**Prof. dr. Davor Eterović**

**Drafts of selected topics:**

# Biophysical grounds of physiology

**Department of Medical Physics and Biophysics/20012/13**
**University of Split – Medical Faculty**

**Original title**:

D. Eterović: Biofizički temelji fiziologije; Katedra za medicinsku fiziku i biofiziku, Sveučilište u Splitu-Medicinski fakultet

**Translated by**:

1. D. Eterović (chapters 6 and 7), editor
2. J. Marinović (chapters 2, 3 and 5)
3. M. Ljubković (chapter 1)
4. D. Kovačić (chapter 4)

**Comments to English edition:**

These drafts are the ground work for a Part 2 of a future textbook. Part 1 is the textbook: *'Physics of diagnostic imaging'*. These texts cover the subjects of the course: *Medical physics and biophysics* for medical students. The remaining task is to cover the missing subjects (Physics of eye and sight, Physics of ear and hearing and Mechanics of body) and to improve the graphics.

Generally, there is a lack of texts on physics of body on graduate level. This is not an overview of high-school physics with medical illustrations. On the contrary, the basic knowledge of physics is assumed and used in mastering the physiology of man. The level is occasionally advanced, matching the subjects covered by standard physiology texts for medical students. Due to limited course size only selected topics are presented. This text does not cover the integrative physiological aspects or histological/anatomical details.

The text has not been peer reviewed or English edited.

**D.E. (January, 2013)**

# Chapter 1: BIOTANSPORTS

## MEMBRANES MAINTAIN THE DIFFERENCES

The cell membrane provides **different** composition of the extracellular and intracellular fluids (plasma and cytoplasm).

In terms of electrical properties of the cell, a key difference is that the plasma contains a lot of sodium ions and cytoplasm a lot of potassium ions. In unexcited cell, these concentration gradients enable its **polarity** (negativity inside), and, like a loaded gun, allow rapid changes of cell's polarity when the cell is stimulated, i.e., when its permeability for these ions is changed. This enables the generation and propagation of neural signals, the so called **action potentials** in specialized nerve cells, the neurons.

In addition to ions, the two compartments differ in concentrations of other molecules. What they do have in common are: 1. **electroneutrality** (molar concentrations of positive and negative particles are equal in both plasma and cytoplasm), and 2. equal **osmotic pressures** (**iso-osmolarity**) in equilibrium, when there is no net flow of water. Specifically, the cell membrane has high permeability for water, and low for ions, thus acting as a semipermeable membrane, and the dissolved particles on both sides of the membrane create osmotic pressures, which are roughly proportional to their molar concentrations.

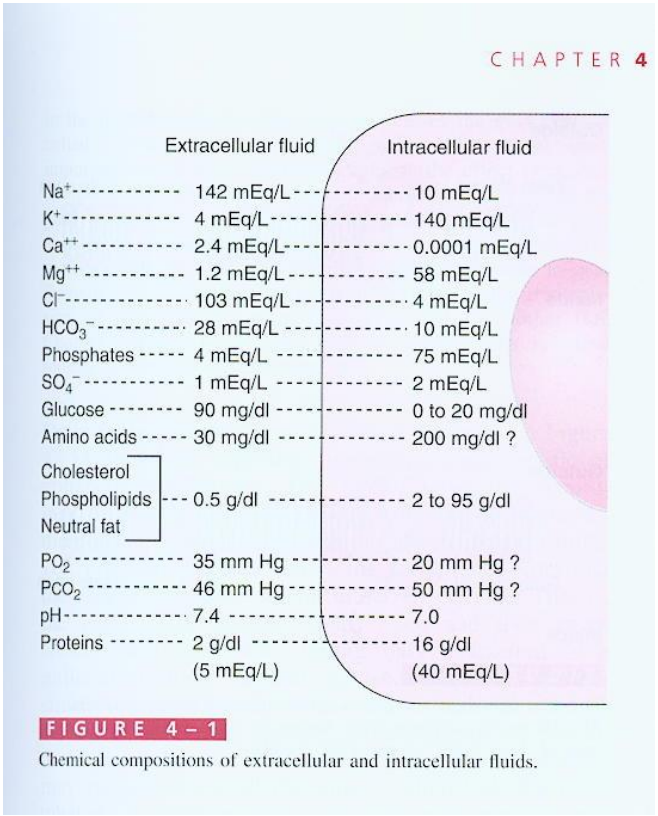



FIGURE 4–1
Chemical compositions of extracellular and intracellular fluids.

SELECTIVE MEMBRANE PERMEABILITY

The cell membrane is composed of lipid bilayer (the wall for polar molecules) continuity of which is disturbed by the membrane proteins, often completely disrupting the bilayer (transmembrane proteins). The role of some of these proteins is being a gate within the lipid wall, through which polar molecules can enter or exit from the cell. These are **transport proteins**.

The lipid bilayer is relatively permeable to small, nonpolar and weakly polar molecules (gasses, alcohol, and urea). The solubility of oxygen is higher in lipids of the membrane than in water. Therefore, oxygen easily passes through the cell membranes as if they were non-existent.

Transport proteins enable passage of water and ions. There are two ways. The so-called **channel proteins** have in their structure a hole through which water and some (not all!) ions dissolved in water freely enter, depending on the shape and size of their hydration shell. Another way is through the **carrier** protein: a molecule binds to a carrier at one of its sides, thereby initiating its conformational change which results in transfer and release of molecules on the opposite side. Depending on whether the conformational change of the carrier requires the expenditure of metabolic energy or not, such a mode of transport is called **active transport**, or **facilitated diffusion,** respectively.

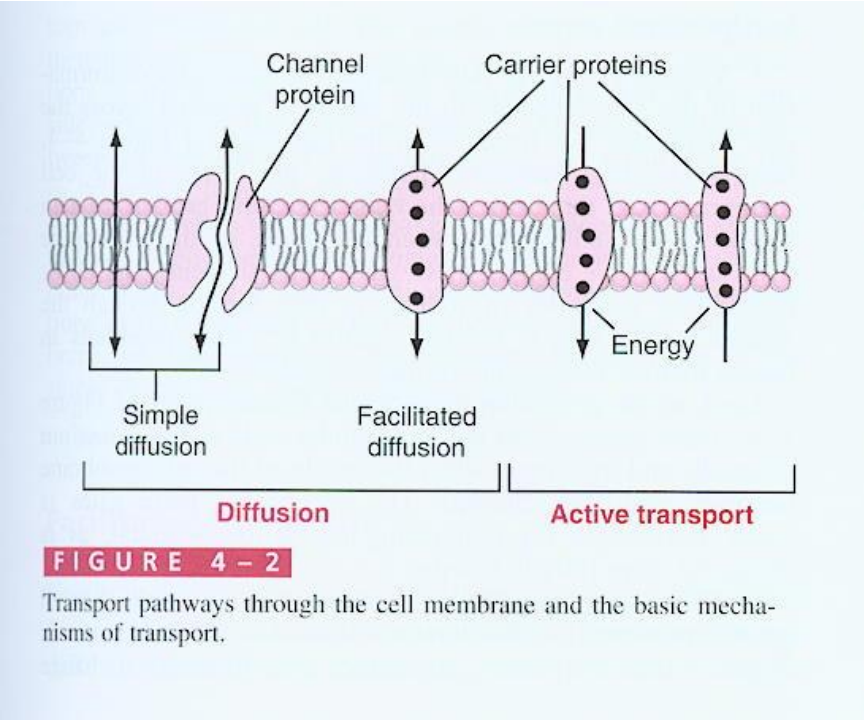



FIGURE 4 – 2

Transport pathways through the cell membrane and the basic mechanisms of transport.

Although water is lipid-insoluble, it readily passes through cell and other subcellular membranes, almost entirely through the channel proteins. A smooth passage of water is enabled by the special proteins **aquaporins**. There are several types of aquaporins, especially in the kidney epithelial cells within the nephron. Unobstructed passage of water molecule is enabled by its small size. For example, in red blood cells, in one second, a 100 times more water than the volume of the cell can pass through the erythrocyte membrane. Only slightly larger molecules pass much harder. Thus, urea, molecular diameter of which is only 20% higher than the water, passes about 100 times slower.

## DIFFUSION THROUGH PROTEIN CHANNELS

Protein channels are often highly selective for water-dissolved particles. Some of them are constantly open, and the others open and close as a result of the external stimulation.

In unstimulated cell, sodium ions, together with potassium ions pass through the cell membrane *via* so-called **Na-K leak channels** that are continuously open. In doing so, potassium ions pass much more easily because their hydration shell is smaller (although the potassium atom is larger then the sodium atom!).

After external stimulation, a selective transport of specific ion is very important. Selectivity of the protein channels is determined by the size, shape and distribution of the effective charge of its inner surface. Accordingly, the **sodium channels** are about 0.4 nm wide and are highly negatively charged. This negative charge quickly releases the sodium ions from their relatively large hydrating shell, so a sodium ion quickly finds itself in the interior of the protein. Another important type of protein channels is **potassium channels**. They are narrower than sodium channels and are not charged inside. Consequently, potassium ions pass through them together with their hydration shell, and sodium ions cannot pass due to their larger shell. Similarly, despite its smaller hydration shell, potassium cannot pass through the sodium channel, because the shape and charge distribution of inside of the channel is designed just for the sodium shell.

Na-K leak channels are always open. In contrast, permeability of specific sodium and potassium channels is regulated. It is said that these channels are **gated**.

