Neuronale Netzwerke

VL Algorithmisches Lernen, Teil 2a

Norman Hendrich

University of Hamburg
MIN Faculty, Dept. of Informatics
Vogt-Kölln-Str. 30, D-22527 Hamburg
hendrich@informatik.uni-hamburg.de

20/05/2009

VL Algorithmisches Lernen (SS 2009)

- Part 1 (Wolfgang Menzel)
  - Lernen symbolischer Strukturen
  - (Instanzenbasierte Verfahren)
  - Probabilistische Methoden
- Part 2 (Norman Hendrich)
  - Lernen mit konnektionistischen Modellen (Neuronale Netze)
  - Dimensionalitätsreduktion, PCA
- Part 3: (Jianwei Zhang, Norman Hendrich)
  - Support-Vektor Maschinen
  - Funktionsapproximation
  - Reinforcement-Lernen
  - Anwendungen in der Robotik

Terminplanung: Part 2

- 20/05/2009 neural networks
- 27-28/05/2009 neural networks

- 10-11/06/2009 dimensionality reduction

Terminplanung: Part 3

- 17-18/06/2009 support vector machines
- 24-25/06/2009 function approximation
- 01-02/07/2009 reinforcement learning (1)
- 08-09/07/2009 reinforcement learning (2)
- 15-16/07/2009 applications in robotics

Disclaimer

Important note: this is the first version of the neural network slides, please report all errors and inconsistencies!

hendrich@informatik.uni-hamburg.de
zhang@informatik.uni-hamburg.de

Outline

Neural Networks

- a.k.a. *Connectionist models*
- a.k.a. *Parallel Distributed Processing*
- in German: *Neuronale Netzwerke*

- A general paradigm for computation
- based on parallel information processing
- by a large number of simple interconnected units

- both bio-inspired and theoretical models
- sub-symbolic processing and representations
- therefore, long-going conflict with *computationalism*

Interdisciplinary research

Research on neural networks combines ideas and results from many different disciplines:

- medical science
- neuroscience
- cognitive science
- psychology
- computer science
- mathematics
- theoretical physics
- . . .

Motivation — Why study neural networks?

- understanding the *principles of biological information processing*
- and especially the human brain

- as an alternative to traditional artificial intelligence (computationalism)
- attempt to mimic the performance of the human brain with artifical neural networks
- apply artificial neural networks to application problems: vision, pattern recognition, associative memory, etc.
- graceful degradation and fault-tolerance

Timeline

- ca. 1700 BC: brain first mentioned in Egyptian papyrus
- 1786: Galvani stimulates frog-muscle with electricity
- 1873: Golgi silver-nitrate stain
- 1906: Ramón y Cajal nobel-prize
- 1909: Brodmann classification of brain areas
- 1943: McCulloch-Pitts model
- 1949: Hebb learning hypothesis
- 1952: Hodgkin-Huxley neuron model
- 1957: Rosenblatt Perceptron
- 1960: Widrow and Hoff Adaline
- 1969: Minsky & Papert Perceptrons book and analysis
- 1982: Kohonen self-organizing maps
- 1982: Hopfield model for associative memory
- 1986: Rumelhart et.al: Backpropagation algorithm

References: textbooks

- Raul Rojas, *Neural Networks — A Systematic Introduction*, Springer (1996)

  page.mi.fu-berlin.de/rojas/neural/
- Simon Haykin, *Neural Networks and Learning Machines*, Pearson International (2009)
- Kandel, Schwarz, Jessell, *Principles of Neural Science*, Prentice Hall (2000)

References: classics

- M.L. Minsky, S.A. Papert, *Perceptrons*, MIT Press (1969)
- D.E. Rumelhart, J.L. McClelland, and the PDP research group, *Parallel Distributed Processing — Explorations in the Microstructure of Cognition* (2 volumes), MIT Press (1986)
- J.A. Anderson (Ed.), *Neurocomputing — Foundations of Research*, MIT Press (1989)
- D.A. Amit, *Modeling Brain Function - the world of attractor neural networks*, Cambridge University Press (1989)
- T. Kohonen, *Self-Organizing Maps*, Springer (2001)

References: software

- SNNS: Stuttgart neural network simulator
  www.ra.cs.uni-tuebingen.de/SNNS/
  www.ra.cs.uni-tuebingen.de/downloads/JavaNNS/
- Matlab neural network toolbox
- UCI Machine Learning Repository
  http://archive.ics.uci.edu/ml/
- suggestions welcome!

Eight major aspects of a PDP model

- a *set of processing units*
- the current *state of activation*
- an *output function* for each unit
- a *pattern of connectivity* between units
- a *propagation rule* for propagating patterns of activities through the network of connectivities
- an *activation rule* for combining the inputs impinging on a unit
- a *learning rule* whereby patterns of connectivity are modified by experience
- the *environment* within the system must operate

(Rumelhard, McClelland, et.al. 1986)

A set of processing units

- the basic elements of any PDP model
- usually called *neurons*
- N units, $u_1 \ldots u_N$

- often useful to distinguish three basic types: *input*, *output*, *hidden* units

Two choices of representation:

- *one-unit-one-concept* representation
- or simply abstract units, and concepts formed by *patterns* of unit activity

The state of activation

- system state at time t
- a vector of N real numbers $a_i(t)$
- the current (internal) *activation* states of each unit u_i

Different choices:

- continuous values $a_i(t)$, bounded or unbounded
- discrete values, e.g. $\{0, 1\}$ or $\{-1, +1\}$

The output function for each unit

- one output function $f_i(a_i(t))$ for every unit u_i
- the actual output signal of the unit at time t
- summarized as *output vector* $o(t) = (f_1(t), \ldots, f_N(t))$

Again, many choices:

- identity function, $f(x) = x$
- usually, f some sort of threshold function so that a unit has no affect on other units unless its internal activation exceeds a certain value
- also possible: f a stochastic function of the internal activation state

The pattern of connectivity

- specifies how units are connected to each other
- represented as a *weight matrix* W
- w_{ij} represents the strength of the connection from unit u_j to unit u_i

- if w_{ij} positive, unit u_j excites u_i
- if w_{ij} negative, unit u_j inhibits u_i
- $w_{ij} = 0$ implies no direct connection from u_j to u_i

