Vorlesung Neuroinformatik

Kapitel 1

Grundlagen

R. Der
Universität Leipzig
Institut für Informatik

16. April 1998

Kapitel 1

Grundlagen

1.1 Neurobiologische Grundlagen

1.1.1 Allgemeines

Gegenstand der Neuroinformatik sind künstliche neuronale Netze (KNN) sowohl (i) als Modelle kognitiver Prozesse im Hirn als auch (ii) als neuartige Medien der Informationsverarbeitung. Wunschziel der Neuroinformatik ist die Entwicklung von Algorithmen oder letztendlich Maschinen (Computer) zur Verarbeitung von Informationen nach dem Vorbild Gehirn. Die Umsetzung dieser weit in die Zukunft reichenden Aufgabe erfordert als unabdingbare Voraussetzung ein hinreichend weit gediehenes Verständnis der Organisations- und Funktionsprinzipien der perzeptuellen und kognitiven Prozesse einschließlich ihrer neuronalen "Implementierung". Davon sind wir aber noch beliebig weit entfernt. Zwar ist einerseits die Anatomie des außerordentlich komplexen Systems Gehirn bis hinab auf die zelluläre Ebene erstaunlich gut bekannt. Andererseits wurde, besonders durch die stürmische Entwicklung bildgebender Verfahren in jüngster Zeit, die funktionelle Architektur in vielen wesentlichen Einzelheiten aufgeklärt. Die neuronale Verankerung der höheren kognitiven Prozesse muß aber bisher als weitgehend unbekannt angesehen werden. Weiterführende Literatur [4, 3].

1.1.2 Das Gehirn als System

Das Gehirn kann als das komplexeste uns bekannte System betrachtet werden. Seine Komplexität hat in der Evolution mit der Herausbildung und Entwicklung des Neocortex bei den Säugern beständig zugenommen. Dieser überdeckt beim Menschen die phylogenetisch früheren Teile des Hirns wie das limbische System

und das Stammhirn und zeichnet sich besonders durch seine Furchungen aus,

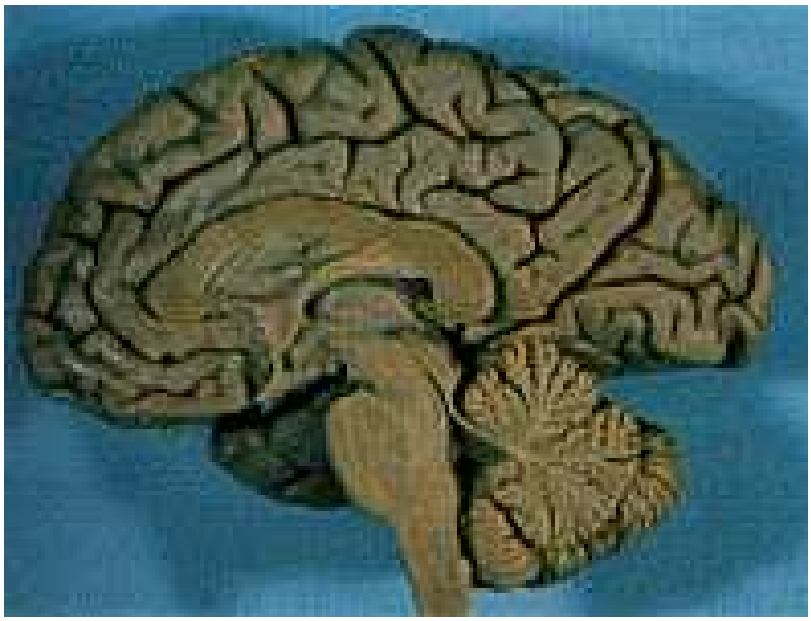


Abbildung 1.1: Senkrechter Schnitt durch das menschliche Hirn mit Hirnstamm, Kleinhirn, limbischem System und Hirnrinde (Neocortex) deutlich erkenntlich an seiner starken Furchung, die der Vergrößerung der Oberfläche dient.

die eine größere Oberfläche für den Neocortex schaffen.

Hinter der Vielzahl hochkomplexer Prozesse, die zielsicher auf dem neuronalen Substrat ablaufen, könnte man eine Vielfalt unterschiedlicher "Bauelemente" vermuten, die, geeignet zusammengeschaltet, die kognitiven Funktionen hervorbringen. Gesicherte neuroanatomische Erkenntnisse belegen aber genau das Gegenteil. In der Tat besteht das Hirn nur aus einigen wenigen Grundtypen von Nervenzellen oder Neuronen, die als die funktionellen Einheiten des Systems zu betrachten sind. Neuere Schätzungen gehen dabei für das menschliche Hirn von ca. 100 Mrd. Nervenzellen (Neuronen) aus. Zusätzlich gibt es im Hirn noch ca. 500 – 1000 Mrd Gliazellen. Die Nervenzellen sind untereinander über die Synapsen verschaltet, wobei jedes Neuron bis zu 1000, teilweise bis zu 10 000 Verbindungen zu anderen Zellen unterhält. [1] In den Abbildungen sind einige Zelltypen zu sehen. Eine ausführliche Zusammenstellung von Daten über das gesamte Nervensystem findet sich im Internet, s. [1].

Neuronen und Synapsen

In der Neuroinformatik werden die biologischen Nervenzellen durch idea-

[1] Purkinjezellen kö nnen im Kleinhirn sogar bis zu 200 000 Verbindungen ausbilden.