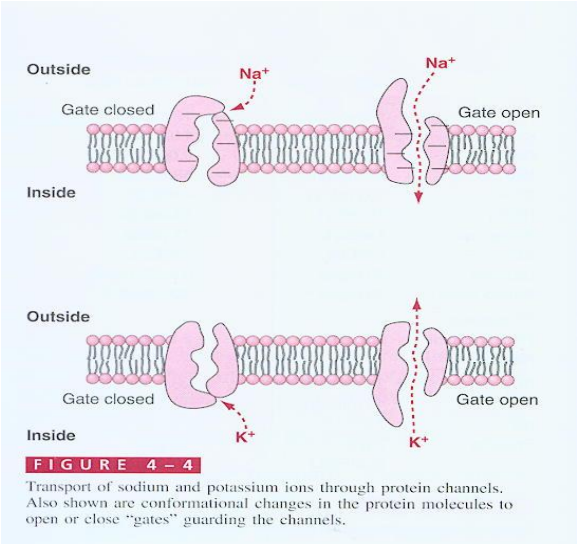



FIGURE 4–4
Transport of sodium and potassium ions through protein channels. Also shown are conformational changes in the protein molecules to open or close "gates" guarding the channels.

Opening and closing of sodium and potassium gates depends on the potential difference between the interior of cells and extracellular fluid, the so called **membrane potential**. That is why it is sad that potassium and sodium channels are **voltage-dependent**. It is believed that, by conformational changes, a special extension of protein molecule closes or opens the entrance to the protein channel. In case of sodium channel there are two, activation and inactivationt, gates, each on one side, while potassium channel has only one gate at the intracellular side.

In addition to voltage-dependent protein channels, there are those whose openness is regulated by direct binding of certain molecule to the protein. These are **ligand-gated channels**. An example is the cholinergic channels which are opened by binding of acetylcholine molecules. They are very important in synaptic transmission of neural signals from one neuron to another, as well as from neurons to muscle cells.

## FACILITATED DIFFUSION

In this case, in contrast to active transport, the energy that ensures the transport is the thermal energy of random motion. Therefore, as in regular diffusion, the net transport occurs from higher towards lower concentration. The difference is the presence of a transport protein that provides to the molecule a passage through the membrane. Given that the number of available transport proteins is limited, and that each transport requires time during which the protein moves to another state and then returns to its original shape ready for acceptance of a new molecule, facilitated diffusion is showing signs of **saturation kinetics**, i.e., the net transmembrane diffusion flow can not exceed a certain maximum size.

Facilitated diffusion is the way the glucose and the majority of amino acids enter the cells. Transport of glucose, through its impact on intra-cellular synthesis of the carrier proteins, is determined to large extent by **insulin**.

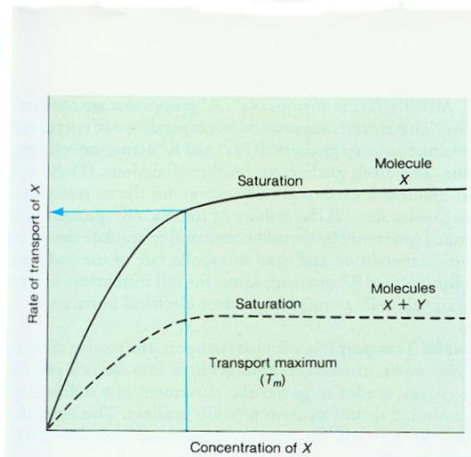



**Figure 5.12** Carrier-mediated transport displays the characteristics of saturation (illustrated by the *transport maximum*) and competition. Molecules *X* and *Y* compete for the same carrier, so that when they are present together the rate of transport of each is less than when either is present separately.



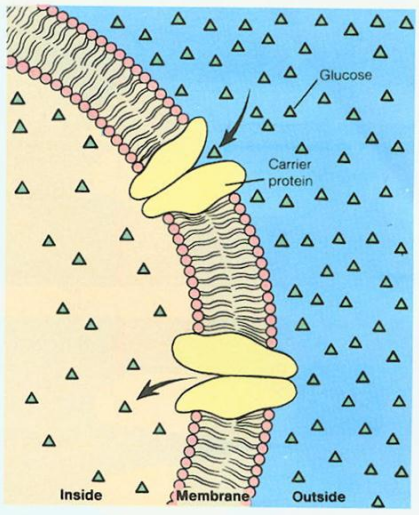



**Figure 5.13** A model of facilitated diffusion, where a molecule is transported across the cell membrane by a carrier protein.

FACTORS OF NET DIFFUSION FLOW

Diffusion is the random, thermal motion of molecules in which the movement direction of a certain molecule is constantly changing as a result of collision with other particles.

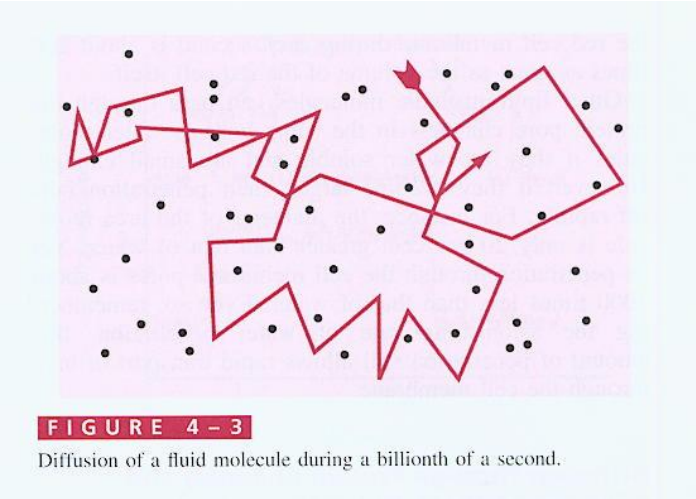


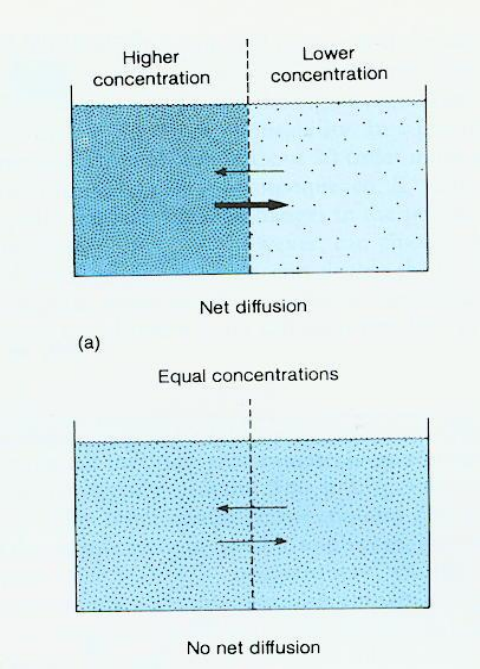
FIGURE 4 – 3

Diffusion of a fluid molecule during a billionth of a second.

If the concentration a substance **c(x)** decreases with distance **x** in the medium of viscosity **η**, the random thermal motion seeks to equalize the concentration, and thus occurs the mass flow **J** through the surface **S** from a larger to a smaller concentration (first Fick's law):

$$J = - D \cdot S \cdot \Delta c / \Delta x$$

**D** is diffusion constant and **Δc/Δx** is the speed (gradient) of concentration change on axis x.



From kinetic molecular theory:

$$\mathbf{D = u \cdot k \cdot T}$$

**k** is Boltzman constant, **T** is absolute temperature, **u** is mobility (diffusibility) of a particle in the medium.

For spherical particles of the radius **a**, Einstein demonstrated the following:

$$\mathbf{u = 1/6\ \Pi\ a\ \eta}$$

Therefore, small particles diffuse faster in the medium that provides little resistance.

During transfer from the fluid to the membrane, diffusion slows down or seizes, depending on the solubility of the particle

$$\mathbf{J = - D \cdot k_p \cdot S \cdot dc/dx}$$

$\mathbf{k_p}$ is the membrane partition coefficient: the ratio of the concentration of particles on the surface of the membrane and close to it ($\mathbf{k_p}$ <1).

The ratio **J/S** is called a unit diffusion flow **j**.