- in *layered networks*, only certain connections exist
- more efficient to use different weight-matrices between layers, instead of a single big matrix with many 0-entries

The propagation rule

- takes the current output values $o(t)$
- and the connectivity matrix W
- to calculate the *net inputs* for each unit u_i

- most popular choice is the vector product:

$$\mathrm{net}(t) = W \cdot o(t)$$

- some models use different propagation rules for excitory and inhibitory connections

The activation rule

- describes the time-evolution of the network
- how to calculate the new activation states $a(t+1)$ from the current states and the propagated inputs
- synchronous or asynchronous update of the units

Again, many choices:

- identity function: $a(t+1) = Wo(t) = \mathrm{net}(t)$
- more general: $a(t+1) = F(a(t), \mathrm{net}(t))$
- most popular are non-decreasing differentiable functions,
  $a(t+1) = F(\mathrm{net}_i(t)) = F(\sum_j w_{ij} o_j)$
  e.g. $F(x) = \mathrm{sgn}(x)$ or $F(x) = \tanh(x)$

The learning rule

- describes how the *connectivity changes*
- as a function of *experience*

- developing of new connections
- loss of existing connections
- modification of connection strength

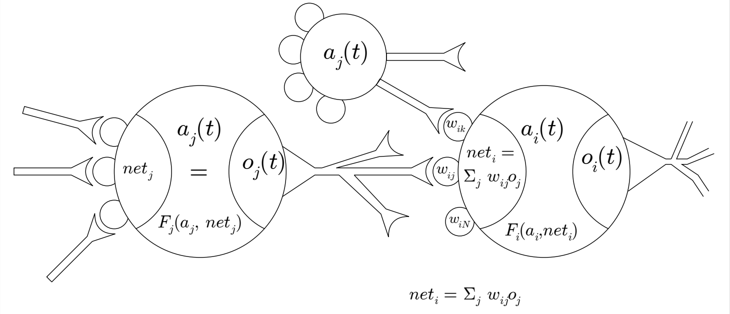
- Hebb learning rule: if units u_j and u_i are both active, strengthen their connection w_{ij}
- unsupervised or supervised (with teacher input)

Representation of the environment

- specify the input patterns to the network

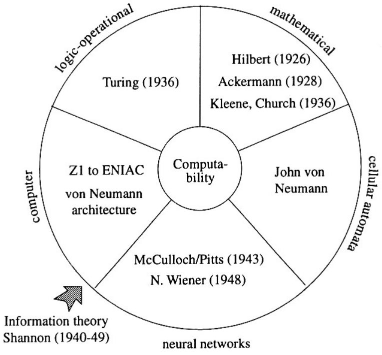
- e.g. time-varying stochastic function over the space of possible input patterns
- many depend on the history of inputs to the system as well as the network outputs
- typically, environment consist of a set of known training and test input patterns.

Putting it all together



- neurons u_i, activation states a_i
- outputs o_i propagated via weights w_{ij}
- net neuron input is $\sum_j w_{ij} o_j(t)$

Five models of computation



- neural networks equivalent to the other classes

(Rojas 1996)

Three models of learning

- *supervised learning* (corrective learning)
  - learning from *labelled examples*
  - provided by a knowledgable external supervisor
  - the correct outputs are known for all *training samples*
  - the type of learning usually assumed in NN studies
- *reinforcement learning*:
  - no labelled examples
  - environment provides a scalar feedback signal
  - combine *exploration* and *exploitation* to maximize the reward
- *unsupervised learning*:
  - no external feedback at all

Outlook

We will look at several PDP/NN architectures:

- McCulloch-Pitts model
- Perceptron
- Multilayer perceptron with backpropagation learning
- Recurrent netwoks: Hopfield-model and associative memory
- Self-organizing maps

- their units and interconnections
- their learning rules
- their environment model
- and possible applications

Overview of the human brain
a bit of neurophysiology

- total brain weight $1.3 \dots 1.4\,\mathrm{kg}$
- power consumption: ≈ 20 Watt
- about 2% of body weight, 20% of body oxygen, 25% of glucose

- estimated 10^{11} neurons, many different types
- up to 10^{5} connections per neuron
- roughly 10^{15} synapses overall
- nerve-pulse propagation roughly 100 m/s

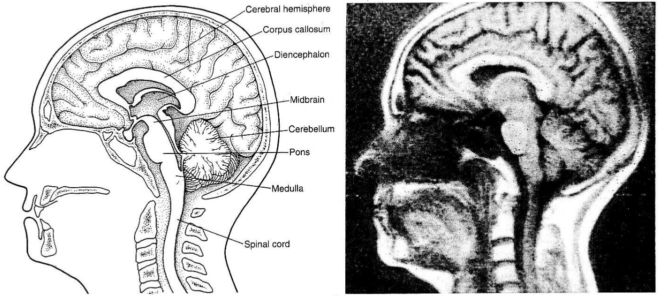
- probably, most complex structure in the universe

Methods of study

- (neuro-) physiology and pathology
- psychology
- optical microscopy (Golgi stain, fluorescent markers)
- electron microscopy
- functional brain imaging (e.g. PET, NMR)
- EEG, SQUID recordings
- experiments on single neurons
- simulation

- but still no complete picture. . .

Lateral view of the human brain



A

B

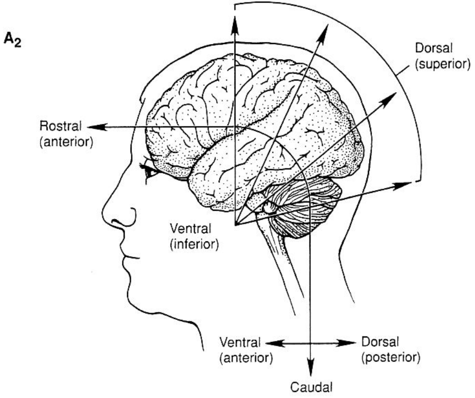
FIGURE 1–3
When the brain is cut between the two hemispheres down the midline (a midsagittal section) the six main divisions illustrated in Figure 1–2 can be seen clearly.

