Introduction to biophysics

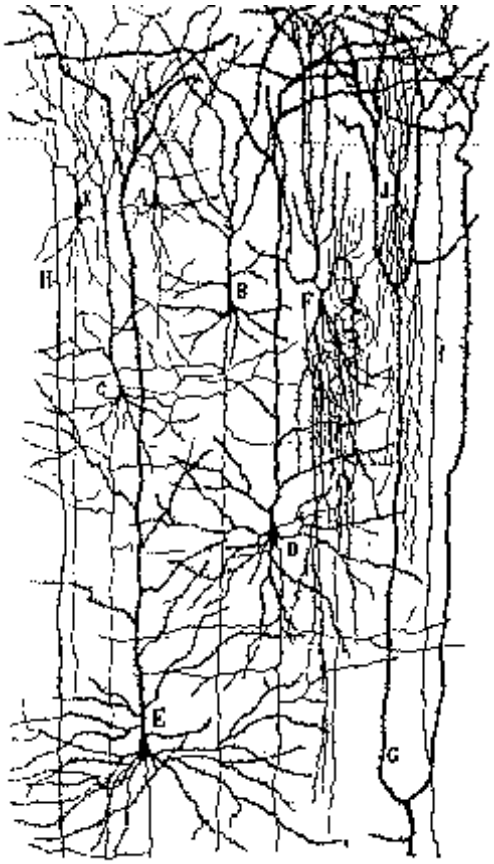
Bert Kappen,
Department of Biophysics
University of Nijmegen

March 26, 2007

Content

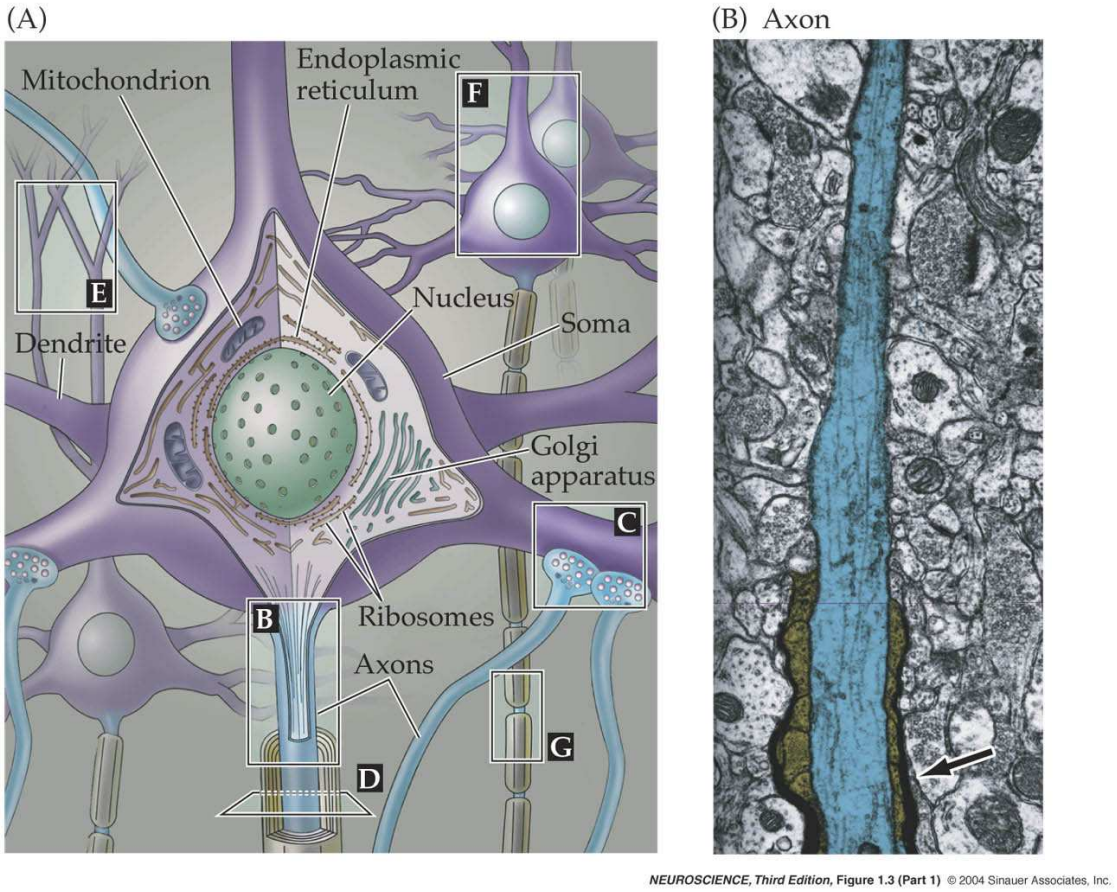
- Introduction to neurons and the brain
- Electrical properties of cells
  - cellular membranes,
  - ionic currents,
  - equilibrium behavior,
  - voltage dependent ionic permeabilities
- The Hodgkin-Huxley model of action potentials
- Synapses and learning
- Perceptrons

Introduction to neurons and the brain



Santiago Ramón y Cajal: nerve cells are discrete entities and communicate by synapses.

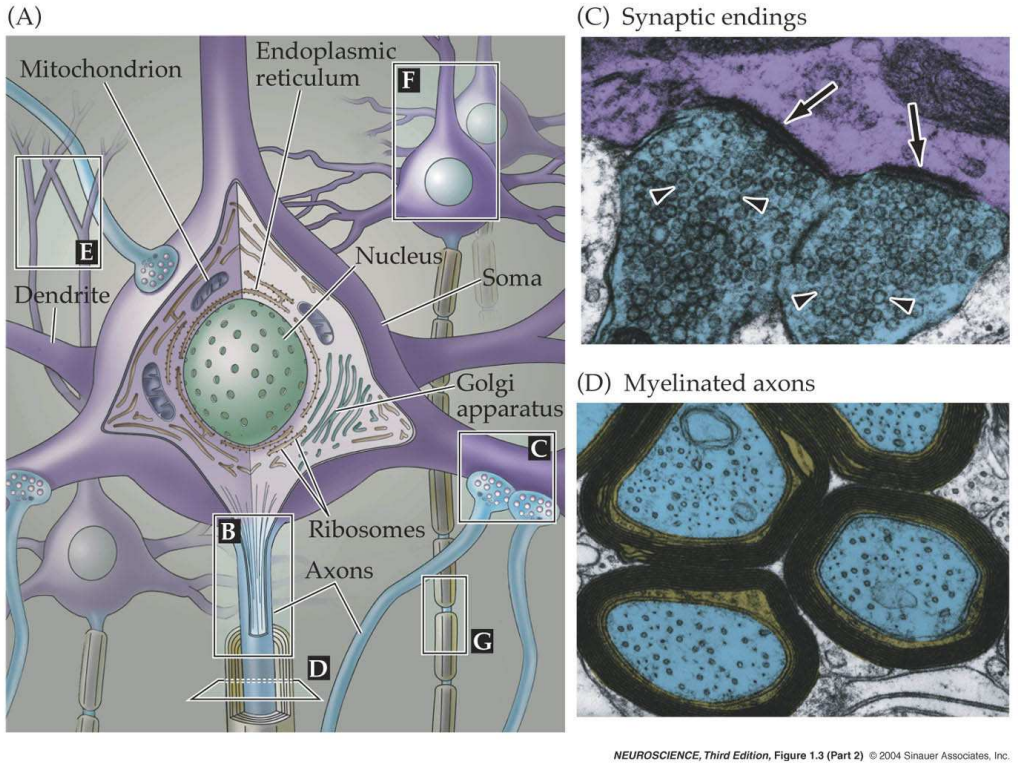
Nerve cells



NEUROSCIENCE, Third Edition, Figure 1.3 (Part 1) © 2004 Sinauer Associates, Inc.

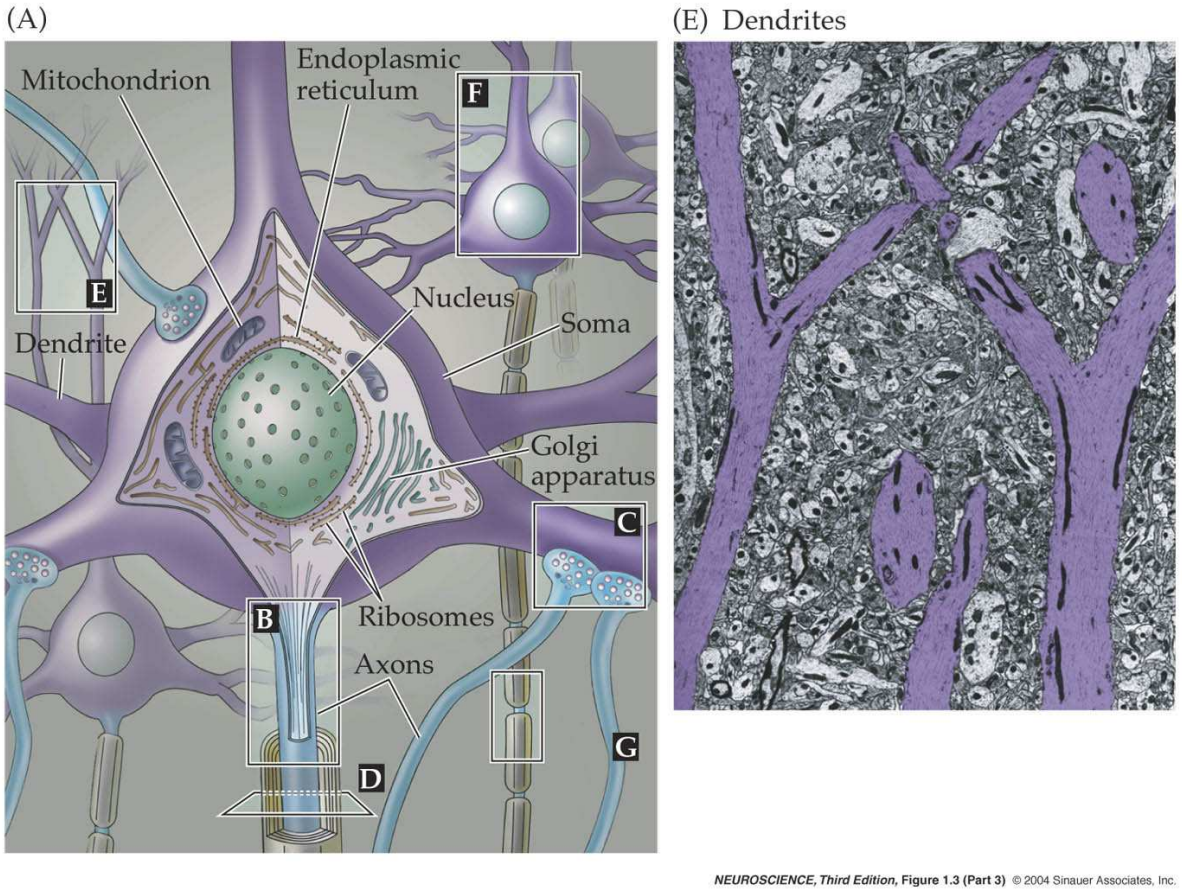
A) Diagram of nerve cells and their component parts. B) Axon initial segment (blue) entering a myelin sheath (gold).

Nerve cells



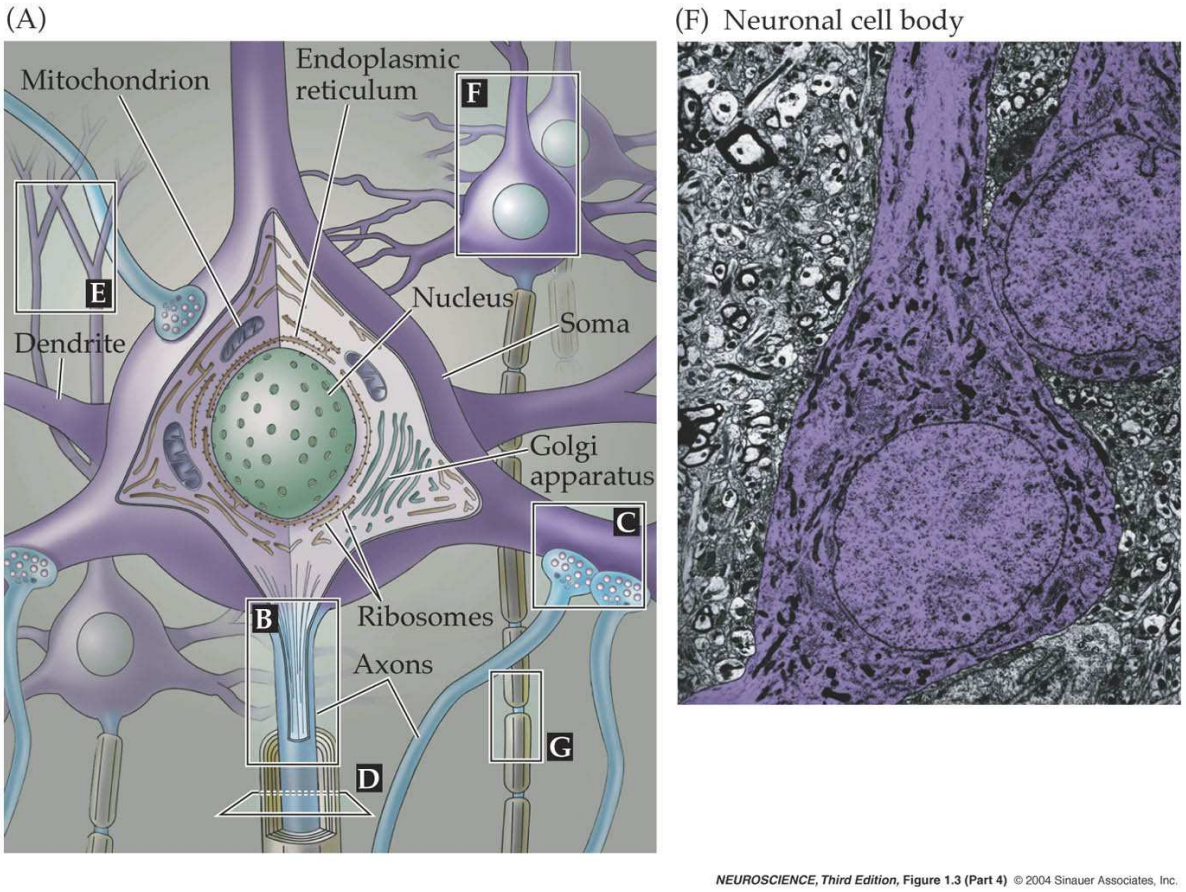
C) Terminal boutons (blue) loaded with synaptic vesicles (arrowheads) forming synapses (arrows) with a dendrite (purple). D) Transverse section of axons (blue) ensheathed by the processes of oligodendrocytes (gold).

Nerve cells



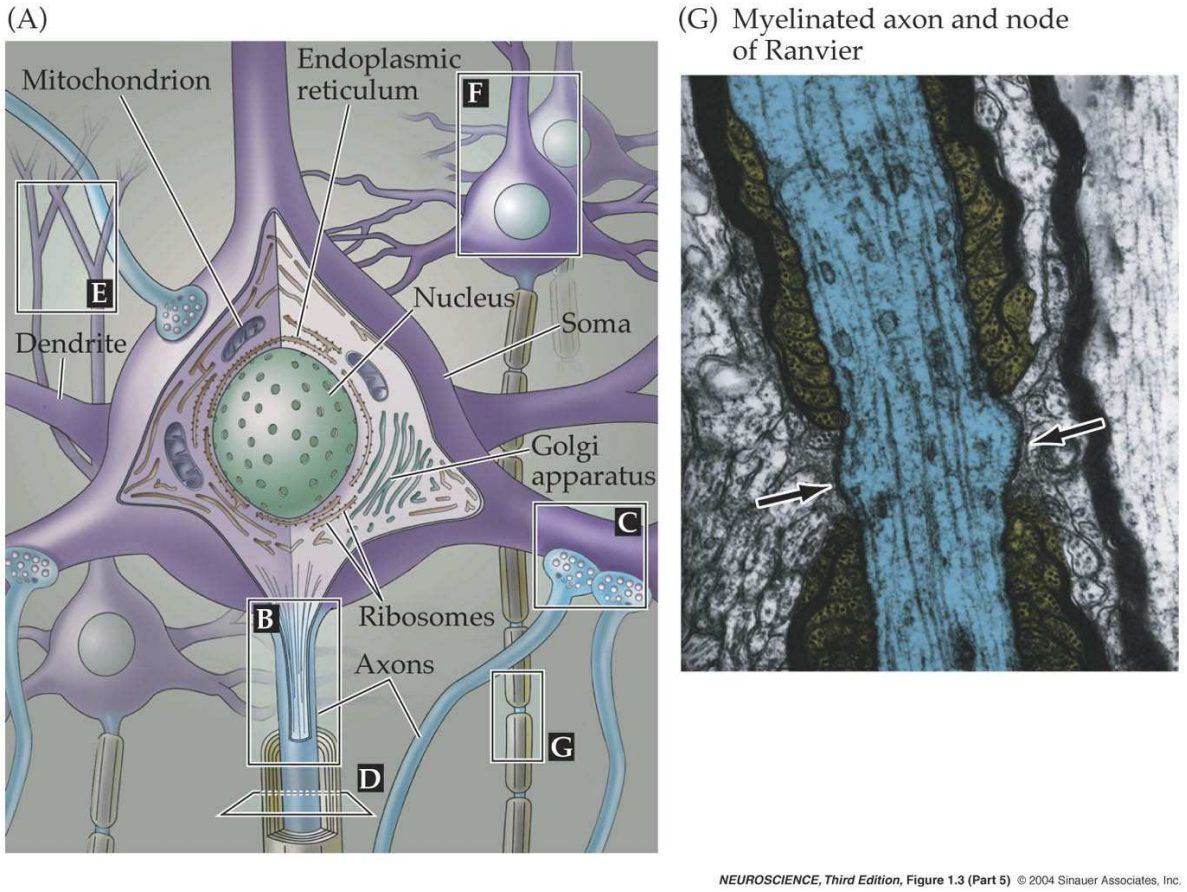
E) Apical dendrites (purple) of cortical pyramidal cells.

Nerve cells

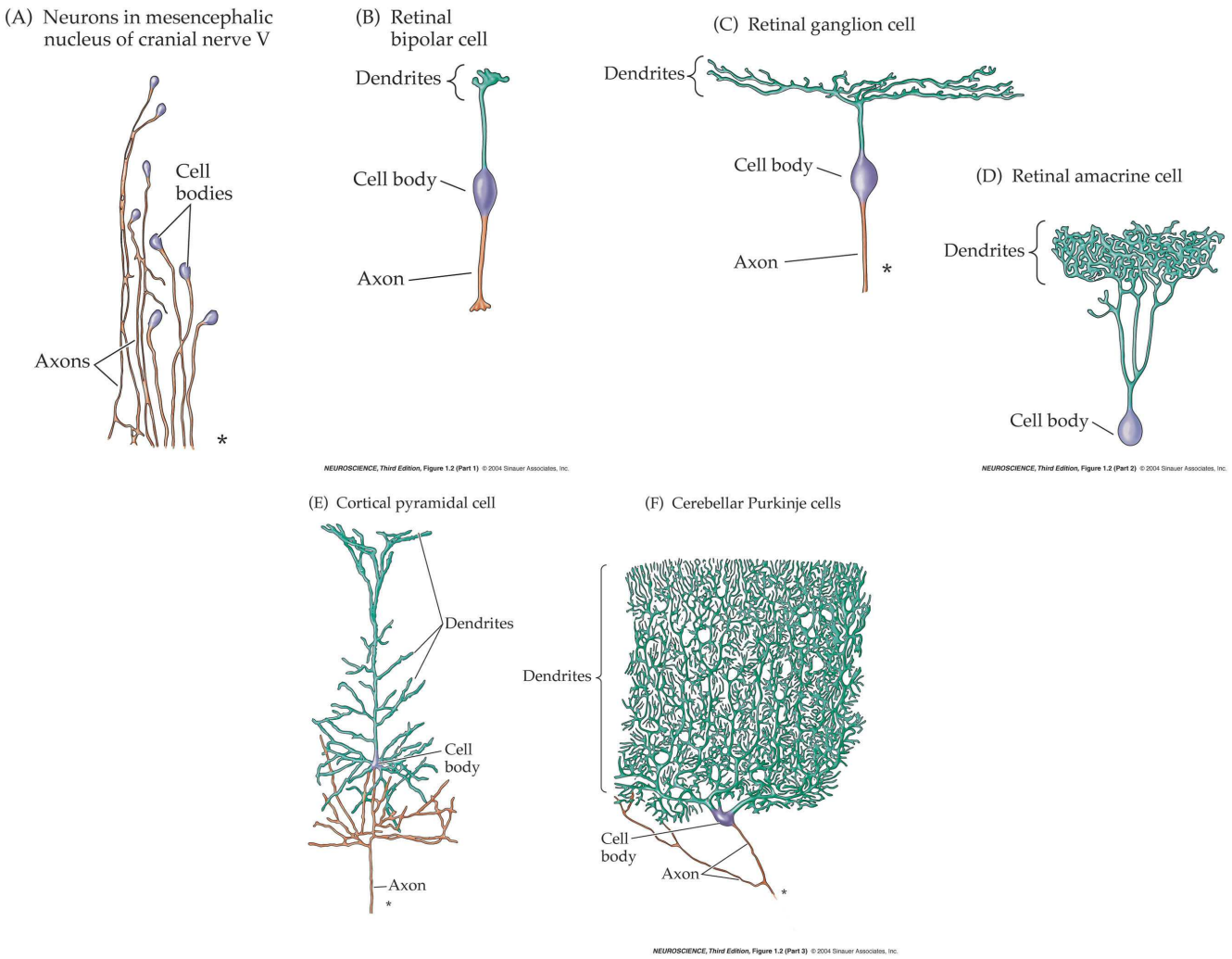


F) Nerve cell bodies (purple) occupied by large round nuclei.

Nerve cells

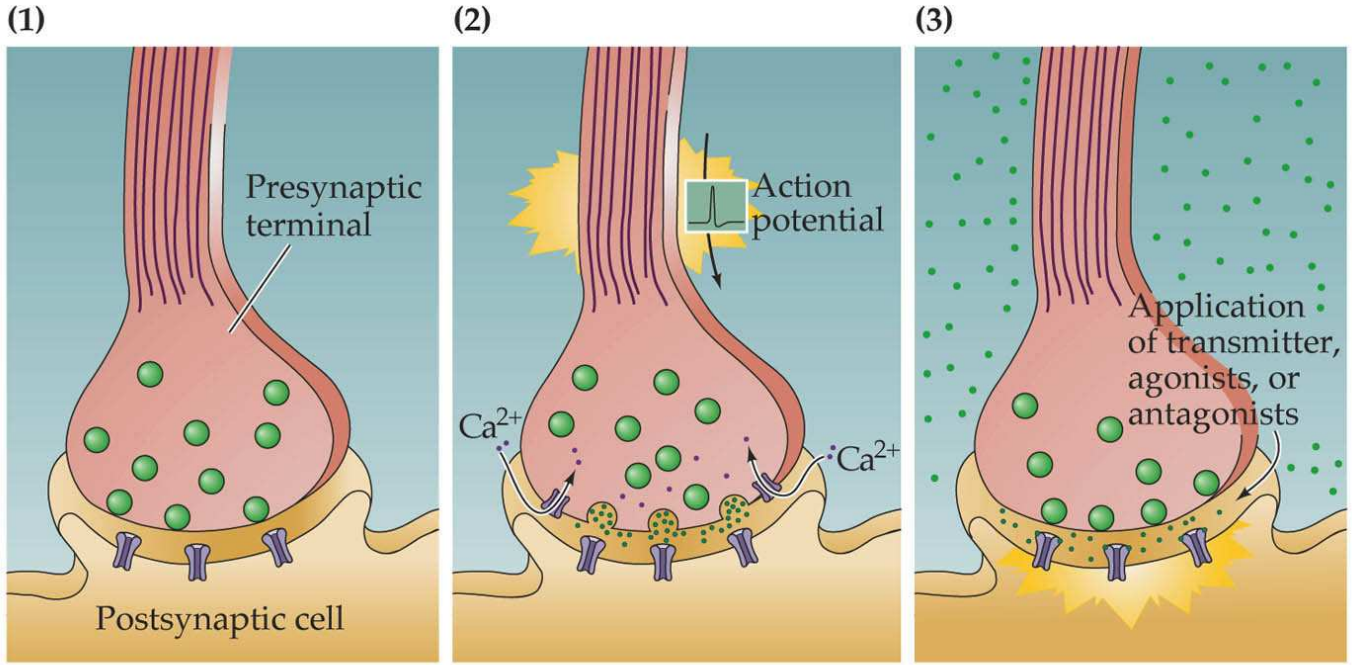


G) Portion of a myelinated axon (blue) illustrating the intervals between adjacent segments of myelin (gold) referred to as nodes of Ranvier (arrows).



a) muscle cell b-d) retinal cell e) Cortical pyramidal cell f) Cerebellar Purkinje cell

Cells communicate through synapses



NEUROSCIENCE, Third Edition, **Chapter 5, Box A**

The nervous system

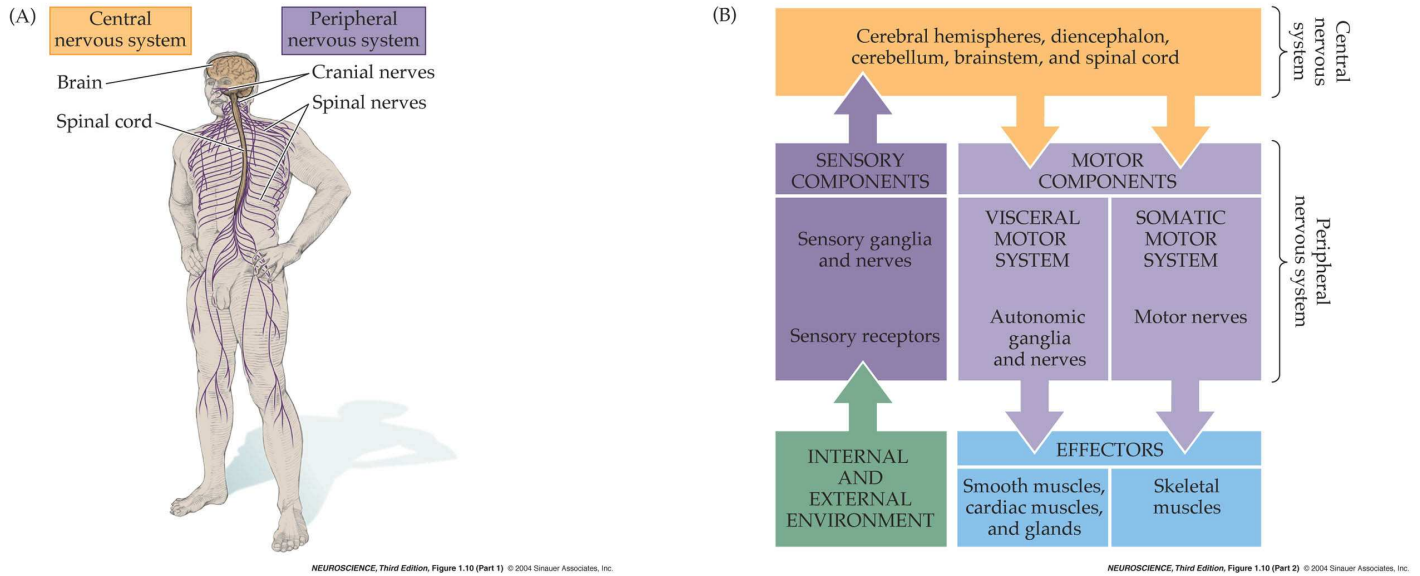


Figure 1: The major components of the nervous system. The peripheral nervous system receives sensory input and outputs motor commands. The central nervous system provides the 'mapping' from sensory input to motor output.

Periphery

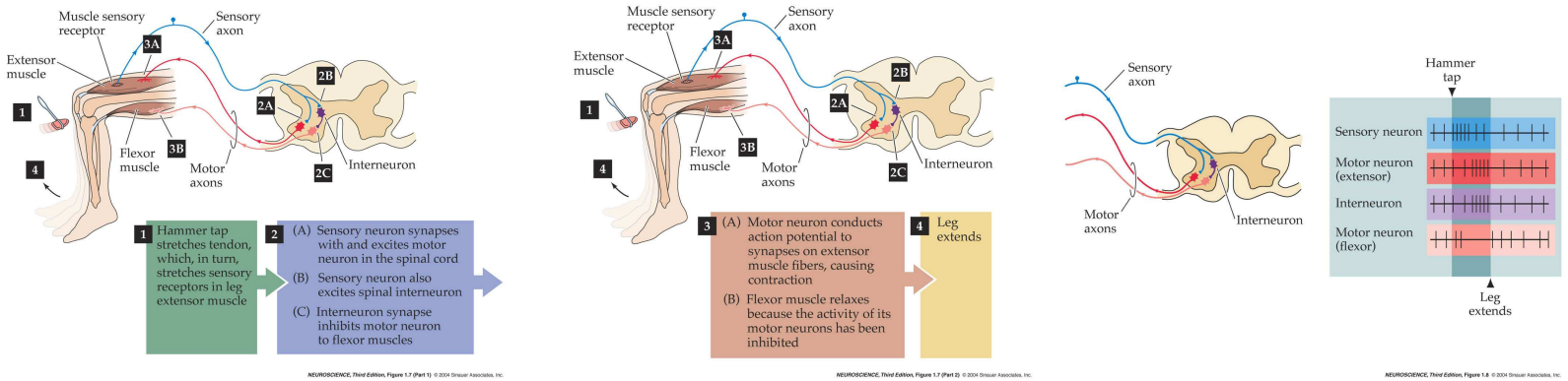


Figure 2: A simple reflex circuit, the knee-jerk response, illustrates several points about the functional organization of neural circuits. Stimulation of a muscle stretch receptor initiates action potentials that travel centrally along the afferent axons of the sensory neurons. This information stimulates spinal motor neurons by means of synaptic contacts. The action potentials generated in motor neurons travel peripherally in efferent axons, giving rise to muscle contraction.