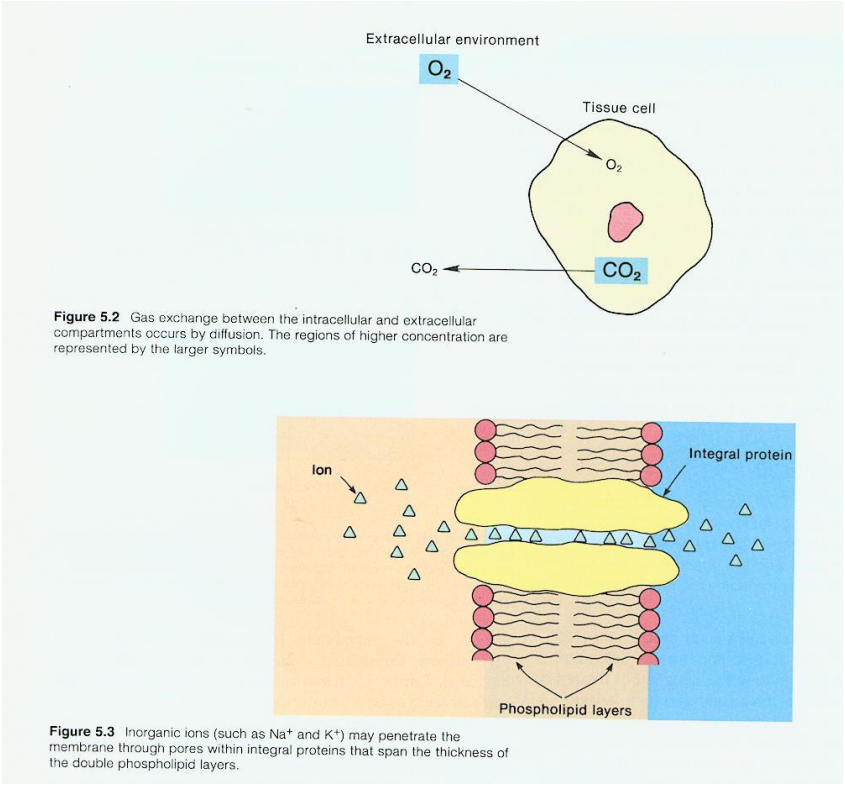



**Figure 5.2** Gas exchange between the intracellular and extracellular compartments occurs by diffusion. The regions of higher concentration are represented by the larger symbols.

**Figure 5.3** Inorganic ions (such as $Na^+$ and $K^+$) may penetrate the membrane through pores within integral proteins that span the thickness of the double phospholipid layers.

OSMOSIS: NET DIFFUSION OF WATER THROUGH SEMIPERMEABLE MEMBRANES

Suppose that a horizontal container is divided by a vertical membrane that is permeable to water, but not to the dissolved particles, and that concentration is higher in one section than the other. The total pressures on both sides of the membrane are equal. A pressure at each side of the membrane consists of a pressure generated by the water molecules and the pressure due to thermal motion of dissolved particles. Therefore, on the side with more dissolved particles, the partial pressure of water will be lower and *vice versa*. Since the dissolved particles cannot pass through the membrane, their partial pressures cannot equalize. In contrast, the water molecules will transfer from the less-dense solution the denser one, until the partial pressures of water are equalized. In doing so, the concentration of dissolved particles will equalize only if the container has infinitely compliant wall, i.e., if it does not resist to increasing its volume.

**Osmosis** is the flow of water through a semi permeable membrane from the compartment where the concentration of dissolved substances is lower to the compartment where it is higher. When the partial pressures of water become equal, a balance is reached and net flow of water stops.

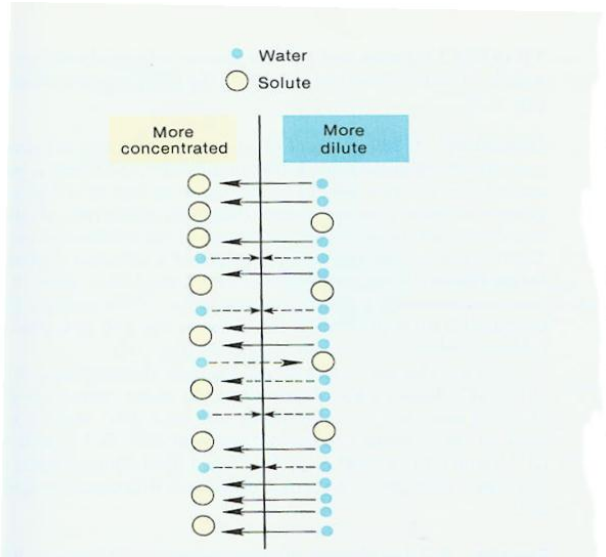



**Figure 5.4** A model of osmosis, or the net movement of water from the solution of lesser solute concentration to the solution of greater solute concentration.

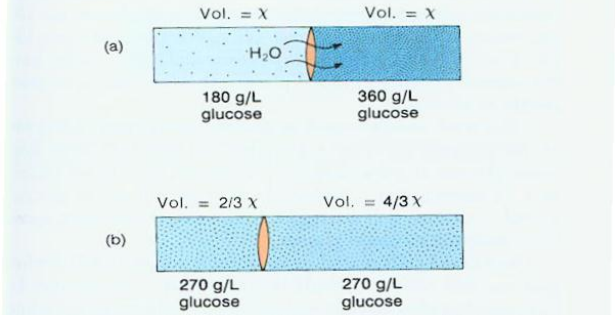



**Figure 5.5** A movable semipermeable membrane (permeable to water but not glucose) separates two solutions of different glucose concentration (*a*). As a result, water moves by osmosis into the solution of greater concentration until (*b*) the volume changes equalize the concentrations on both sides of the membrane.

For the osmotic pressure of dissolved particles (**π**, lack of partial pressure of water), the gas laws are approximately valid (van't Hoff's law): it is proportional to the **molar** concentration (**c**) of dissolved particles and temperature (**T**):

$$\pi = iRTc$$

**i** is the number of ions formed by dissociation of the molecule, and **R** the gas constant.

The product **ic** is called **osmolar concentration** or **osmolarity** and is measured in osmoles per liter (Os/L). Van't Hoff's law applies approximately for diluted solutions. Its more accurate form contains correction factor, the osmotic coefficient **θ**:

$$\pi = RT\,\theta ic$$

the product θic is called **effective osmolar concentration**. Osmotic coefficient can be greater or lower than 1. It is lower than 1 for physiologically important electrolytes. For all dissolved particles it approaches 1 as their concentration decreases. In addition to the concentration, it depends on the chemical properties of the solute.

Protein solutions deviate from van't Hoff's law substantially, and the degree of deviation is different for different proteins. As a rule, the osmotic pressure of proteins is larger than that predicted by van't Hoff's relation ($\theta > 1$). So for albumin, the most common blood protein, $\theta$ in blood is about 1.5.

In humans, normal body fluid osmolarity is about 300 mOs/L (in equilibrium, it is equal in plasma and cytoplasm!), which would cause osmotic pressure of 5790 mm Hg! However, the measured value is somewhat lower and is about 5500 mm Hg. By using effective osmolar concentration, we would approach this value.

In fact, a deviation from reality of van't Hoff law is less if, instead of *molar* concentration (amount of substance/volume of solution), we use the *molality* (amount of substance/mass of solvent). Therefore, in addition to osmolarity, a solution is defined by its **osmolality**. However, the osmolarity is measured more easily, and the difference between the two measures is less than 1% for fluids in the human body.

The table demonstrates the values of osmotic coefficient for concentrations of solutes in physiological range.

■ **Table 1-2** Osmotic coefficients ($\phi$) of certain solutes of physiological interest

| *Substance* | *i* | *Molecular weight* | $\phi$ |
|---|---|---|---|
| NaCl | 2 | 58.5 | 0.93 |
| KCl | 2 | 74.6 | 0.92 |
| HCl | 2 | 36.6 | 0.95 |
| $NH_4Cl$ | 2 | 53.5 | 0.92 |
| $NaHCO_3$ | 2 | 84.0 | 0.96 |
| $NaNO_3$ | 2 | 85.0 | 0.90 |
| KSCN | 2 | 97.2 | 0.91 |
| $KH_2PO_4$ | 2 | 136.0 | 0.87 |
| $CaCl_2$ | 3 | 111.0 | 0.86 |
| $MgCl_2$ | 3 | 95.2 | 0.89 |
| $Na_2SO_4$ | 3 | 142.0 | 0.74 |
| $K_2SO_4$ | 3 | 174.0 | 0.74 |
| $MgSO_4$ | 2 | 120.0 | 0.58 |
| Glucose | 1 | 180.0 | 1.01 |
| Sucrose | 1 | 342.0 | 1.02 |
| Maltose | 1 | 342.0 | 1.01 |
| Lactose | 1 | 342.0 | 1.01 |

Reproduced with permission from Lifson N, Visscher MB: *Osmosis in living systems*. In Glasser O, editor: *Medical physics*, vol 1, 1944, Chicago, Year Book Medical Publishers, Inc.

Osmotic pressures are rarely measured directly, but rather using the fact that the presence of a certain substance lowers the freezing point of the solution.

If two solutions have the same osmotic pressures, it is said that they are **isoosmotic**. If their osmotic pressures are not equal, the solution with higher pressure is **hyperosmotic** and the one with lower pressure is **hypoosmotic** with respect to other.

## OSMOTIC SWELLING AND SHRINKAGE OF THE CELL

Membranes of most somatic cells are almost impermeable to most of particles dissolved in interstitial fluid, and are highly permeable to water molecules. Therefore, when the osmotic pressure in the interstitium increases, the water leaves the cell by osmosis and the cell resultantly shrinks. Consequently, the concentration of particles in the cytoplasm increases, until its osmotic pressure is balanced with the interstitial one. On the other hand, if the osmotic pressure in the interstitium decreases, the water enters

the cell, which swells, thus reducing and eventually evening the difference between osmotic pressures.

If in a solution we suspend the cells, after which we do not notice their volume change, we say that the solution is **isotonic**. If the cells shrink, we say that it is **hypertonic**, and, if they swell, **hypotonic**.

