wohl niemand konnte vor sechs Jahren vorhersagen, daß aus der Festkörperforschung ein nützliches Konzept für die Neurobiologie und die Informatik entstehen würde. In der theoretischen Physik beginnt sich ein neues Gebiet zu entwickeln – die neuronalen Netzwerke –, und die Universität Gießen ist ganz vorn mit dabei.

### Zum Autor

**Prof. Dr. Wolfgang Kinzel**, Jahrgang 1949, ist seit 1986 Professor für Theoretische Physik an der Universität Gießen. Er hat an der Universität Köln Physik studiert und wurde dort 1978 promoviert. Danach war er Wissenschaftlicher Mitarbeiter am Institut für Festkörperforschung der Kernforschungsanlage Jülich. Zu Forschungsaufenthalten war er an der University of Washington in Seattle, am Weizmann Institut of Science in Israel und bei den Bell Laboratorien in New Jersey. Zur Zeit ist er als jüngster Professor Dekan des Fachbereichs Physik.