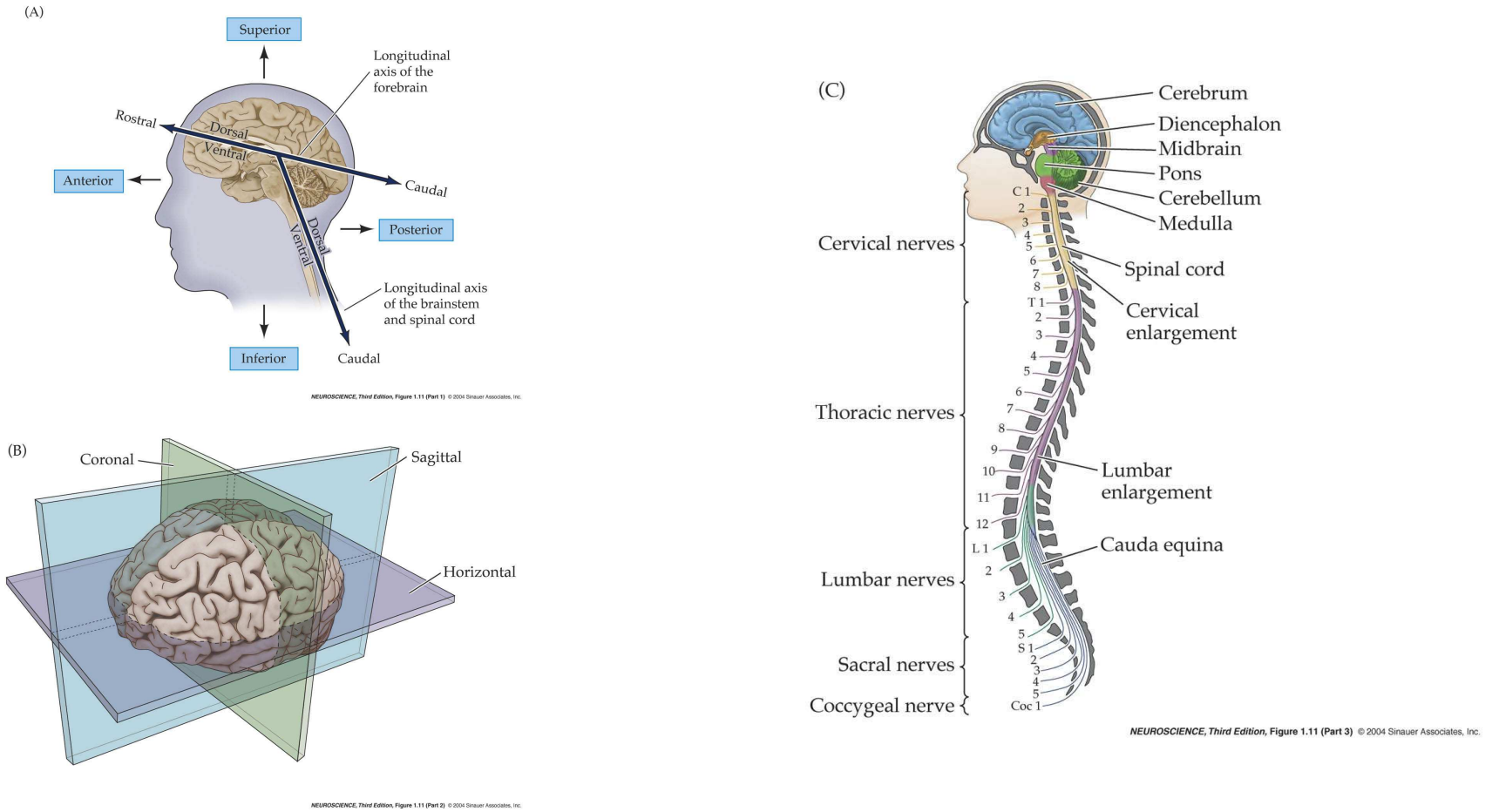
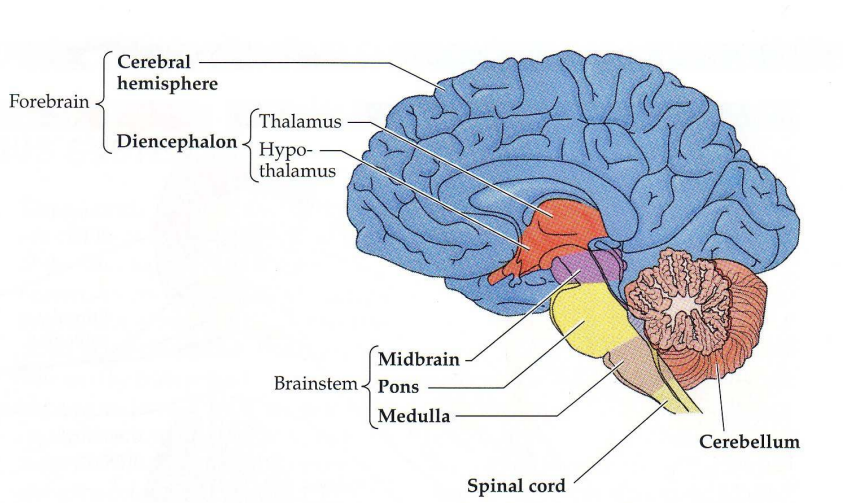


Figure 3: A) The terms anterior, posterior, superior, and inferior refer to the long axis of the body. B) The major planes of section used in cutting or imaging the brain. C) The subdivisions and components of the central nervous system.



Thalamus:
relay to and from cortex
motor functions, reproduction and hormone secretion

Brainstem contains structures, such as the *superior colliculus* that is involved in eye movement.

Cerebellum involved in coordination of motor activity, posture and equilibrium.

Hippocampus is involved in the storage of episodic memories.

Some features of the cortex

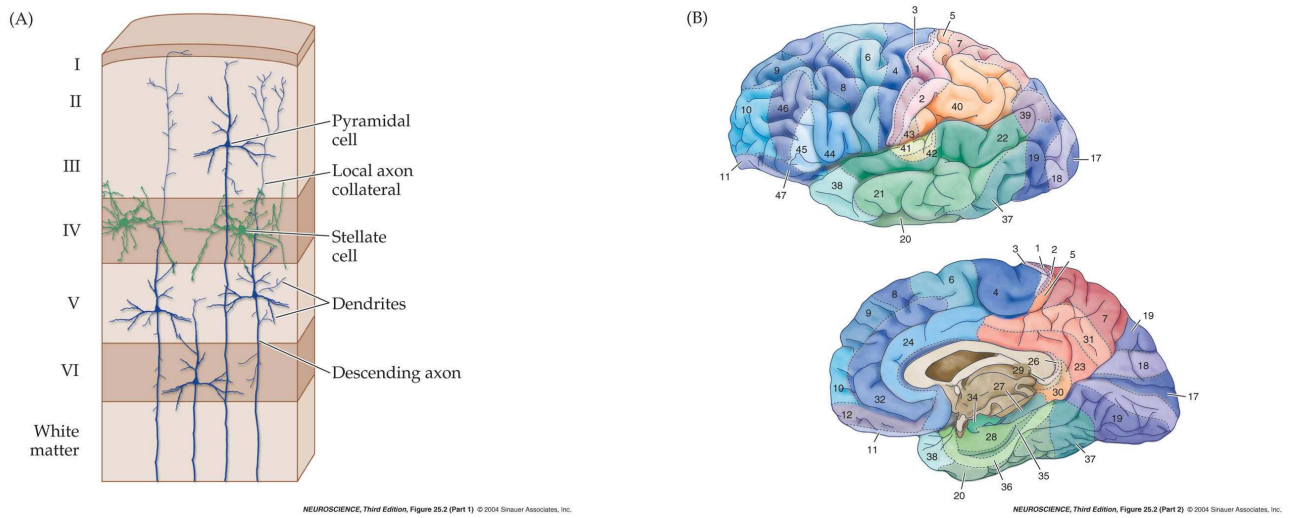


Figure 4: A) Cellular composition of the six layers of the neocortex. B) Brodmann areas. Red: primary motor cortex, blue: the primary somatic sensory cortex, green: the primary auditory cortex and yellow the primary visual cortex. All other Brodmann areas are considered association cortex.

The cortical tissue consists for about 80 % of pyramidal cells and the remainder are so called inter-neurons.

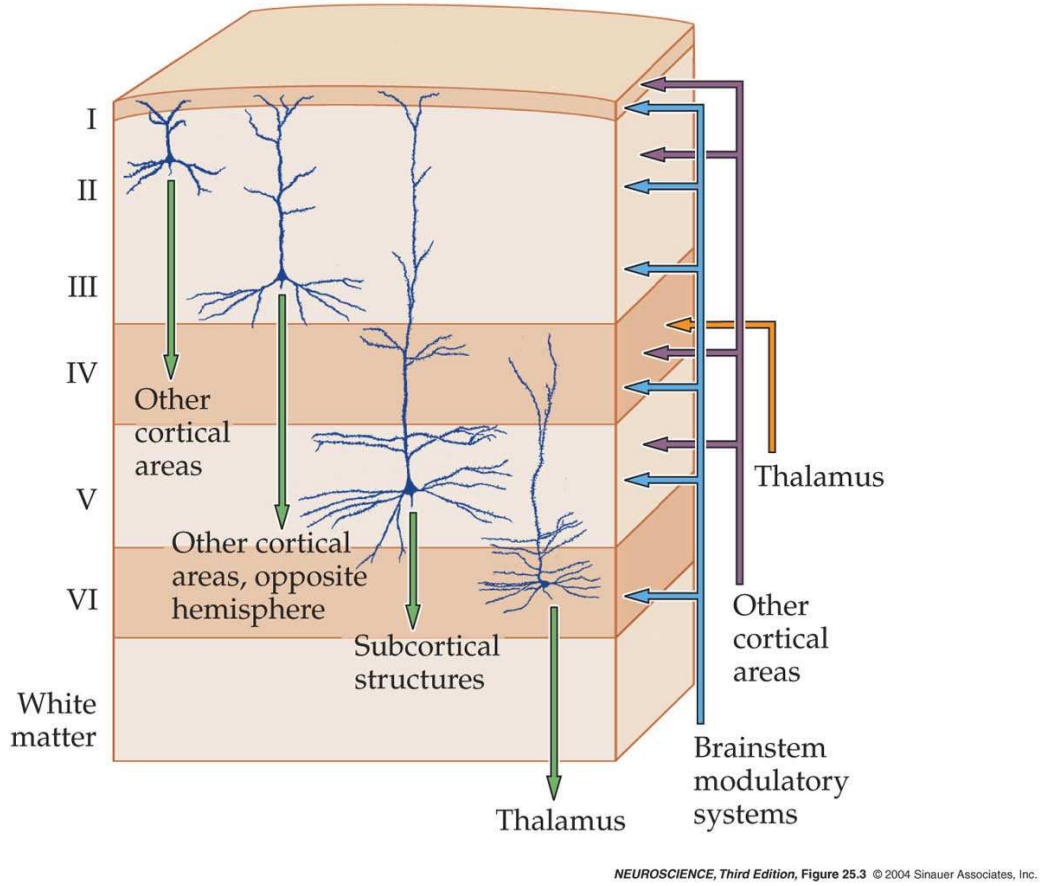


Figure 5: Canonical neo-cortical circuitry. Green arrows indicate outputs to the major targets of each of the neo-cortical layers in humans; orange arrow indicates thalamic input (primarily to layer IV); purple arrows indicate input from other cortical areas: and blue arrows indicate input from the brainstem to each layer.

Receptive field

The receptive field of a neuron is the collection of all stimuli that elicit an electrical response in that neuron.

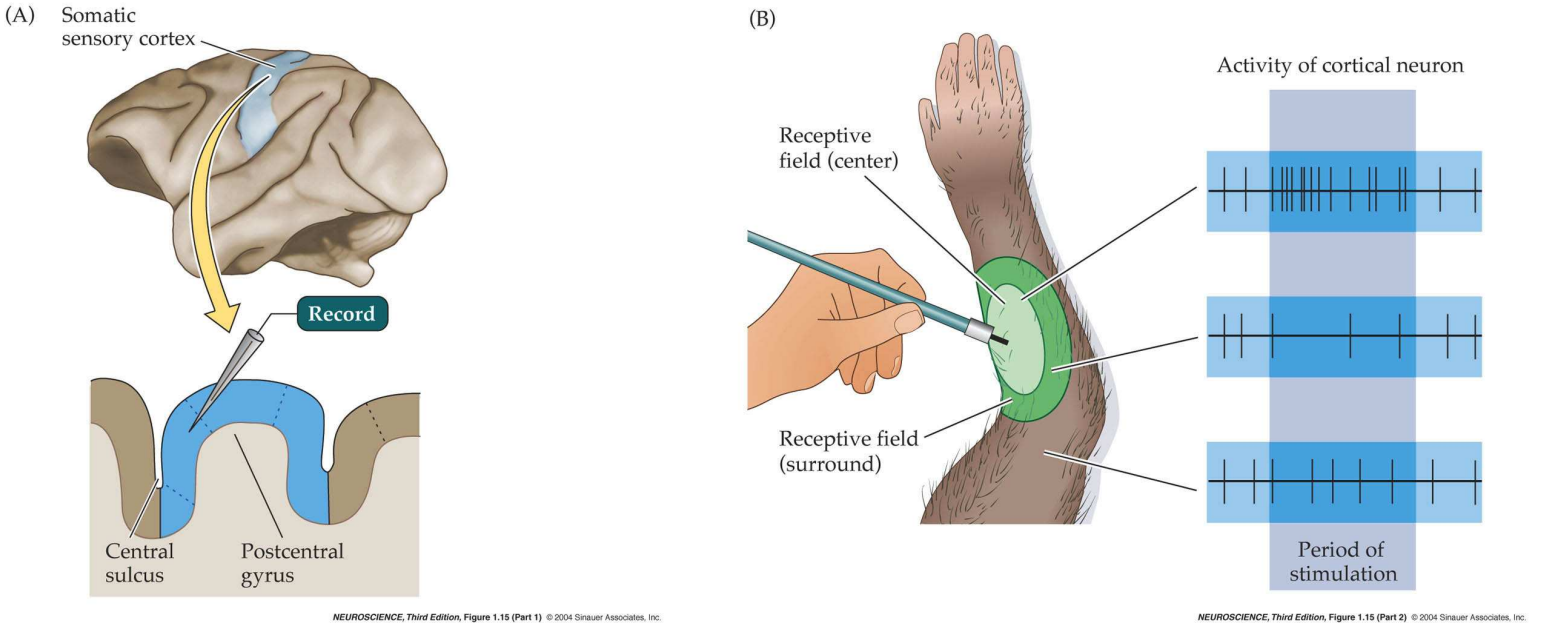


Figure 6: Single-unit electrophysiological recording from cortical pyramidal neuron, showing the firing pattern in response to a specific peripheral stimulus. A) Typical experimental set-up. B) Defining neuronal receptive fields.

Orientation selective cells

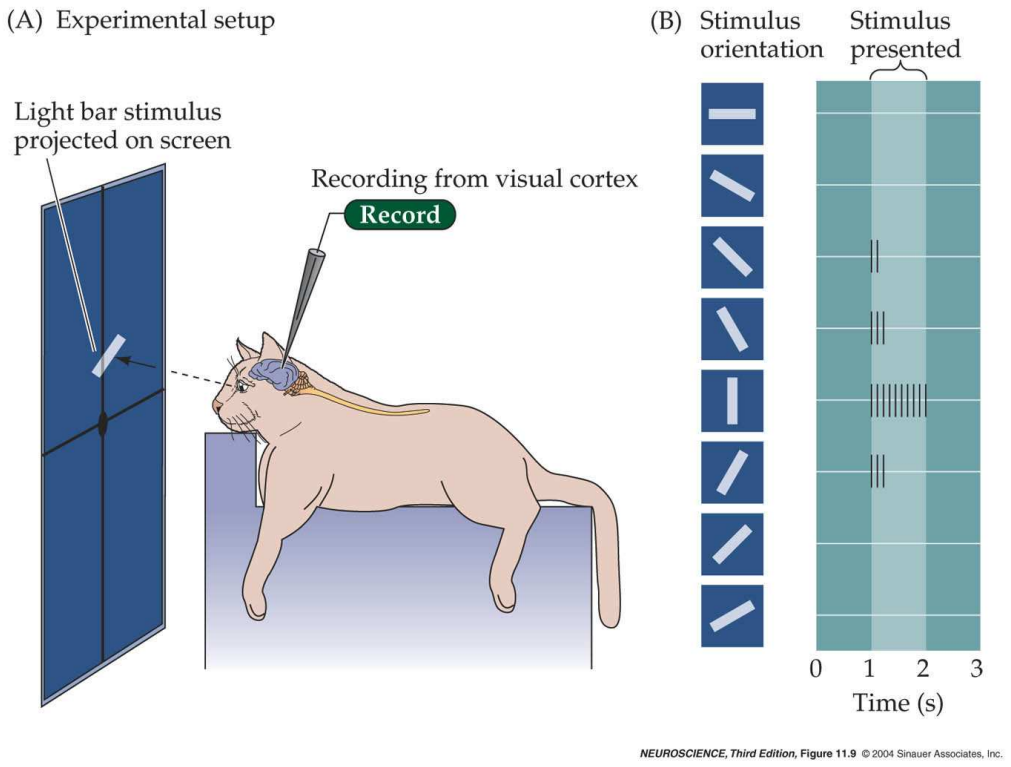


Figure 7: Neurons in the visual cortex respond selectively to oriented edges. A) An anesthetized cat focuses on a screen, where images can be projected; an extracellular electrode records the responses of neurons in the visual cortex. B) Orientation selectivity of neurons in the visual cortex.

Cortical columns

Nearby pyramidal cells make synaptic connections and have correlated activity.

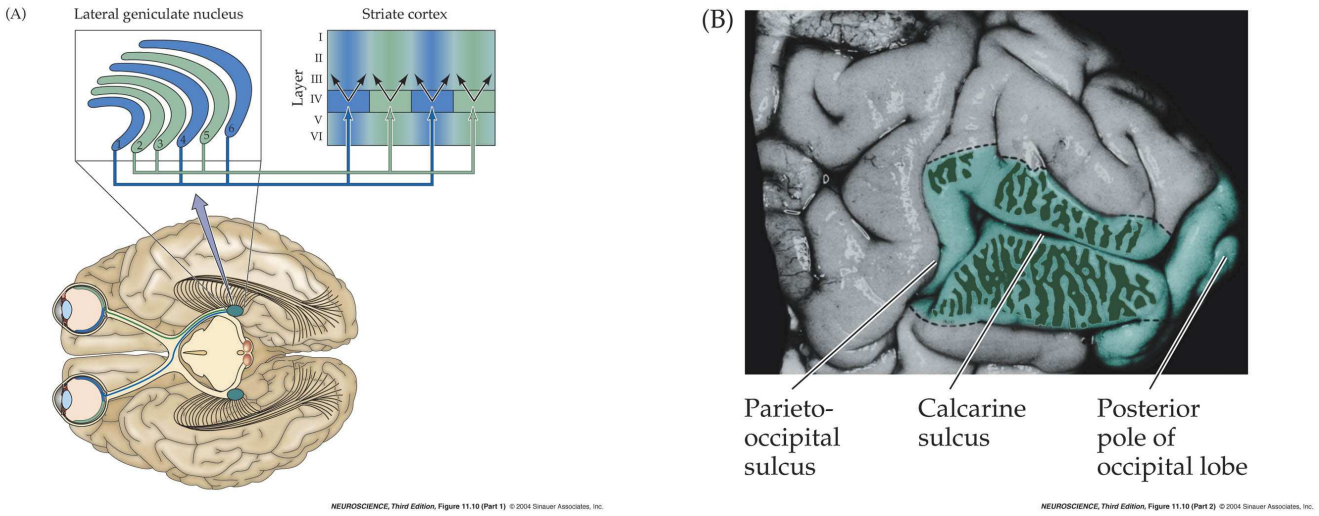


Figure 8: The LGN receives inputs from both eyes but in separate neuron populations. The inputs from the two eyes remain segregated in the ocular dominance columns of layer IV. A) Ocular dominance stripes in LGN and layer IV primary visual cortex. B) Pattern of ocular dominance columns in human striate cortex.

Cortical maps

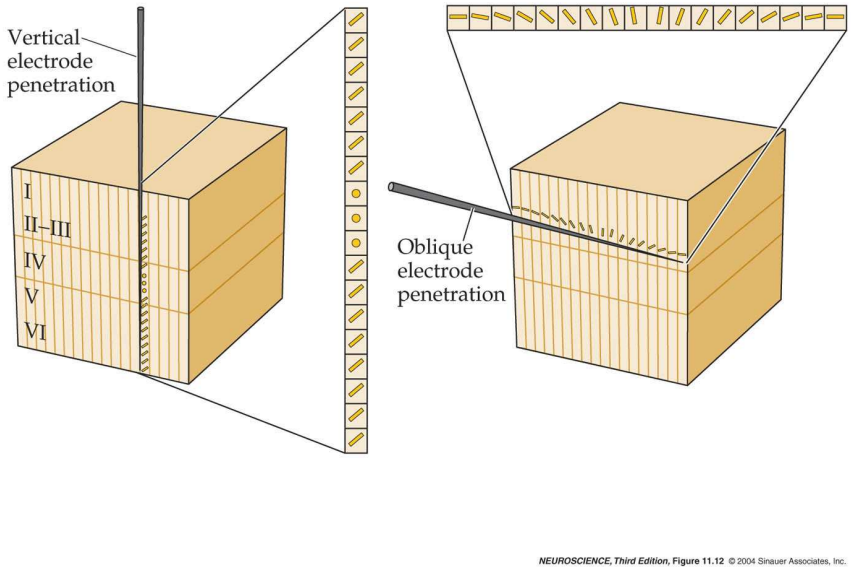
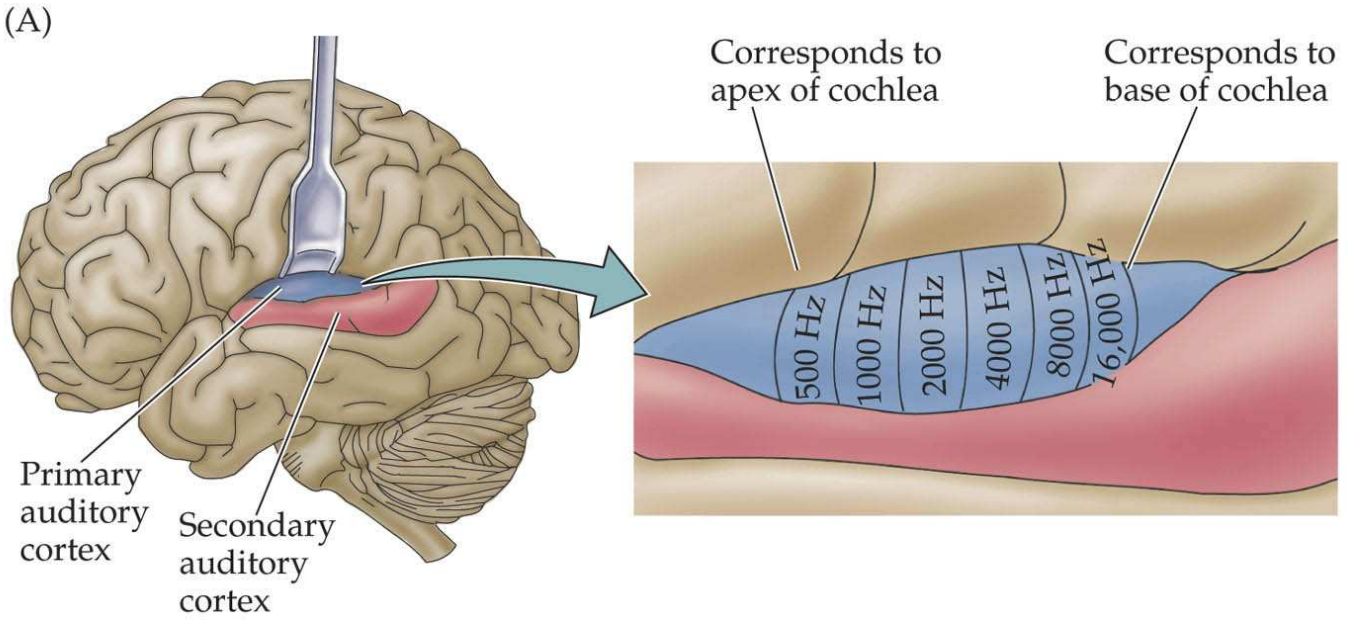


Figure 9: Columnar organization of orientation selectivity in the monkey striate cortex. Vertical electrode penetrations encounter neurons with the same preferred orientations, whereas oblique penetrations show a systematic change in orientation across the cortical surface.

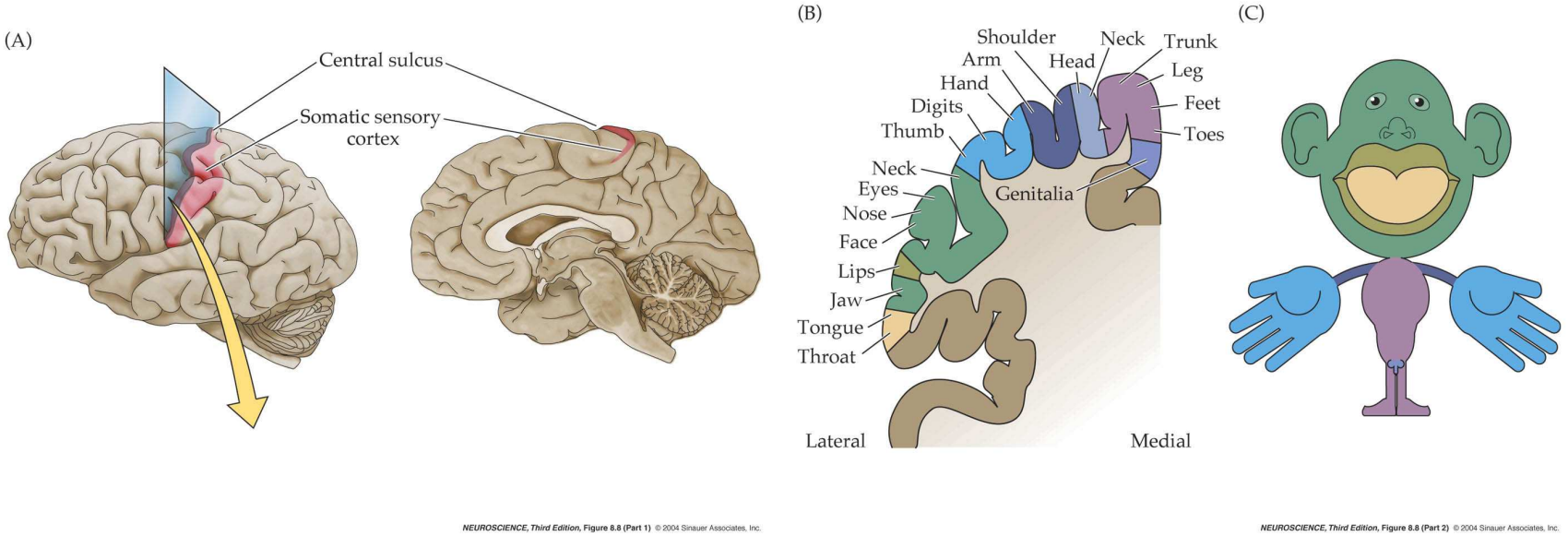
Topographical map: stimulus features are mapped continuously onto the spatial location in the cortex.

Example: Auditory cortex



NEUROSCIENCE, Third Edition, **Figure 12.15 (Part 1)** © 2004 Sinauer Associates, Inc.

Example: Somatosensory cortex



NEUROSCIENCE, Third Edition, Figure 8.8 (Part 1) © 2004 Sinauer Associates, Inc.

NEUROSCIENCE, Third Edition, Figure 8.8 (Part 2) © 2004 Sinauer Associates, Inc.

Typically, maps are deformed representations.

Summary

The brain contains a large number of systems and subsystems.

Information processing is electrical.