It may seem that the solution is isotonic if the intracellular and extracellular fluids are isoosmotic. However, this is only true if the cell membrane is completely impermeable to all dissolved particles (in the cytoplasm and the external fluid). Particles that freely pass will not cause (except briefly, at the beginning) osmotic pressure, while the influence of those particles that pass with difficulty will be longer-lasting, but also transient. This means that, for example, isotonicity is not synonymous with isoosmolarity (as well as hyper- or hipo- tonicity to the hyper- and hipo-osmolarity, respectively). The membrane permeability for specific solutes must also be taken into account. **Reflection coefficient** is a dimensionless measure ranging from 0 (complete permeability) to 1 (completely impermeable). It is the ratio between the osmotic water flow for certain particle and a completely impermeable particle, for the same membrane and same difference in osmotic pressures.

## ACTIVE TRANSPORT

Sometimes the cells needs to ensure a high concentration of certain particles, in spite they are relatively sparse in the extracellular fluid. An example is potassium ions. The opposite is the case for sodium ions. It is clear that the spontaneous process of diffusion can not perform this task, indeed it opposes it. Therefore, there must be special, energy-dependent processes of specific purpose. They are called **active transports**.

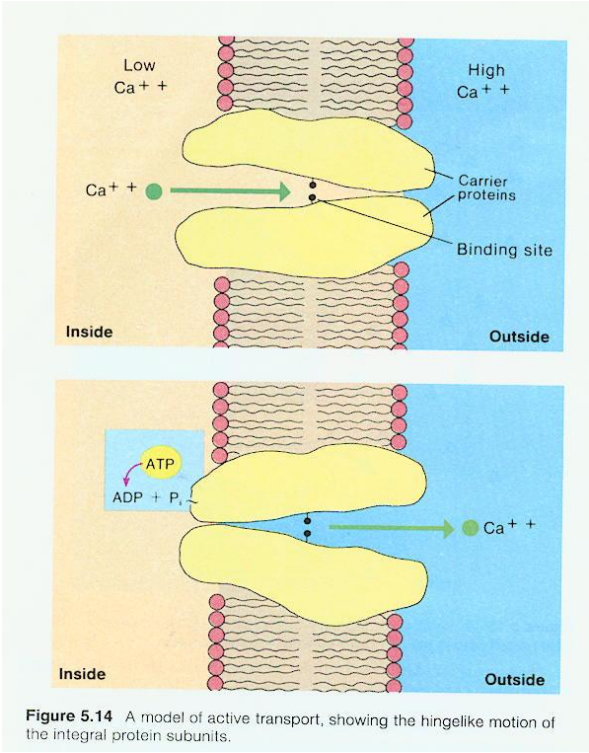



**Figure 5.14** A model of active transport, showing the hingelike motion of the integral protein subunits.

Various particles that are being actively transported through at least some cellular and intracellular membranes are ions of sodium, potassium, calcium, iron, hydrogen, chloride, iodine, urate ions, some sugars and most amino acids.

There are **primary** and **secondary active transports**. In primary active transport the energy is supplied by the breakdown of adenosine triphosphate (ATP) or some other high-energy phosphate compound. In the secondary active transport, the energy is provided *indirectly* through the ion concentration gradients, which had previously been established by the primary active transport. In both cases, the transport is carried out by the transmembrane protein carrier, as in the facilitated diffusion.

The most studied active transport is that by the Na-K pump. It is a protein complex of two distinct globular proteins. Larger unit has three receptor sites for sodium ions in the intracellular part, and two sites for potassium ions on the part that protrudes out into extracellular space. Part of the protein near the binding sites for sodium ions possesses an ATP-ase activity and is activated after binding of sodium ions from the intracellular fluid and extracellular potassium ions. ATP breaks down to ADP, while the released energy, by the still not completely known mechanism, participates in the conformational change of the carrier, so that sodium ions are expelled out of the cell and potassium ions are pushed in it.

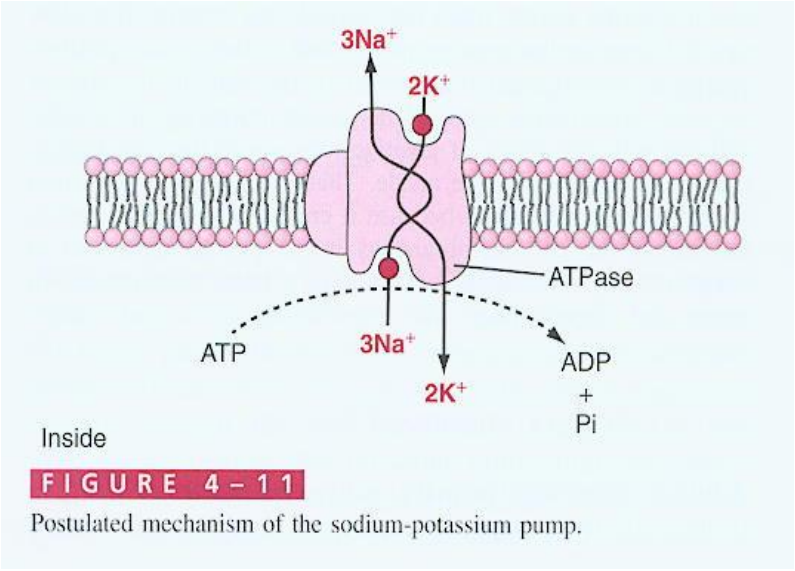



FIGURE 4 – 11

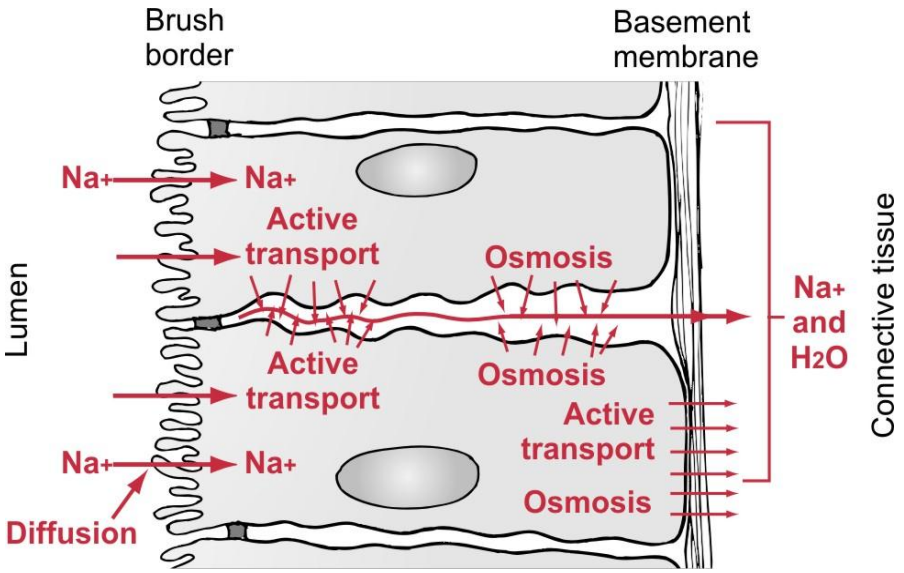
Postulated mechanism of the sodium-potassium pump.

Notice the two effects of Na-K pump activity: 1. reduction of the number of positive ions in the cell, and 2. reduction of the total number of ions in the cell. By opposing the constant entry of sodium ions and exit of potassium ions through leak channels, Na-K pump ensures their concentration gradients, which, together with its electrogenic effect, enables existence of the membrane potential and maintains the conduction of nerve signals (action potentials). In addition, Na-K pump is crucial in controlling the cell volume. Cells continuously synthesize proteins and other, mostly negative particles, which then gather positive ions around themselves. All these particles increase the osmotic pressure and draw water into the cell. Na-K pump, with its permanent reduction in the number of particles in the cytoplasm, is acting opposite and preventing bursting of the cell. Under conditions when the osmotic pressure increases in the extracellular fluid, its activity decreases.

Additionally, Na-K pump indirectly provides energy for secondary active transports, which are divided into **co-transports** and **counter-transport**s. In both cases, the driving force is the spontaneous tendency of sodium ions to enter the cell. In the co-transport, the sodium ions and certain molecule (e.g. glucose or amino acids in the renal tubules) bind together at the extracellular part of the transport molecule. Their binding starts a conformational change that ends with an entrance of sodium ions and that molecule into the cell. In counter-transport, a certain molecule needs to be extruded out of the cell. Therefore, the transport protein possesses a binding site for this molecule in

the part that protrudes into the cell, while the sodium ions bind to the extracellular side. This way is used by the cells to extrude calcium ions and hydrogen.

At many places in the body the particles must be transferred through a layer of cells, not just enter the cell or be extruded out of the cell. This means that the particles must enter the cell at one side, and, at the other side, exit the cell. This necessarily implies that the corresponding parts of the cell membrane have different composition and have different functions. Most commonly, the particles are actively transported into the cell at one side, and, on the other side, exit the cell by simple or facilitated diffusion. This is the mechanism of the transport of certain substances in the intestinal epithelium, kidney tubules, exocrine glands, gall bladder, as well as elsewhere.



INFLUENCE OF MEMBRANE POTENTIAL ON IONIC TRANSPORT

So far we have neglected that the interior of the cell is at negative potential as compared to the extracellular fluid (the cause of these phenomenon will be discussed in the following section: membrane potential). This means that there is a potential difference across the membrane (membrane potential, although, it would be more accurate to say, transmembrane voltage), or an electric field that attracts positive ions into the cell and pushes negative ions out. This field causes an *oriented* (not chaotic, as in diffusion) motion of ions- the ionic current. If, at the same time, there is a concentration gradient, there is also a net flow of ions diffusing from greater towards lower concentration and directed movement in or out of the cells, depending on the charge. These two flows are added algebraically, which determines the net flow of ions. (Remember that the chaotic, diffusion, temperature-dependent motion of ions is always present, regardless of the concentration difference.)