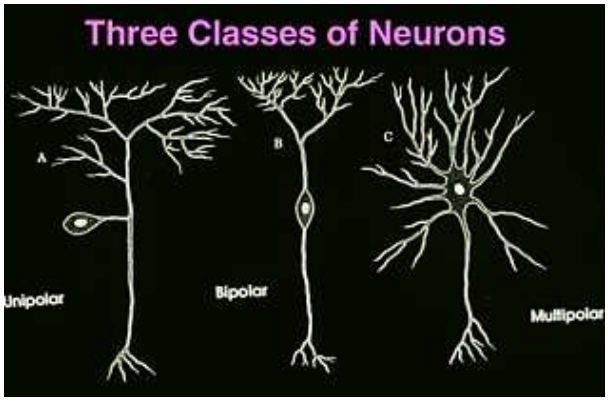


Abbildung 1.2: Einteilung der Zellen nach der Zahl der von ihnen ausgebildeten Pole. Unipolare haben nur einen Fortsatz, der vom Soma ausgeht. Sie überwiegen bei den Avertebraten (Wirbellosen). Bipolare Zellen sind vor allem in der Retina anzutreffen, wo sie die Schaltstelle zwischen den Photorezeptoren und den nachgeschalteten amakrinen und Ganglien–Zellen bilden. Multipolar heißen Zellen mit mehreren Fortsätzen. Überwiegender Zelltyp im Zentralnervensystem der Vertebraten.

lisierte mathematische Modelle repräsentiert. Vorbild für diese Modelle sind Zellen des in Abbildung 1.3 dargestellten Typs (Pyramidenzelle).

Diese Zellen wirken erregend aufeinander ein. Dieser Selbststimulierung der neuronalen Aktivität muß eine ausreichend starke Hemmung entgegenstehen. Diese wird vornehmlich von den Interneuronen aber auch z. B. von den sog. Purkinje–Zellen ausgeübt, s. Abb. 1.4.

Für die Visualisierung der zellularen Struktur des Hirngewebes stehen eine Reihe von Färbungsverfahren zur Verfügung, s. Abb. 1.5 und 1.6.

Von einem funktionell-neuroanatomischen Standpunkt aus gesehen besteht das Neuron aus dem Zellkörper (Soma) mit seinen dendritischen Fortsätzen und dem Axon. Das Neuron empfängt über synaptische Kontakte entweder direkt über sein Soma oder über seine Dendriten elektrische Signale x_i von anderen Neuronen $i = 1, 2, ...$ oder den Muskeln bzw. Sinnesorganen. Diese werden aufsummiert und sobald ihre summarische Wirkung einen gewissen Schwellwert überschreitet, sendet das Neuron selbst ein Aktionspotential y (s.u.) über sein Axon ab. Dieses kann die Signale über eine vergleichsweise lange Strecke weiter-

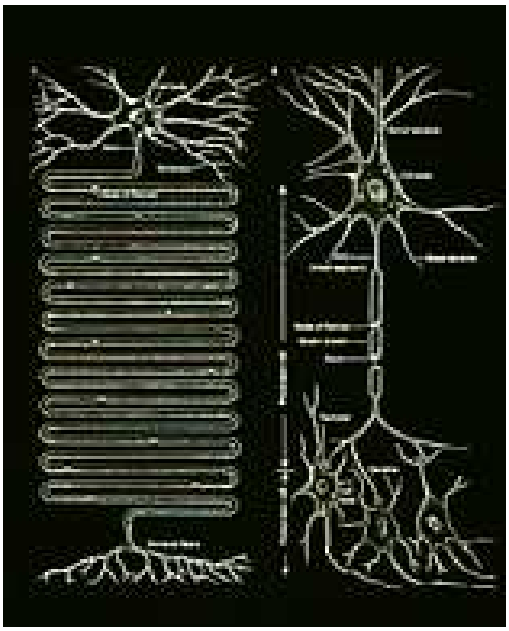


Abbildung 1.3: Schematische Darstellung einer typischen multipolaren Nervenzelle des Zentralnervensystems (links). Zellkörper und dendritische Fortsätze sind deutlich zu erkennen. Zur Veranschaulichung der natürlichen Längenverhältnisse ist das Axon in seiner vollen Länge dargestellt. Die Endverzweigungen des Axons dienen der Kontaktierung anderer Neuronen oder von Muskelzellen (im Fall von Motorneuronen). Rechts im Bild ist die Verschaltung mit weiteren Nervenzellen schematisch dargestellt.

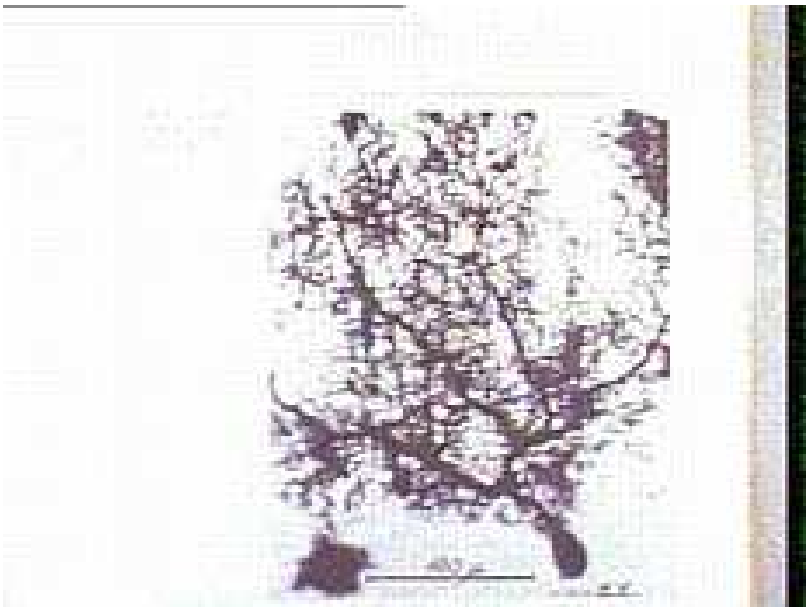


Abbildung 1.4: Eine Purkinje–Zelle. Dieser Zelltyp ist ähnlich wie die Interneurone für hemmende Schaltkreise wichtig. Purkinje–Zellen werden unabhängig von ihrer eigenen Erregung durch Korbzellen gehemmt (Vorwärtshemmung). Andererseits bilden die Purkinje–Zell–Axone hemmende Synapsen auf den Kleinhirnkernen aus. Die ausgelöste Hemmung ist direkt der Aktivität der Purkinje–Zellen proportional (tonische Hemmung).