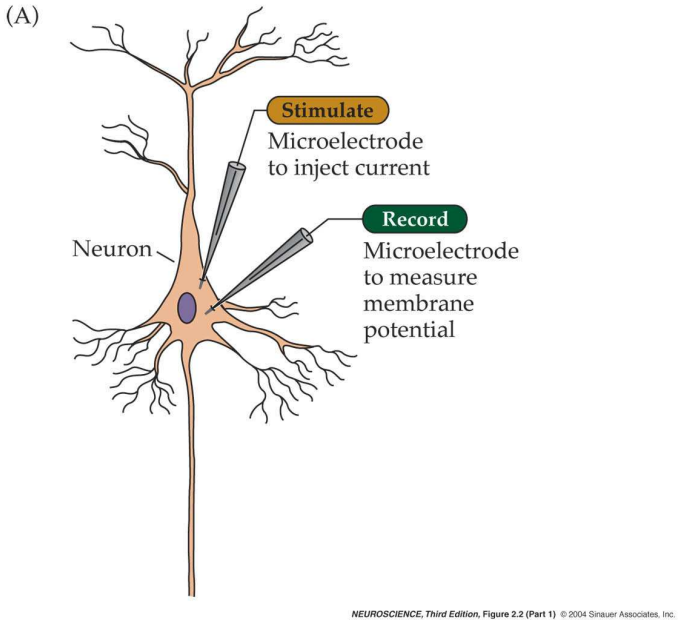
Neurons have specialized function: receptive fields

Cortex is organized in columns and maps.

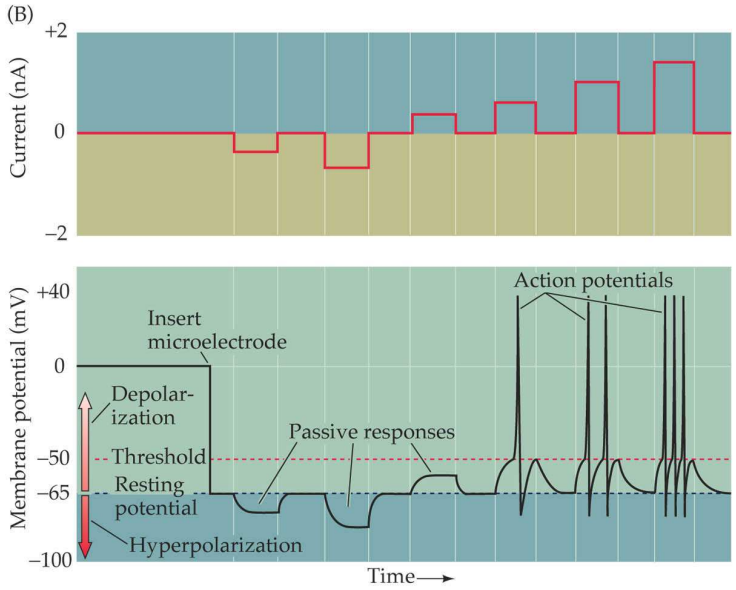
Electrical properties of cells

Nerve cells generate electrical signals that transmit information.

Neurons are not good conductors of electricity



NEUROSCIENCE, Third Edition, Figure 2.2 (Part 1) © 2004 Sinauer Associates, Inc.



NEUROSCIENCE, Third Edition, Figure 2.2 (Part 2) © 2004 Sinauer Associates, Inc.

The resting potential and the action potential can be understood in terms of the nerve cell's selective permeability to different ions and the relative concentrations of these ions inside and outside the cell.

Ion channels

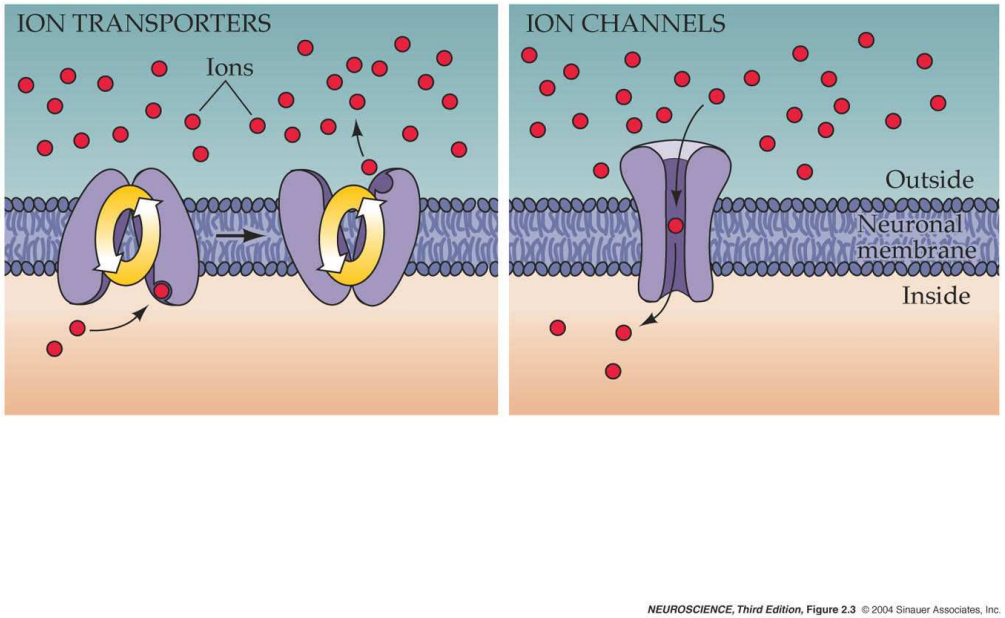


Figure 10: Ion pumps and ion channels are responsible for ionic movements across neuronal membranes.

Electrical potentials across membrane results from concentration differences and selective permeability.

Membrane channels can open or close in response to changes in their direct vicinity (membrane potential, concentration of neurotransmitters, sensory input).

Channels open and close rapidly in a stochastic manner.

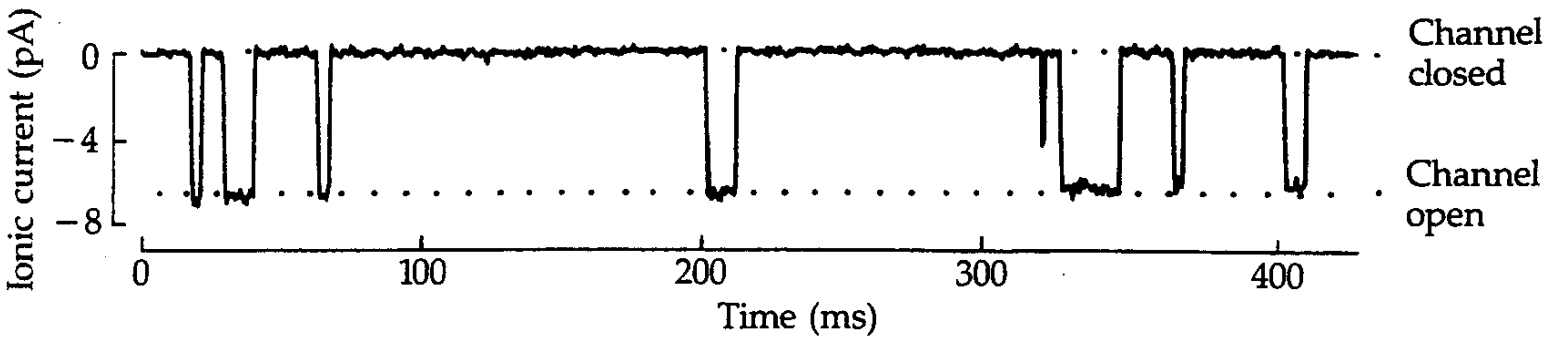
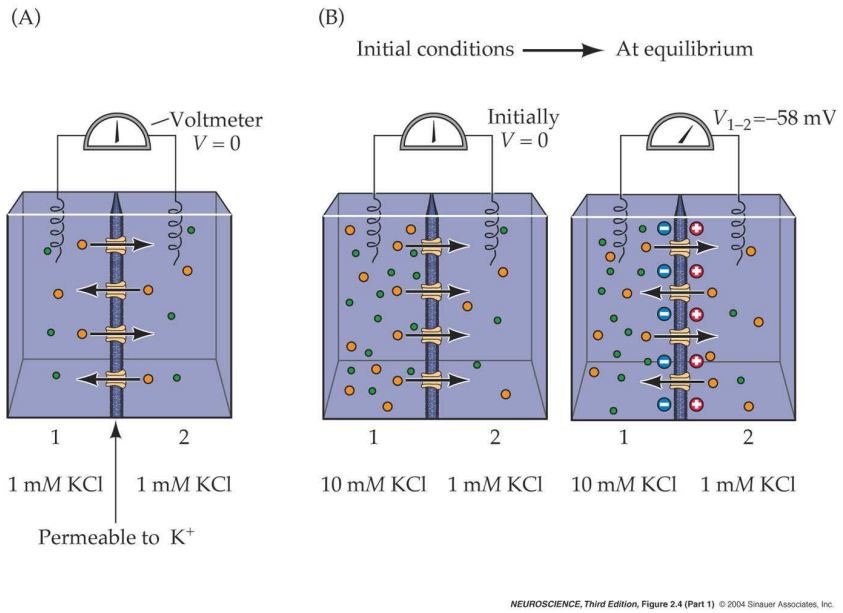


Figure 11: Open-shut gating of an ionic channel showing 8 brief openings. The probability of opening depends on many factors. At -140 mV applied membrane potential, one open channel passes 6.6 pA, corresponding to a flow of 4.1×10^7 ions per second.

The macroscopically observed permeability of the membrane is related to the probability that the channel is open.

The Nernst equation



Membrane is only permeability to K^+ ions.
Concentration gradient drives K^+ ions to the right establishing a charge difference.
Equilibrium: diffusion balances the electrical potential difference.
The equilibrium potential will be an increasing function of concentration ratio of K^+.

Boltzmann statistics: the probability P to encounter a system in equilibrium in a state with energy u is proportional to

$$P \propto \exp\left(-\frac{u}{kT}\right) = \exp\left(-\frac{U}{RT}\right)$$

with k is the Boltzmann constant and T is the absolute temperature.

$U = N_A u$ is the potential energy per mole
$R = kN_A = 8.314$ J mol^{-1} K^{-1} is the Gas constant.

The potential energy for potential V is given by

$$U = z_S e N_A V = z_S F V$$

$F = eN_A = 9.648 \times 10^4$ C mol^{-1} is called the Faraday constant.

Thus,

$$\begin{aligned} \frac{[S]_l}{[S]_r} &= \frac{P_l}{P_r} = \exp\left(-\frac{z_S F}{RT}(V_l - V_r)\right) \\ V_{\text{Nernst}} &= V_l - V_r = \frac{RT}{z_S F}\ln\frac{[S]_r}{[S]_l} \end{aligned}$$

The Nernst equation and describes

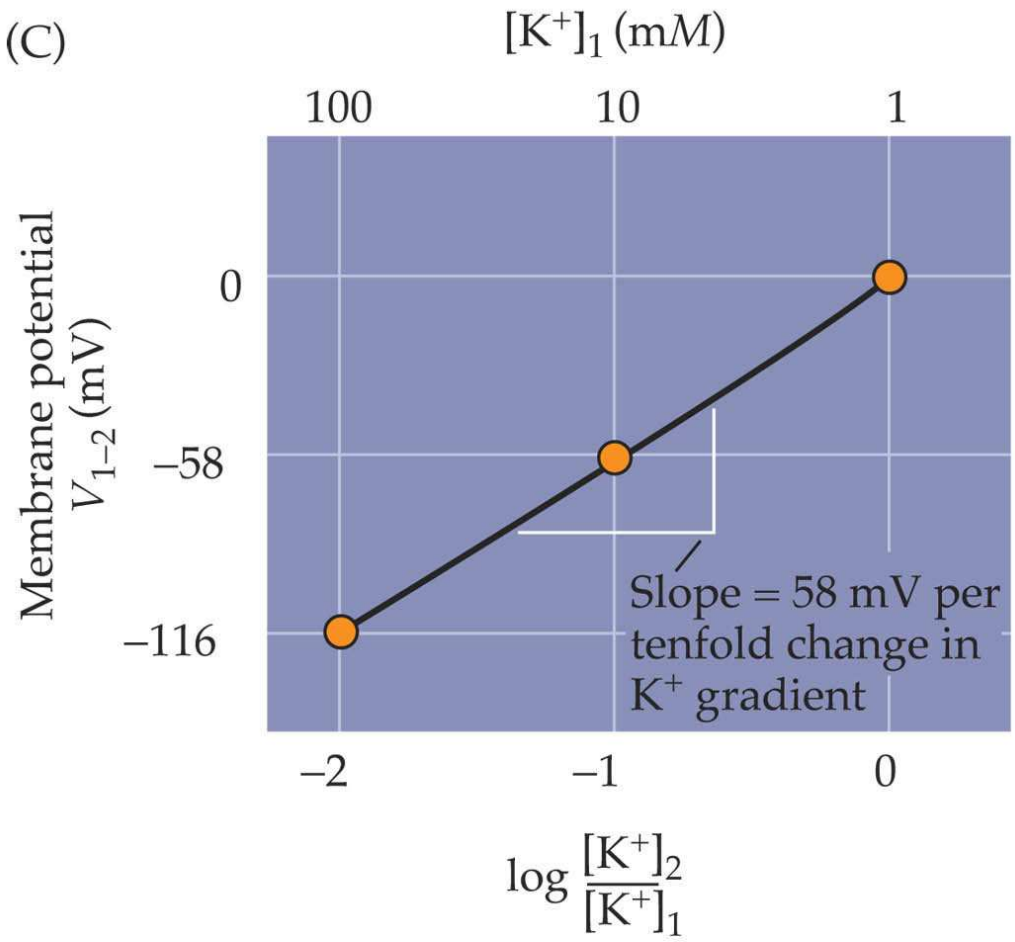
- the equilibrium potential difference when we fix the concentration differences

- the equilibrium concentration ratio when we apply an external potential difference

The Nernst potential depends on temperature. At $T = 20^0$C

$$V_{\text{Nernst}} = \frac{1}{z_S} \log_{10} \left(\frac{[S]_r}{[S]_l} \right) \times 58\text{mV}$$

For biological membranes, $[\text{K}^+]_{\text{out}} : [\text{K}^+]_{\text{in}} = 1 : 10$, yielding a Nernst potential of -58 mV.



NEUROSCIENCE, Third Edition, **Figure 2.4 (Part 2)**

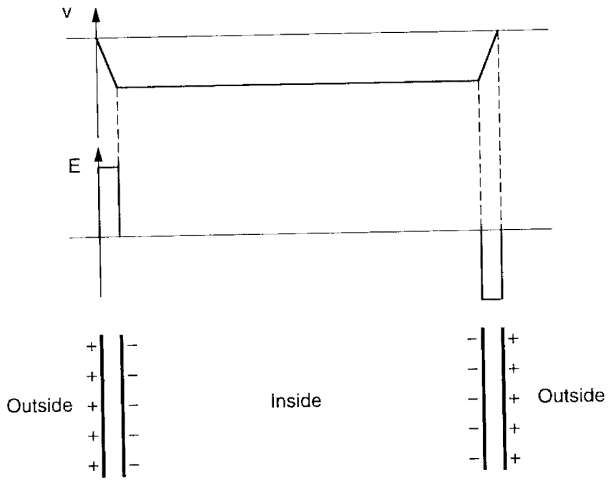
The Goldman equation

In reality, not one but many types of ions are present each of which has its own permeability.

Ion	Intracellular	Extracellular
Squid axon		
K^+	400	20
Na^+	50	440
Cl^-	40-150	560
Ca^{2+}	0.0001	10
Mammalian neuron		
K^+	140	5
Na^+	5-15	145
Cl^-	4-30	110
Ca^{2+}	0.0001	1-2

$$V = \log_{10}\left(\frac{P_{K^+}[K^+]_r + P_{Na^+}[Na^+]_r + P_{Cl^-}[Cl^-]_l}{P_{K^+}[K^+]_l + P_{Na^+}[Na^+]_l + P_{Cl^-}[Cl^-]_r}\right) \times 58\text{mV}$$

The Nernst-Planck equation



Membrane of thickness a :

$$V_{\text{out}} = V(0) = 0 \qquad V_{\text{in}} = V(a) = V$$

Different ions, each with its own concentration $C_i(x)$ and valence z_i:

$$[C_i]_{\text{out}} = C_i(0) \qquad [C_i]_{\text{in}} = C_i(a)$$

Electrical and diffusive forces

The electric force per ion of type i is

$$-z_i e \frac{dV(x)}{dx}.$$

The number of ions per liter is $N_A C_i(x)$, with $C_i(x)$ in units of mol per liter. Therefore, the electric force per unit volume is

$$-z_i C_i(x) F \frac{dV(x)}{dx}$$

The diffusive force on ion i per unit volume is proportional to the concentration gradient as well as the absolute temperature.

$$-RT \frac{dC_i(x)}{dx}$$

in units of Newton per liter.

Current voltage relation

$$I_i = u_i z_i F_i = u_i z_i \left(-RT \frac{dC_i(x)}{dx} - z_i C_i(x) F \frac{dV(x)}{dx} \right)$$

u_i is the mobility, related to permeability.

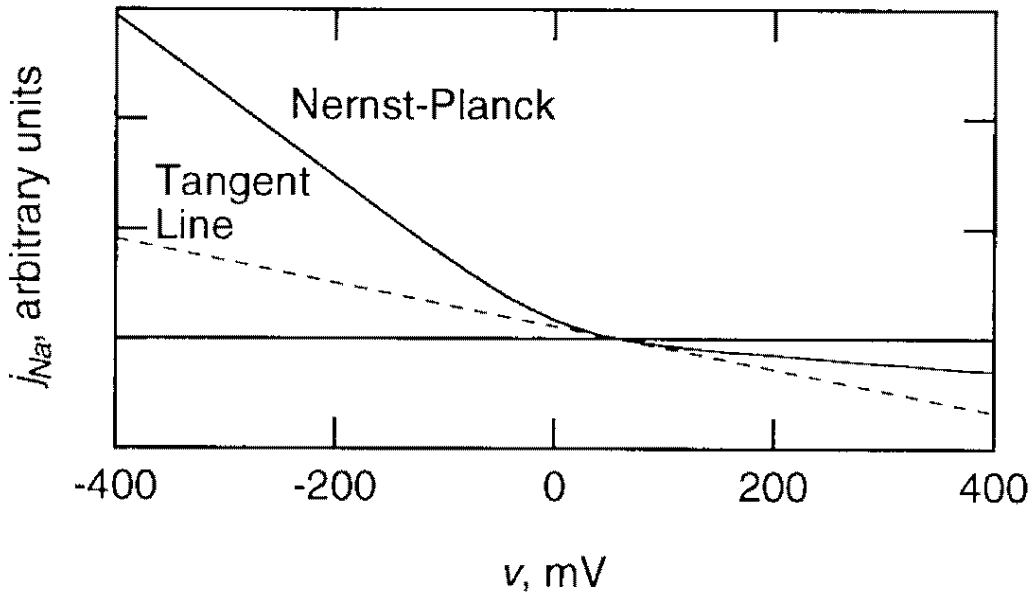
Given I_i and $V(x)$, this is a differential equation in $C_i(x)$ (Nernst-Planck equation).

Note, the current is independent of x.

Assume: $dV/dx = V/a$ for $0 < x < a$.

Then

$$I_i = \frac{-u_i z_i^2 FV}{a} \frac{[C_i]_{\text{out}} - [C_i]_{\text{in}} \exp\left(z_i \frac{FV}{RT}\right)}{1 - \exp\left(z_i \frac{FV}{RT}\right)}$$



Relation between current and voltage in the presence of ionic concentration difference shows rectification.

When $[C_i]_{\text{out}} = [C_i]_{\text{in}}$, the current voltage relation is linear as in Ohms law.

For $i = \mathsf{K}^+, \mathsf{Na}^+, \mathsf{Cl}^-$,

$$I = I^+_{\mathsf{K}} + I^+_{\mathsf{Na}} + I^-_{\mathsf{Cl}} = \frac{-FV}{a}\frac{w - ye^{FV/RT}}{1 - e^{FV/RT}}$$

with

$$\begin{aligned} w &= u_{\mathsf{K}^+}[\mathsf{K}^+]_{\mathsf{out}} + u_{\mathsf{Na}^+}[\mathsf{Na}^+]_{\mathsf{out}} + u_{\mathsf{Cl}^-}[\mathsf{Cl}^-]_{\mathsf{in}} \\ y &= u_{\mathsf{K}^+}[\mathsf{K}^+]_{\mathsf{in}} + u_{\mathsf{Na}^+}[\mathsf{Na}^+]_{\mathsf{in}} + u_{\mathsf{Cl}^-}[\mathsf{Cl}^-]_{\mathsf{out}} \end{aligned}$$

The form is particularly simple because $z_i = \pm 1$.

In the stationary case, there will be no net movement of charge and $I = \frac{dQ}{dt} = 0$:

$$\frac{FV}{RT} = \log\left(\frac{w}{y}\right)$$

This is the Goldman equation.

Linearized system

For small currents, we can linearize the I-V relation around the stationary solution.

$$\begin{aligned} G &= \frac{dI}{dV}_{V=V_0} = \frac{F}{a}\frac{wy}{y-w}\log\frac{y}{w} \\ I &\approx G(V-V_0) + \mathcal{O}((V-V_0)^2) \end{aligned}$$

G has units of Ω^{-1} per liter and is called the conductance and is the inverse resistance.

V_0 the equilibrium membrane potential given by the Goldman equation.

The Hodgkin-Katz experiment

Squid axon Nernst potentials:

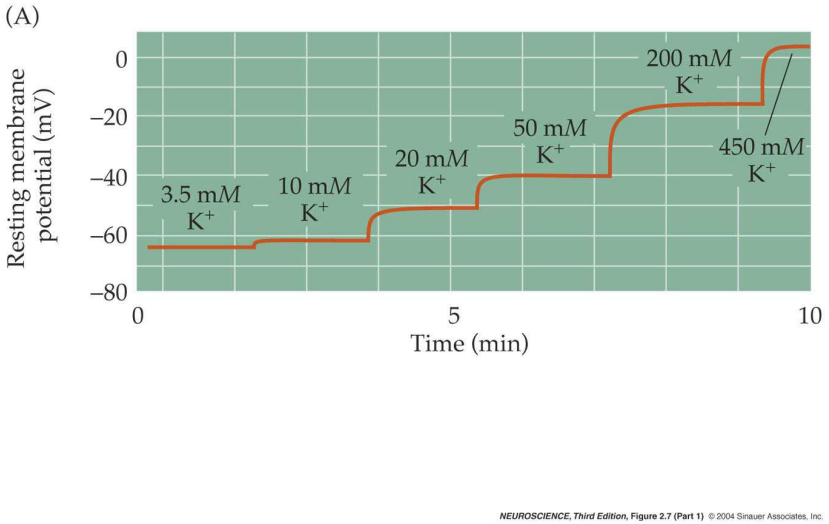
$$V_{\mathsf{K}^+} \approx -58mV, \quad V_{\mathsf{Na}^+} \approx +55mV$$

V_{K^+} is close to the resting membrane potential (-65 mV).

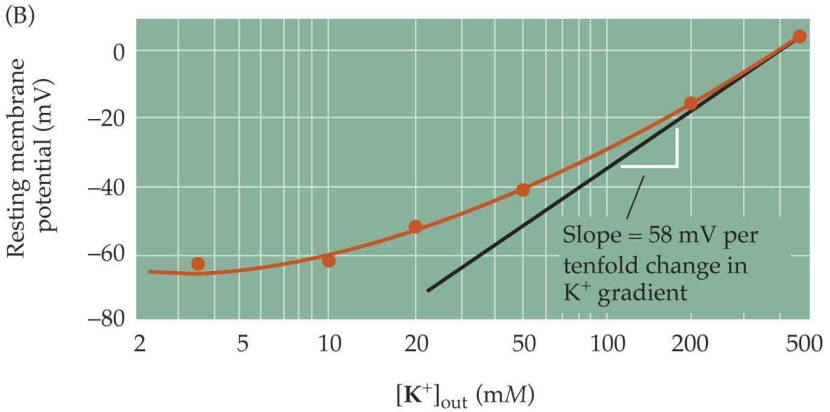
This fact suggests that the resting membrane is more permeable to K^+ than to the other ions.

This implies that membrane rest potential depends on (external) K^+ concentration.

The Hodgkin-Katz experiment



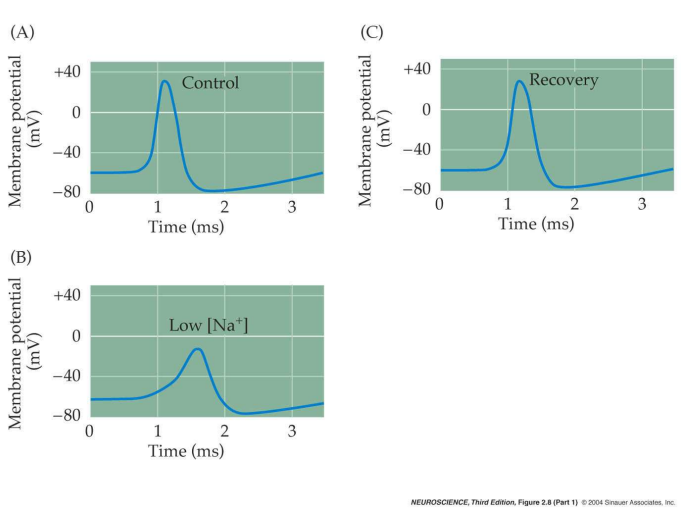
NEUROSCIENCE, Third Edition, Figure 2.7 (Part 1) © 2004 Sinauer Associates, Inc.

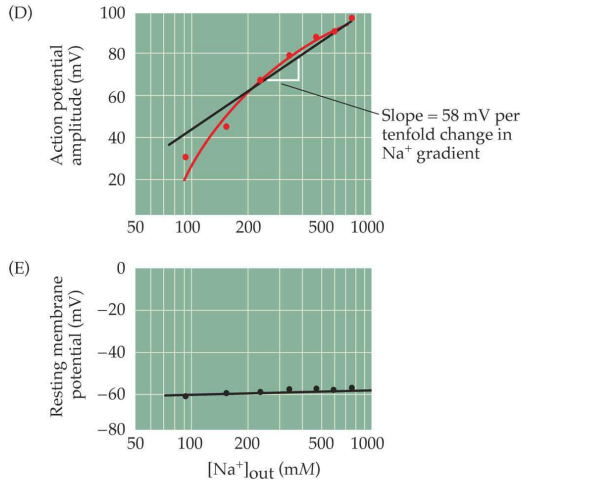


NEUROSCIENCE, Third Edition, Figure 2.7 (Part 2) © 2004 Sinauer Associates, Inc.

The role of Na^+

During an action potential the membrane resting potential reverses from negative to positive, near the Na^+ Nernst potential.





Changing the external sodium concentration
affects the size and rise time of the action potential.
does not affect the resting membrane potential.
Membrane becomes membrane is very permeable to Na^+ during action potential.

Permeability changes during action potential

We can use the Goldman equation to investigate the change in the membrane potential when we change the concentrations of the ions outside the cell.

$$\frac{FV}{RT} = \log\left(\frac{u_{\mathsf{K}^+}[\mathsf{K}^+]_{\text{out}} + u_{\mathsf{Na}^+}[\mathsf{Na}^+]_{\text{out}} + u_{\mathsf{Cl}^-}[\mathsf{Cl}^-]_{\text{in}}}{u_{\mathsf{K}^+}[\mathsf{K}^+]_{\text{in}} + u_{\mathsf{Na}^+}[\mathsf{Na}^+]_{\text{in}} + u_{\mathsf{Cl}^-}[\mathsf{Cl}^-]_{\text{out}}}\right)$$

$$[\mathsf{K}^+]_i = 345mM, \quad [\mathsf{Na}^+]_i = 72mM, \quad [\mathsf{Cl}^-]_i = 61mM$$

Varying the ion concentrations outside the cell we find membrane potential in agreement with the Goldman equation:

- resting membrane potential: $P^+_{\mathsf{K}} : P^+_{\mathsf{Na}} : P^-_{\mathsf{Cl}} = 1 : 0.04 : 0.45$
- the peak of the action potential: $P^+_{\mathsf{K}} : P^+_{\mathsf{Na}} : P^-_{\mathsf{Cl}} = 1 : 20 : 0.45$
- the refractory period: $P^+_{\mathsf{K}} : P^+_{\mathsf{Na}} : P^-_{\mathsf{Cl}} = 1.8 : 0 : 0.45$

TABLE 7.