A. This schematic midsagittal section shows the position of the six major brain structures in relation to external landmarks. The corpus callosum is a large fiber bundle that interconnects the left and right hemispheres.

B. The same section in **A** is illustrated in this magnetic resonance image of the living brain.

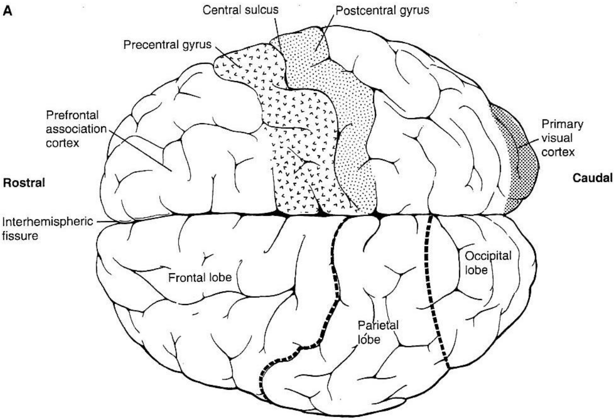
(Kandel et.al. 1991)

Sections of the human brain: views



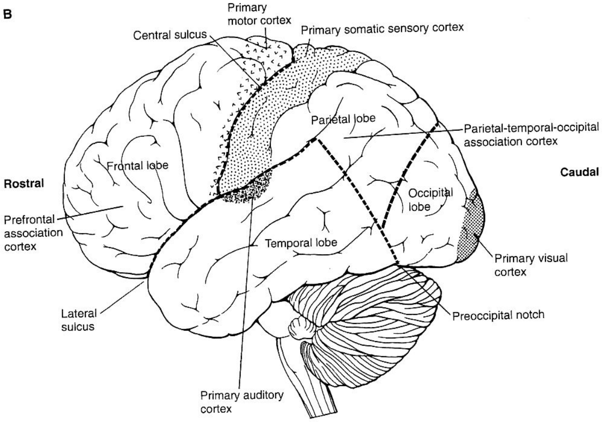
(Kandel et.al. 1991)

Sections of the human brain: dorsal



(Kandel et.al. 1991)

Sections of the human brain: lateral



(Kandel et.al. 1991)

Sections of the human brain: pathways

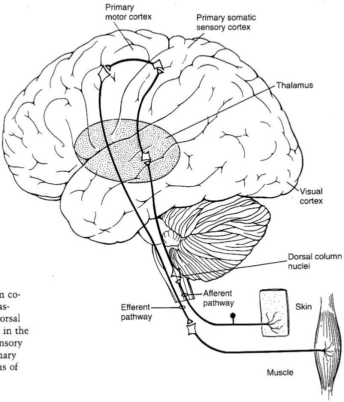
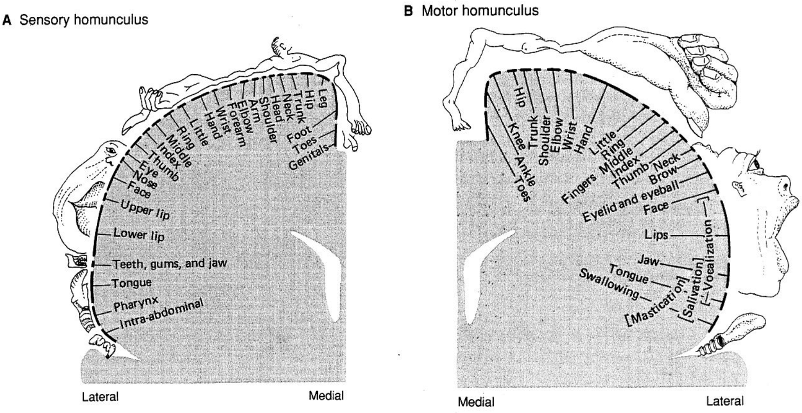


FIGURE 19–5
The major somatic sensory systems and the motor system cooperate to carry out most behavioral acts. Sensory input ascends through the spinal cord to a synaptic relay in the dorsal column nuclei of the brain stem, then to a synaptic relay in the thalamus, and eventually reaches the primary somatic sensory cortex. The direct motor pathway descends from the primary motor cortex through the brain stem to the motor neurons of the spinal cord, and from there to the muscle.

(Kandel et.al. 1991)

Sensory and motor homunculus



(Kandel et.al. 1991)

Importance of body regions

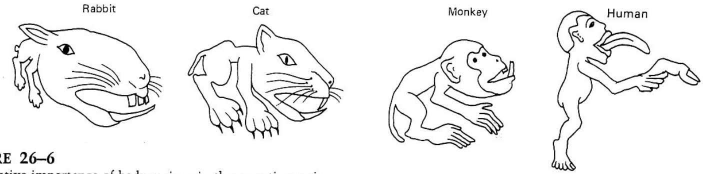


FIGURE 26–6
The relative importance of body regions in the somatic sensibilities of different species are shown in these drawings, which were based on studies of evoked potentials in the thalamus and cortex.

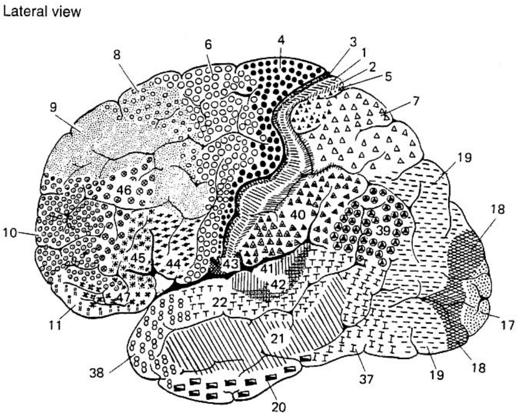
(Kandel et.al. 1991)

Brodmann classification

FIGURE 20–11
The human cerebral cortex was divided into about 50 discrete cytoarchitectonic areas more than 80 years ago by Korbinian Brodmann. Distinct areas are represented by different symbols and numbered as shown (there is no rationale for the numbering of the different fields). Brodmann's areas have consistently been found to correspond to distinctive functional fields, each of which has a characteristic pattern of connections. Area 4, the primary motor cortex, occupies most of the precentral gyrus. The primary somatic sensory cortex includes areas 1, 2, and 3 in the postcentral gyrus. Area 17 is the primary visual cortex. Areas 41 and 42 comprise the primary auditory cortex. The prefrontal association cortex and the parietal–temporal–occipital association cortex are also composed of a number of distinct cytoarchitectonic areas.