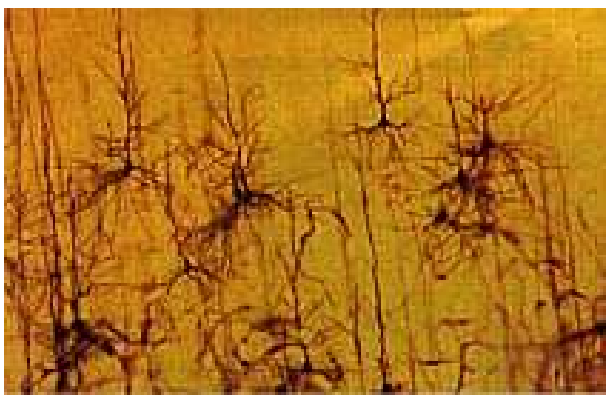


Abbildung 1.5: Durch Golgi-Färbung sichtbar gemachte Nervenzellen.

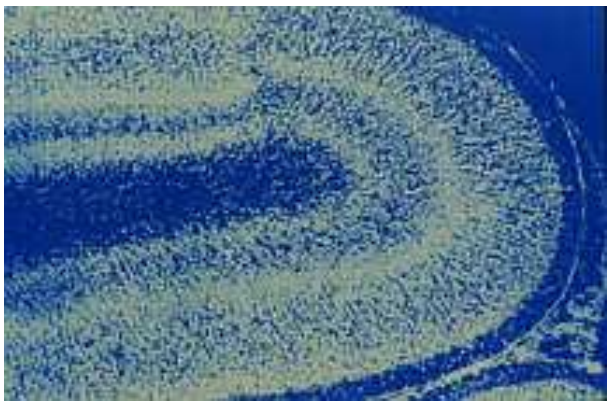
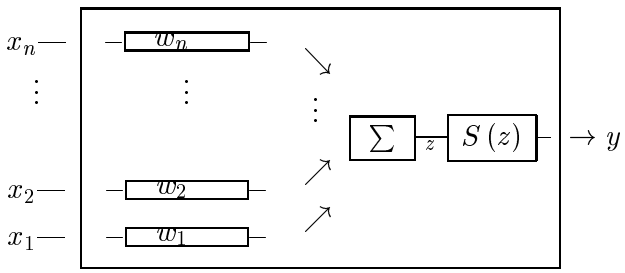


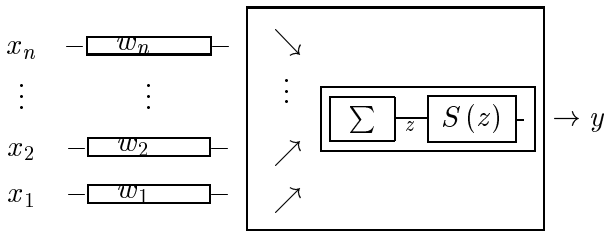
Abbildung 1.6: Die Nissl-Färbung macht die Zytoarchitektur des visuellen Kortex sichtbar.

leiten und kontaktiert dann durch seine Endverzweigungen andere Neuronen:



(1.1)

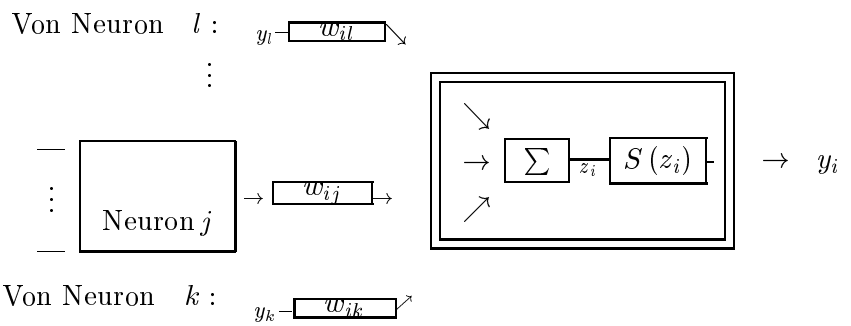
wobei die w_i die Effektivitäten bzw. Leitfähigkeiten der Synapsen (Kontakte) symbolisieren. Wenn wir stärker das Soma als eine Einheit betrachten wollen, dann stellen wir das Neuron einschließlich seiner synaptischen Kontakte so dar:



(1.2)

Funktionell (und stark vereinfachend) kann das Neuron also wie ein elektrisches Bauelelement mit einem definierten wenn auch komplexen Input–Output–Verhalten betrachtet werden. Das System Gehirn ist in diesem Sinne ein komplexes Netzwerk bestehend aus Neuronen, die über die synaptischen Kontakte

miteinander wechselwirken können, ausschnittsweise



(1.3)

Bei der Einordnung dieses Systems ist aber zu beachten, daß wir es hier mit einem offenen, aktiven System zu tun haben, das nicht nur von der Umwelt stark beeinflußt wird sondern das auch gezielt auf seine Umwelt einwirken und diese modifizieren kann, was zu hochkomplexen Rückkopplungseffekten führt. Eine weitere Besonderheit gegenüber den etwa in der Physik betrachteten komplexen Systemen ist die Tatsache, daß das System Gehirn (bzw. das Teilsystem KNN) im Prozeß des Lernens (d. h. der Adaption an die Umwelt) die Wechselwirkungen (Effektivitäten der synaptischen Kontakte) zwischen seinen Elementen (Neuronen) selbst verändern kann, so daß sich die Systemeigenschaften dauernd ändern. Damit ist das Hirn der Prototyp eines adaptiven Systems in seiner höchstentwickelten Form.