State of nerve	Solution	Composition of test solution K mM.	Na mM.	Cl mM.	Change in membrane potential on substituting test solution for sea water or artificial sea water: Observed mV.	Calculated mV.	Permeability coefficients used in calculation: P_K	P_{Na}	P_{Cl}
Resting	A	0	465	587	+ 3	+ 5	1	0·04	0·45
	B	15	450	587	− 2	− 2			
	C	20	445	587	− 4	− 4			
	D	7	324	384	0	+ 1			
	E	5	227	270	+ 2	+ 2			
	F	3	152	180	+ 2	+ 2			
	G	2	91	108	+ 4	+ 3			
	H	10	573	658	+ 1	0			
	I	10	711	796	− 2	0			
Active (peak of spike)	A	0	465	587	+ 1	− 1	1	20	0·45
	B	15	450	587	+ 1	0			
	C	20	445	587	+ 5	+ 1			
	D	7	324	384	+ 9	+ 8			
	E	5	227	270	+21	+16			
	F	3	152	180	+44	+25			
	G	2	91	108	+59	+38			
	H	10	573	658	− 3	− 5			
	I	10	711	796	− 9	−10			
Refractory (maximum of positive phase)	A	0	465	587	+13	+12	1·8	0	0·45
	B	15	450	587	− 6	− 5			
	C	20	445	587	−10	− 9			
	D	7	324	384	+ 1	+ 1			
	E	5	227	270	+ 4	+ 2			
	F	3	152	180	+ 4	+ 3			
	G	2	91	108	0	+ 3			
	H	10	573	658	+ 1	+ 1			
	I	10	711	796	0	+ 3			
Membrane potential at rest in sea water					+48 + J	+59	As above		
Membrane potential at height of activity in sea water					−40 + J	−38			
Membrane potential at maximum of positive phase					+62 + J	+74			
Action potential in sea water					88	97			
Positive phase in sea water					14	15			

Solutions A, B and C were tested against an artificial sea-water solution containing 10 mM-K 455 mM-Na, 587 mM-Cl. Solutions D–I were tested against sea water containing approximately 10mM-K, 455 mM-Na, 540 mM-Cl. Calculated potentials were obtained from equation 4 using values of $(K)_i = 345$ mM., $(Na)_i = 72$ mM., $(Cl)_i = 61$ mM. J is the liquid junction potential between the axoplasm and the sea water in the microelectrode.

Summary

The resting membrane potential is the result of
different ions concentrations inside and outside the cell
the specific permeability of the membrane to different ions

The relation between ionic concentrations and equilibrium membrane potential is described by the Nernst equation for single ions and by the Goldman equation for multiple ions.

At rest, the nerve cell is mainly permeable to K^+ ions resulting in a negative resting membrane potential.

During the action potential, the Na^+ permeability dominates and the membrane potential reverses sign. The increased Na^+ permeability is short, resulting in a short voltage spike.

As we will see, the change in the neural membrane potential itself affects the membrane permeability.

The voltage clamp technique

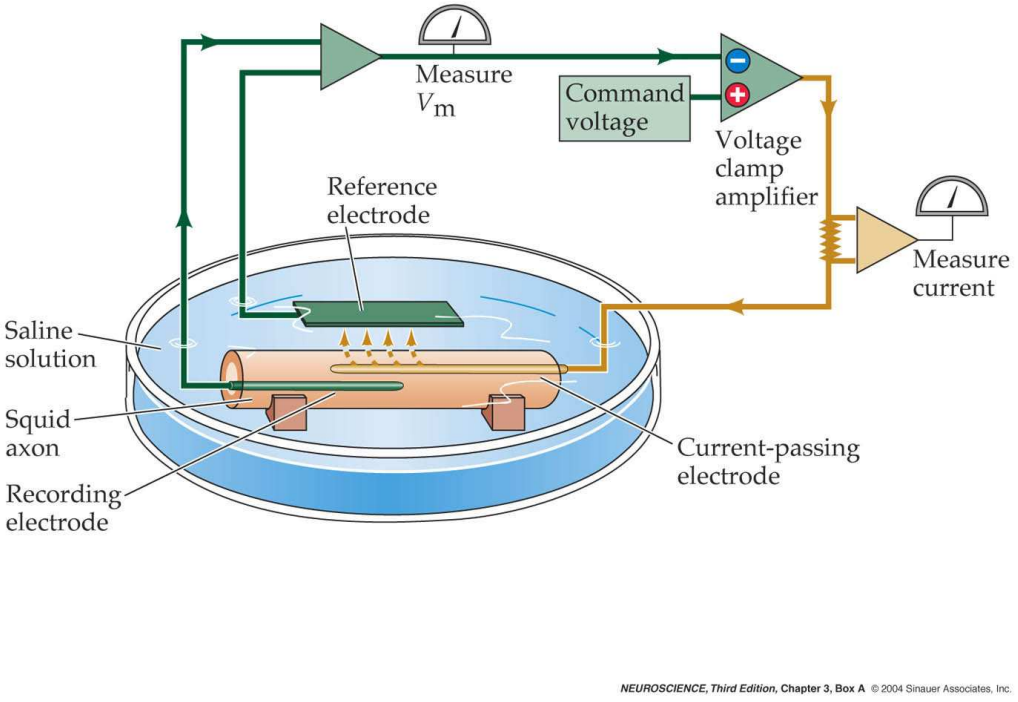


Figure 12: 1) One internal electrode measures membrane potential and is connected to the voltage clamp amplifier. 2) Amplifier compares membrane potential to the desired potential. 3) When different, the amplifier injects current into the axon through a second electrode, and is measured (4).

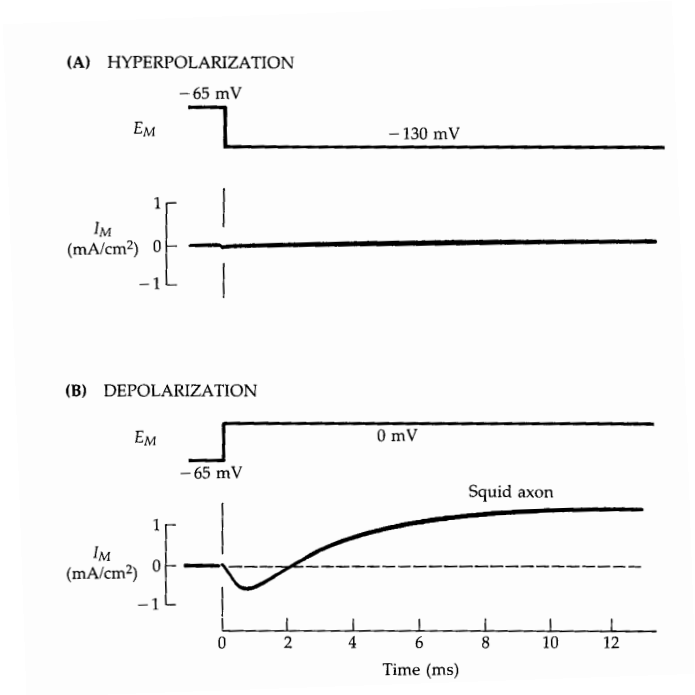
Two types of voltage dependent ionic currents

In the late 1940s, Alan Hodgkin and Andrew Huxley used the voltage clamp technique to work out the permeability changes underlying the action potential.

Giant neuron of the squid because its large size (up to 1 mm in diameter) allowed insertion of the electrodes necessary for voltage clamping.

To investigate the voltage dependent permeability of the membrane they asked whether ionic currents flow across the membrane when its potential is changed.

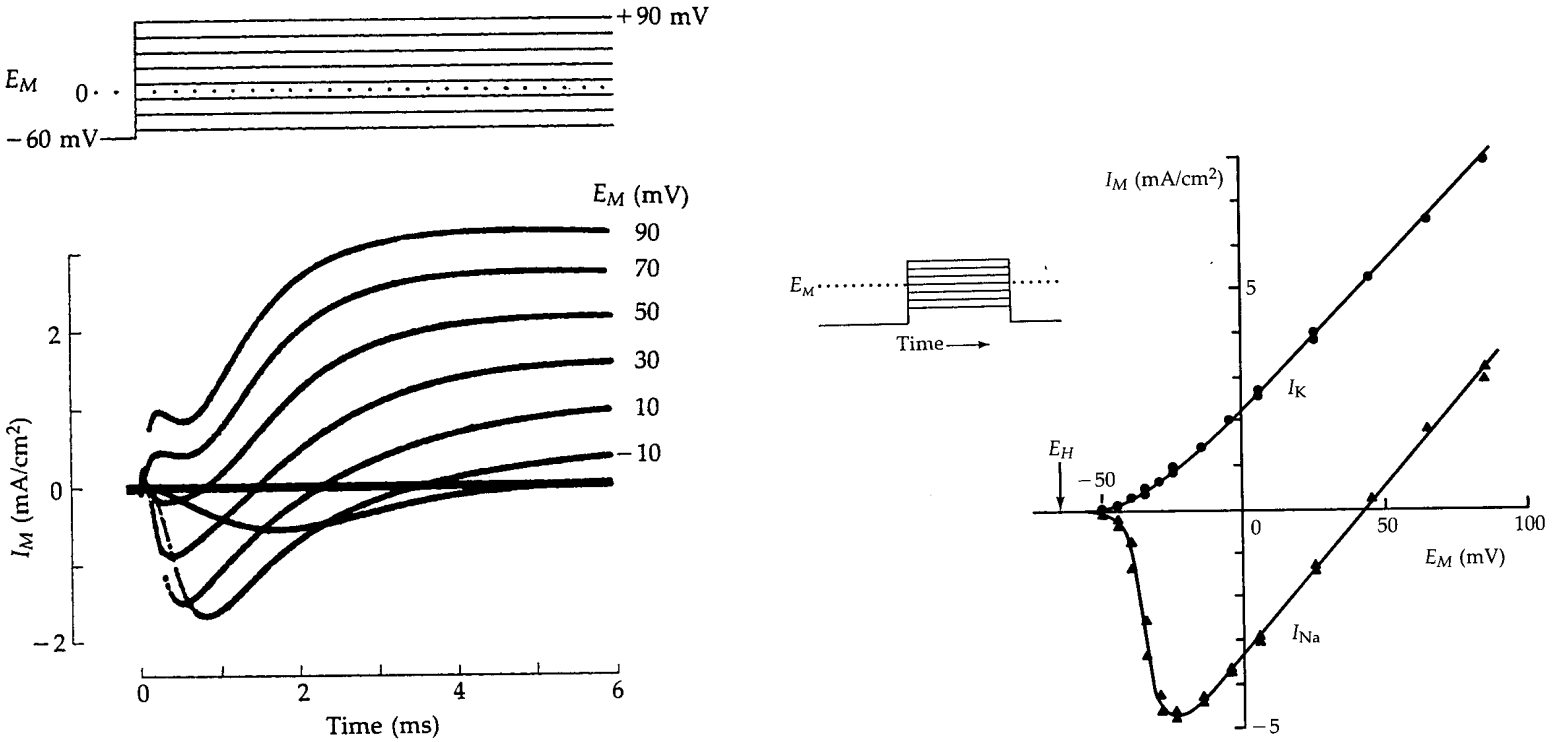
Two types of voltage dependent ionic currents



Hyper-polarization to -130 mV yields very little current.

Depolarization of the membrane potential to 0 mV produces a rapidly rising inward ionic current, which later changes into an outward current.

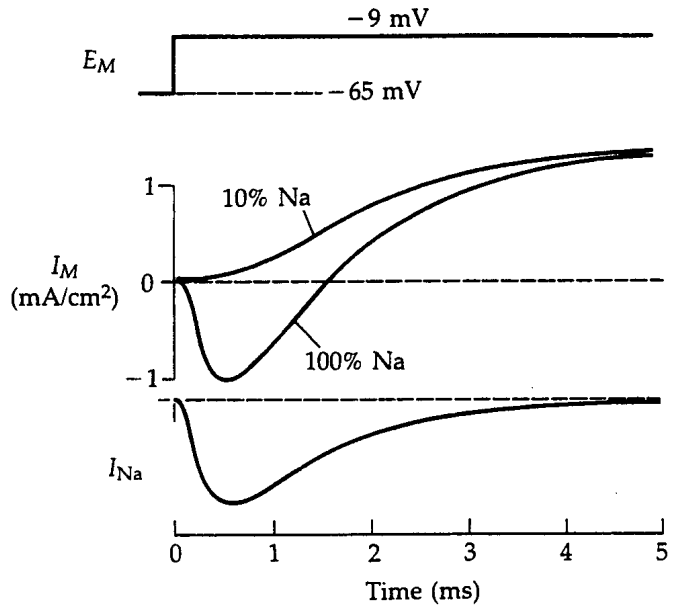
Dependence of currents on voltage



Early current first decreases then increases with voltage. Changes sign at $\approx 50mV$. Suggest that Na^+ is involved.

The late current increases monotonically with increasingly positive membrane potentials.

Separating the currents



Reducing the external Na^+ concentration by factor 10 removes early current.

In this case, both internal and external Na^+ concentrations are approximately equal and the Na^+ Nernst potential is close to 0 mV.

The reduction of external Na^+ has no effect on the outward current.

Pharmacological separation

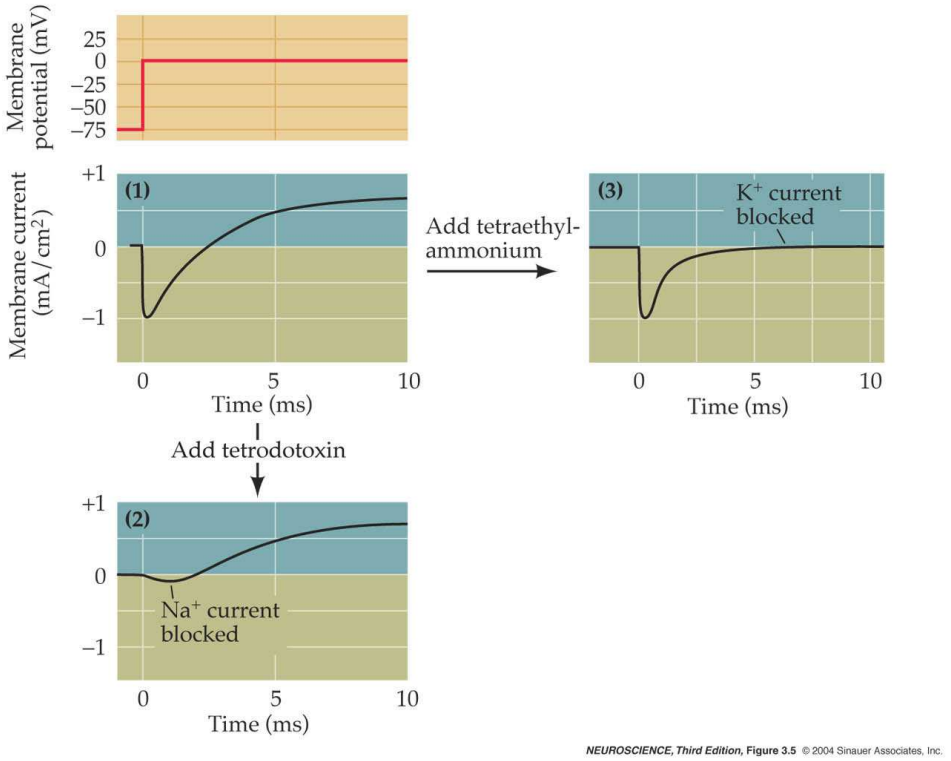
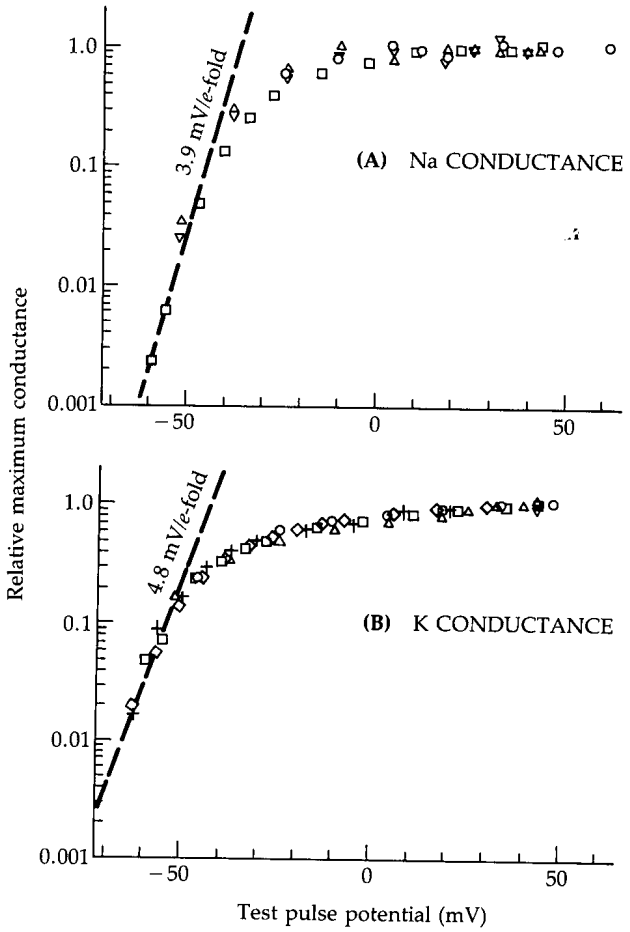


Figure 13: Panel 1 shows the current that flows when the membrane potential of a squid axon is depolarized to -10 mV in control conditions. 2) Treatment with tetrodotoxin causes the early Na^+ currents to disappear but spare the late K^+ currents. 3) Addition of tetraethyl-ammonium blocks the K^+ currents without affecting the Na^+ currents.

Voltage dependent conductance



Voltage dependent conductance

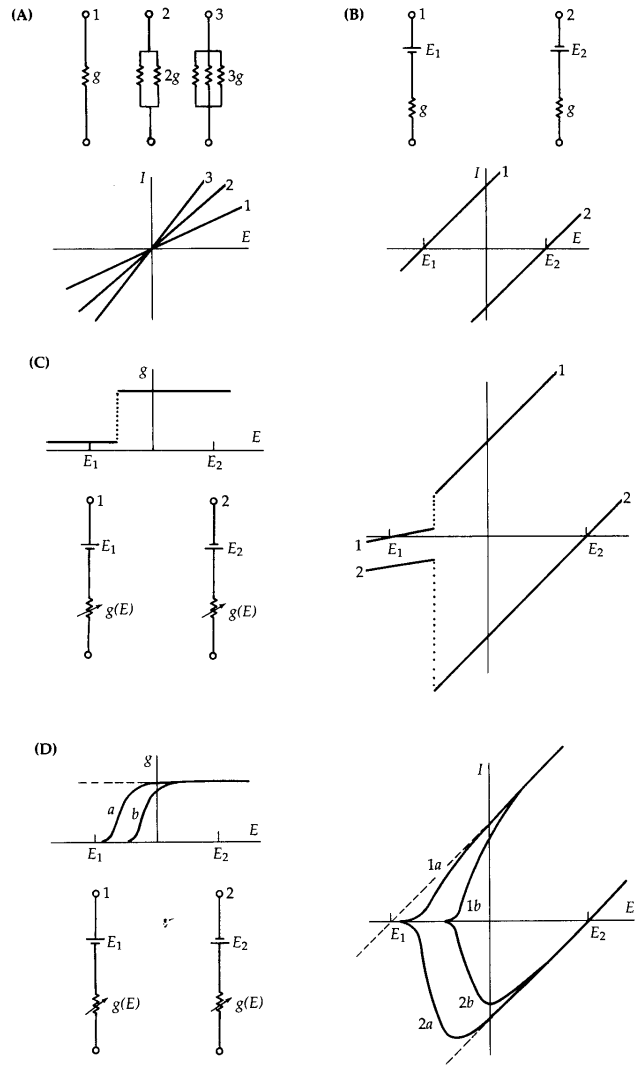
From the I-V measurements and

$$I_i = g_i(V,t)(V - V_i), \quad i = \mathrm{K}^+, \mathrm{Na}^+$$

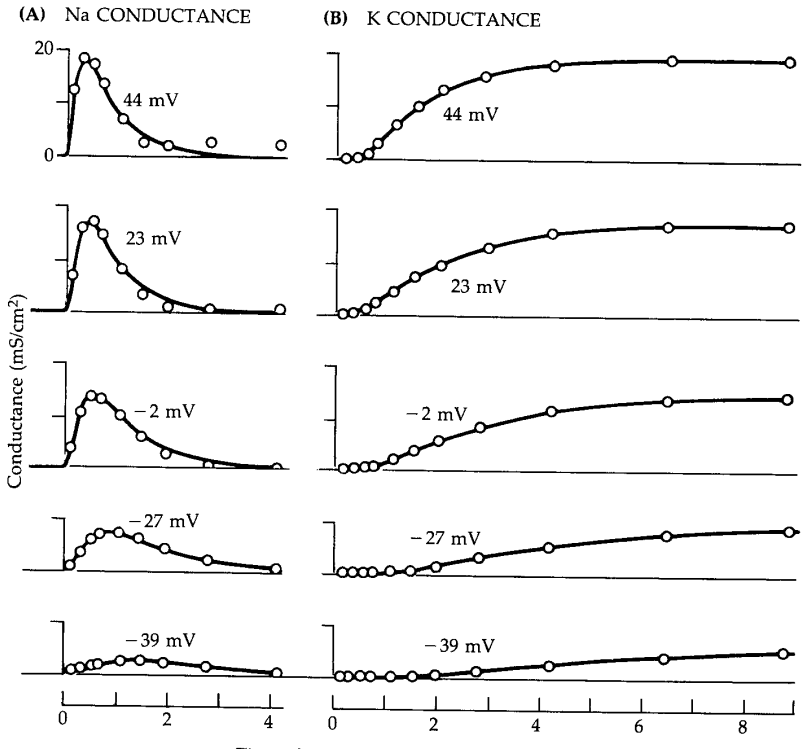
one can compute the peak conductance $g_i(V)$ as a function of voltage.

The peak magnitude of Na^+ conductance and steady-state value of K^+ conductance both increase steeply as the membrane potential is depolarized.

Illustration of voltage dependent conductance



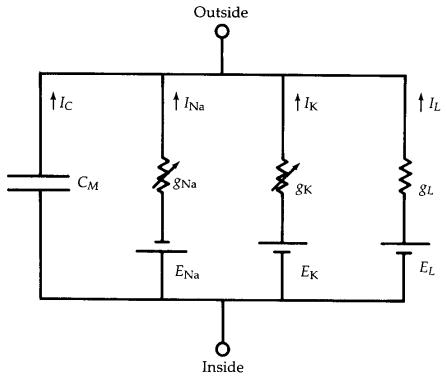
Conductance dynamics



The conductances change over time.

Na^+ is fast. Activates and inactivates.
K^+ is slower.

The Hodgkin-Huxley model



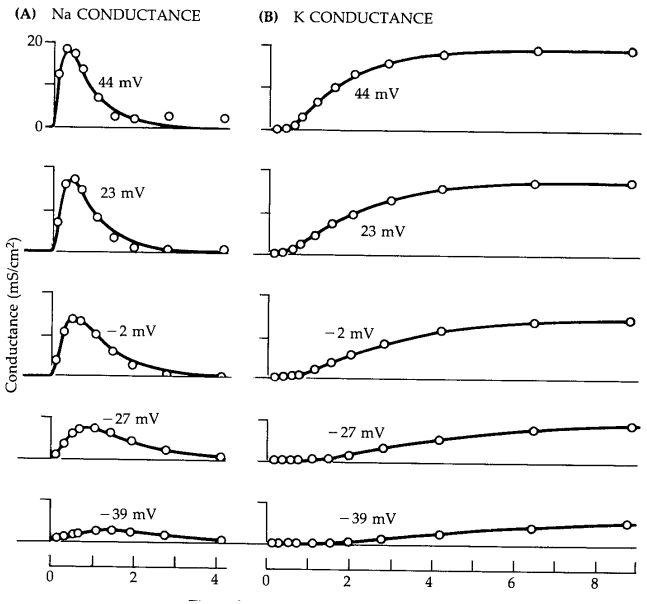
Ionic currents:

$$I_i = g_i(V,t)(V - V_i), \quad i = \mathsf{K}^+, \mathsf{Na}^+, \mathsf{leak}$$

Capacitative current:

$$I_c = C\frac{dV}{dt}, \qquad C = \frac{\epsilon\epsilon_0 A}{d}$$

The K^+ conductance

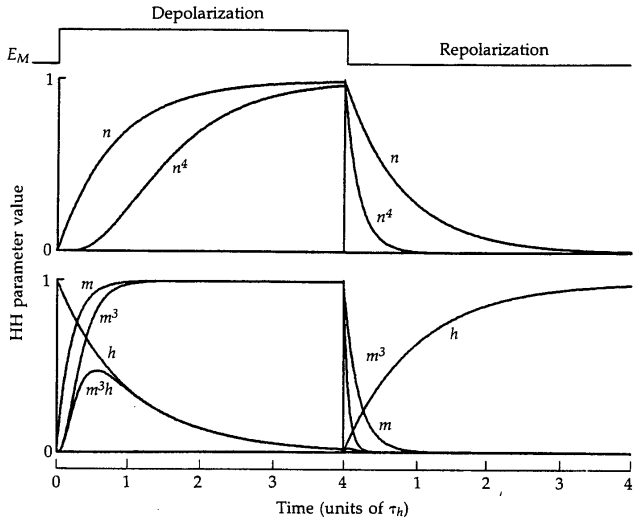


$$g_{\mathsf{K}}^{+} = \bar{g}_{\mathsf{K}^+} n^4$$

$$\tau_n(V)\frac{dn}{dt} = n_\infty(V) - n$$

Conductance modeled as a dynamic quantity.

The K^+ conductance

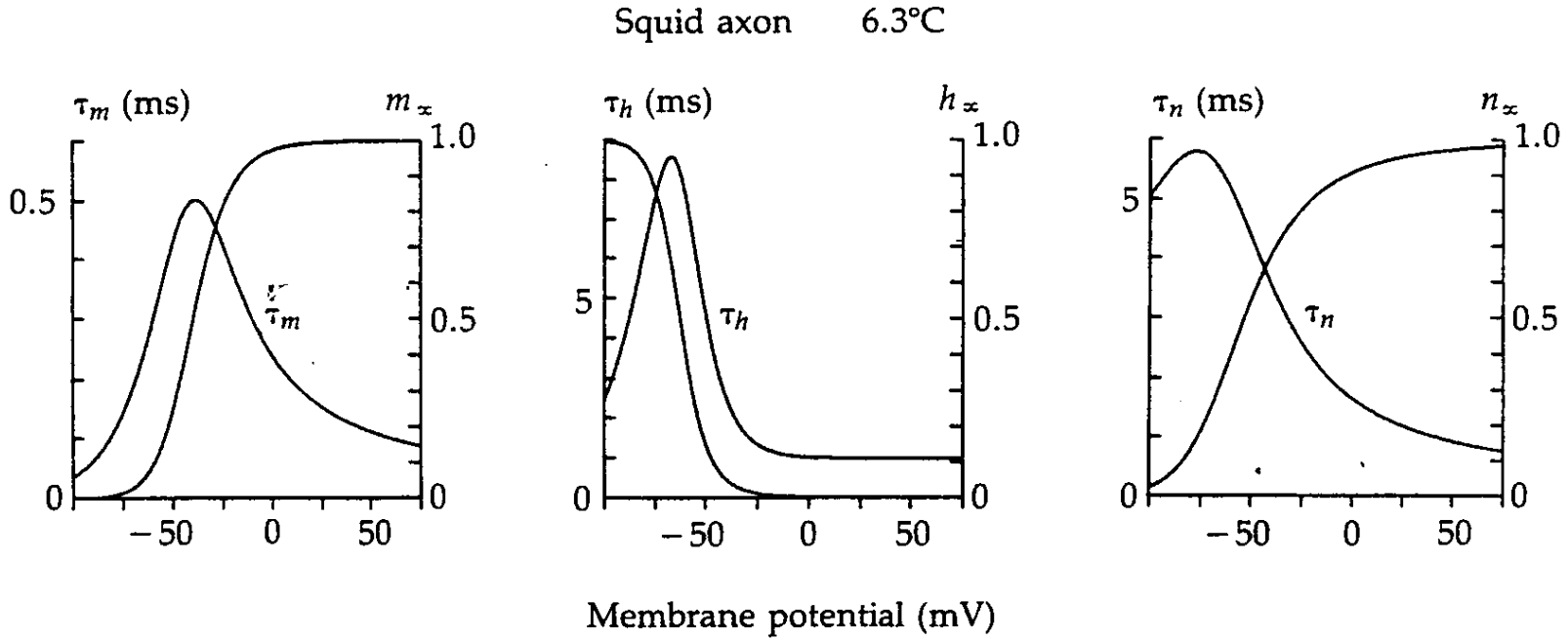


$$g_{\mathsf{K}}^{+} = \bar{g}_{\mathsf{K}^+} n^4$$

$$\tau_n(V)\frac{dn}{dt} = n_\infty(V) - n$$

Conductance modeled as a dynamic quantity.

Voltage dependence of the constants



The Na^+ conductance

The Na^+ conductance depends on independent activation and inactivation processes.