(Kandel et.al. 1991)(Brodmann 1909)

Brodmann classification: lateral



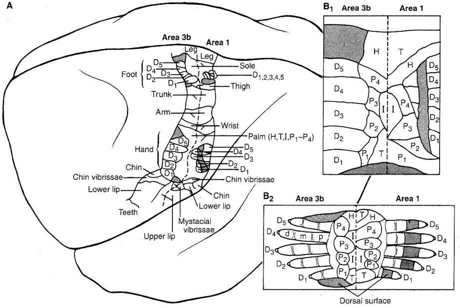
(Kandel et.al. 1991)(Brodmann 1909)

Brodmann classification: hand and foot

FIGURE 26–7
Each of the four subregions of the primary somatic sensory cortex (Brodmann's areas 3a, 3b, 1, and 2) has its own complete representation of the body surface. This figure illustrates the representation for the hand and the foot in areas 3b and 1. (Adapted from Kaas et al., 1983.)

A. Somatosensory maps in areas 3b and 1 are shown in this dorsolateral view of the brain of an owl monkey. The two maps are roughly mirror images. The digits of the hand and foot are numbered D_1 to D_5.

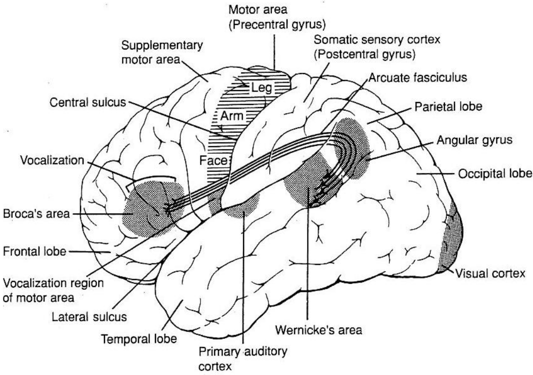
B. 1. A more detailed illustration of the representation of the glabrous pads of the palm in areas 3b and 1. These include the palmar pads (numbered in order, P_4 to P_1), two insular pads (**I**), two hypothenar pads (**H**), and two thenar pads (**T**). **2.** An idealized map of the hands based on studies of a large number of monkeys. The distorted representations of the palm and digits reflect the extent of innervation of each palmar area in the cortex. The five digital pads (D_1 to D_5) include distal, middle, and proximal segments (**d**, **m**, **p**).



(Kandel et.al. 1991)

Sections of the human brain: language processing

FIGURE 1–4
This lateral view of the cerebral cortex of the left hemisphere shows some of the areas involved in language. Wernicke's area, near the primary auditory cortex, is important to the understanding of spoken language. Wernicke's area lies near the angular gyrus, which combines auditory input with information from other senses. The arcuate fasciculus is a fiber tract that connects Wernicke's area to Broca's area. Broca's area initiates grammatical speech. It, in turn, lies near the vocalization region of the motor area, which issues the specific commands that cause the mouth and tongue to form words. (Adapted from Geschwind, 1979.)



(Kandel et.al. 1991)

Observations

- brain structure specific for a species
- cortex composed of regions with specific functions
- e.g., visual-, auditory-, somatic sensory-, motor cortex

- microstructure of cortex still not understood
- information encoding? (spikes or averages?)
- function of the associative cortex?
- how does consciousness emerge in humans?

- no method to get high-res data from live brains

Neuron types and the model neuron

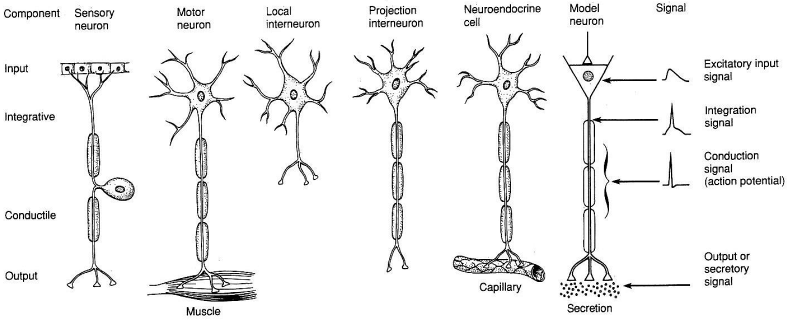


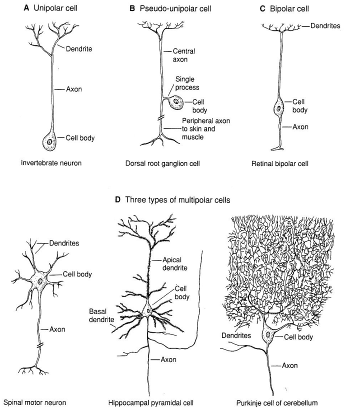
FIGURE 2–7
Most neurons, whether they are sensory, motor, interneuronal, or neuroendocrine, have four functional components in common: an input component, an integrative component, a conductile component, and an output component. On the basis of these common features, the functional organization of neurons in general can be represented by a model neuron. The functional components of the neuron are represented in distinct regions, with unique shapes and properties, and each produces a characteristic signal. Not all neurons share all of these features; for example, local interneurons often lack conductile components.