An ion is in **electrochemical equilibrium** if its transmembrane net flow is zero. This means that the flow due to electric field is equal in size and opposite in direction to the diffusion flow, which is a consequence of the concentration gradient.

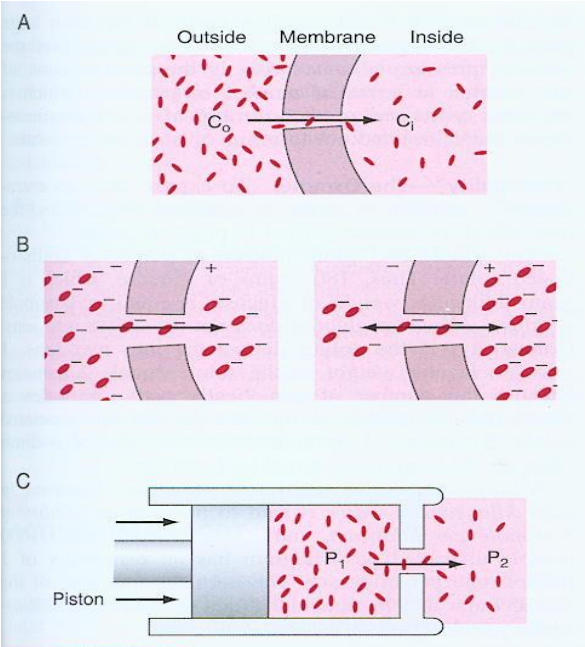



FIGURE 4 – 8

Effect of concentration difference (*A*), electrical potential difference (*B*), and pressure difference (*C*) to cause diffusion of molecules and ions through a cell membrane.

The following example shows how an electrochemical equilibrium of ions can be established between the two compartments with different concentrations of electrolytes, separated by a membrane. Let's say the membrane is permeable to one ion, a cation but not anion. In the beginning, a resulting diffusion net movement of both ions occurs from the direction of their higher concentration towards lower concentration. However, the membrane stops the anions, which remain on membrane surface at the same side. Cations pass into another compartment, but, attracted by the opposite negative charges, remain adherent to the other side of the membrane. This creates an electrical capacitor

whose field soon stops further diffusion of cations, and the steady state is achieved with no net flow of ions, i.e. the electrochemical equilibrium.

An important and seemingly unusual fact is that the quantity of ions that create a capacitor field that is sufficient to reach the equilibrium is negligibly small. This means that the initial concentrations of the electrolytes during the establishment of equilibrium did not change measurably. It is the same in our cells. Approximately is very true that the cells are electrically neutral. However, if we were to be completely precise, the cell has an immeasurably small excess of negative charge. But this extremely small surplus is sufficient to create a capacitor with a very strong electric field.

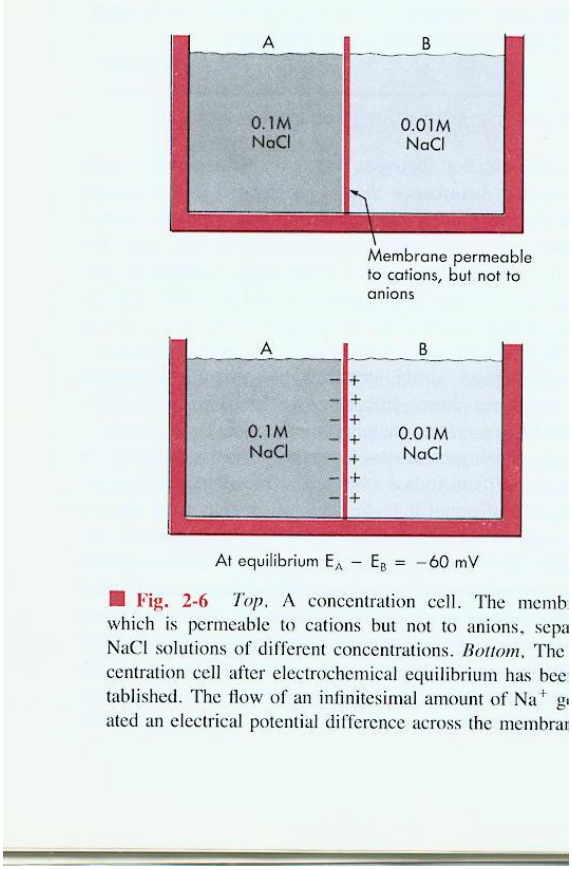



**Fig. 2-6** *Top*, A concentration cell. The memb which is permeable to cations but not to anions, sepa NaCl solutions of different concentrations. *Bottom*, The centration cell after electrochemical equilibrium has bee tablished. The flow of an infinitesimal amount of $Na^+$ g ated an electrical potential difference across the membra

The greater the ion concentration difference between inside and outside of the cell, the greater will be the potential difference required to balance its diffusion flow through the membrane. Let us show exactly the kind of relations that are valid. Suppose that an ion is in electrochemical equilibrium. Let’s associate its diffusion flow with the current due to electric field and acknowledge the condition of existence of electrochemical equilibrium.

**1. Ions move due to existence of electrical field**

Ionic current **I** per unit of surface **S** is called current density **j**:

$$\mathbf{j} = \mathbf{I}/\mathbf{S}$$

According to Ohm's law (microscopic form), the current density is proportional to the field intensity **E** (force per unit charge):

$$\mathbf{j}\ (\mathrm{A/m^2}) = \boldsymbol{\lambda}\mathbf{E}$$

**λ** is ion conductivity, which is proportional to the product of ion mobility, **u**, and its concentration, **c**. The constant of proportionality is the unit charge, **e**:

$$\boldsymbol{\lambda} = \mathbf{c}\,\mathbf{u}\,\mathbf{e}$$

Given that the field intensity **E** is associated with the spatial gradient of the potential **V**:

$$\mathbf{E} = -\,\mathrm{d}\mathbf{V}/\mathrm{d}\mathbf{x}$$

it follows:

$$\mathbf{j}\ (\mathrm{A/m^2}) = -\,\mathbf{c}\,\mathbf{u}\,\mathbf{e}\,\mathrm{d}\mathbf{V}/\mathrm{d}\mathbf{x}$$

**2. Ions move due to concentration difference**

According to the first Fick's law: $\mathbf{j}\ (\mathrm{mol/m^2 s}) = -\,\mathbf{u}\,\mathbf{k}\,\mathbf{T}\,\mathrm{d}\mathbf{c}/\mathrm{d}\mathbf{x}$

**3. Equilibrium**

In equilibrium, the unit diffusion flow and current density due to the electric field are equal and oppositely directed:

$$\mathbf{j}\ (\mathrm{A/m^2}) = -\,\mathbf{j}\ (\mathrm{mol/m^2 s})$$

$$\mathbf{u}\,\mathbf{k}\,\mathbf{T}\,\mathrm{d}\mathbf{c}/\mathrm{d}\mathbf{x}\ (\mathrm{mol/m^2 s}) = \mathbf{c}\,\mathbf{u}\,\mathbf{e}\,\mathrm{d}\mathbf{V}/\mathrm{d}\mathbf{x}\ (\mathrm{A/m^2})$$

After integration and adjustment of units we obtain the Nernst equation:

$$\mathbf{V_1}-\mathbf{V_2} = -\frac{\mathbf{RT}}{\mathbf{ZF}}\ln\frac{\mathbf{C_1}}{\mathbf{C_2}}$$

**R** is the gas constant (8.314 $\mathrm{Jmol^{-1}K^{-1}}$), **F** is Faraday constant (charge of one mole of unit charge, $9.65\cdot10^4$ $\mathrm{Cmol^{-1}}$), **Z** is the ion valence, and **T** is absolute temperature (K).

If the intracellular potential and ion concentration are marked by $\mathbf{V_1}$ and $\mathbf{c_{in}}$, respectively, and $\mathbf{V_2}$ and $\mathbf{c_{out}}$ denote the corresponding extracellular variables, the difference $\mathbf{V_1}$-$\mathbf{V_2}$ is called the membrane potential, $\mathbf{U_m}$; for a normal temperature of human body (T = 310 K) and monovalent cations (Z=1), and after converting to the decimal logarithm, we obtain the final form of the well-known Nernst equation:

$$\mathbf{U_m\ (mV) = -\ 61\times \log \frac{c_{in}}{c_{out}}}$$

Of course, in the case of anion, the sign of the right side of the equation is a plus. Also, for multivalent ions, the number 61 should be divided with the valence number.

Let's notice several important things:

1. Membrane potential does not depend on the difference, but the ratio of the concentrations of ion inside and outside of cells;

2. this relation is logarithmic, therefore a large difference in ion concentration ratio corresponds to a small difference in membrane potential, and

3. dependency on the ion valence is not weak, but direct and inversely proportional.

A slight dependence of membrane potential on ratios of concentrations of ions can be understood even without analysis of mathematical derivation. The point is that any change in ion concentration, in addition to change in diffusion flow, immediately changes the current due to electric field, even without it (field) changing. This is because the current is proportional to the ion conductivity and conductivity changes with the concentration of ions.