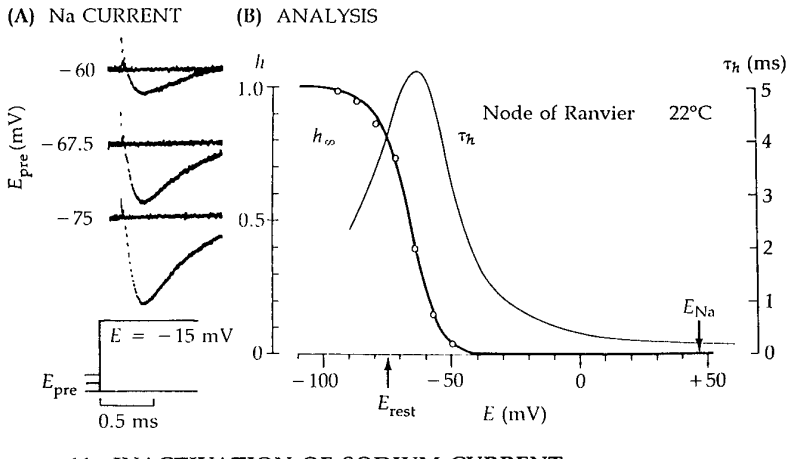
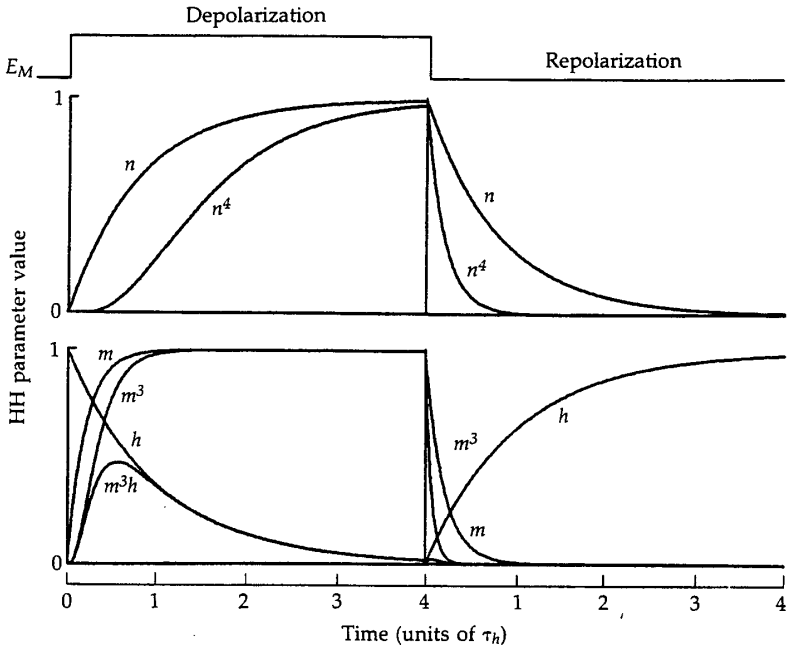


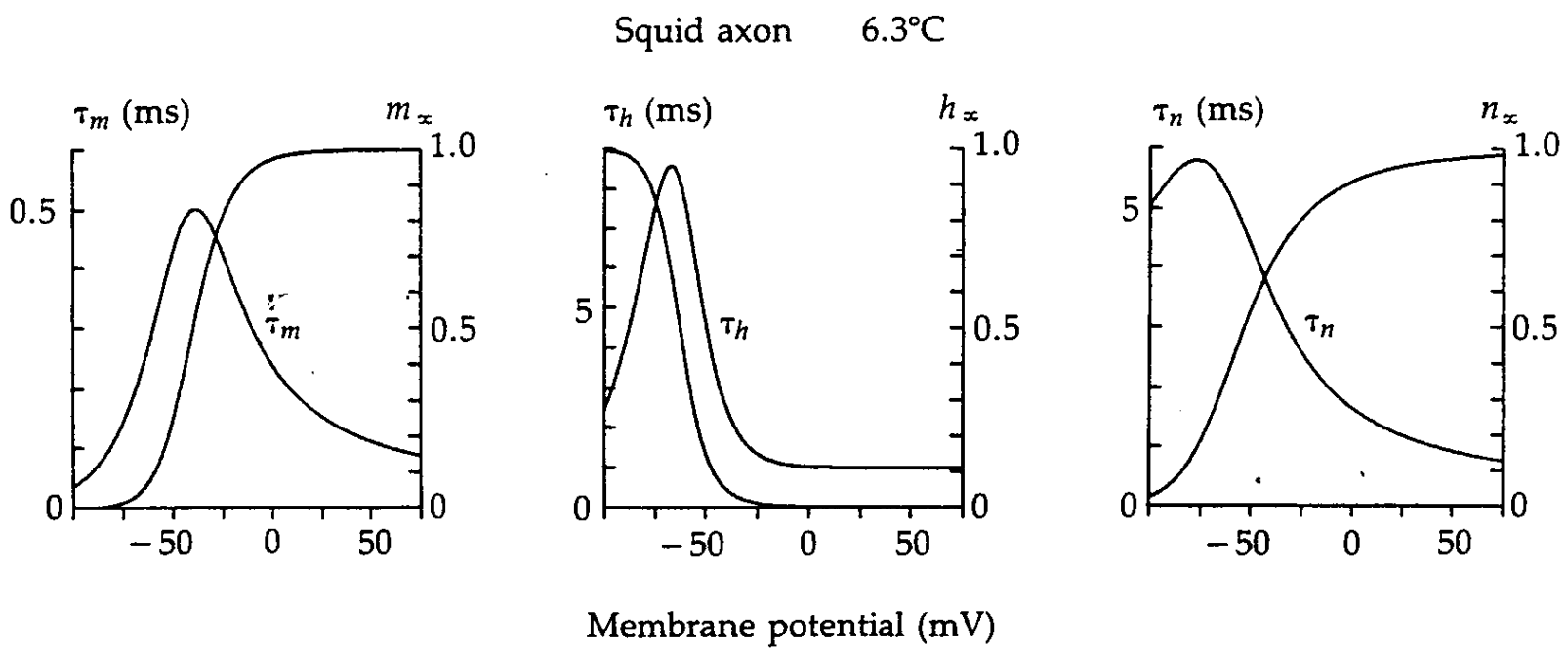
Figure 14: Sodium currents elicited by test pulses to -15 mV after 50 milliseconds pre-pulses to three different levels. The peak value of the Na^+ current decreases with increasing pre-step membrane potential due to Na^+ inactivation. B) The inactivation $1 - h_\infty$ is an increasing function of the membrane potential.

The Na^+ conductance

$$
\begin{aligned}
g_{\mathsf{Na}^+} &= \bar{g}_{\mathsf{Na}^+} m^3 h \\
\tau_m(V)\frac{dm}{dt} &= m_\infty(V) - m \\
\tau_h(V)\frac{dh}{dt} &= h_\infty(V) - h
\end{aligned}
$$



The Na^+ conductance



Action potentials

Summary of the Hodgkin Huxley model:

$$I_c + I_{\mathsf{Na}^+} + I_{\mathsf{K}^+} + I_{\mathsf{leak}} + I_{\mathsf{ext}} = 0$$

leads to 4 non-linear first order differential equations:

$$\begin{aligned} C\frac{dV}{dt} &= -m^3 h \bar{g}_{\mathsf{Na}^+}(V - V_{\mathsf{Na}^+}) - n^4 \bar{g}_{\mathsf{K}^+}(V - V_{\mathsf{K}^+}) \\ &\quad -\bar{g}_{\mathsf{leak}}(V - V_{\mathsf{leak}}) - I_{\mathsf{ext}}(t) \\ \tau_n \frac{dn}{dt} &= n_\infty - n \quad \tau_m \frac{dm}{dt} = m_\infty - m \quad \tau_h \frac{dh}{dt} = h_\infty - h \end{aligned}$$

V_i and $\bar{g}_i, i = \mathsf{K}^+, \mathsf{Na}^+, \mathsf{leak}$ are constants

$\tau_{n,m,h}$ and n_∞, m_∞ and h_∞ are voltage dependent

Spike wave form

The HH equations not only describe the voltage clamp experiments, but also correctly generate the form and time course of the action potential

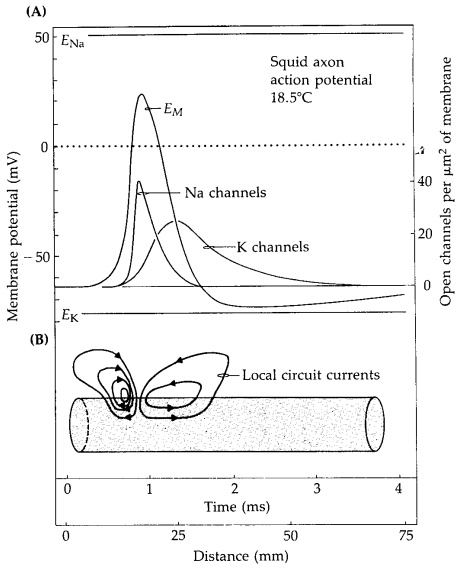
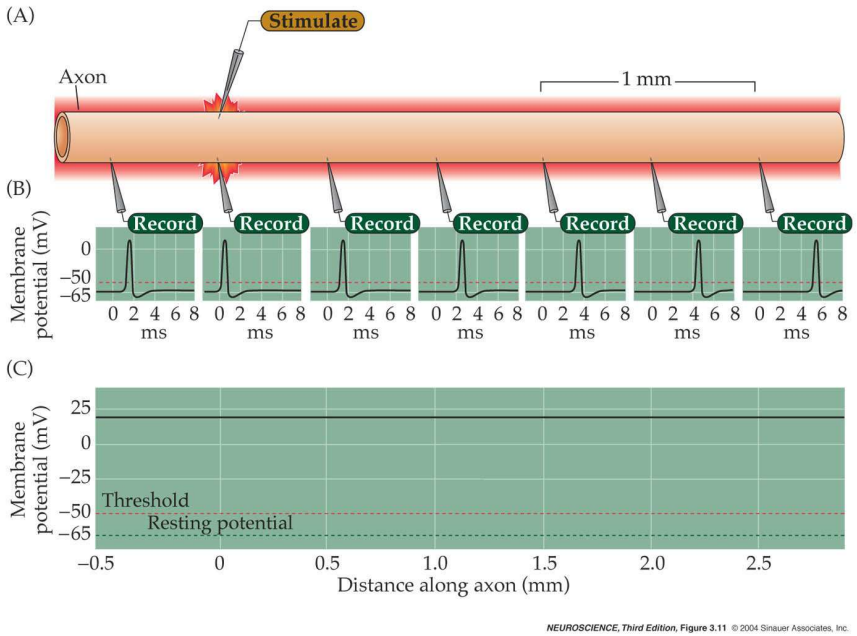


Figure 15: A) The solution of the Hodgkin-Huxley eqs. for the membrane potential V and the conductances $g_{\mathrm{K}+}$ and $g_{\mathrm{Na}+}$ as a function of time.

Spike propagation

Stimulus $\longrightarrow$ Membrane depolarization $\longrightarrow$ rapid opening of Na^+ channels $\longrightarrow$ inrush of Na^+ ions $\longrightarrow$ further depolarization of the membrane potential.

Slower inactivation of Na^+ and activation of K^+ channels restores the membrane potential below its resting value.



NEUROSCIENCE, Third Edition, Figure 3.11 © 2004 Sinauer Associates, Inc.

Spike velocities

Neural information processing depends on spike propagation from one cell to the next. The action potential lasts about 1 μsec and travels at 1-100 m/sec.

Table 1. Conduction Velocities in Nerve and Muscle

Tissue	*Temperature* °C	*Myelinated* (M) *or unmyelinated* (U)	*Fibre Diameter* μ	*Velocity* m/sec.	*Notes*
Cat myelinated nerve fibres	38	M	2–20	10–100	a
Cat unmyelinated nerve fibres	38	U	0·3–1·3	0·7–2·3	a
Frog myelinated nerve fibres	24	M	3–16	6–32	b
Prawn myelinated nerve fibres	20	M	35	20	c
Crab large nerve fibres	20	U	30	5	d
Squid giant axon	20	U	500	25	d
Frog muscle fibre	20	U	60	1·6	d

(a) References, particularly Hursh (1939) in Patton (1960).
(b) Tasaki (1953).
(c) Holmes, Pumphrey & Young (1942).
(d) References and data in Katz (1948).
For myelinated fibres the figure given is the external diameter of the myelin.

Spike propagation is an active process, more like burning a fuse than electrical signaling in copper wire.

Axon longitudinal resistance is exceedingly high due to small diameter.

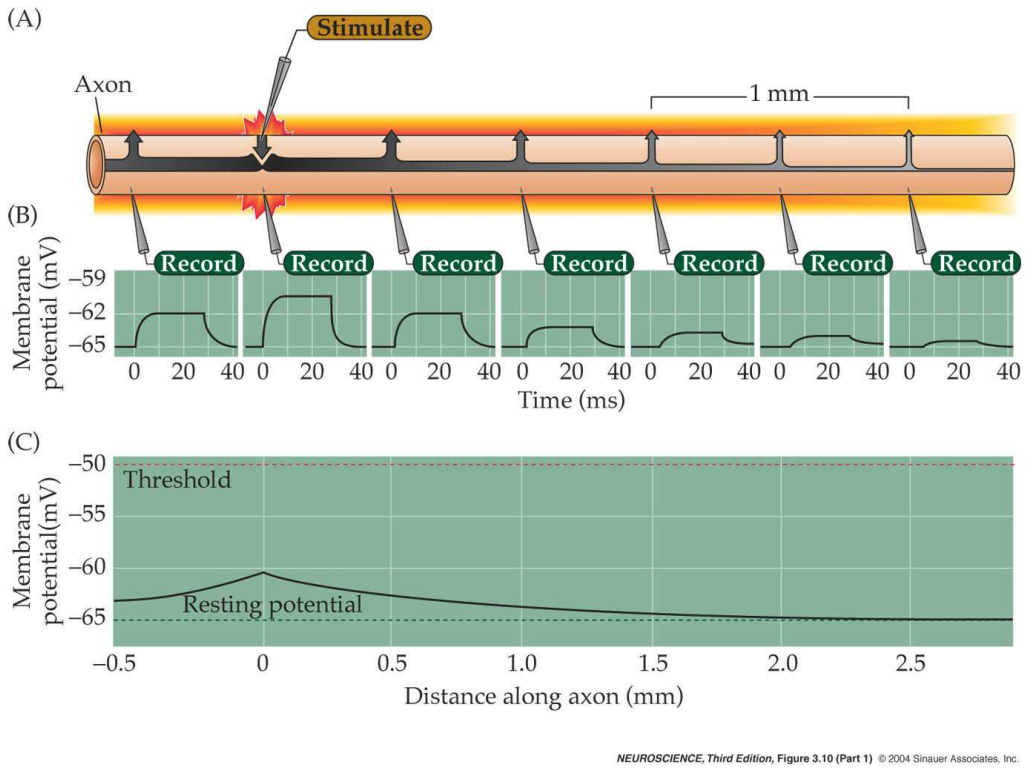
$$R = \frac{\rho l}{A}, \quad A = \frac{\pi d^2}{4}$$

Axoplasm resistivity $\rho = 100\Omega cm$, diameter $d = 1\mu m$, $l = 1m$ yields $R = 10^{12}\Omega$. This is the resistance of about $10^{10}km$ copper wire.

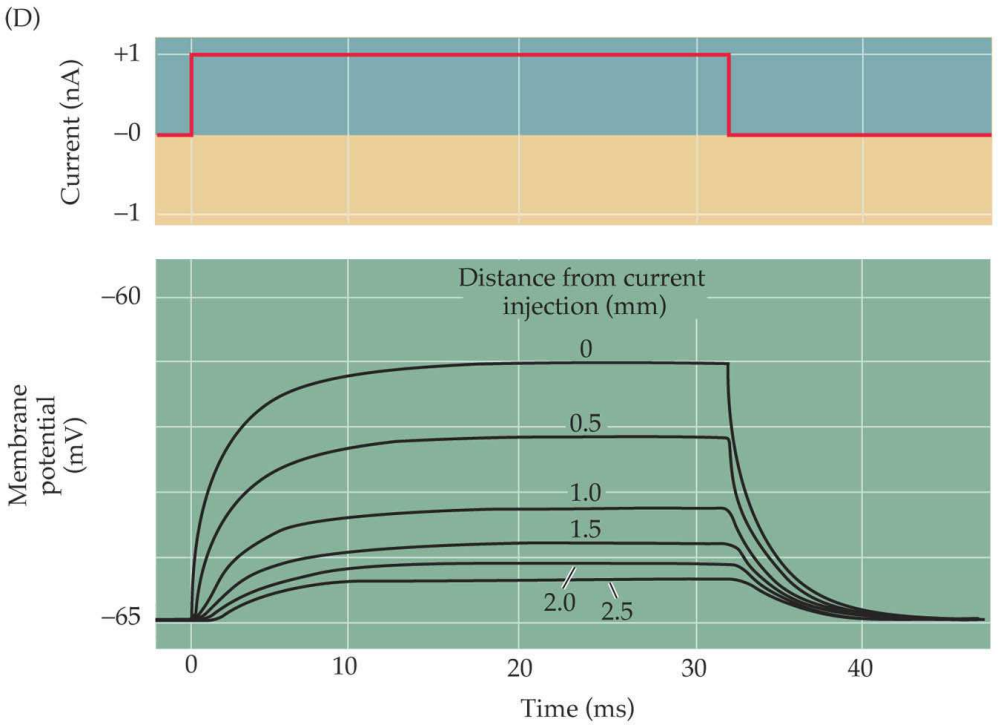
Therefore, one needs repeated amplification along the axon. Binary nature of spike facilitates this.

Passive versus active flow

Membrane potential in passive axon attenuates with distance.

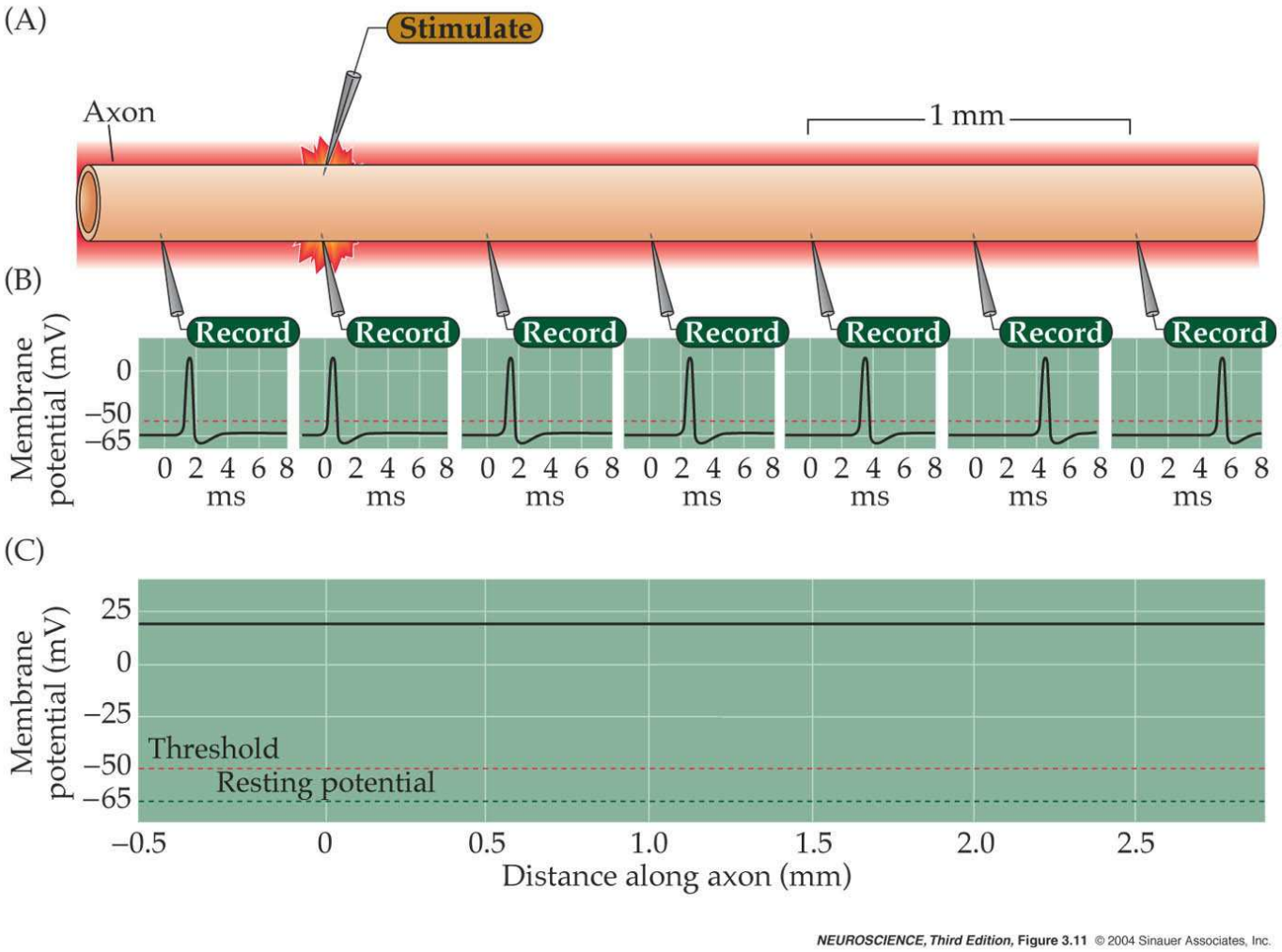


NEUROSCIENCE, Third Edition, Figure 3.10 (Part 1) © 2004 Sinauer Associates, Inc.



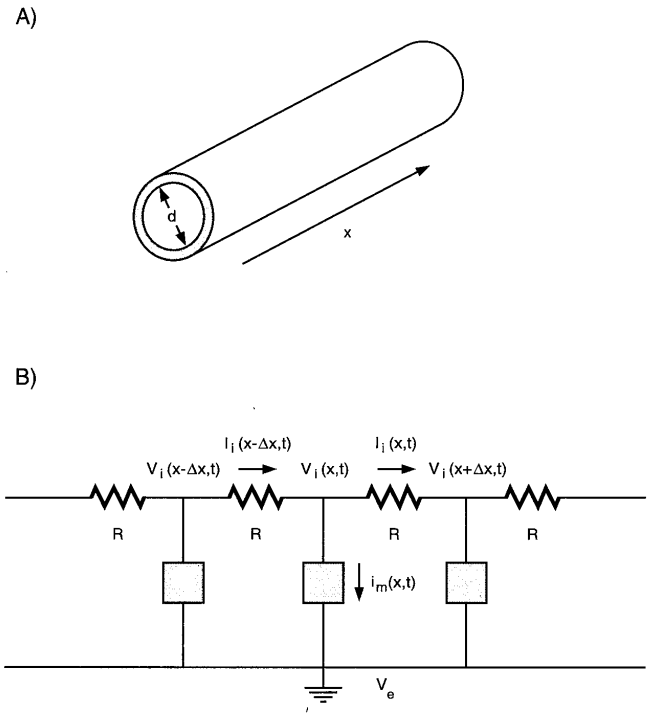
NEUROSCIENCE, Third Edition, **Figure 3.10 (Part 2)**

Membrane potential in active axon does not attenuate with distance.



NEUROSCIENCE, Third Edition, Figure 3.11 © 2004 Sinauer Associates, Inc.

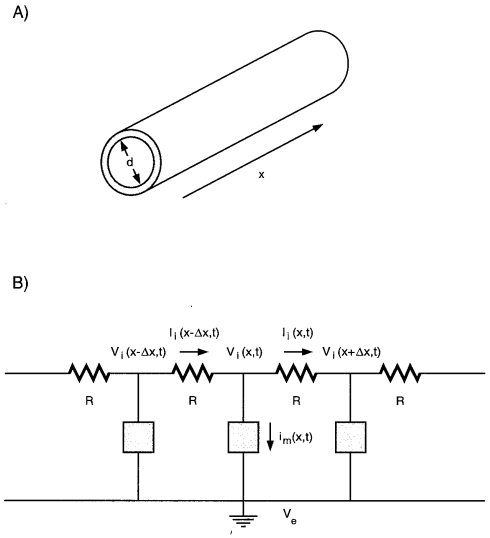
Linear cable theory



Assumptions: Currents result from ion movement, which is due to the electric field as well as due to diffusion. One can safely ignore the contribution due to diffusion, ie. Ohms law is valid.

We assume a constant external potential independent of space and time.

Linear cable theory



$$V_m(x,t) - V_m(x+\Delta x,t) = RI_i(x,t) \qquad R = \frac{R_i \Delta x}{A} = r_a \Delta x$$

$$\frac{\partial V_m}{\partial x}(x,t) = -r_a I_i(x,t)$$

with R the axial resistance of a cylinder of length Δx, with R_i the intracellular resistivity.

Linear cable theory

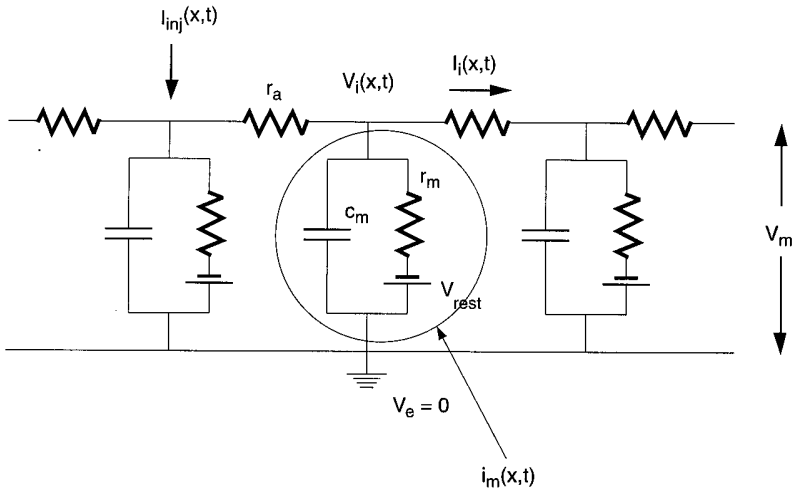
Kirchhoff's law:

$$i_m(x,t)\Delta x + I_i(x,t) - I_i(x-\Delta x,t) = 0 \qquad i_m(x,t) = -\frac{\partial I_i}{\partial x}(x,t)$$

Combining these results yields

$$\frac{1}{r_a}\frac{\partial^2 V_m}{\partial x^2}(x,t) = i_m(x,t)$$

Linear cable theory



The membrane current through length Δx is

$$
\begin{aligned}
i_m(x,t)\Delta x &= \frac{V_m(x,t) - V_{\text{rest}}}{R} + C\frac{\partial V(x,t)}{\partial t} - I_{\text{inj}}(x,t) \\
R &= \frac{\rho_m d_m}{A_m} = \frac{R_m}{\pi d \Delta x} = \frac{r_m}{\Delta x} \\
C &= \frac{\epsilon\epsilon_0 A_m}{d_m} = \frac{\epsilon\epsilon_0 \pi d \Delta x}{d_m} = c_m \Delta x
\end{aligned}
$$

Linear cable theory

Thus,

$$\begin{aligned} i_m(x,t) &= \frac{V(x,t)}{r_m} + c_m \frac{\partial V(x,t)}{\partial t} - i_{\text{inj}}(x,t) \\ \lambda^2 \frac{\partial^2 V}{\partial x^2} &= \tau_m \frac{\partial V}{\partial t} + V - r_m i_{\text{inj}} \end{aligned}$$

with $\lambda^2 = r_m/r_a$ and $\tau_m = r_m c_m$ the space and time constants and $V = V_m - V_{\text{rest}}$.

Steady-state solution

Suppose, a constant current is injected at $x = 0$. The membrane potential reaches the steady-state solution satisfying

$$\begin{aligned} \lambda^2 \frac{d^2 V(x)}{dx^2} &= V(x) - r_m i_{\text{inj}}(x) \\ i_{\text{inj}}(x) &= I_0 \delta(x) \end{aligned}$$

The solution is given by

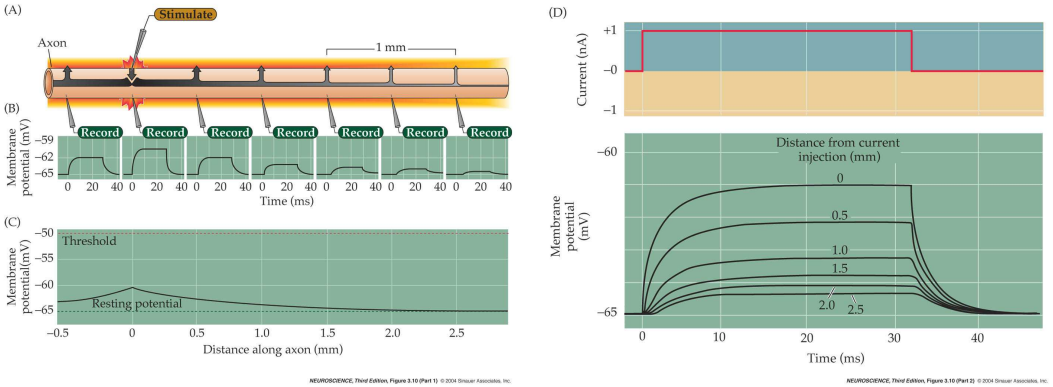
$$V(x) = V_0 \exp(-|x|/\lambda)$$

We can compute V_0 by observing that

$$\frac{dV}{dx}_{x=0} = -\frac{V_0}{\lambda} = -r_a I_i(x=0) = \frac{r_a I_0}{2} \Rightarrow V_0 = \frac{r_a I_0 \lambda}{2} = \frac{\sqrt{r_a r_m}}{2} I_0$$

Space constant

The space constant, $\lambda = \sqrt{r_m/r_a}$ controls the decay of the voltage away from the site of current injection.