(Kandel et.al. 1991)

Basic classification of neurons

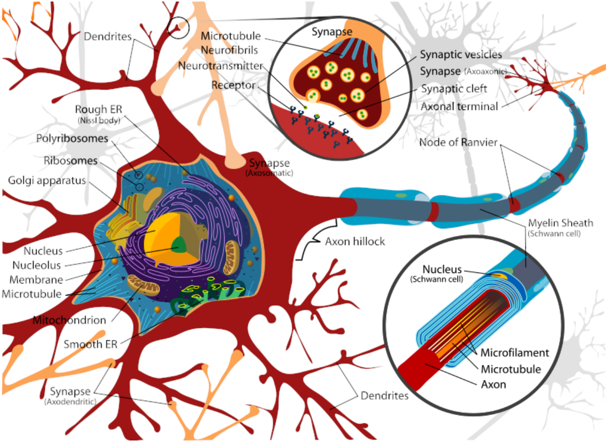
- cell body
- one axon (bare or myelinated)
- tree of dendrites

- unipolar (invertebrates)
- bipolar (e.g. spinal cord, retina)
- multipolar (mammalian)



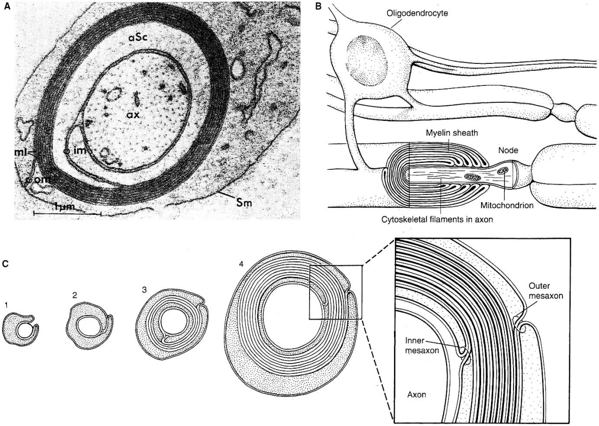
(Kandel et.al. 1991) (Ramon y Cajal 1933)

Neuron schema



(Wikipedia)

Schwann cell and myelin sheath



(Kandel et.al. 1991)

Hodgkin-Huxley model

- based on studies on *squid giant neurons*
- measurements of potentials inside the cells
- results accepted to apply to other types of neurons

- main neural activity is electrical
- rest-potential slowly built-up by ion-pump proteins

- membrane channels open quickly after stimulation
- when total activation (=internal potential) exceeds a threshold
- spike (*action potential*) propagates along the cell axon
- neuron inactive during recovery interval

Membrane rest potential

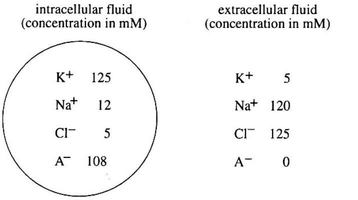


Fig. 1.5. Ion concentrations inside and outside a cell

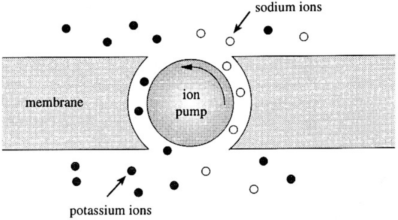


Fig. 1.8. Sodium and potassium ion pump

- rest potential of about 60 mV against the extracellular fluid
- generated and maintained by ion pump molecules
- high K^+ and low Na^+ concentration in the cell

(Kandel et.al. 1991) (Rojas 1996)

Ion channels

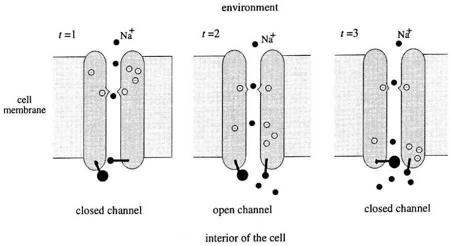
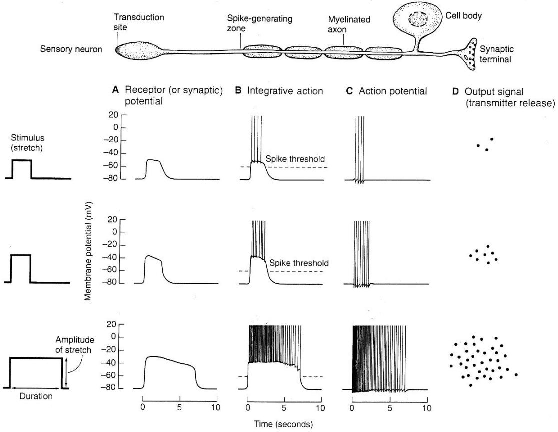


Fig. 1.7. Electrically controlled ionic channels

- separate voltage-controlled ion-channels
- when open, fast inflow of Na ions
- changed potential opens neighbor ion-channels
- channel closes and ion-pumps slowly restore rest potential

Neural activity: action potential and spiking



(Kandel et.al. 1991)

Typical action potential

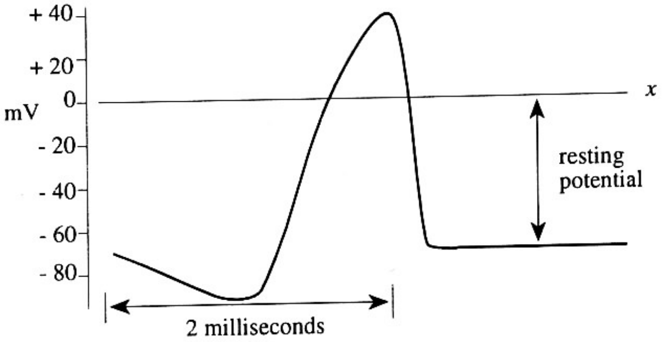


Fig. 1.9. Typical form of the action potential

Note: action potential propagates left-to-right along the axon. Read right-to-left for time-signal $V(t)$ at constant x

(Rojas 1996)

Spike frequency depends on cell activation

Example: receptive fields in the visual cortex

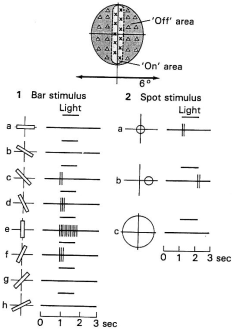


FIGURE 29–10
Receptive field of a simple cell in the primary visual cortex.

(Kandel et.al. 1991)

Repetitive firing

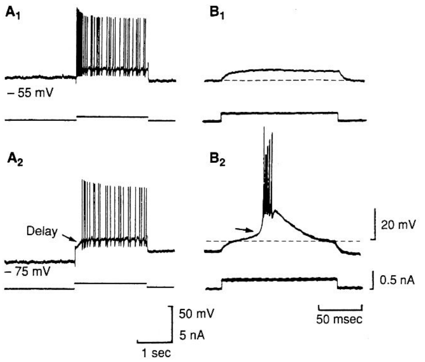


FIGURE 8–8
Repetitive firing properties vary widely among different types of neurons.