## SELF ASSESSMENT

1. The gas diffuses from capillary blood to erythrocytes. The partition coefficient of erythrocyte membrane for that gas is 0.5. After short time (while erythrocyte is still in the capillary), the steady state is established (provided by steady alveolar gas concentration and fast binding of the gas to hemoglobin in erythrocyte). The concentration gradient of gas concentration from blood to erythrocyte cytoplasm is 2 mmol/l. Which is the concentration gradient between outer and inner surface of erythrocyte membrane?

2. Two solutions, that have the same volume, contain 2g NaCl (the first one) and 3 g KCl (the second one). Which solution has greater osmotic pressure?

3. Describe all ways sodium ions can enter or leave cells; distinguish between bidirectional and unidirectional ways! Compare this with potassium ions!

4. Explain whether the positive ion can have greater extracellular than intracellular concentration and still be in electrochemical equilibrium!

# Chapter 2: MEMBRANE POTENTIAL

Measurements indicate that the majority of cells are polarized, i.e., that their interior is at a negative potential to the extracellular fluid, usually in the range of -30 to -90 mV, depending on the function. Although, compared to artificial capacitors, the voltage of this biological capacitor is not large, the corresponding electric field $\mathbf{E = U_m/d}$ is enormous, because the **d**, the thickness of the cell membrane, is very small.

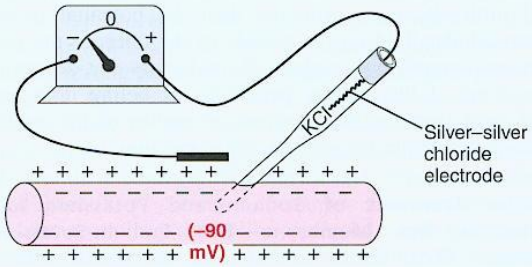



FIGURE 5 – 2

Measurement of the membrane potential of the nerve fiber using a microelectrode.

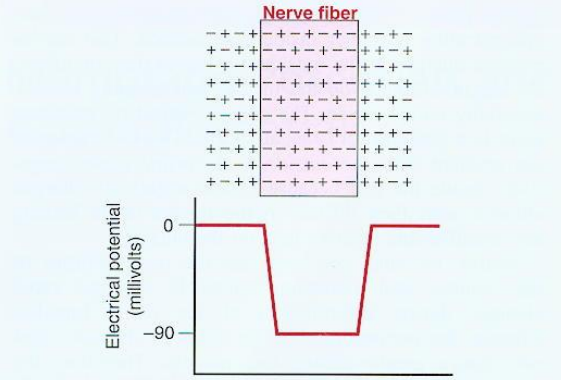



FIGURE 5 – 3

Distribution of positively and negatively charged ions in the extracellular fluid surrounding a nerve fiber and in the fluid inside the fiber; note the dipolar alignment of negative charges along the inside surface of the membrane and positive charges along the outside surface. In the lower panel are displayed the abrupt changes in membrane potential that occur at the membranes on the two sides of the fiber.

Contrary to expectations of the early researchers, it was quickly demonstrated that the majority of ions are not in electrochemical equilibrium, even in unstimulated neurons. Potassium is closest to its equilibrium, but the cell is always slightly more positive (less negative) than predicted by the Nernst equation for this ion. Therefore, potassium ions continuously, spontaneously exit the cell (electrical attraction is not sufficient to balance the diffusion losses). Sodium ions are driven into the cell by both electric field and concentration gradient. Therefore, the sodium is completely out of its equilibrium; in the cells with the membrane potential of -85 mV, the sodium equilibrium potential is about +60 mV. Sodium ions tend to swarm into the cell, but their passage through the leak channels is much harder than for potassium ions (due to the larger hydration shell).

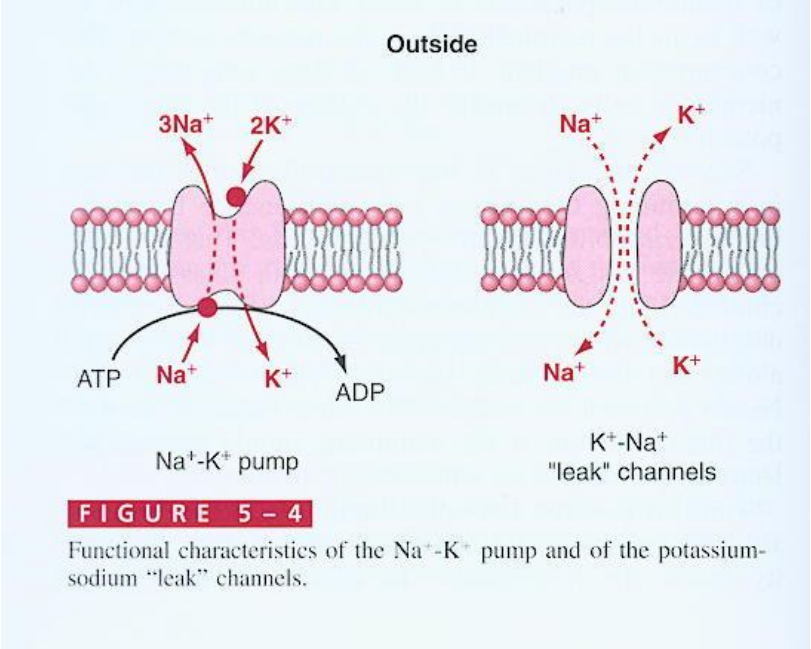



FIGURE 5 – 4

Functional characteristics of the $Na^+$-$K^+$ pump and of the potassium-sodium "leak" channels.

Therefore, through these always open leak channels there is a continuous net flow of ions. This constant moving away from the equilibrium is compensated by ion pumps. If a cell does not have a pump for an ion, it means that this ion is in its electrochemical equilibrium in this cell, i.e., that, even if it can pass through the ion channels, it does it to the same extent in both directions (in and out of the cell). For such ions the Nernst equation is exactly applicable. For example, some cells do not have (or are not yet proven to have) a chloride ion pump.

The question is raised about the purpose of leak channels which would justify the ongoing energy consumption, since active transport must return the ions that are continuously leaking through them trying to reach their equilibrium potential. The answer will soon be clear.

Cells are always electroneutral, i.e., their total charge is zero. This means that there must be as much positive charge as the negative (we neglect the immeasurable excess of negative charge in unstimulated cell, as well as the reverse case, the excess of positive charge during the formation of nerve signals). Electroneutrality of the cell entails that efflux of certain positive (negative) ions is accompanied by the influx of other positive (negative) ions. Another possibility would be that loss (increase) of positive ions is followed by an equal loss (increase) of negative ions, but then the cells would not be in osmotic balance. Thus, a constant efflux of potassium ions from the cell must be accompanied by an equal influx of sodium ions into the cell. Although sodium ions pass a lot harder through leak channels than potassium ions, they are driven by the greater 'force', and the transmembrane fluxes of these two ions are equal and opposite in direction. Permanent loss of potassium ions and influx of sodium ions is compensated by the Na-K pump. Since the pump ejects sodium ions sodium and injects potassium ions at a ratio of 3:2, and that, on the other hand, opposite transport of these ions through the leak channels is equal, the question remains as to why concentration of sodium ions is not continuously reduced. The answer is that sodium ions enter the cell also *via* secondary active transport, and that Na-K pump must eject them as well (along with those sodium ions that are constantly entering through the leak channels).

When passing through the membrane, cations drag behind ions of opposite charge (anions). This creates a moving **dipole**. Dipole is even stronger as the opposite charges are more apart. The distance between the opposite ions is higher, the greater the difference in their mobility. In general the mobility of cations is greater than the mobility of anions, which lag behind them. This creates moving potassium and sodium

dipoles which are oriented opposite and moving in opposite directions. Since the transmembrane fluxes of potassium and sodium ions are approximately equal, there are approximately equally much their dipoles, and their effects partially cancel. However, the mobility of potassium ions is much larger than sodium ions, so that the potassium dipoles are stronger. Because of this, constant, equal and opposite in direction fluxes of potassium and sodium ions through the cell membrane effectively produce an electric field (like in the capacitor) with the negative side being inside of the membrane.

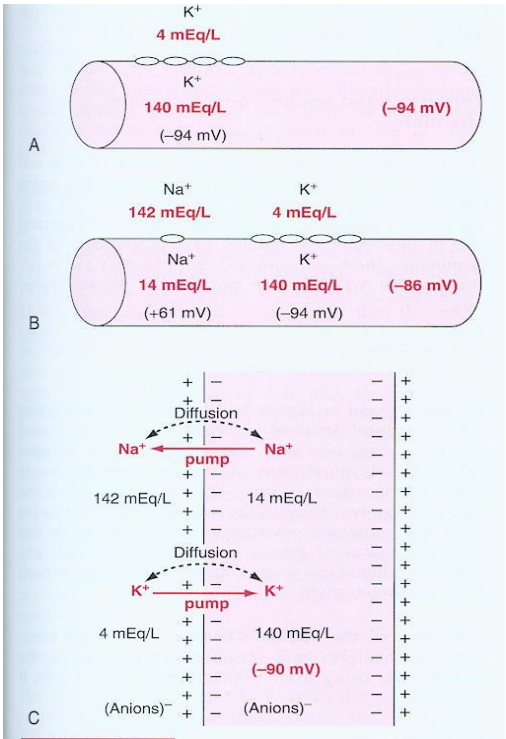



FIGURE 5–5

Establishment of resting membrane potentials in nerve fibers under three conditions: *A*, when the membrane potential is caused entirely by potassium diffusion alone; *B*, when the membrane potential is caused by diffusion of both sodium and potassium ions; and *C*, when the membrane potential is caused by diffusion of both sodium and potassium ions plus pumping of both these ions by the $Na^+$-$K^+$ pump.