Large r_m reduces membrane current and thus yields larger λ.

Since

$$\lambda^2 = \frac{r_m}{r_a} = \frac{R_m}{\pi d}\frac{\frac{1}{4}\pi d^2}{R_i} = \frac{R_m}{R_i}\frac{d}{4},$$

$\lambda \propto \sqrt{d}$. Thick axons spread current over larger distance.

Typical values: $R_i = 200\Omega$cm, $R_m = 20000\Omega\text{cm}^2, d = 4\mu$m, $\lambda = 1$mm.

Propagation velocity

The capacitance of the ring is $c_m \Delta x$.

Approximating the membrane as two parallel plates, the capacitance is proportional to the area of the ring:

$$c_m \propto \frac{A}{d_m} \propto d.$$

Thus, the characteristic time scale τ_m depends on the diameter as

$$\tau_m = r_m c_m \propto d^{-1} d = \text{constant}$$

The propagation velocity depends on the diameter of the axon as

$$v \propto \sqrt{d}$$

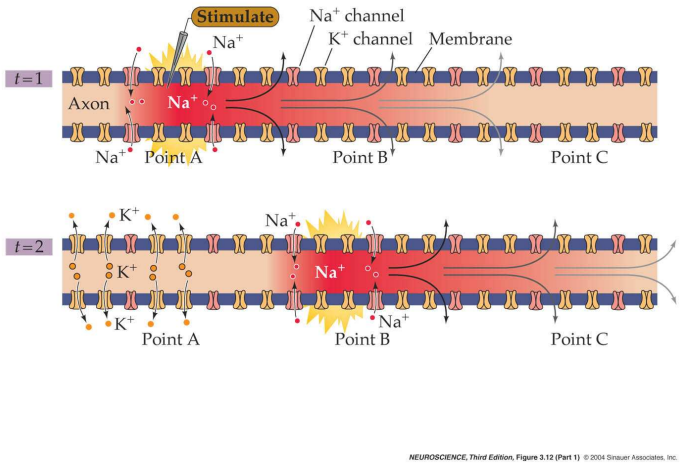
This requires very thick axons for fast propagation (eg. squid giant axon).

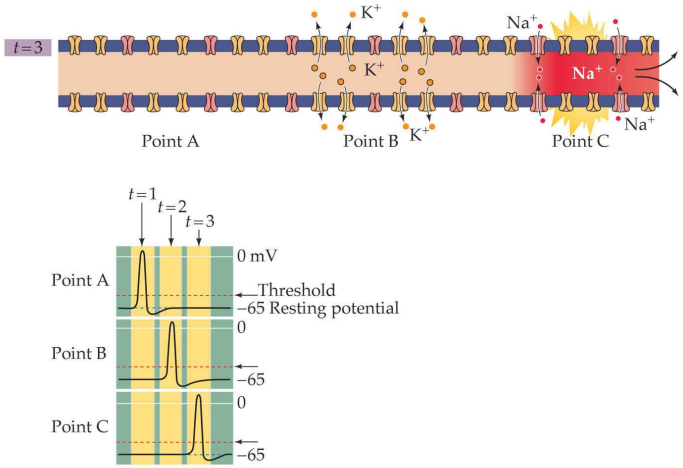
Propagation velocity

Examples of unmyelinated axons:

	$d(\mu)$	v (m/sec)	$v/\sqrt{d}$
Cat	0.3-1.3	0.7-2.3	1.3-2.0
Crab	30	5	0.91
Squid giant axon	500	25	1.1

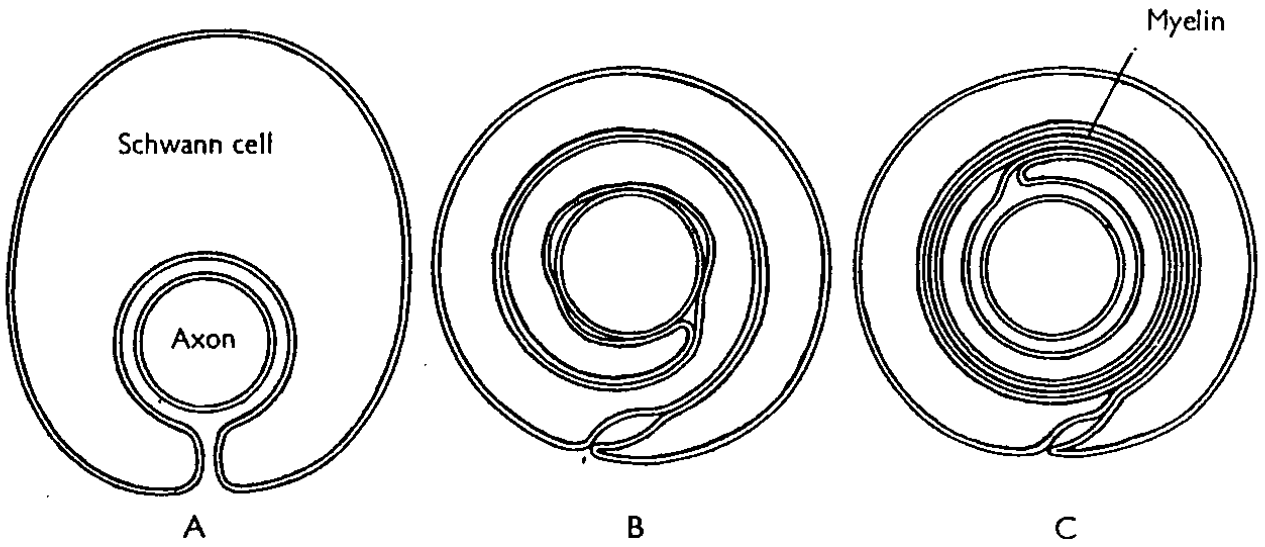
This requires very thick axons for fast propagation (eg. squid giant axon).





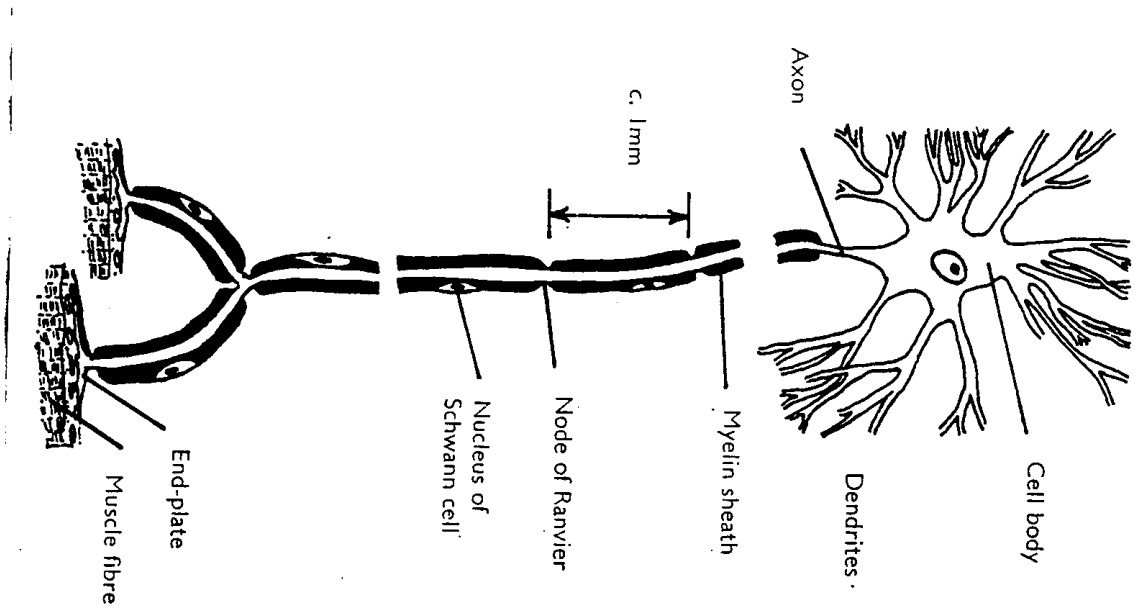
Myelin

Another way to increase the propagation velocity of spikes is by insulating the axon with myelin.



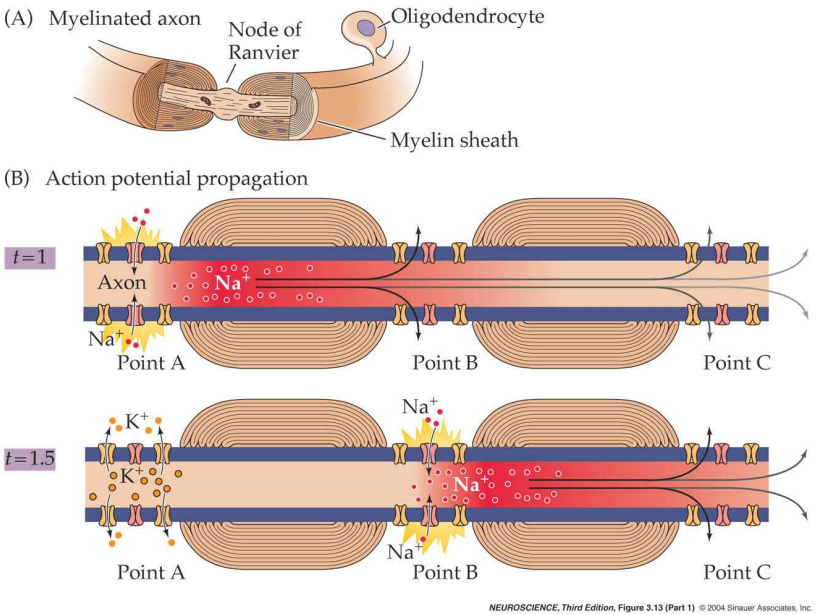
Myelin

The sheath is interrupted at intervals of about 1 mm by short gaps called nodes of Ranvier.

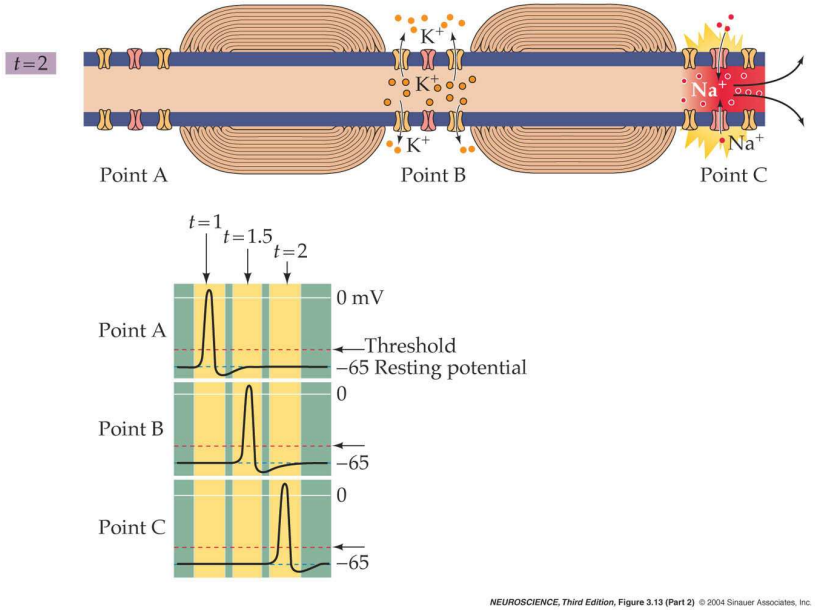


Axon length: ranges from less than a mm to several meters.

Propagation in myelinated axons



NEUROSCIENCE, Third Edition, Figure 3.13 (Part 1) © 2004 Sinauer Associates, Inc.



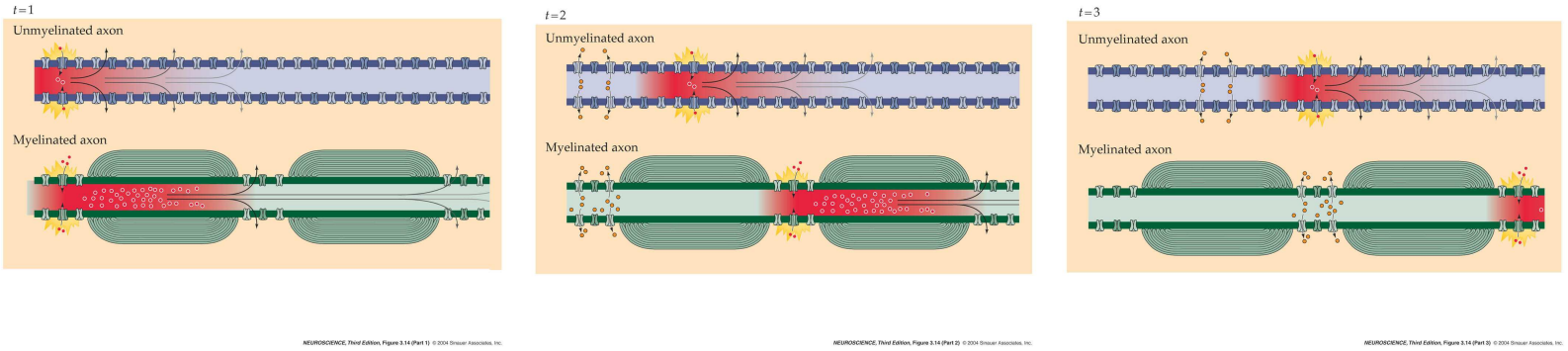
NEUROSCIENCE, Third Edition, Figure 3.13 (Part 2) © 2004 Sinauer Associates, Inc.

The effect of 250 layers of myelin is to increase the membrane resistance and at the same time decrease the capacitance. The membrane time constant is unaffected and the space constant is increased.

Propagation in myelinated axons

Thus, myelin increases the velocity of passive spread from 0.5-10 m/sec up to 150 m/sec.

Action potentials are only generated at the nodes of Ranvier. This process is called saltatory conduction.



Summary

Membrane permeability is sensitive function of membrane potential and time and is ion specific. Permeabilities increase (K^+), or increase and decrease (Na^+), as the membrane potential depolarizes.

For most types of axons, permeability changes consist of a rapid and transient rise in the sodium permeability, followed by a slower but more sustained rise in the potassium permeability. The HH model a complete explanation of action potential generation.

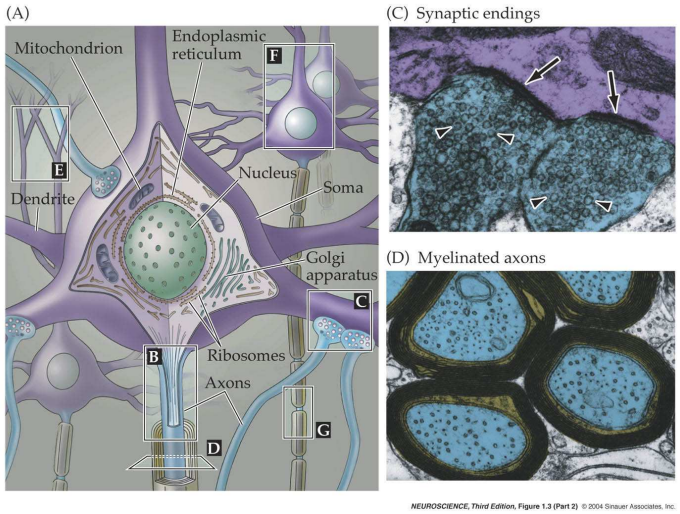
Spike propagation results from passive spread, combined with active action potential generation.

- The velocity of passive spread is $\propto \sqrt{d}$ and thus requires thick axons for fast propagation.

- By covering the axon with myelin the velocity of passive spread can be increased more effectively.

Synapses

Nerve cells interact through synapses. They come in two flavors:
electrical synapses, also called gap junctions;
and chemical synapses.



Terminal boutons (blue) loaded with synaptic vesicles (arrowheads) forming synapses (arrows) with a dendrite (purple).

Chemical synapses come in a large variety and have a complex internal dynamics.

Some facts

Synapses typically connect the axon of one cell to the dendrite of another cell.

Dendro-dendritic synapses occur, but are rare.

Autapses, an axon making connection onto its own dendritic tree, are rare on pyramidal cells, but occur more frequently on some classes of cortical inhibitory inter-neurons.

Synapses are small: about 0.5-1.0 μm in diameter.

If their size is 1 μm, one mm^3 full of synapses would contain 10^9 synapses. In fact, the experimental estimate is very close to this: 8×10^8 synapses/mm^3 in mouse cortex.

In addition, one mm^3 of brain tissue contains 100000 neurons, 4.1 **km** of axon ($d \approx 0.3\mu$) and 456 m of dendrite ($d \approx 0.9\mu$).

More facts

Thus, the average neuron in the mouse cortex is connected to 8000 other neurons and uses 4 mm of dendrite wire and 4 cm of axon wire to connect to other cells.

Since the total cortical surface in humans is about 100000 mm^2 and 2 mm thick, there are about 2×10^{10} neurons and 2×10^{14} synapses in the human brain.

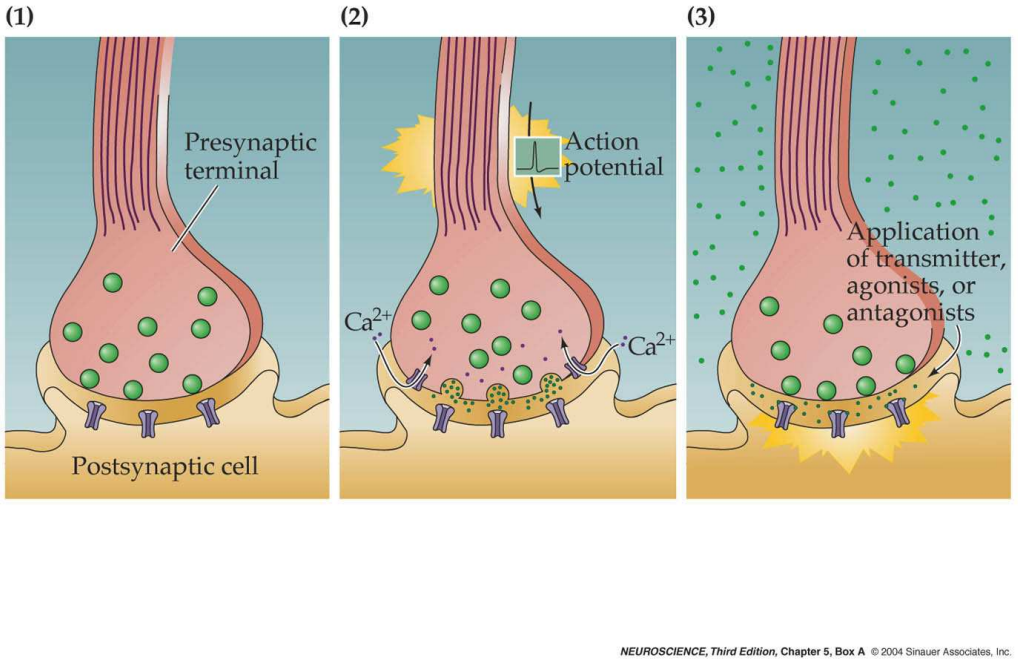
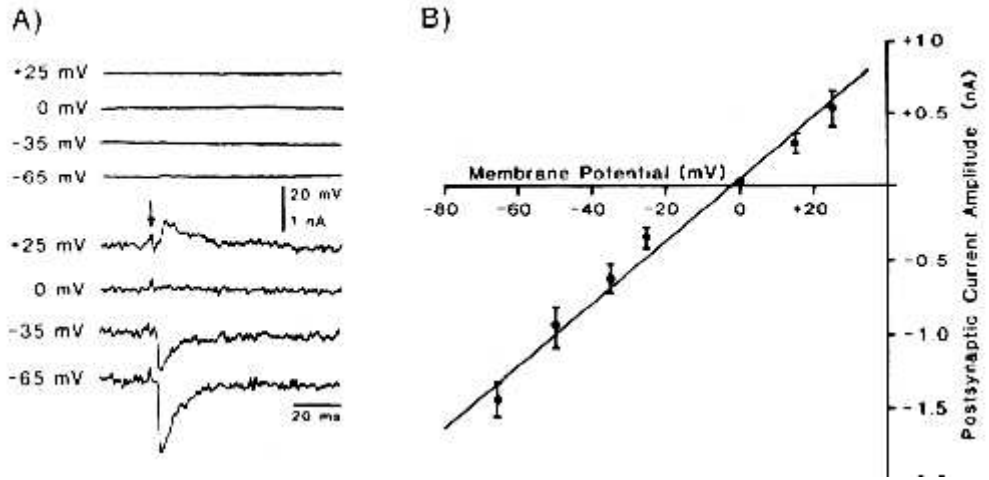


Figure 16: Canonical chemical synapse model. Action potential causes an inrush of Ca^{2+} ions via voltage dependent Ca channels. 2) Elevated Ca^{2+} concentration allows one or more vesicles to fuse with the presynaptic neuron membrane, releasing its neurotransmitter. 3) The neurotransmitter binds to postsynaptic receptors, increasing the permeability of post-synaptic ion channels. An in- or out-rush of current temporarily changes the post-synaptic potential (PSP).

Upon activation of a fast chemical synapse one can observe a rapid and transient change in the postsynaptic potential. The response can be either excitatory (EPSP) or inhibitory (IPSP). These EPSPs and IPSPs are caused by excitatory and inhibitory post-synaptic currents (EPSCs and IPSCs).

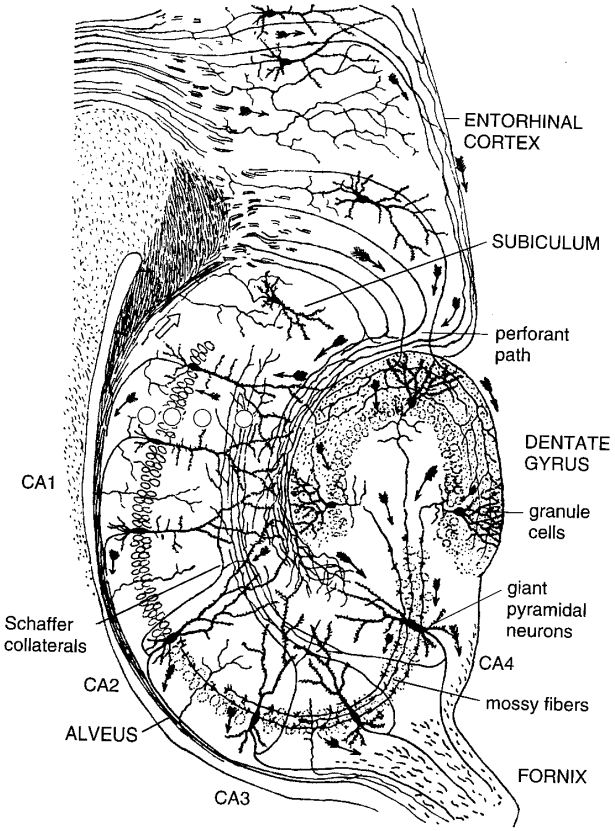


Activation of synapses made by the mossy fibers onto CA3 pyramidal cells in the rodent hippocampus. A) The pyramidal cell is voltage clamped to different values and the clamp current is recorded.

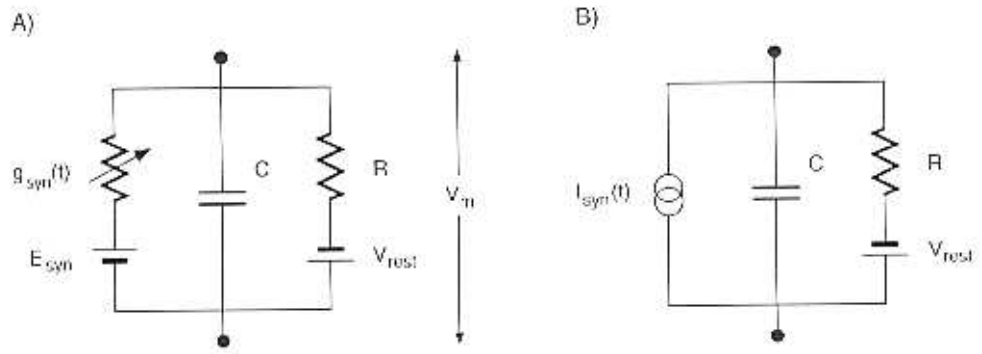
The synaptic current rises fast and decays to zero in 20-30 msec. The peak current is linearly related to the membrane potential. This suggest that:

$$I_{\text{syn}}(t) = g_{\text{syn}}(t)(V_m - V_{\text{syn}})$$

The post-synaptic current is caused by a temporary increase in the membrane conductance, modeled by $g_{\text{syn}}(t)$.



Hippocampal formation is involved in transfer from short- to long-term memory. Granule cells in the dentate gyrus send their output axons, so-called *mossy fibers*, to pyramidal cells in the CA3 region. These pyramidal cells project, with so-called *Schaffer collaterals* onto pyramidal cells in the CA1 region.

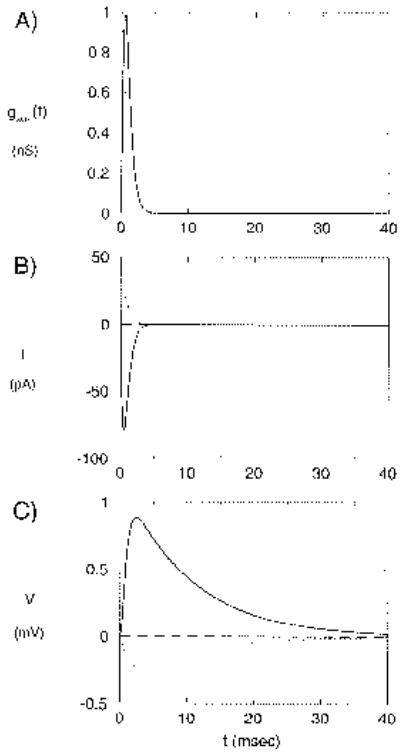


Inserting this synapse in a patch of membrane:

$$
\begin{aligned}
C\frac{dV_m}{dt} &+ g_{\text{syn}}(t)(V_m - V_{\text{syn}}) + \frac{V_m - V_{\text{rest}}}{R} = 0 \\
\tau\frac{dV}{dt} &= -\left(1 + Rg_{\text{syn}}\right)V - Rg_{\text{syn}}\left(V_{\text{rest}} - V_{\text{syn}}\right)
\end{aligned}
$$

with $V = V_m - V_{\text{rest}}$ and $\tau = RC$.

When $V_{\text{syn}} > V_{\text{rest}}$ the current will depolarize the membrane. An example is the excitatory synapse using the neurotransmitter glutamate with $V_{\text{syn}} - V_{\text{rest}}$=80mV.



A) $g_{\text{syn}} = g_{\text{peak}} \frac{t}{t_{\text{peak}}} e^{1-t/t_{\text{peak}}}$, t_{peak} =0.5 msec, $g_{\text{peak}} = 10^{-9}\Omega^{-1}$. B) Synaptic current for two different values of the synaptic reversal potential $V_{\text{syn}} - V_{\text{rest}}$ =80 mV (solid line), -20 mV(dotted line). C) Associated EPSP (solid line) and IPSP(dotted line). $R = 100M\Omega$, τ =10msec.