1.1.3 Unterschiede Hirn – Computer

Der Unterschied zwischen einem herkömmlichen Computer und dem Hirn folgt aus dem Vergleich von Verarbeitungsprinzipien und Leistungen. Interessant ist etwa ein Vergleich zwischen dem Hirn eines Spatzen und einem heutigen Supercomputer. Wir betrachten ersteres als ein elektronisches System mit ca. 1 Milliarde Bauelementen (Neuronen), die bestenfalls alle 10 ms einen verarbeitungsrelevanten Schaltvorgang vornehmen können (Übergang von einer hohen zu einer niedrigen Feuerrate oder umgekehrt). Dann führt dieses System maximal 10^{11} identifizierbare Schaltvorgänge pro Sekunde aus, liegt also noch unterhalb des Bereiches, in dem moderne Supercomputer operieren.

Obgleich in der reinen Rechenleistung durchaus vergleichbar, sind die kognitiven Leistungen eines Spatzenhirns denen eines jeden Supercomputers weit überlegen. Das betrifft insbesondere alle Fähigkeiten zur Mustererkennung und –Verarbeitung, Lernfähigkeit, die zielgerichtete Steuerung der Bewegungen und

allgemein die Überlebensfähigkeit in einer komplexen Umgebung, die immer neue Anpassungsleistungen erfordert.

Wichtige Charakteristika, die das Hirn vom Computer unterscheiden und die zumindest teilweise den trotz der außerordentlich geringen Geschwindigkeit der neuronalen Schaltkreise erheblichen Performanzvorteil des "Neurocomputers" Gehirn erklären, sind

- Hochgradige Parallelverarbeitung.
- Integration von "Spezialhardware" in Form funktionsspezifischer Neuromodule.
- Keine Trennung von Hard– und Software. Die Funktion des "Neurocomputers" folgt aus seiner Architektur und wird großenteils in der Auseinandersetzung mit der Umwelt erworben.
- Weitere unbekannte Architektur– und Funktionsprinzipien.

Der Unterschied zwischen Hirn und Computer kommt aber insbesondere auch in der Robustheit des Hirns gegenüber Störungen aller Art zum Ausdruck. Eines der markantesten Beispiele sind Hirnverletzungen. Besonders bekannt geworden ist der Fall des Phineas Gage, dessen Verletzungen durch einen Arbeitsunfall in der Computerrekonstruktion in Abb. 1.7 deutlich werden.

Trotz seiner furchtbaren Verletzungen genas Gage nach etwa zwei Monaten und erlangte seine grundlegenden geistigen, sprachlichen und beruflichen Fähigkeiten zurück. Allerdings war nach den Worten seines behandelnden Arztes Dr. Harlow "das Gleichgewicht zwischen seinen geistigen Fähigkeiten und seinen animalischen Neigungen" entscheidend gestört. Diese fatale Charakteränderung bewirkte, daß er auf seinen wechselnden Arbeitsstellen nicht lange geduldet wurde, so daß er im Verlaufe seines immer noch fast zwanzig Jahre währenden Lebens in große persönliche Schwierigkeiten geriet.

1.2 Grundlegende elektrophysiologische Mechanismen

Physiologische Flüssigkeiten (wässrige Salzlösungen) wie $NaCl$ und KCl liegen intra– und extrazellulär in dissoziierter Form, d. h. in Form der anionischen Cl^- und der kationischen Na^+ bzw. K^+ Ionen vor. Die Zellmembran, die die

Abbildung 1.7: Der berühmt gewordene Fall des Phineas Gage zeigt besonders eindringlich die Unvergleichbarkeit von Hirn und Computer. Im Bild eine Rekonstruktion der Verletzungen, die er durch einen Arbeitsunfall erlitt. Obwohl die eingedrungene Eisenstange große Teile seines Hirns zerstörte, erlangte er nach seiner physischen Genesung auch seine grundlegenden geistigen, sprachlichen und beruflichen Fähigkeiten zurück.

äußere Umhüllung des Neurons einschließlich seiner Dendriten und des Axons darstellt, trennt elektrisch diese beiden gut leitenden Medien voneinander. Das System

$$\begin{array}{c} \underline{\quad intrazelluläres\ Medium \quad} \\ \times\times\times\times Membran \times\times\times\times\times\times \\ \overline{\quad extrazelluläres\ Medium \quad} \end{array} \tag{1.4}$$

bildet folglich einen Kondensator mit der Membran als Dielektrikum. Diese ist allerdings teilweise für die Ladungen (Ionen) durchlässig – feinste, kompliziert gebaute Poren in der Zellmembran ermöglichen sowohl passive als auch aktive Transportprozesse durch die ca. 6 nm (≈ 60 Å) dicke Zellmembran hindurch. Für die Na^+ Ionen wirken spezielle Poren wie eine Art Ionenpumpe, die diese Ionen in eine Vorzugsrichtung (von innen nach außen) auch gegen elektrische Kräfte durch die Membran hindurchbefördert. Die erforderliche Energie wird durch Stoffwechselvorgänge bereitgestellt. Die passiven Kanäle sind nur für die K^+ und bei Nervenzellen in geringerem Maße auch für die Cl^-– Ionen durchlässig, deren Hydrathüllen kleiner als die der Na^+ Ionen sind.