This so-called, diffusion potential is the largest component of the membrane potential. The remaining factor is the excess of negatively charged particles inside the cells (mostly proteins and phosphates), which partly results from the electrogenic effect of Na-K pump. These negative ions are grouped along the inner side of the membrane and attract positive ions from the outside of the membrane. Thus, membrane potential is only partially a result of excess non-moving negative charges within the cell; it is more the effect of orientation of the moving dipoles. Now we observe the physiological justification for the existence of leak channels. Without them, the membrane potential would be much smaller. Consequently, it could not change quickly and greatly, which is the basis for the formation and conduction of nerve signals.

Let's emphasize once again the overwhelming impact of Na-K-pump. Specifically, in addition to its direct electrogenic action, more important is the indirect effect, because the pump maintains the ion concentration gradients. Termination of its activity results in shutting down the cell's membrane potential to zero.

It is possible to precisely connect the membrane potential of the cell with the intra- and extra-cellular concentrations of those ions that can pass through the cell membrane. It is necessary to know the permeability of the membrane to these ions. Depending on how you define the permeability, as the conductivity or the mobility, there are two models: a **cable model** and a **constant field model**.

### Cable model

If the membrane potential of cell is equal to the equilibrium Nernst potential for the ion, the ion's net flow through the membrane with thickness d is equal to zero. If not, the difference between the membrane potential of the cell $\mathbf{U_m}$ and equilibrium potential of this ion $\mathbf{U_x}$ determines the effective field $\mathbf{E_x=(U_m-U_x)/d}$, which causes an ion current directed in or out of the cell, depending on charge sign and direction of force. According to Ohm's law (microscopic form) the density of this current ($\mathbf{j_x}$) is proportional to the effective field, and constant of proportionality is the conductivity of this ion in the membrane ($\boldsymbol{\lambda_x}$):

$$\mathbf{j_x = \lambda_x (U_m - U_x)/d}$$

Electronegativity condition implies that the total ion current through the membrane is zero.
Suppose, for simplicity, that the membrane is permeable only for ions of potassium and sodium. Then it is valid:

$$\mathbf{j_K + j_{Na} = 0}$$

or

$$\mathbf{\lambda_K (U_m - U_K) + \lambda_{Na} (U_m - U_{Na}) = 0}$$

Solving by the $\mathbf{U_m}$ we get so-called **cable equation**:

$$\mathbf{U_m} = \frac{\lambda_K}{\lambda_K + \lambda_{Na}} \mathbf{U_K} + \frac{\lambda_{Na}}{\lambda_K + \lambda_{Na}} \mathbf{U_{Na}}$$

Therefore, the equilibrium potential of the cell is the weighted mean between the equilibrium (Nernst) potentials of potassium ions and sodium ions. Weighting factors are membrane ion conductances. As in all cells, including the non-excited neurons, potassium ion conductivity is much higher than the conductance of sodium ions, the equilibrium potential of the cell is closer to the Nernst potential for potassium (about -90 mV) than the Nernst potential for sodium (about +60 mV). This relation is only transiently changed in the excited neurons and associated sensory cells.

In addition to potassium and sodium, other ions, depending on the type of the cell, can pass through the membrane in an effort to reach their equilibrium potential. Effect of chloride and other ions can be taken into account by simply adding the appropriate term in the above equation.

Since in different cells various ions affect the membrane potential, and additionally, membrane conductances for individual ions are not equal, the membrane potential in our body varies from -7 mV in erythrocytes, through -30 mV in some smooth muscles, -90 mV in the cardiac muscle to -150 mV in the sensory cells of the ear.

The cable equation correctly predicts the membrane potentials and is a simple, plausible and instructive as a starting point for an explanation of the inequilibrium states during generation of action potentials. Disadvantage is that it uses ionic conductivity. Ion conductivity is proportional to the product of its mobility and concentration ($\mathbf{\lambda = c\ u\ e}$). Therefore, the ion conductivity depends not only on the intrinsic properties of the membrane for the ion (mobility), but also on the concentrations of ions at each side of the membrane. This sometimes makes it difficult for its implementation, in comparison with the equation that uses only the mobility of ions.

### Constant field model

Assuming that the electric field within the membrane does not change spatially, for monovalent ions, similarly as we derived the Nernst equation, scientists Goldman and Katz have shown (for simplicity, we present only contributions of potassium, sodium and chloride)

$$U_m = -\frac{RT}{F}\ln\frac{u_K(c_K)_{in} + u_{Na}(c_{Na})_{in} + u_{Cl}(c_{Cl})_{out}}{u_K(c_K)_{out} + u_{Na}(c_{Na})_{out} + u_{Cl}(c_{Cl})_{in}}$$

The index **in** denotes the intracellular and **out** extracellular concentrations of ions.
Goldman-Katz equation and the equation of cables give similar results. And here it is clear that the ions whose mobility is high are "running the show".

**Why we say that the membrane is impermeable to ions**

At the end of this detailed presentation of transmembrane transport of ions and neutral particles, we need to know to answer one important question. When we talk about the distribution of substances in our body, we must first identify two compartments: intracellular and extracellular. The latter is subdivided into the intravascular and extravascular. For example, most of the blood cells and, to some extent, proteins can not pass through the capillary walls, and for them the latter two compartments are separate entities. In contrast, there are no obstacles for water and a glass of water you drink is distributed evenly in all body areas. Electrolytes from the blood easily (together with water) cross to interstitium and *vice versa*, so that the whole extracellular space for them is one compartment. However, the salt that you ingest will not end up in our cells. We say that cells are almost impermeable to ions. How is it that, when we have just seen that various ions can enter the cell and come out through the transport channels, which they constantly do and not only equally in both directions, but also there are constant net fluxes of potassium and sodium ions (which are compensated by Na-K pump)?

The answer is this: The net flux of potassium and sodium is compensated by Na-K pump, so that the number of ions inside the cell will not change measurably. Furthermore, each input or loss of ions in the extracellular liquid (diet, renal excretion) does not result in change in the number of ions within the cells (number, not concentration!). For example, if a person drinks a tablet of potassium chloride (as compensation, since the person is taking a diuretic that eliminates the potassium ion), the concentration of potassium (and chloride) ions in the extracellular fluid will increase. This increase means that the concentration of potassium ions does not any more correspond to Goldman-Katz equation (the ion is even more thrown out of equilibrium Nernst potential). As there is more potassium ions outside the cell than in the previous steady state (at the level of whole cell), a net-entry of potassium ions into the cell begins. However, as we have seen in the case of semi-permeable membrane, an immeasurably minute influx of ions is sufficient for a substantial change in membrane potential, which, rapidly achieving a new steady state, prevents further net influx of ions. We can, seemingly paradoxical, state that the entrance of an ion into the cell causes its impermeability to this ion.

This does not mean that significant changes in the concentration of electrolytes in the interstitial fluid have no effects on cell function. Indeed, in the case of potassium, the change of the membrane potential (decrease of polarity - depolarization in the case of hyperkalaemia, and, conversely, hyperpolarization in the case hypokalaemia) is the phenomenon that can have serious, even fatal consequences by compromising the

conductance of nerve signals. Sodium ions pass through the cell membrane harder than potassium ions, and it is much more difficult to significantly change its extracellular concentration (it is about 35 times higher). Therefore, a perturbation in its extracellular concentration is primarily manifested osmotic.

**EXAMPLE. Let's say that in a human cell the potassium ions are in equilibrium, and their cytoplasmic concentration of 140 mmol / L and plasma 4 mmol / L. After an overdosing with potassium-chloride tablets and resulting plasma potassium increases to 6 mmol / L, what will happen to the cell membrane potential?**

By applying the Nernst equation for potassium ions, we obtain the equilibrium potential of the cells before taking the tablets:

$$U_{m1}\,(\text{mV}) = -61\times \log \frac{\mathbf{c_{in}}}{\mathbf{c_{out}}} = -61\times \log \frac{140}{4} = -94\ \text{mV}$$

It is important to note that the new membrane potential, after an increase in plasma potassium, $U_{m2}$, is obtained by the inclusion of new plasma concentration $c_{out}$ in the Nernst equation, because we know that the cytoplasmic concentration $c_{in}$ will not change. So:

$$U_{m2}\,(\text{mV}) = -61\times \log \frac{\mathbf{c_{in}}}{\mathbf{c_{out}}} = -61\times \log \frac{140}{6} = -83\ \text{mV}$$

The conclusion is that hyperkalaemia depolarizes the cell and that this change in membrane potential prevents entry of a measurable amount of excess ions from the extracellular fluid into the cell. However, a consequence is that cell depolarization, in the case of neurons, can compromise its function, i.e. generation and conduction of action potentials. Specifically, depolarization of neurons brings them to their firing threshold. The result is hypersensitivity, which, in case of heart muscle, causes rhythm disorder. Similar will happen in the case of reduced extracellular concentration of potassium, only then the cell will be hyperpolarized. And that will also compromise the neuronal function and thus the function of muscle cells that are stimulated by the neurons.

# Chapter 3: ACTION POTENTIAL

Almost all cells are polar that is, they have a membrane potential. However, only some neurons have the ability to rapidly, but reversibly change their polarity at the site of excitation. This event is propagated away from the excitation point in a self-renewing manner, without losses, along the nerve fiber (axon). This is a **nervous signal** or **action potential**.