(Kandel et.al. 1991)

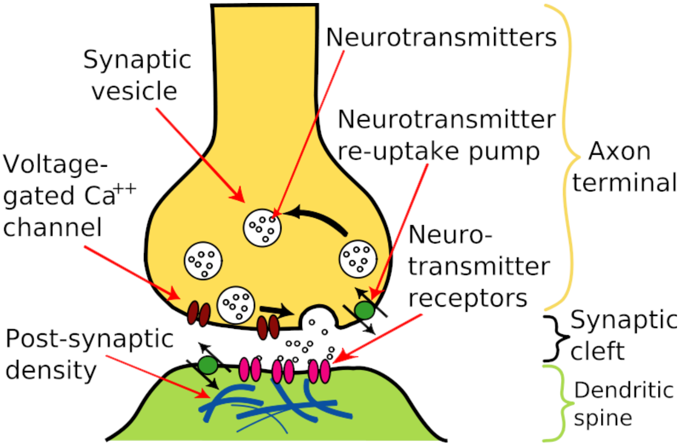
Synapses: connections between neurons

- between axon and dendrites, but also axon to soma/axon
- special membrane channels and also mechanical connection

- synaptic vesicles filled with neurotransmitters generated in the axon
- different substances: excitory or inhibitory
- on arrival of an action-potential, vesicles are released into the synaptic cleft via voltage-gated channels
- transmitters reach channel-proteins on receiving neuron,
- ion-inflow changes the activiation of the receiving neuron

- the actual memory device in the brain
- evidence for synapse changes during learning

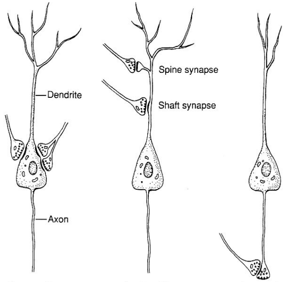
Synapse



(Wikipedia)

Synapse

FIGURE 11–13
Synaptic contact can occur on the cell body, the dendrites, or the axon. The synapse names—axosomatic, axodendritic and axo-axonic—identify the contacting regions of both the presynaptic and postsynaptic neurons (the presynaptic element is identified first). Note that axodendritic synapses can occur on either the main *shaft* of a dendrite branch or on a specialized input zone, the *spine*.



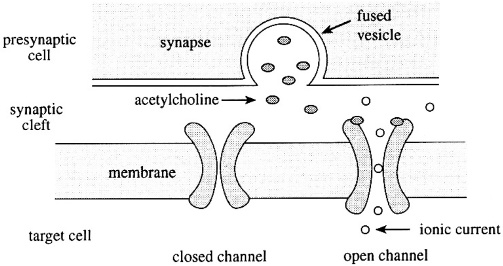


Fig. 1.12. Chemical signaling at the synapse

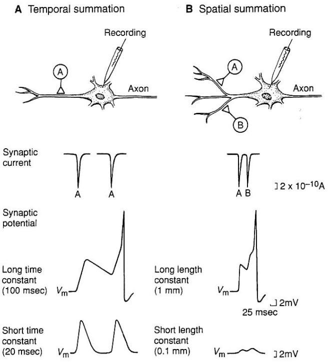
(Kandel et.al. 1991) (Rojas 1996)

Spatial and temporal summation of activation

FIGURE 11–12
The effects of temporal and spatial summation on neuronal integration.

A. Temporal summation of two EPSPs produced consecutively by a single presynaptic neuron A. The synaptic current flow, I_{EPSP}, generated by the action of the presynaptic neuron is illustrated at the cell body. This same synaptic current will give rise to very different synaptic potentials depending on whether the postsynaptic cell has a long or a short time constant. In a cell with a long time constant the first EPSP will not decay totally by the time the second EPSP is triggered. Therefore the depolarizing effects of both potentials are additive, bringing the membrane potential above the threshold and triggering an action potential. In a cell with a short time constant the first EPSP decays to the resting potential before the second EPSP is triggered. The second EPSP alone does not cause enough depolarization to trigger an action potential.

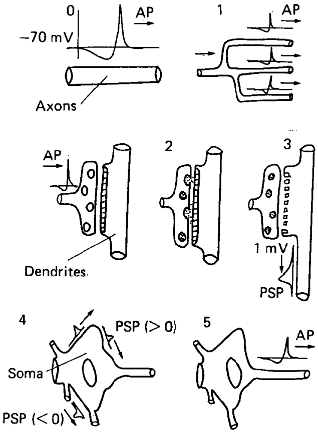
B. Spatial summation of two EPSPs produced by two presynaptic neurons (A and B) assuming two different length constants for the postsynaptic cell. In this hypothetical experiment the current (I_{EPSP}) produced by each of these synaptic contacts is assumed to be the same. Both synapses are the same distance from the postsynaptic trigger zone, but in one case the postsynaptic cell has a long length constant, the other a short length constant. In the cell with a long length constant, the initial segment is only one length constant away from the site of the synaptic contacts. Therefore, the EPSPs produced by each of the two presynaptic neuron will decrease only 37% before reaching the trigger zone. This results in enough depolarization to exceed threshold, triggering an action potential. For the cell with a short length constant, the distance between the synapse and the trigger zone in the



initial axon segment is equal to three length constants. Therefore, each synaptic potential is barely detectable when it arrives in the postsynaptic cell body, and even the summation of two potentials is not sufficient to trigger an action potential.