1.2.1 Das Ruhepotential

Das Ruhepotential wird durch die ungleiche Konzentration der An– und Kationen auf beiden Seiten der Membran erzeugt. Träger des Potentials ist vornehmlich das $K^+ - Cl^-$ System gemeinsam mit auschließlich in der Zelle vorhandenen kationischen Eiweißmolekülen A^- ($\ominus$), die im Gegensatz zu den K^+ Ionen ($\oplus$) **nicht** hinausdiffundieren können. Unter Vernachlässigung der Cl^- Ionen ergibt sich schematisch folgendes Bild

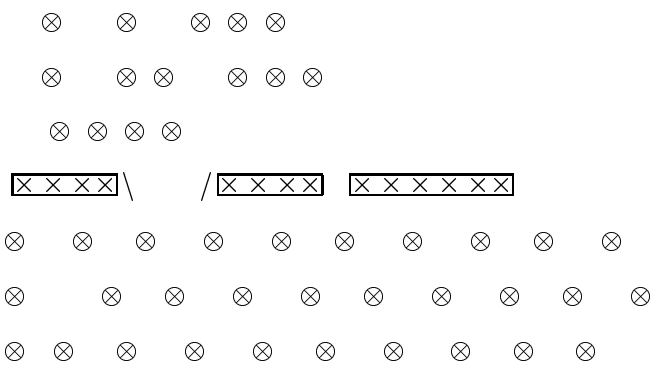
$$\begin{array}{ll} \ominus\ \oplus\ \ominus\ \oplus\ \ominus\ \oplus\ \ominus\ \oplus\ \ominus\ \oplus & \\ \ominus\ \oplus\ \ominus\ \oplus\ \ominus\ \ominus\ \oplus\ \ominus\ \oplus\ \ominus & \text{(intrazellulär)} \\ \oplus\ \ominus\ \oplus\ \ominus\ \oplus\ \ominus\ \ominus\ \oplus\ \ominus\ \oplus & \\ \boxed{\times\times\times\times\times\times} \quad \boxed{\times\times\times\times\times\times\times\times\times} & \\ \text{Pore} \ \nearrow & \\ \oplus & \text{(extrazellulär)} \end{array} \tag{1.5}$$

Die Diffusion entfernt positive Ionen aus dem Inneren, was zu einem Überwiegen der negativen Ladungen und damit zu einem Absenken (Negativierung) des intrazellulären Potentials führt. Allerdings würde dieses Potential durch die ebenfalls diffusionsfähigen Cl^- Ionen wieder aufgehoben. Die Stabilisierung des

Ruhepotentials auf seinen beobachteten Wert bewirken letztendlich die aktiven Transportprozesse durch die Ionenpumpen, die Na^+ Ionen aus dem Inneren der Zelle entfernen. Im Ergebnis der so erzeugten ungleichen Ladungsverteilung[2] erscheint die Zelle **polarisiert**, das Innere der Zelle liegt typischerweise auf einem Potential von ca. -70 mV. (Potentialangaben sind immer relativ zum extrazellulären Raum, dessen Potential also als Null angenommen wird).

1.2.2 Aktionspotential und Reizleitung

Die Wirkung der Ionenpumpen bzw. der Permeabilität der Membran für bestimmte Ionen hängt entscheidend von der anliegenden Potentialdifferenz ab. Wird etwa durch einen externen, überschwelligen Reiz (von einem anderen Neuron) die Membran lokal depolarisiert (Potentialdifferenz unter einen Schwellwert von ca. -50 mV vermindert), dann wird zunächst die Membran kurzzeitig für die Na^+ Ionen durchlässig, diese strömen also massiv solange in das Innere der Zelle ein, bis die negative Ladung lokal kompensiert ist. Kurz danach bewirken einströmende K^+ Ionen einen weiteren **positiven** Potentialaufbau. Diese Abweichung vom Ruhepotential wird als Aktionspotential bezeichnet. In 1.6 ist schematisch das Eindringen der Kationen gezeigt, wobei die dadurch verursachte Depolarisation auch der Umgebung gut zu erkennen sein sollte (dargestellt sind nur die Na^+ Ionen $\otimes$)



(1.6)

Dieser Transportvorgang ist außerordentlich schnell, seine Zeitkonstane liegt bei 0.2 mS. Die Depolarisation strahlt in die Umgebung aus, denn der Zusammenbruch des Potentialgefälles erhöht auch die Na^+ Leitfähigkeit der umliegenden Regionen, was auch dort zu einem Einströmen von Na^+ Ionen führt. Somit

[2] Das Konzentrationsverhältnis (intra:extrazellulär) der K^+ Ionen ist typischerweise $50:1$, das der Na^+ Ionen etwa $1:10$. Die Konzentrationsverteilung der Cl^- Ionen ist ziemlich genau umgekehrt gleich dem der K^+ Ionen.

pflanzt sich die ursprüngliche lokale Potentialstörung über die gesamte Neuronenmembran fort.

Die durch die Depolarisation ausgelöste Durchlässigkeit der Membran ist nicht irrreversibel. Komplizierte Prozesse in der Membran sorgen nach wenigen Millisekunden für deren Inaktivierung und führen damit zur baldigen Wiederherstellung der alten Verhältnisse. Einmal ausgelöste Aktionspotentiale leben lokal also nur eine bestimmte Zeit, so daß die Zelle nach der sog. Refraktärphase (ca. $1 - 3$ mS) für die Aufnahme eines neuen Reizes bereit ist. Die Aktionspotentiale breiten sich über das Axon mit einer Geschwindigkeit von ca. 100 m/S aus. Die Zeitspanne für das Aktionspotential an einer festen Stelle ist i. a. nicht größer als 1/2 mS, d. h. das Aktionspotential hat einen impulsförmigen Zeitverlauf und wird deshalb i. a. als **Spike** bezeichnet. Bei der Ausbreitung behalten diese Impulse ihre Form bei. Folgen von solchen Spikes stellen den Output des Neurons dar.

Spikes

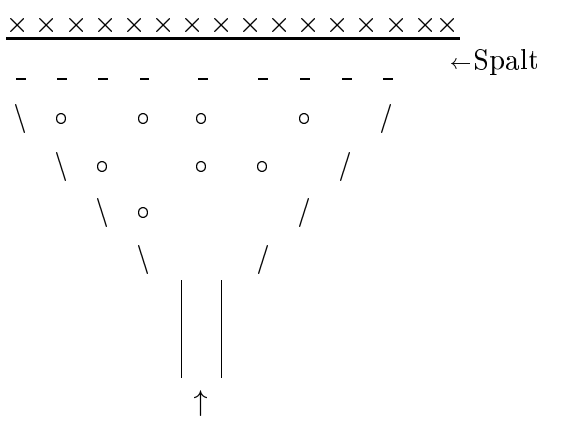
Neben dieser Form der Ausbreitung gibt es noch die passive Ausbreitung elektrischer Reize, die dabei aber abgeschwächt und in ihrem Zeitverlauf verzerrt werden. Diese Art der Reizleitung bestimmt im wesentlichen die Ausbreitung der Inputsignale. Im Gegensatz zu den Spikes haben diese lokalen Potentiale beliebige Stärken, so daß sie auf diese Weise die Stärke der Inputs (z. B. sensorische Reizung) kodieren können. Von dieser Natur sind auch die postsynaptischen Potentiale (s. u.).