Reversal potentials

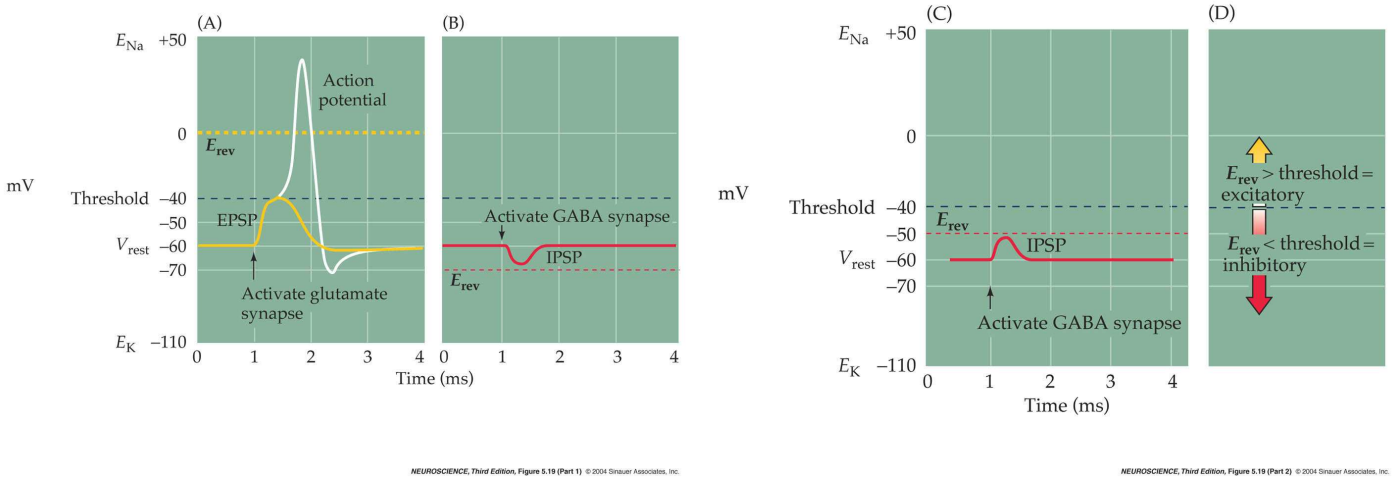
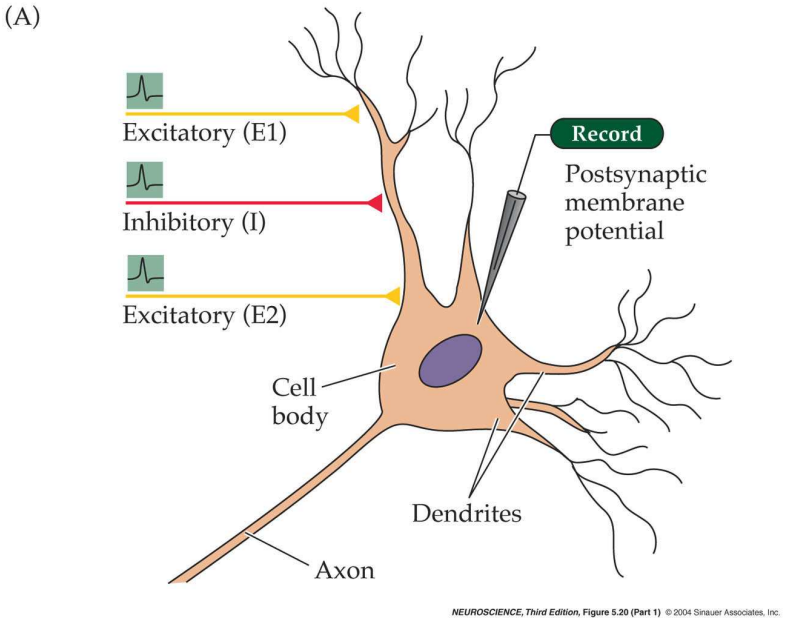
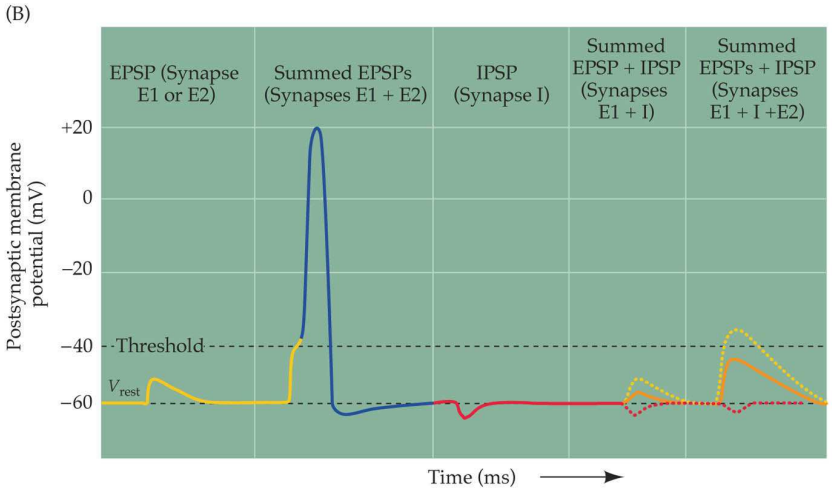


Figure 17: Reversal potentials and threshold potentials determine postsynaptic excitation and inhibition. A, C) If the reversal potential for a PSP (E_{rev}) is more positive than the action potential threshold (-40 mV), the effect of a transmitter is excitatory, and it generates an EPSP. B) If the reversal potential for a PSP is more negative than the action potential threshold, the transmitter is inhibitory and generates IPSPs.

Summation of postsynaptic potentials

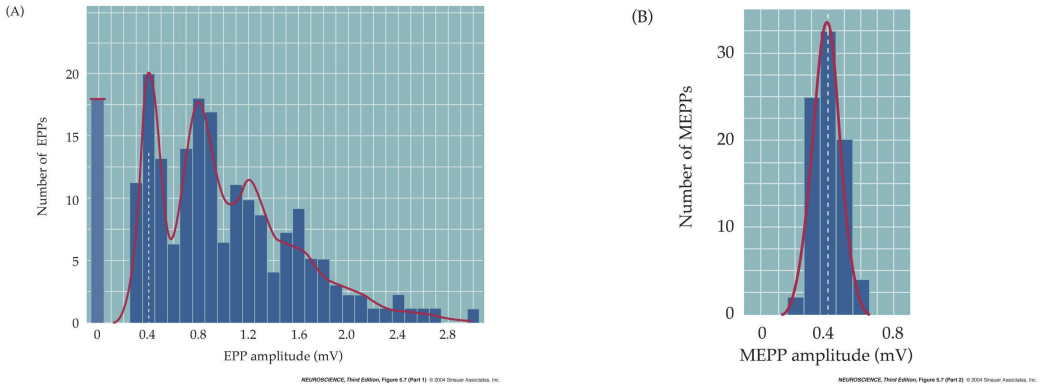


NEUROSCIENCE, Third Edition, Figure 5.20 (Part 1) © 2004 Sinauer Associates, Inc.



NEUROSCIENCE, Third Edition, Figure 5.20 (Part 2) © 2004 Sinauer Associates, Inc.

Probabilistic response

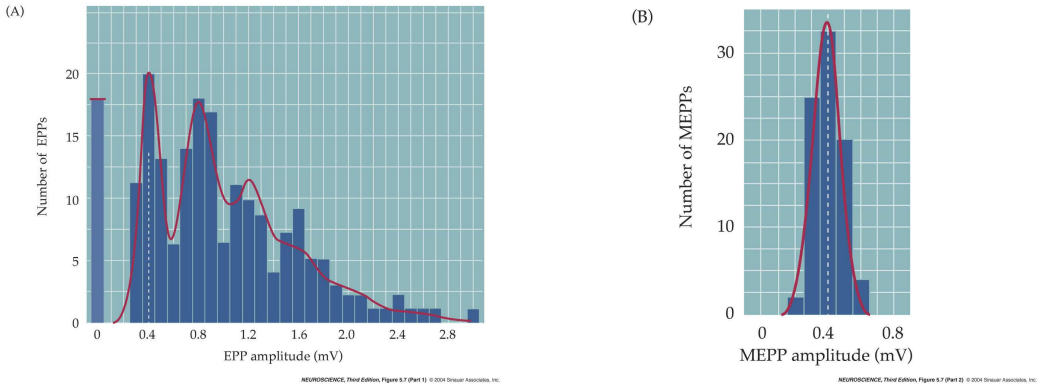


Chemical synapses were first studied at the neuromuscular junction.

Stimulation of presynaptic motoneurons leads under normal condition to post synaptic action potentials and contraction of the muscle.

In the absence of stimulation, miniature EPSPs are observed as a result of spontaneous neurotransmitter release, which are alway more or less of the same size (inset). This suggest quantal (=constant) release of neurotransmitter with EPSP of ≈ 0.4 mV.

Probabilistic response



Katz studied the neuromuscular junction at low external Ca^{2+} concentration. Stimulation of presynaptic motoneurons leads to subthreshold post-synaptic response in multiples of 0.4 mV.

Katz model of quantal release

Assume the junction has n release sites, each with an independent probability p to release a vesicle after pre-synaptic stimulation. p depends of course on the Ca^{2+} concentration.

The probability that the synapse releases k quanta

$$\begin{aligned} p_n(k) &= \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \\ m &= \langle k \rangle = np \\ \sigma^2 &= \left\langle k^2 \right\rangle - m^2 = np(1-p) \end{aligned}$$

In the limit, $p \to 0, n \to \infty$ with $m = pn$ constant, the binomial distribution can be approximated by the Poisson distribution

$$p_n(k) \to p_m(k) = \frac{m^k}{k!} \exp(-m)$$

With $m = 2.33$, the expected and observed results for each k agree very well.

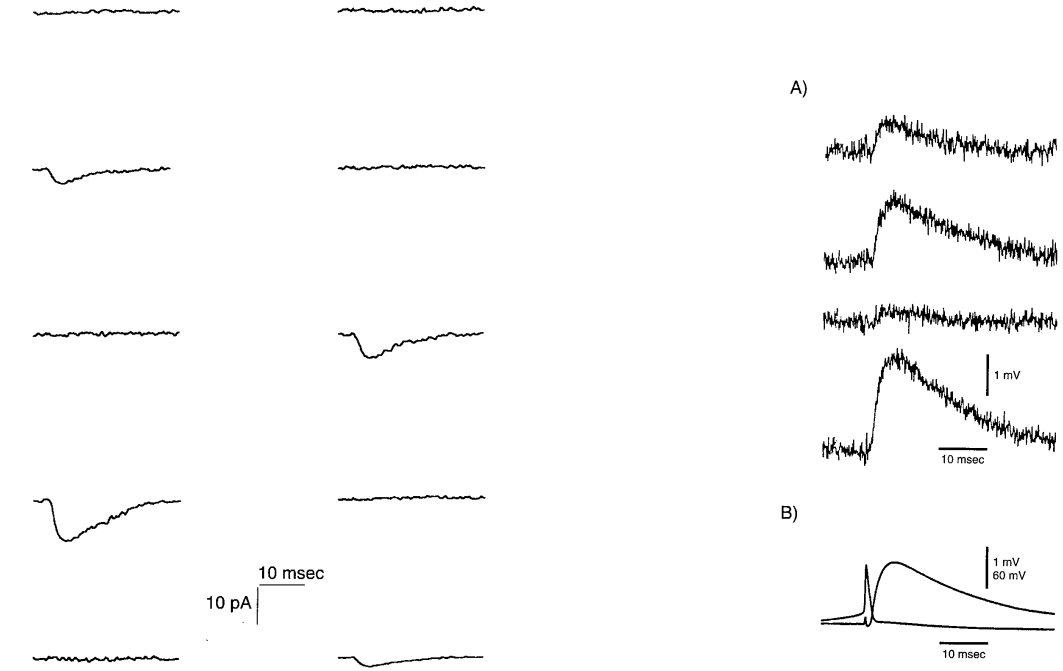
k	$np(k)$	Observed
0	19	18
1	44	44
2	52	55
3	40	36
4	24	25
5	11	12
6	5	5
7	2	2
8	1	1
9	0	0

Table 1: Numerical comparison between observed quantal response of synapse in neuro-muscular junction and prediction from binomial distribution. $n = 198, m = 2.33$.

$m = 2.33 \ll n$ due to low external Ca^{2+}. In normal operation $m = \mathcal{O}(n)$.

Central synapses are unreliable

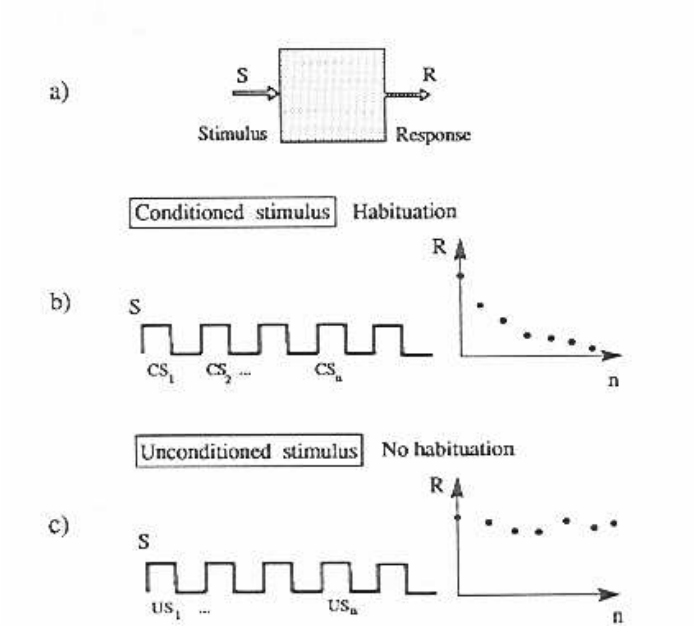
In cortex mostly mono-synaptic connections and $0.1 < p < 0.9$.



Left. EPSC in CA 1 pyramidal cells. Only 3 out of 9 presynaptic stimuli produce a response. In addition, the response is variable in strength. Right. Rat visual cortex. 4 EPSPs resulting from identical stimulation.

Habituation

Classical experiments on learning involve Pavlov's dog.



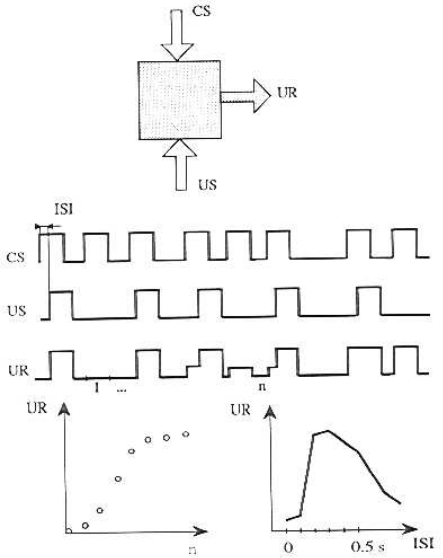
Habituation: dog turns head (CR) on bell (CS). After repeated presentation, the dog no longer responds.

No habituation: dog salivates (UR) on sight of meat (US). No matter how often the stimulus is presented, the dog will always respond.

Classical conditioning

Consider two stimuli CS (bell) and US (meat). By itself CS does not yield a response, but US does.

Classical conditioning is the phenomenon that when CS and US are presented together for some time, the dog will start responding to CS alone.



Classical conditioning

$$R = \Theta(J^u S^u + J^c S^c - \theta)$$

Before: $J^c < \theta$ and $J^u > \theta$

After: $J^c > \theta$ and $J^u > \theta$ through Hebbian learning

$$\Delta J \propto RS$$

Long term potentiation

Synaptic plasticity is the basis for learning, memory and permanent changes in behavior.

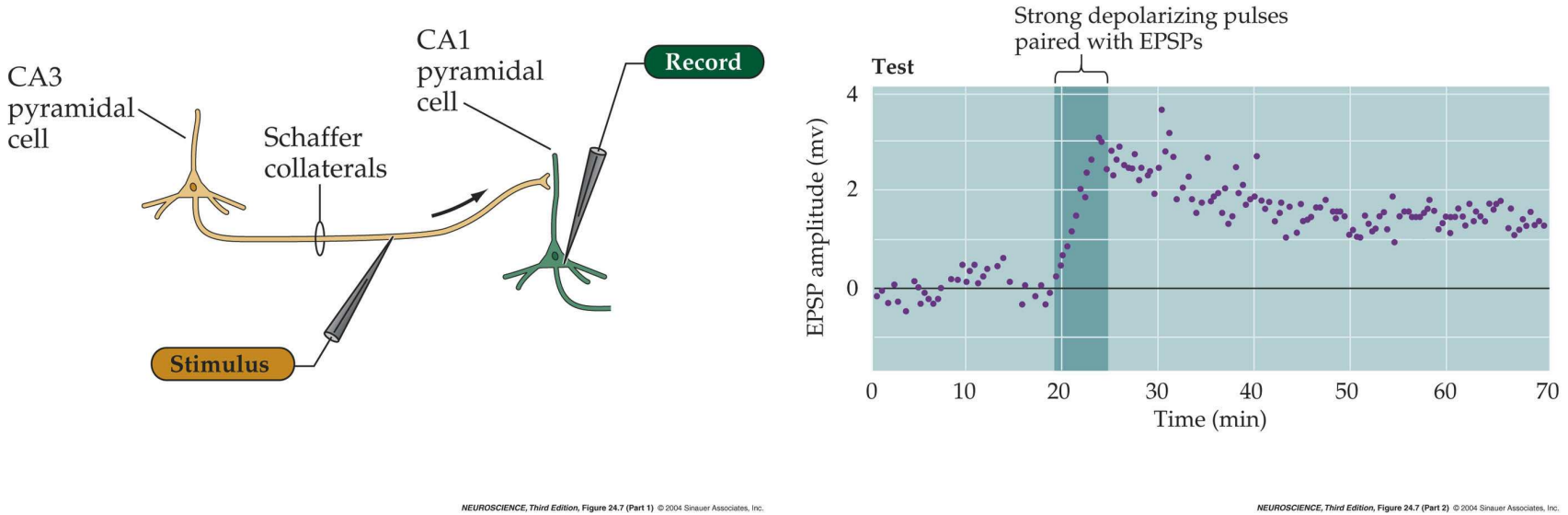


Figure 18: LTP in hippocampus: Single stimuli applied to a Schaffer collateral evokes EPSPs in the post-synaptic CA1 neuron, but no change in synaptic strength. When the CAI neuron's membrane potential is depolarized in conjunction with the Schaffer collateral stimuli, there is a persistent increase in the EPSPs, which can last for hours or days.

LTP as a cellular analog of classical conditioning

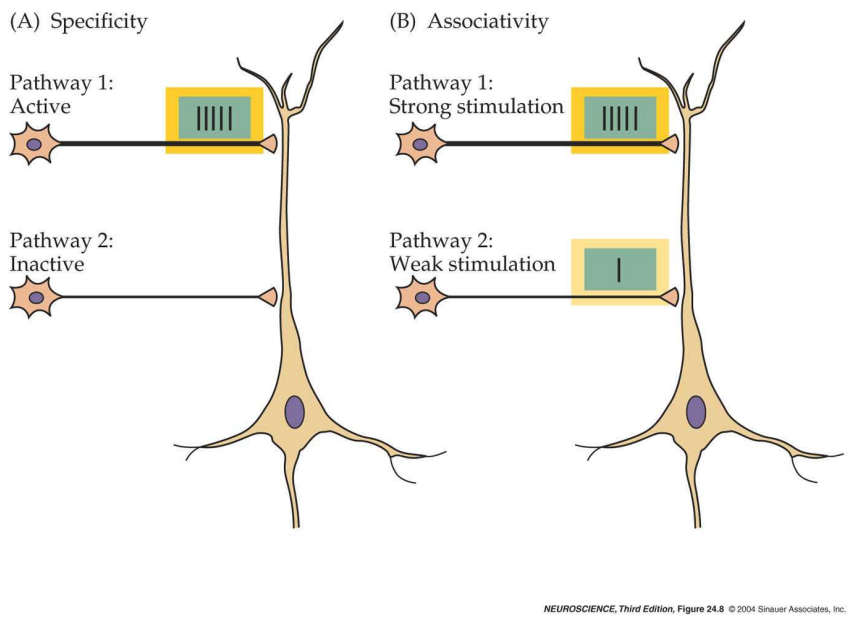


Figure 19: A) Strong activity initiates LTP at active synapses without initiating LTP at nearby inactive synapses. B) Weak stimulation of pathway 2 alone does not trigger LTP. However, when the same weak stimulus to pathway 2 is activated together with strong stimulation of pathway 1, both sets of synapses are strengthened.

Hebbian learning

The simplest plasticity rule that follows the spirit of Hebb's conjecture takes the form

$$\tau_w \frac{dw}{dt} = vu - \lambda w$$

In the absence of neural activity ($u = v = 0$) the weight decays to zero.

Assume that u and v are randomly drawn from a probability distribution $p(u, v)$. Average synaptic weight satisfies

$$\tau_w \frac{dw}{dt} = \langle vu \rangle - \lambda w$$

Ocular dominance

A single layer 4 cortical neuron that receives input from just two LGN neurons with activity $u_i, i = 1, 2$.

Two synaptic weights $w_i, i = 1, 2$ describe the synaptic connection strengths of the LGN neurons with the cortical neuron.

The output activity we assume simply linear:

$$v = \sum_{i=1}^{2} w_i u_i$$

Ocular dominance

$$\begin{aligned} \tau_w \frac{dw_i}{dt} &= \sum_j Q_{ij} w_j - \lambda w_i \\ Q_{ij} &= \langle u_i u_j \rangle \end{aligned}$$

$Q_{11} = Q_{22} = q_s$ and $Q_{12} = Q_{21} = q_d$.

$$\begin{aligned} \tau_w \frac{d(w_1 + w_2)}{dt} &= (q_s + q_d - \lambda)(w_1 + w_2) \\ \tau_w \frac{d(w_1 - w_2)}{dt} &= (q_s - q_d)(w_1 - w_2) \end{aligned}$$

For λ sufficiently large, the first equation will yield the asymptotic solution $w_1 + w_2 = 0$.

Ocular dominance

Since $q_s - q_d > 0$, $w_1 - w_2$ will grow indefinitely. Non-linearities in the system will prevent this indefinite growth.

The final solution is then

$$w_1 = -w_2 = w_\infty$$

with w_∞ a positive or negative value depending on the sign of the initial value $w_1(0) - w_2(0)$.

For $w_\infty > 0$, the cortical neuron will be sensitive to eye 1 and insensitive to eye 2, and vise versa.

Summary

There are chemical and electrical synapses. Chemical synapses are thought to be implied in learning.

Synapses can be excitatory, inhibitory or shunting depending on the reversal potential of the synapse relative to the membrane resting potential.

Synapses are stochastic elements: a presynaptic action potential yields a postsynaptic response with a certain probability.

The most important mechanism for learning is called Hebbian learning. The strength of a synapse increases when pre- and postsynaptic cell fire at the same time. This is in agreement with the psychological phenomenon of classical conditioning and also found as a mechanism for synaptic plasticity in the brain.

Hebbian learning can be used to explain the receptive field properties of many neurons, such as for instance ocular dominance.

Neural networks: Perceptrons

- single layer perceptrons
  - binary
  - linear
- multi layer perceptrons

Perceptrons

Perceptrons are feed-forward neural networks. Examples are given in Fig. 20.

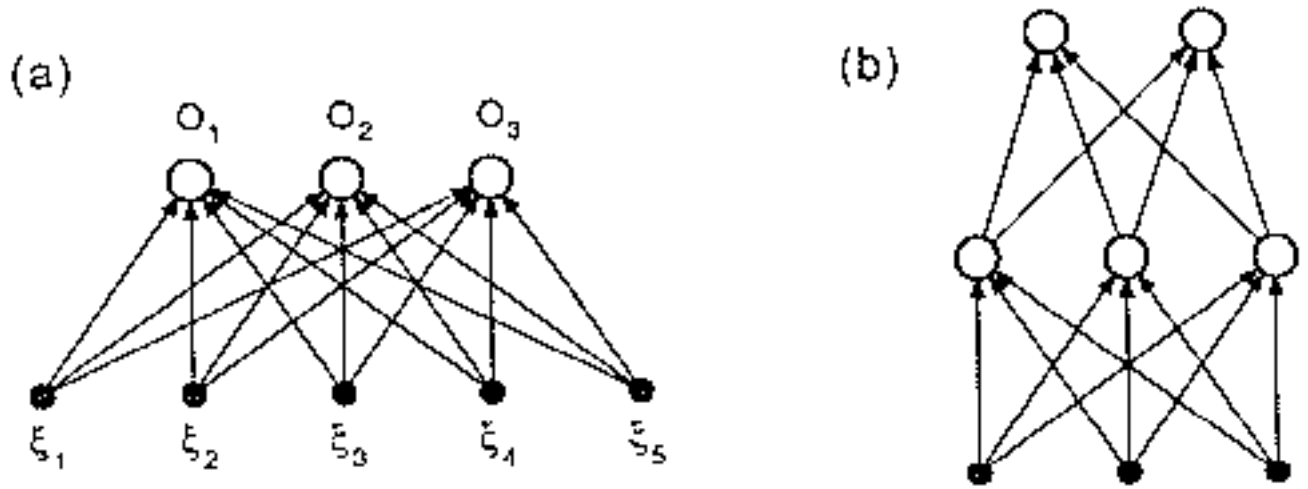


Figure 20: A) Simple Perceptron B) Multi-layered Perceptron

Consider a simple perceptron with one output:

$$o = g(h) = g\left(\sum_{j=1}^{n} w_j \xi_j - \theta\right) = g\left(\sum_{j=0}^{n} w_j \xi_j\right)$$

with weights w_j and inputs ξ_j. $\xi_0 = -1$ and $\theta = w_0$. g is a non-linear function.

Learning

Given a number of input-output pairs $(\xi_j^\mu, \zeta^\mu), \mu = 1, \ldots, P$, find w_j such that the perceptron output o for each input pattern ξ^μ is equal to the desired output ζ^μ:

$$o^\mu = g\left(\sum_{j=0}^{n} w_j \xi_j^\mu\right) = \zeta^\mu, \quad \mu = 1, \ldots, P$$

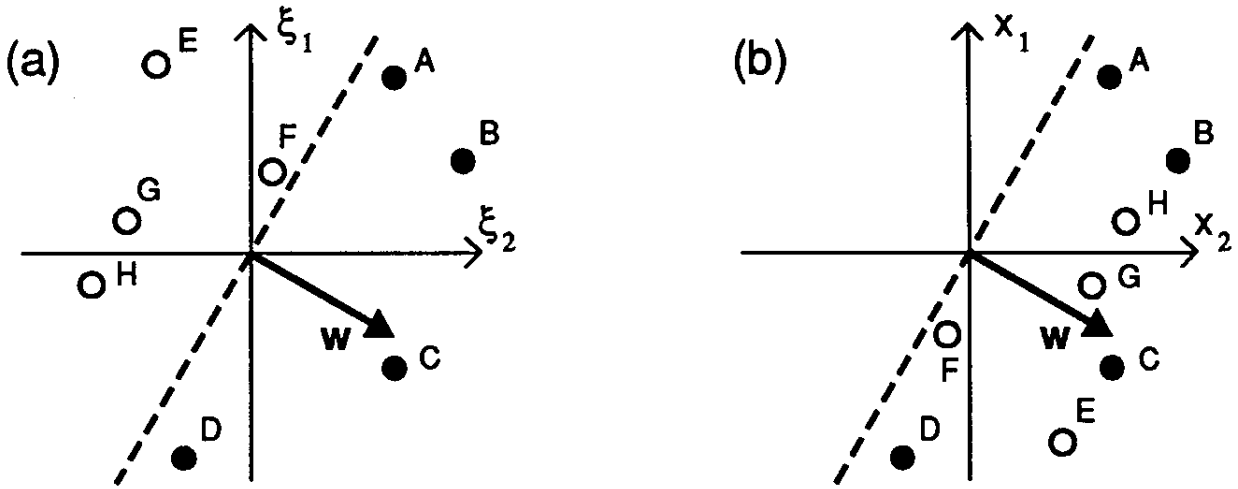
Threshold units

Consider the simplest case of binary threshold neurons:

$$g(h) = \text{sign}(h) \qquad \zeta = \pm 1$$

Then, the learning condition becomes

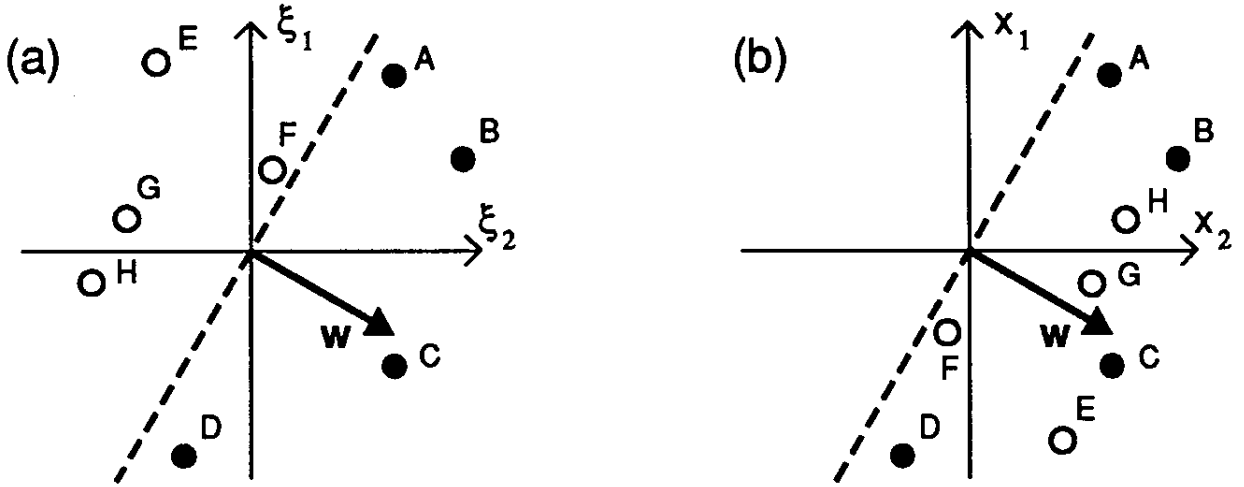
$$\text{sign}(w \cdot \xi^\mu) = \zeta^\mu, \quad \mu = 1, \ldots, P$$



Since $\zeta^\mu = \pm 1$, we have

$$\text{sign}(w \cdot \xi^\mu \zeta^\mu) = 1 \quad \text{or} \quad w \cdot x^\mu > 0$$

with $x_j^\mu = \xi_j^\mu \zeta^\mu$.