(Kandel et.al. 1991)

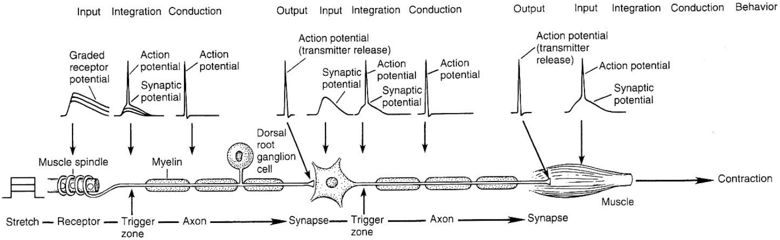
Summary of neural activity



(Amit, Modeling brain function)

Example: a reflex

FIGURE 2–9
This diagram summarizes the sequence of signals that produces a reflex action. Graded stretching of a muscle produces a graded (proportional) *receptor potential* in the terminal fibers of the sensory neuron (the dorsal root ganglion cell). This potential then spreads passively to the integrative segment, or trigger zone, at the first node of Ranvier. If the receptor potential is sufficiently large, it will trigger an *action potential* at the integrative segment, and the action potential will propagate actively and without change along the axon to the terminal region. At the terminal of the afferent neuron the action potential leads to an output signal: the release of a transmitter substance. The transmitter diffuses across the synaptic cleft and interacts with receptor molecules on the external membranes of the motor neurons that innervate the stretched muscle. This interaction initiates a synaptic potential in the motor cell. The synaptic potential then spreads passively to the axon hillock or initial segment of the motor neuron axon, where it may initiate an action potential that propagates actively to the terminal of the motor neuron. At the terminal the action potential causes transmitter release, which triggers a synaptic potential in the muscle. This signal produces an action potential in the muscle, causing contraction of the muscle fiber.



(Kandel et.al. 1991)

Interconnection patterns: retina

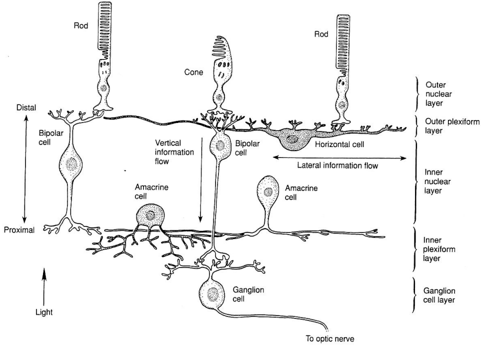


FIGURE 28–6
The retina has five major classes of neurons arranged into three nuclear layers: photoreceptors (rods and cones), bipolar cells, horizontal cells, amacrine cells, and ganglion cells. Photoreceptors, bipolar, and horizontal cells make synaptic connections with each other in the outer plexiform layer. The bipolar, amacrine, and ganglion cells make contact in the inner plexiform layer. Bipolar cells bridge the two layers. Details of these connections are illustrated in Figure 28–11. Information flows vertically from photoreceptors to bipolar cells to ganglion cells. Information also flows laterally, mediated by horizontal cells in the outer plexiform layer and amacrine cells in the inner plexiform layer. (Adapted from Dowling, 1979.)

(Kandel et.al. 1991)

Interconnection patterns: visual cortex

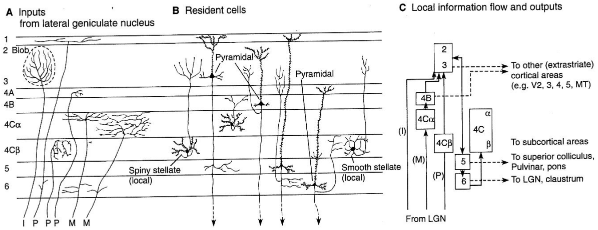


FIGURE 29–8
The primary visual cortex has distinct anatomical layers, each with characteristic synaptic connections.

A. Most afferent fibers from the lateral geniculate nucleus terminate in layer 4. Axons of type P cells (in the parvocellular layers) terminate primarily in layer 4Cβ, with minor inputs to 4A and 1, while axons from type M cells (in the magnocellular layer) terminate primarily in layer 4Cα. Collaterals of both types of cells also terminate in layer 6. Cells of the intralaminar regions of the lateral geniculate nucleus terminate in layers 2 and 3.

B. Several types of resident neurons make up the primary visual cortex. Spiny stellate and pyramidal cells, both of which have spiny dendrites, are excitatory. Smooth stellate cells are inhibitory. Pyramidal cells project out of the cortex, whereas both types of stellate cells are local neurons.

C. Afferents from M and P cells in the lateral geniculate nucleus end on spiny stellate cells in layer 4C, and these cells project axons to layer 4B and the upper layers 2 and 3. Cells from the interlaminar zones (I) in the lateral geniculate nucleus project directly to layers 2 and 3. From there, pyramidal cells project axon collaterals to layer 5 pyramidal cells, whose axon collaterals project both to layer 6 pyramidal cells as well as back to cells in layers 2 and 3. Axon collaterals of layer 6 pyramidal cells then make a loop back to layer 4C onto smooth stellate cells. Each layer, except for 4C, has different outputs. The cells in layers 2, 3, and 4B project to higher visual cortical areas. Cells in layer 5 project to the superior colliculus, the pons, and the pulvinar. Cells in layer 6 project back to the lateral geniculate nucleus and the claustrum. (Adapted from Lund, 1988.)

(Kandel et.al. 1991)

Interconnection patterns: cerebral cortex

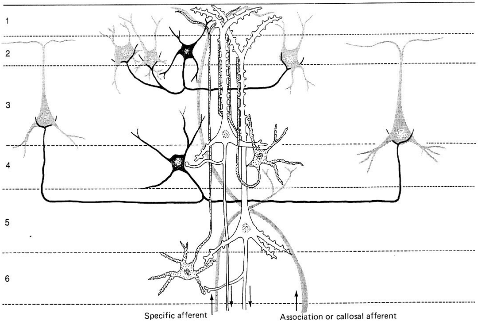


FIGURE 50–5
The principal neuron types and their interconnections are similar in the various regions of the cerebral cortex. Note that the two large pyramidal cells **(white)** in layers 3 and 5 receive multiple synaptic contacts from the star-shaped interneuron (stellate cell, **stippled**) in layer 4. The inhibitory action of the basket cells (**black**) is directed to the cell bodies of cortical neurons (**gray**). Major input to the cortex derives from specific thalamic relay nuclei (specific afferents) and is directed mostly to layer 4; association and callosal input (association and callosal afferents) is, in large part, directed to more superficial layers. (Adapted from Szentágothai, 1969.)