1.2.3 Synaptische Kopplungen

Die Übertragung der elektrischen Signale zwischen den Neuronen erfolgt über die Synapsen. Zum synaptischen Bereich rechnen wir den präsynaptischen Endkolben, den ca. $0,1\,\mu m$ schmalen synaptischen Spalt und die sub– oder postsynaptische Membran als den von der Synapse überdeckten Bereich der Zellmembran des kontaktierten Neurons. Die Synapsen haben i. a. eine Ventilfunktion in dem Sinne, daß die Übertragung der Signale immer nur in der Richtung auf

Ventilfunktion der Synapsen

das kontaktierte Neuron hin erfolgen kann (Pfeilrichtung in (1.7)).



(1.7)

Die sog. synaptischen Bläschen oder Vesikel ∘ (Durchmesser ca. 50 nm) im präsynaptischen Endkolben enthalten jeweils eine ziemlich genau definierte Menge[3] (Quantum) einer speziellen chemischen Substanz, den Neurotransmitter[4]. Ein afferenter (ankommender) elektrischer Reiz bewirkt *via* Depolarisation der präsynaptischen Membran die Freisetzung und in Folge die Ausschüttung in den synaptischen Spalt dieses Überträgerstoffes. Einmal im Spalt bewirken die Transmitterstoffe ihrerseits eine Depolarisation der subsynaptischen Membran und lösen dadurch ein Aktionspotential in der kontaktierten (postsynaptischen) Zelle aus. Der gesamte Prozeß ist sehr schnell, die synaptische Latenz beträgt ca. 0.2 mS. Die Menge der freigesetzten Quanten (genauer die Wahrscheinlichkeit für ihre Freisetzung) hängt durchaus von der Stärke des präsynaptischen Aktionspotentials ab, so daß diese Information weitergegeben werden kann. Ähnlich der automatischen Inaktivierung der Zellmembran werden auch die Transmitter nach erfolgter Signalübertragung innerhalb kürzester Zeit wieder von dem synaptischen Endkolben resorbiert, wodurch die Synapse in ihren Ausgangszustand zurückkehrt.

Neurotransmitter

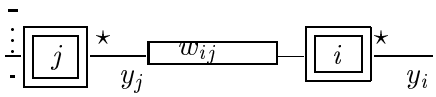
Entscheidend für die Funktion neuronaler Netze ist die Tatsache, daß sich die synaptische Effektivitäten, d. h. die Übertragungseigenschaften der Synapsen, modifizieren lassen. Sie übertragen bei häufiger Benutzung besser als bei seltener. Das ist auch der wesentliche Inhalt der Hebbschen Lernregel. Diese führt das Lernen auf eine erregungsproportionale Verstärkung der synaptischen Verbindungen zwischen gleichzeitig aktiven Neuronen zurück. Betrachten wir

Synaptische Effektivitäten

[3] Man schätzt, daß ein Vesikel ca. 1000 bis 10 000 Transmittermoleküle (z. B. Acetylcholin) enthält und daß pro Synapse 200 bis 2000 Quanten freigesetzt werden.

[4] Wir besprechen hier nur die chemischen Synapsen. Die selteneren elektrischen Synapsen funktionieren anders.

zwei Neuronen wie in (1.3)



(1.8)

Hebbsches Lernen

Die Veränderung Δw_{ij} der synaptischen Effektivität [5] w_{ij} ist nach Hebb

$$\Delta w_{ij} = \epsilon y_i y_j \tag{1.9}$$

Sind nun, wie in 1.8 durch den Stern $\star$ symbolisiert, die Outputs beider Neuronen i, j auf einem hohen Niveau, so wird die synaptische Verbindung w_{ij} gemäß 1.9 deutlich verstärkt. Der typische Wertebereich für die Lernrate ist $0.01 < \epsilon < 0.1$.

Auf der Plastizität der Synapsen beruht allgemein die Lern– und Gedächtnisfunktion neuronaler Netze. Trotz ihrer Einfachheit und ihrer nur lokalen Einflußnahme (Δw_{ij} ist nur von den unmittelbar beteiligten Neuronen abhängig) ist die Hebbsche Lerndynamik ein grundlegendes Wirkprinzip für das Erlernen globaler Fähigkeiten durch das neuronale Netz.

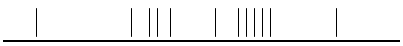
1.3 Das mathematische Modell des Neurons

1.3.1 Neuronale Informationskodierung

Die impulsförmigen Aktionspotentiale anderer Neuronen erreichen die Dendriten oder das Soma (Zellkörper) eines Empfängerneurons über die synaptischen Kontakte. Diese elektrischen Signale werden passiv über die Dendriten und aktiv nach dem in Abschnitt 1.2 beschriebenen Mechanismus über die gesamte Zellmembran fortgeleitet. Das Neuron integriert diese Potentiale mit einer sehr kleinen Zeitkonstanten auf. Überschreitet deren summarischer Effekt am Axonhügel eine bestimmte Schwelle, so sendet das Neuron über sein Axon ebenfalls ein Aktionspotential (Spike) aus, das seinerseits andere Neuronen aktivieren kann. Systemtheoretisch gesehen ist das Neuron also durch ein bestimmtes Input–Output–Verhalten gekennzeichnet, das eine komplexe Zeitstruktur aufweist. Als Element des informationsverarbeitenden Systems Gehirn betrachtet ist zu fragen, in welcher Weise die Information in einer solchen funktionellen Architektur kodiert ist und wie das Neuron diese Information lesen und verarbeiten kann.