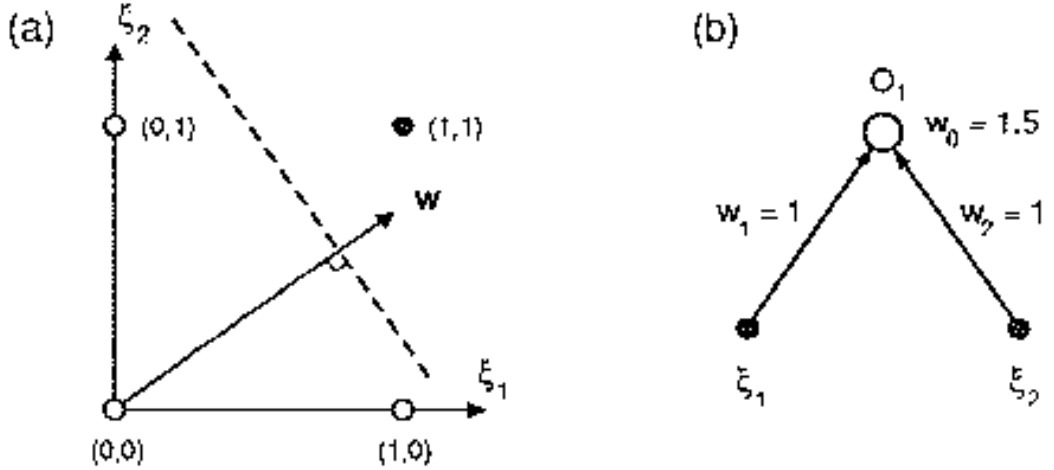


Linear separation

Classification depends on sign of $w \cdot \xi$. Thus, decision boundary is hyper plane:

$$0 = w \cdot \xi = \sum_{j=1}^{n} w_j \xi_j - \theta$$

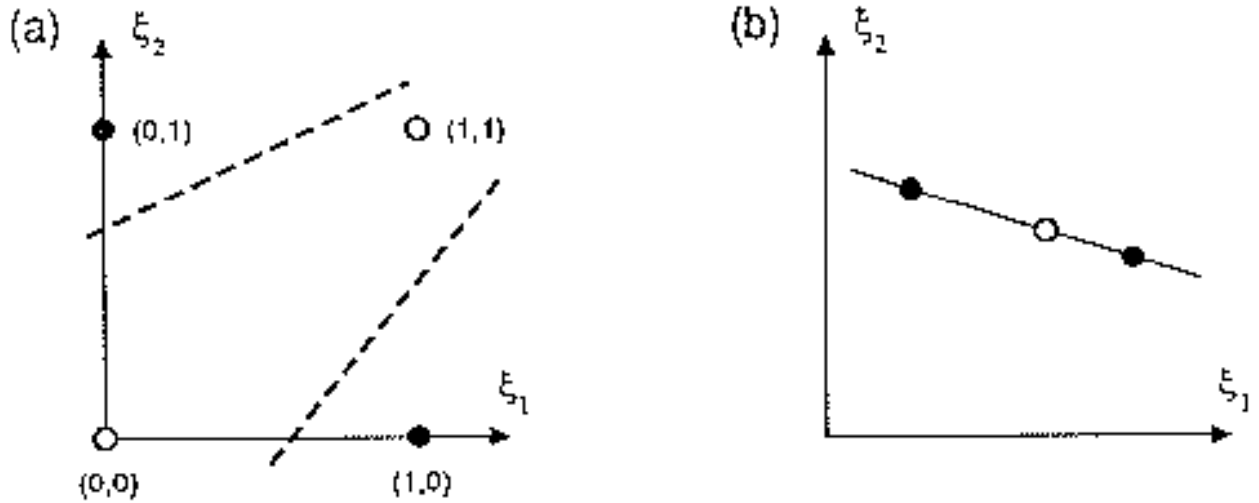
Perceptron can solve linearly separable problems. An example of a linearly separable problem is the AND problem: The output of the perceptron is 1 if all inputs are 1, and -1 otherwise



Linear separation

By definition, problems that are not linearly separable need more than one separating hyper plane to separate the two classes.

Examples are the XOR problem and linearly dependent inputs.



Perceptron learning rule

We have seen that the desired weight vector satisfies

$$w \cdot x^\mu > 0, \quad \text{all patterns}$$

We define the following perceptron learning rule:

$$\begin{aligned} w_j^{\text{new}} &= w_j^{\text{old}} + \Delta w_j \\ \Delta w_j &= \eta \Theta(-w \cdot x^\mu) \xi_j^\mu \zeta^\mu = \eta \Theta(-w \cdot x^\mu) x^\mu \end{aligned}$$

η is the learning rate.

This learning rule is Hebbian.

Perceptron learning rule

Illustration on dataset consisting of three data patterns x^1, x^2 and x^3 and $\eta = 1$.

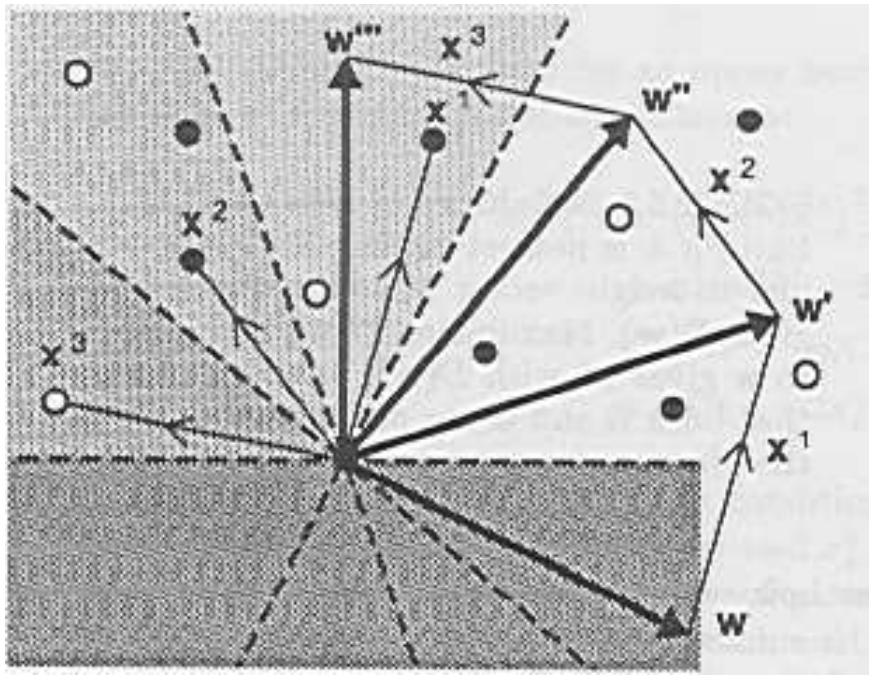


Figure 21: Learning rule is applied to all patterns in some random or given order. Learning stops, when a weight configuration is found that has positive inner product with all training patterns.

Quality of the solution

Depending on the data, there may be many or few solutions to the learning problem, or non at all!

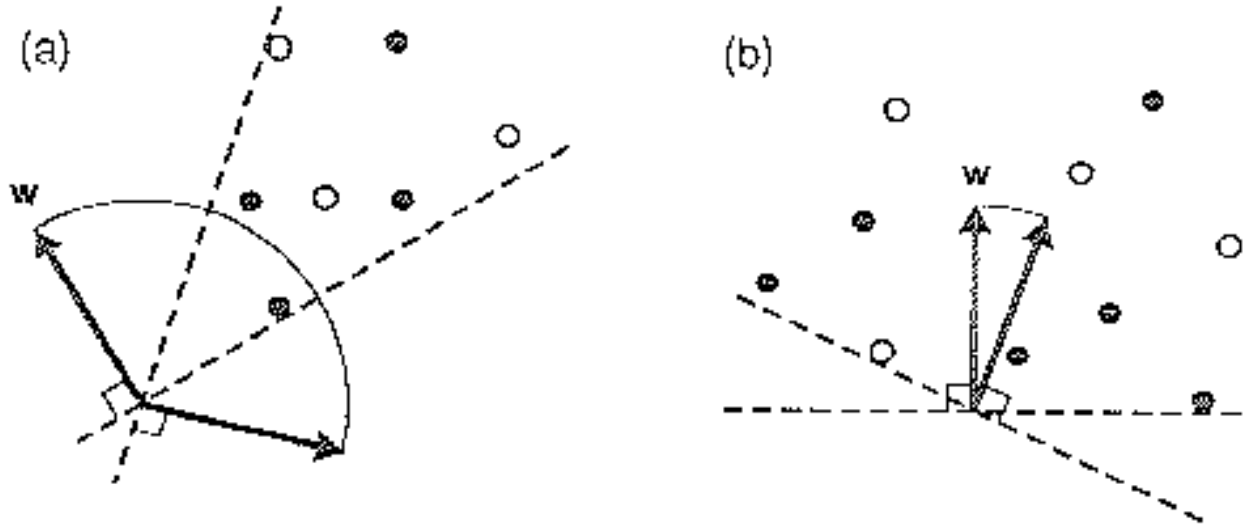


Figure 22: A) Many solutions B) Few solutions.

Quality of the solution

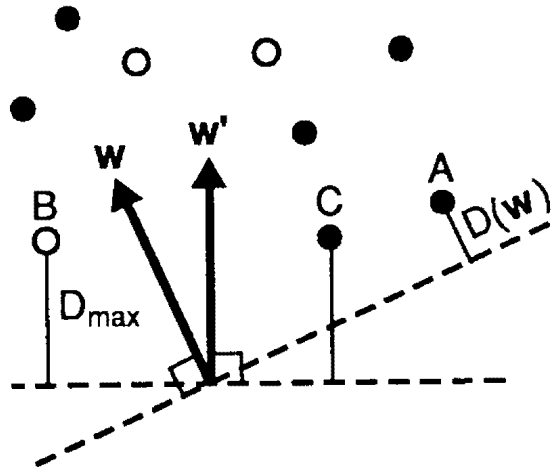


Figure 23: Two admissible solutions w and w' and their values of D respectively. Since $D(w') > D(w)$, w' is the preferred solution.

Quality of the solution

We define the quality of the solution w by the pattern that has the smallest inner product with w.

$$D(w) = \frac{1}{\|w\|} \min_{\mu} w \cdot x^{\mu}$$

The best solution is given by $D_{\max} = \max_w D(w)$.

If we can find a w such that $D(w) > 0$ the problem is linearly separable and learnable by the perceptron learning rule. If the problem is not linearly separable not such solution exists.

Convergence of Perceptron rule

If the problem is linearly separable, the perceptron learning rule converges in a finite number of steps.

We start with initial value $w = 0$. At each iteration, w is updated only if $w \cdot x^\mu < 0$. M^μ denote the number of times pattern μ has been used to update w. Thus,

$$w = \eta \sum_\mu M^\mu x^\mu$$

$M = \sum_\mu M^\mu$ is the total number of iterations in which the weight vector is updated. If the learning rule converges, it means that M is finite and does not grow indefinitely.

Convergence of Perceptron rule

The proof goes as follows. Assume that the problem is linearly separable, so that there is a solution w^* with $D(w^*) > 0$. We will show that

$$\mathcal{O}(\sqrt{M}) \leq \frac{w \cdot w^*}{\|w\|\|w^*\|} \leq 1$$

Thus, M can not grow indefinitely and the perceptron learning rule converges in a finite number of steps.

Convergence of Perceptron rule

The proof of the first inequality is elementary:

$$\begin{aligned} w \cdot w^* &= \eta \sum_\mu M^\mu x^\mu \cdot w^* \geq \eta M \min_\mu x^\mu \cdot w^* = \eta M D(w^*) \|w^*\| \\ \Delta \|w\|^2 &= \|w + \eta x^\mu\|^2 - \|w\|^2 = 2\eta w \cdot x^\mu + \eta^2 \|x^\mu\|^2 \leq \eta^2 \|x^\mu\|^2 = \eta^2 N \end{aligned}$$

$$\|w\|^2 \leq \eta^2 N M$$

Combining these two inequality, we obtain Thus,

$$\frac{w \cdot w^*}{|w||w^*|} \geq \sqrt{M} \frac{D(w^*)}{\sqrt{N}}$$

Note, that the proof makes essential use of the existence of w^* with $D(w^*) > 0$. If $D(w^*) < 0$ the bound Eq. becomes a trivial statement and does not yield a bound on M.

Convergence of Perceptron rule

If the problem is linearly separable, we can in conclude that the number of weight updates:

$$M \leq \frac{N}{D^2(w^*)}$$

where N is some trivial constant. We see that convergence takes longer for harder problems (for which $D(w^*)$ is closer to zero).

Linear units

We now turn to a possibly simpler case of linear units:

$$o^\mu = \sum_j w_j \xi_j^\mu$$

Desired behavior is that the perceptron output equals the desired output for all patterns: $o^\mu = \zeta^\mu, \mu = 1, \ldots, P$. In this case, we can compute an explicit solution for the weights. It is given by

$$w_j = \frac{1}{N} \sum_{\rho\nu} \zeta^\rho \left(Q^{-1}\right)_{\rho\nu} \xi_j^\nu, \quad Q_{\rho\nu} = \frac{1}{N} \sum_j \xi_j^\rho \xi_j^\nu$$

Q is a matrix with dimension $P \times P$ and contains the inner products between the input patterns.

Linear units

To verify that Eq. solves the linear perceptron problem, we simply check for one of the input patterns (ξ^μ) whether it gives the desired output:

$$\begin{aligned}\sum_j w_j \xi_j^\mu &= \frac{1}{N} \sum_{\rho,u,j} \zeta^\rho \left(Q^{-1}\right)_{\rho\nu} \xi_j^u \xi_j^\mu \\ &= \sum_{\rho,u} \zeta^\rho \left(Q^{-1}\right)_{\rho\nu} Q_{\nu\mu} \\ &= \sum_\rho \zeta^\rho \delta_{\rho\mu} = \zeta^\mu\end{aligned}$$

Q must be invertible. Therefore, the input patterns must be linearly independent. Therefore $P \leq N$.

Linear units

When $P < N$ the solution $w_j = \frac{1}{N} \sum_{\rho\nu} \zeta^\rho \left(Q^{-1}\right)_{\rho\nu} \xi_j^u$ is not unique. In fact, there exists a linear space of dimension $N - P$ of solutions w. Namely, let

$$\begin{aligned} w_j^0 &= \frac{1}{N} \sum_{\rho\nu} \zeta^\rho \left(Q^{-1}\right)_{\rho\nu} \xi_j^u \\ w_j &= w_j^0 + \xi^\perp \end{aligned}$$

with $\xi^\perp$ an n-dimensional vector that is perpendicular to all training patterns: $\xi^\perp \perp \{\xi^\mu\}$. Then the output of the perceptron is unaffected by $\xi^\perp$:

$$\zeta^\mu = \sum_j w_j \xi_j^\mu = \sum_j (w_j^0 + \xi_j^\perp) \xi_j^\mu = \sum_j w_j^0 \xi_j^\mu$$

Gradient descent learning

Often $P > N$, and thus patterns are linearly dependent.

General strategy is to define a learning rules through a cost function, such as the quadratic cost:

$$E(w) = \frac{1}{2}\sum_{\mu}\left(\zeta^{\mu} - \sum_{j} w_j \xi_j^{\mu}\right)^2$$

Gradient descent learning

The cost function can be minimized by the so-called gradient descent procedure. We start with an initial random value of the weight vector w and we compute the gradient in this point:

$$\frac{\partial E}{\partial w_i} = -\sum_{\mu} \left(\zeta^{\mu} - \sum_{j} w_j \xi_j^{\mu} \right) \xi_i^{\mu}$$

We change w according to the 'learning rule

$$w_i = w_i + \Delta w_i \qquad \Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

and repeat this until the weights do not change any more.

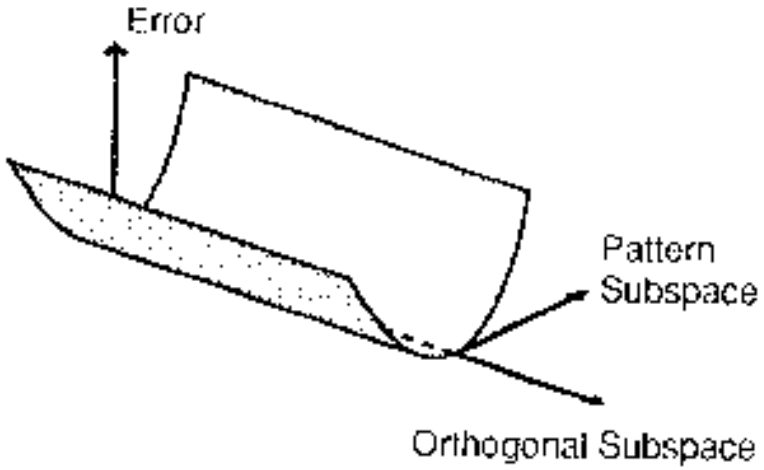
When η is sufficiently small, the gradient descent procedure converges.

The value of η

When η is very small, convergence is guaranteed, but may take long.

If η is large, however, convergence is no longer guaranteed.

The optimal choice of η is different for different components of the weight vector w.



The value of η

We can analyze the problem, by assuming that the energy has the form

$$E(w) \quad = \quad \frac{1}{2}\sum_i a_i \left(w_i - w_i^*\right)^2 + E_0$$

with w^* the location of the minimum, and a_i the curvatures in the two directions $i = 1, 2$. Eq. becomes

$$\Delta w_i \quad = \quad -\eta \frac{\partial E}{\partial w_i} = -2\eta a_i \left(w_i - w_i^*\right) = -2\eta a_i \delta w_i$$

with $\delta w_i = w_i - w_u^*$. The effect of learning step on δw_i is

$$\delta w_i^{\text{new}} \quad = \quad w_i^{\text{new}} - w_i^* = w_i^{\text{old}} - 2\eta a_i \delta w_i^{\text{old}} - w_i^* = (1 - 2\eta a_i)\delta w_i^{\text{old}}$$

thus, δw_i converges asymptotically to zero iff for all i

$$|1 - 2\eta a_i| < 1.$$

The value of η

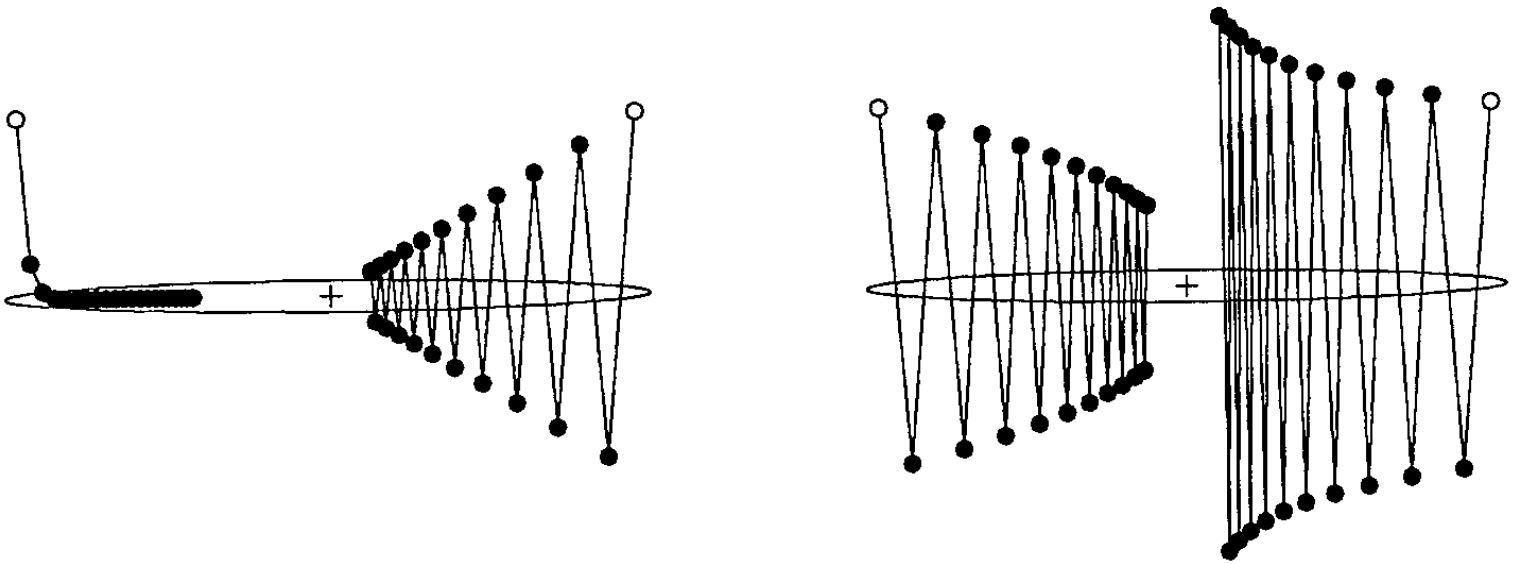


Figure 24: Behavior of the gradient descent learning rule Eq. for the quadratic cost function $E(w_1, w_2) = w_1^2 + 20w_2^2$ for $\eta = 0.02, 0.0476, 0.049, 0.0505$.

Non-linear units

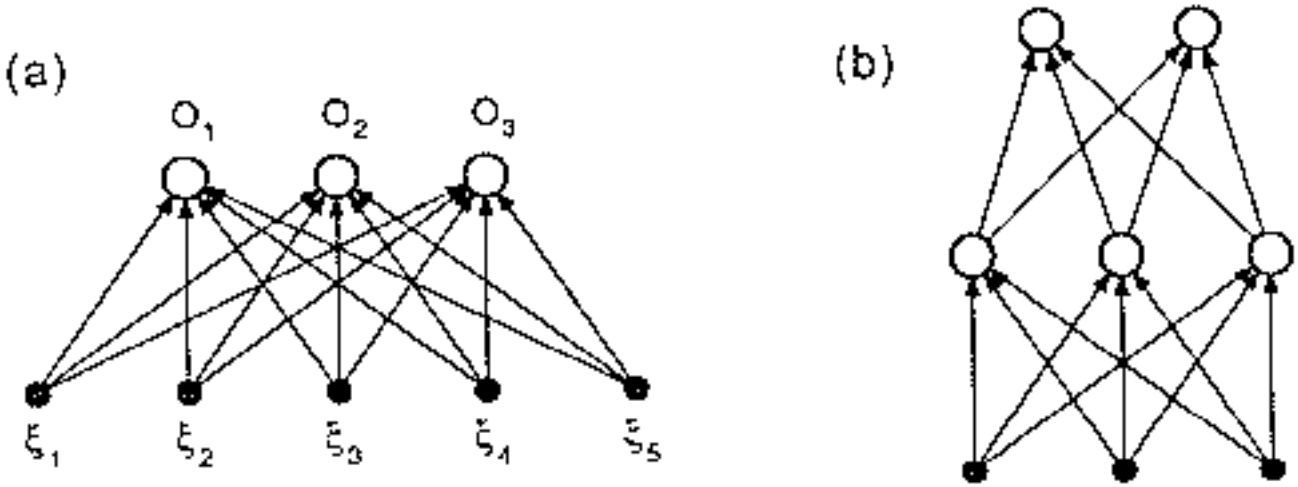
We can extend the gradient descent learning rule to the case that the neuron has a non-linear output:

$$o^\mu = g(h^\mu), \quad h^\mu = \sum_j w_j \xi_j^\mu$$

We use again the quadratic cost criterion:

$$\begin{aligned} E_1(w) &= \frac{1}{2} \sum_\mu \left(\zeta^\mu - o^\mu\right)^2 \\ \Delta w_i &= -\eta \frac{\partial E}{\partial w_i} = \sum_\mu \left(\zeta^\mu - o^\mu\right) g'(h^\mu) \xi_i^\mu \end{aligned}$$

Multi-layered perceptrons



Input variables ξ_k, output variable o_i, hidden variables v_j.

$$o_i(\xi, w) \quad = \quad g\left(\sum_j w_{ij} v_j\right) = g\left(\sum_j w_{ij} g\left(\sum_k w_{jk}\xi_k\right)\right)$$

Multi-layered perceptrons

Given a set of P training patterns $(\xi_k^\mu, \zeta_i^\mu), \mu = 1, \ldots, P$, we again use the gradient descent procedure to find the weights that minimize the total quadratic error:

$$E(w) = \frac{1}{2} \sum_i \sum_\mu (o_i^\mu - \zeta_i^\mu)^2$$

with $o_i^\mu = o_i(w, \xi^\mu)$.

Generalization: prediction on novel data

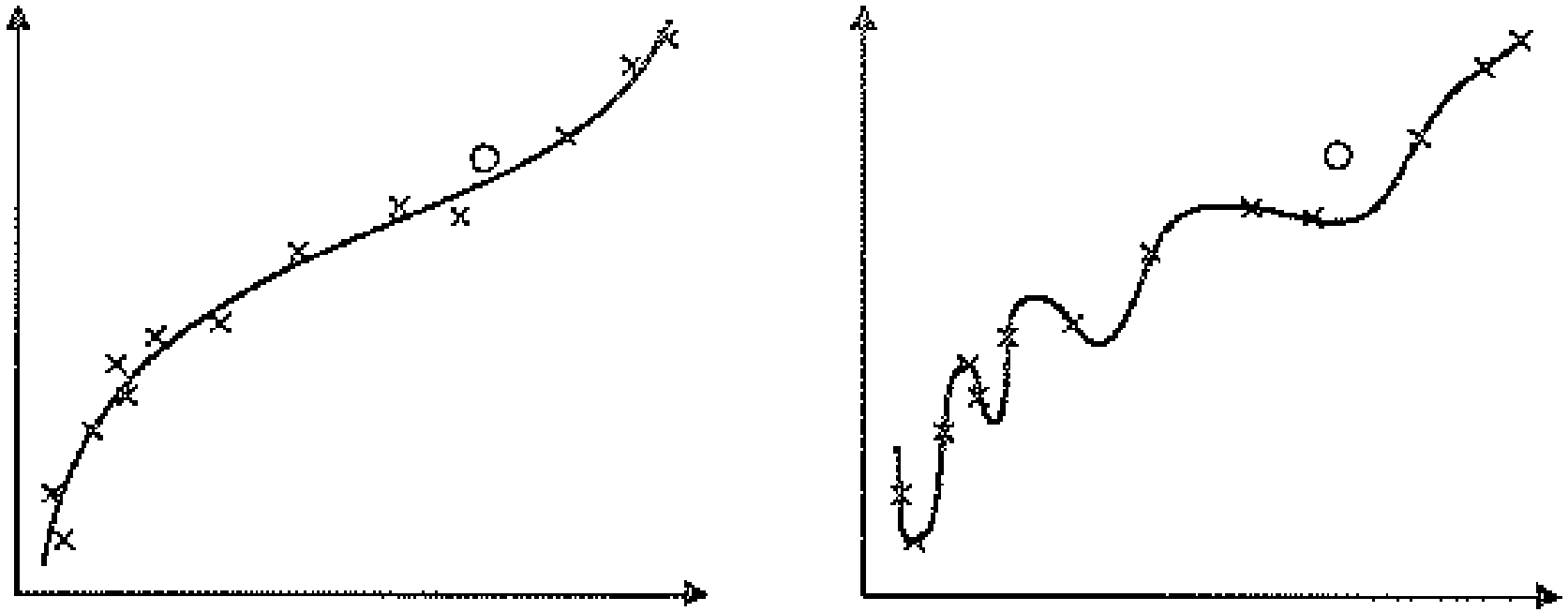


Figure 25: Network output versus network input. A) Network with a small number of hidden units. B) Network with a large number of hidden units. Networks with more hidden units can implement more complex functions and can better fit a given training set. However, more complex networks do not necessarily generalize better on novel data. Training data (x) and test data (o).

Summary

Perceptrons are simple models of feed-forward computation in a network of neurons. Binary perceptrons can be used for classification problems. Learning is done using the perceptron learning rule. The learning rule converges in a finite number of iterations if and only if the problem is linearly separable.

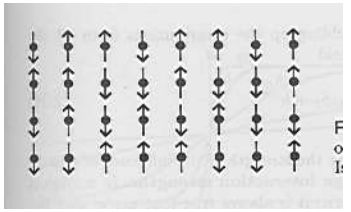
Perceptrons can also be constructed with continuous output, either using a linear or non-linear transfer function. These perceptrons can be learned using the gradient descent method. Gradient descent converges asymptotically for any data set.

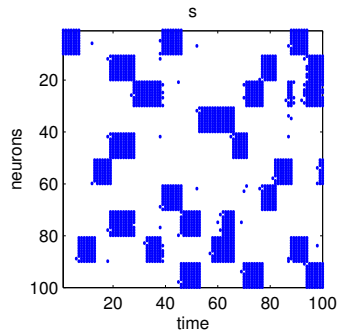
The quality of the perceptron can be significantly improved by using multiple layers of hidden units. The multi-layered perceptron can learn any function by using a sufficiently large number of hidden units. However, prediction quality on novel data does not generally increase with the number of hidden units. Optimal generalization is obtained for a finite number of hidden units.

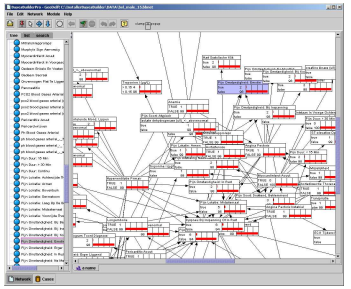
AI and neural networks research @ Biofysica

Stat. physics of:
- neural networks
- probability models

Applied to:
- brain
- AI







Courses:
Neural computation; Computational Neuroscience
Intro pattern recognition; Machine learning