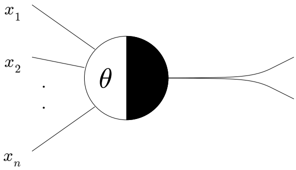
(Kandel et.al. 1991)

Neurons

- Hodgkin-Huxley model
- rest potential of the neuron built by ion-pumps
- voltage-changes due to cell activation
- action-potential generated when voltage exceeds trigger value
- action-potentials are all-or-nothing and identical
- but duration/repetition specific to the cell type

- neurons interconnected by synapses
- the actual memory and learning devices
- note: about 10^9 synapses/mm^3 in the human cortex

The McCulloch-Pitts model



- network built from binary neurons
- output values is $\{0, 1\}$
- internal threshold θ

- unweighted edges of two types
- excitory connections $x_1, x_2, \ldots x_n$
- inhibatory connections $y_1, y_2, \ldots y_m$ (marked with small circle)

- neuron output is zero if any inhibatory input is active
- otherwise, compute total excitation $x = x_1 + x_2 + \ldots + x_n$
- output $a = (x \geq \theta)$: neuron fires if excitation larger than the threshold

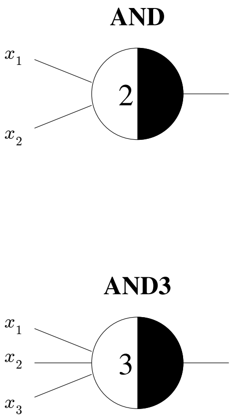
(Rojas 1996), (McCulloch and Pitts, 1943)

McCulloch-Pitts neuron

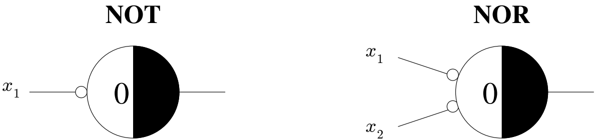
- neuron fully inactivated by any inhibitory input:
  also the case for some real neurons

- otherwise acting as a threshold gate: capable of implementing many logical functions of n inputs
- in particular, *monotonic* functions
- some examples on the next slides

Elementary functions: AND and OR

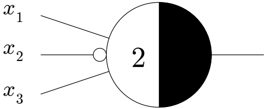


Negated functions: NOT



- uninhibited McCulloch-Pitts units can only implement monotonic logical functions
- inhibited connections allow realization of NOT, NOR, etc.

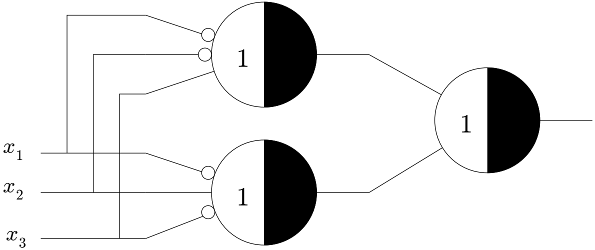
Decoder



$$(x_1,\ x_2,\ x_3) = (1,\ 0,\ 1)$$

- unit fires only for input-vector $(1, 0, 1)$
- a decoder for a specific bit-pattern (minterm)

Construction of arbitrary functions



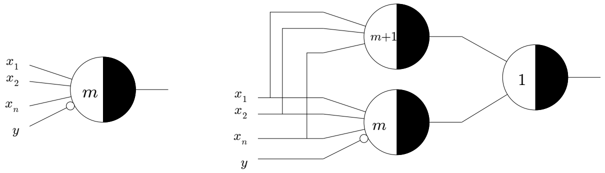
- equivalent to DNF (disjunctive normal form) logic synthesis
- any logical function $F : \{0,1\}^n \mapsto \{0,1\}$ can be realized with a two-layer McCulloch-Pitts network

Arbitrary logical functions

- two-layer construction implements any given function
- but very many units for most functions (one per minterm)

- use multi-layer network to reduce the costs (number of neurons and inputs per neuron)
- logic-minimization problem and algorithms, e.g. Karnaugh-Veitch diagrams, Quine-McCluskey, etc.

Absolute and relative inhibition



- absolute inhibition used in McCulloch-Pitts model
- use above network to simulate relative inhibition, any input reduces the excitation by one
- to replace relative inhibition with weight w, use $m + w$ in the upper unit

Weighted and unweighted networks

- left model uses weight neurons, computes
- $0.2x_1 + 0.4x_2 + 0.3x_3 \geq 0.7$

- rescale to integer factors
- $2x_1 + 4x_2 + 3x_3 \geq 7$
- corresponding McCulloch-Pitts neuron shown on the right
- positive rational weights can be simulated by neurons with fan-out

Recurrent networks

- feed-forward networks can implement any logical function
- so far, no memory or internal state of the network

- assume discrete time-steps $t = (t_0, t_1, t_2, \ldots)$
- assume that output calculation of a unit takes one time-step
- connect (some) outputs back to inputs of other units

- any finite automata can be implemented with a network of McCulloch-Pitts units

Fault-tolerance

- systems with large number of components must consider faults
- e.g. defective units and connections

- distributed representations

- duplicated units to overrule defective units
- multiple parallel connections for reliable transmission
- the neuron threshold operation inherently is a majority/voting operation
- networks can be designed to tolerate faulty components

- see Rojas, section 2.5.3

McCulloch-Pitts model

- network with simple binary units
- excitory weightless connections ($w_{ij} \in \{0, 1\}$)
- absolutely inhibatory connections can inactivate the unit
- if not inhibited, simple threshold calculation
- proposed in 1943

- single unit can calculate many monotonic functions
- non-monotonic functions possible with inhibitory connections
- two-layer network can implement any logic function
- feedback-connections and time-delays allow information storage
- rational weights can be simulated

- but we have no learning, yet

Coming soon

- Rosenblatt Perceptron model
- Perceptron learning
- limitations of the perceptron

- multi-layer feed-forward networks
- backpropagation
- techniques to speed-up the learning

Summary: Neural Networks (1)

- motivation
- connectionism as a paradigm for computation
- biological inspiration
- non-linear processing elements
- lots of them, interconnected in interesting ways
- fully parallel processing

- overview of the human brain
- basic function of single neurons
- Hodgin-Huxley model and the standard neuron
- McCulloch-Pitts model