[5] Die betrachtete Synapse kontaktiert das Neuron i (Zielneuron). Wir schreiben w_{ij} für die Kopplungsstärke (Effektivität) der Synapse, d.h. der erste Index bezieht sich immer auf das Zielneuron.

Es gibt hier eine Vielzahl von Ansätzen, von denen wir nur die zwei grundlegenden nennen wollen. Der erste geht davon aus, daß die Zeitstruktur der Signale wesentlich ist. Im Extremfall kann man annehmen, daß die Information in einer Spikefolge durch den zeitlichen Abstand zwischen aufeinanderfolgenden Spikes (die alle etwa die gleiche Amplitude haben) kodiert ist, schematisch



(1.10)

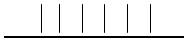
Dem entspricht in der Nachrichtentechnik die Puls**phasen**modulation. Für deren Realisierung als neuronale "Sprache" gibt es bisher aber relativ wenig empirische Anhaltspunkte, da sie eine recht genaue Messung der zeitlichen Abstände voraussetzt und sie demzufolge nicht die für biologische Systeme erforderliche Robustheit aufweist. Andererseits ermöglicht diese Zeitstruktur der neuronalen Signale Synchronisationsphänomene zwischen verschiedenen Neuronen, was als ein Lösungsansatz zum Verständnis des Binding–Problems allgemein akzeptiert wird.

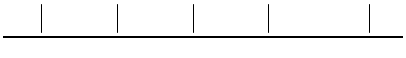
Pulsphasen - Modulation

Grundlegend für die in dieser Vorlesung zu besprechenden künstlichen neuronalen Netze ist der zweite Ansatz, daß nämlich die Information, ähnlich der Puls**frequenz**modulation in der Nachrichtentechnik, einzig durch die Feuerrate (das ist die Frequenz mit der die Spikes abgesendet werden) kodiert ist. Damit die Feuerrate als robuster Informationsträger fungieren kann, müssen diese zeitlichen Abstände dann über einen gewissen Minimalzeitraum hinreichend konstant bleiben (der Minimalzeitraum ist frequenzproportional). Die Feuerraten lassen sich durch kontinuierliche Zahlen abbilden,

Ratenkodierung

Kodierung ≈ 1

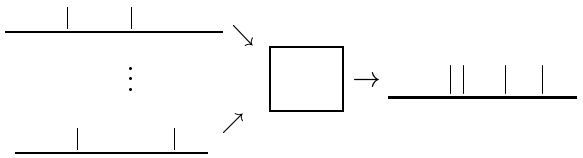
 Kodierung $\approx 1/2$

 Kodierung ≈ 0.1

die geeignet skaliert werden, so daß sie beispielweise zwischen 0 (Neuron feuert nicht) und 1 (Neuron feuert mit maximaler Frequenz) liegen. Diese Neuronen werden auch als stationäre Neuronen bezeichnet.

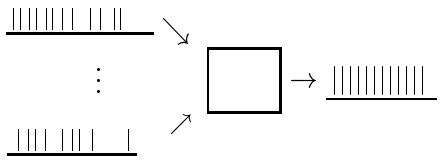
Im Extremfalle des McCulloch–Pitts Neurons (s.u.) werden nur zwei Situa-

tionen unterschieden, der Fall niedriger Aktivität



(1.11)

wird durch den Output $y = 0$, der Fall hoher Aktivität

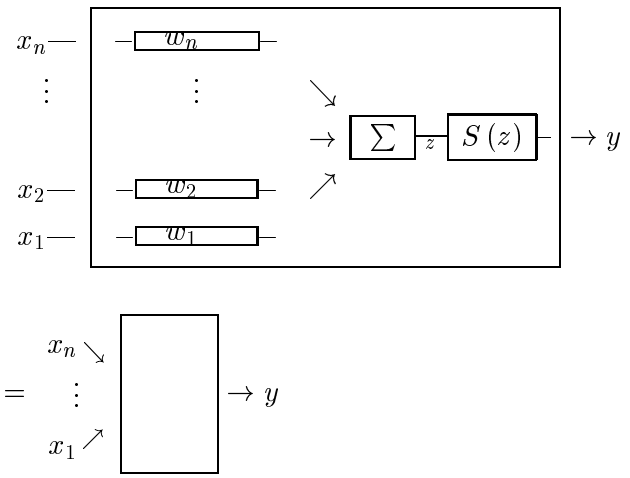


(1.12)

durch $y = 1$ kodiert. In 1.11 und 1.12 bezeichnet $|||\ |||\ ||$ wieder eine zeitliche Abfolge von Spikes.

1.3.2 Funktionelles Schema des stationären Neurons

Für die Untersuchung künstlicher neuronaler Netze fassen wir das in 1.1 eingeführte Neuron als ein kompaktes Bauelement auf,



(1.13)

das mathematisch durch seine Transferfunktion modelliert wird. Wie in Abschnitt 1.3.1 besprochen, sind dabei die Inputs und Outputs als statische Größen aufzufassen (Ratenkodierung), von deren elektrophysiologischer Rolle oder Interpretation abstrahiert wird.

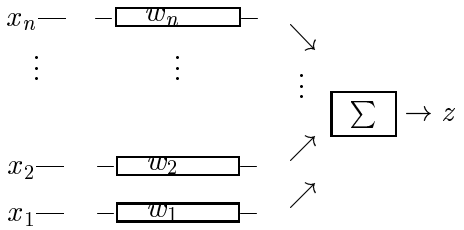
1.3.3 Das mathematische Modell des stationären Neurons

Das stationäre Neuron als funktionelle Einheit künstlicher neuronaler Netze wird modellhaft durch seine Transferfunktion

$$y = f(\vec{x}) \tag{1.14}$$

beschrieben, diese bildet die Inputs $\vec{x} = (x_1, ..., x_N)$ auf den Output y des Neurons ab. Wir setzen immer eine Ratenkodierung der neuronalen Information voraus, d.h. sowohl die Inputs x_i als auch der Output y kodieren jeweils eine Feuerrate, s. Abschnitt 1.3.1. Explizit berechnet das Neuron in idealisierter Form zunächst sein postsynaptisches Potential (PSP) z

Postsynaptisches Potential

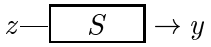


(1.15)

oder

$$z = \sum_{i=1}^{n} w_i x_i = \vec{w} \cdot \vec{x} \tag{1.16}$$

als mit den synaptischen Effektivitäten w_i gewichtete Summe seiner Inputs, vergleicht diese mit einer Schwelle U und bildet das Ergebnis dann nichtlinear auf das Intervall $[0, 1]$ ab, letzteres besorgt das Element



in 1.13, das die biologische Funktion des Axonhügels mathematisch modelliert, explizit

$$y = S\,(z - U) \tag{1.17}$$

Die Transferfunktion

Transferfunktion

$$y = f\,(\vec{x}) = f\,(\vec{x} \mid \vec{w}) \tag{1.18}$$

lautet folglich in expliziter Form

$$\boxed{y = S\,(\vec{w} \cdot \vec{x} - U)} \tag{1.19}$$

Für die feed–forward Netze ist S eine sigmoide Funktion, z. B.

$$S\,(z) = \frac{1}{1 + e^{-\beta z}} \tag{1.20}$$

wobei β die Steilheit des Anstieges parametrisiert, s. Abbildung 1.8. Die Ausgabefunktion S reflektiert den nichtlinearen Zusammenhang zwischen Input und Output, insbesondere in Form des Sättigungsverhaltens des biologischen

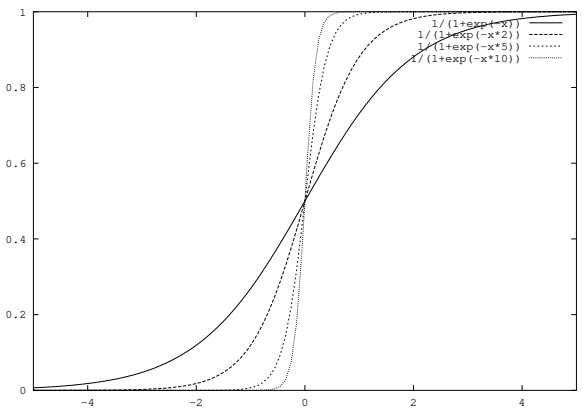


Abbildung 1.8: Die Ausgabefunktion für Parameter $\beta = 1$, $\beta = 2$, $\beta = 5$ und $\beta = 10$.

Neurons: Für sehr große PSP ist die Feuerrate praktisch nicht mehr vom PSP abhängig.

Der Trick mit der Schwelle

Die Schwelle U kann durch die Einführung eines fiktiven Inputs $x_0 = 1$ und $w_0 = -U$ und Definition des Skalarproduktes $\vec{x} \cdot \vec{w} = \sum_{i=0}^{n} w_i x_i$ berücksichtigt werden. Dann ist die allgemeine Transferfunktion $y = f(\vec{x})$

$$\boxed{f(\vec{x}) = f(\vec{x} \mid \vec{w}) = S(\vec{w} \cdot \vec{x})} \tag{1.21}$$

Wenn die Abhängigkeit von den Parametern explizit zum Ausdruck gebracht werden soll, schreiben wir $f(\vec{x})$ als $f(\vec{x} \mid \vec{w})$.

Ein Spezialfall der sigmoiden Ausgabefunktion ist die Sprungfunktion

$$\theta(z) = \begin{cases} 1 & \text{für} \quad z \geq 0 \\ \\ 0 & \text{für} \quad z < 0 \end{cases} \tag{1.22}$$

Binäres Neuron

wobei $\theta(z) = \lim_{\beta \to \infty} S(z)$. Dieses binäre Neuron wurde zuerst von McCulloch und Pitts eingeführt[2]. Für eine Reihe von Netzwerkmodellen hat es sich als günstig erwiesen, den Zustand minimaler Erregung des Neurons statt mit 0 mit -1 zu kodieren, d. h.

$$S(z) = \tanh(\beta z) = \frac{e^{\beta z} - e^{-\beta z}}{e^{\beta z} + e^{-\beta z}} \tag{1.23}$$

Bipolares Neuron

Dem McCulloch–Pitts Neuron entspricht dann das bipolare Neuron mit

$$S(z) = sign(z) = \begin{cases} 1 & \text{für} \quad z \geq 0 \\ \\ -1 & \text{für} \quad z < 0 \end{cases} \tag{1.24}$$

Bemerkungen: Für die Anwendungen mit feed–forward Netzwerken (s.u.) wird meist die Kodierung gemäß 1.20 gewählt. Die Ausgabefunktion 1.23 ist oft nützlich, wenn das Netz auch negative Werte ausgeben soll. Für die Steuerung eines mobilen Roboters kann das beispielweise interessant sein, wenn das Ausgabeneuron direkt die Radansteuerung vorgibt (negative Werte bei Rückwärtsfahrt).

Literaturverzeichnis

[1] E. H. Chudler. Brain facts and figures. http://weber.u.washington.edu/ chudler/facts.html, 1996.

[2] W. S. McCulloch and W. Pitts. A logical calcus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[3] J. G. Nicholis, A. R. Martin, and B. G. Wallace. *From neuron to brain.* Sinauer Associates, Sunderland, Massachusetts, 1992. 3rd edition.

[4] K. R. Popper and J. C. Eccles. *Das Ich und sein Gehirn.* Piper, München, 1982.