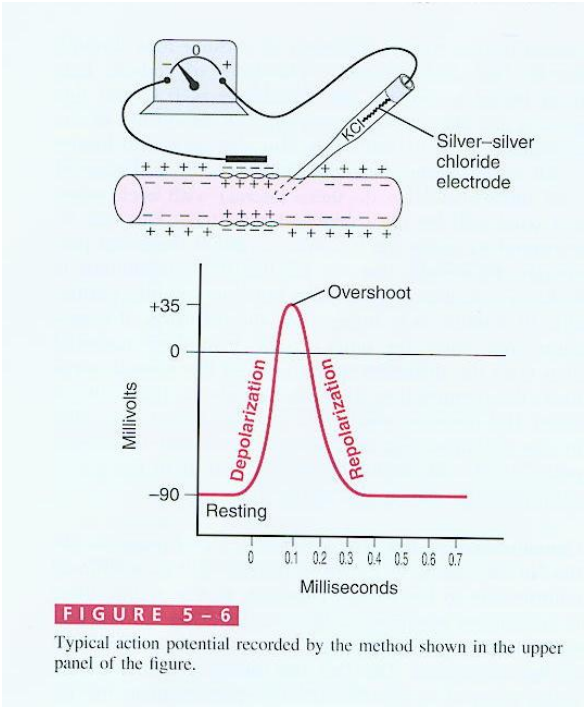
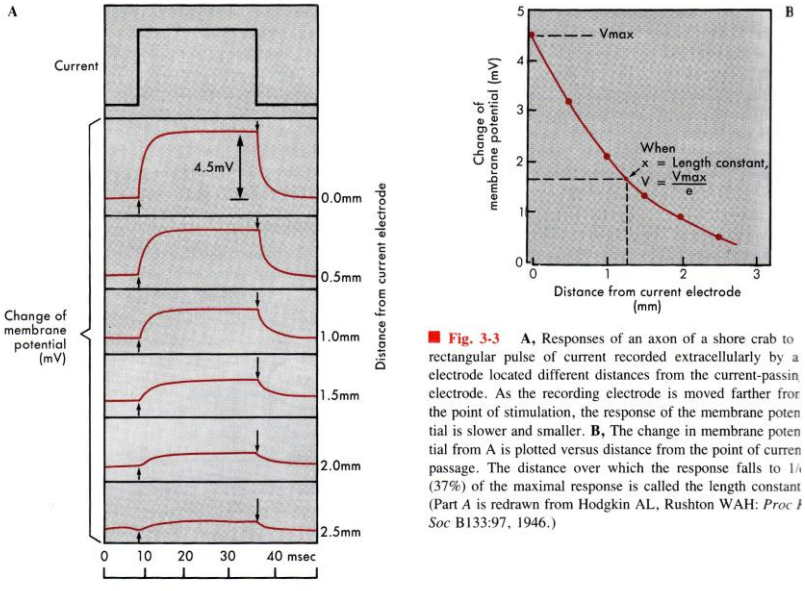



FIGURE 5–6

Typical action potential recorded by the method shown in the upper panel of the figure.

Recall that in non-stimulated steady state, potassium and sodium ions pass through the neuronal membrane (potassium net outward, sodium inward) through *leak channels* so that none are in electrochemical equilibrium (these changes are balanced by Na-K pump, so that in the long-term concentrations of both ions do not change). In doing so, sodium ions have a strong tendency to enter into the cells, because they are forced by both the diffusion gradient and the electric field. However, the channel permeability is much smaller for sodium ions, so that the trans-membrane fluxes of both ions are equal.




■ Fig. 3-3 **A,** Responses of an axon of a shore crab to rectangular pulse of current recorded extracellularly by a electrode located different distances from the current-passin electrode. As the recording electrode is moved farther fror the point of stimulation, the response of the membrane poten tial is slower and smaller. **B,** The change in membrane poten tial from A is plotted versus distance from the point of curren passage. The distance over which the response falls to 1/ (37%) of the maximal response is called the length constant (Part *A* is redrawn from Hodgkin AL, Rushton WAH: *Proc F Soc* B133:97, 1946.)

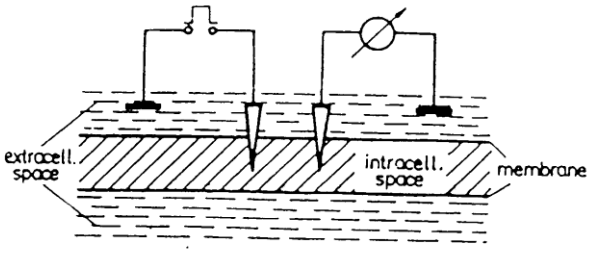



*Fig. 6.3.* Diagram relating to the change of the resting potential

On the left the exciting electrode pair; on the right the recording electrode pair. The distance between the two electrode pairs is of the order of a tenth of a mm

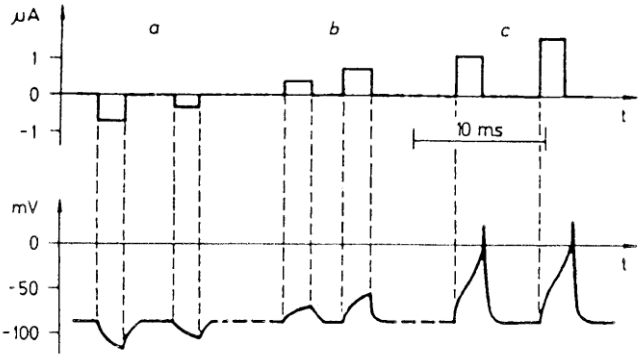



*Fig. 6.4.* The effect of square-wave current pulses (upper diagram) on the membrane potential (lower diagram)

The upper ordinate shows the amplitude of the current pulses, and the lower one the membrane potential. The abscissae give the time

**Local potential** (stimulus below the threshold, acute or subliminal stimulus) occurs when an outside stimulus (physical deformation of the membrane, electrical current, binding of ligands to the post-synaptic neurons) induces a change of the membrane potential of only a few mV on one part of the membrane.

This change can be **hyperpolarization** (increase in negativity of intracellular potential) or **depolarization** (reduction in negativity of intracellular potential) and it induces local potentials that travel with decreasing intensity as far as couple of millimeters from the stimulation point. The greater the stimulus, the greater the local potential will be.

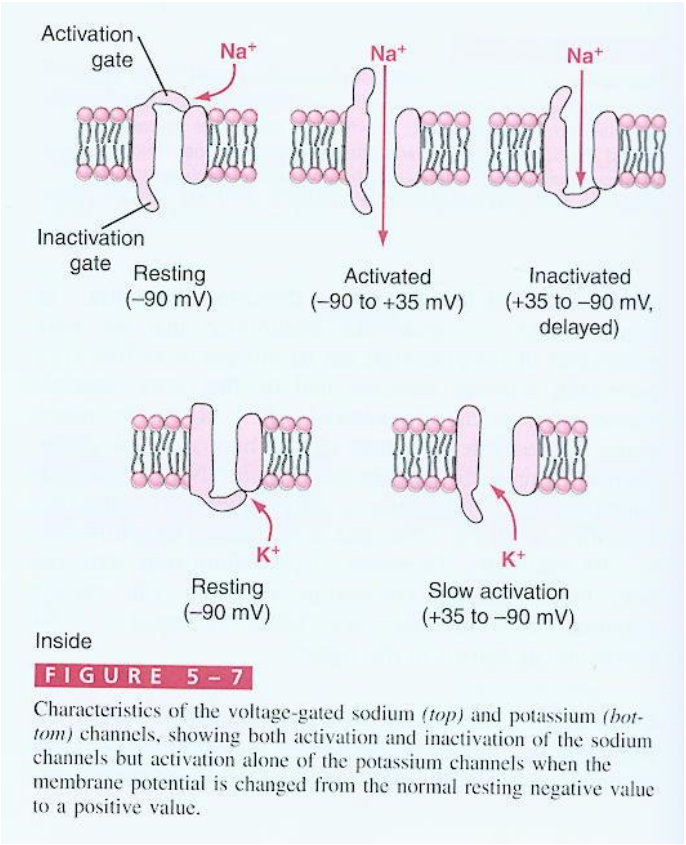



FIGURE 5–7

Characteristics of the voltage-gated sodium *(top)* and potassium *(bottom)* channels, showing both activation and inactivation of the sodium channels but activation alone of the potassium channels when the membrane potential is changed from the normal resting negative value to a positive value.

When the initial change is a **sufficiently large depolarization** (for example 20 mV), a locally-induced voltage change takes a typical form, **independent from further increasing the stimulus** and begins to propagate along the axon **without losses**, i.e. an **action potential** occurs. Membrane potential at which a neuron fires an action potential is called the **firing threshold** or **triggering threshold**.

A starting point is a steady state, when the cell is polarized (more negative inside). Sodium ions tend to enter the cell, but it is difficult for them to pass

through the leak channels. Increase in permeability for sodium ions causes a rapid depolarization because the sodium ions (positive ions) rush into the cell after the opening of their, previously closed, channels and so reduce the intracellular potential negativity. After that, before the sodium reaches its equilibrium potential, the permeability for potassium increases, and potassium comes out of the cell and thus the cell is **repolarized.** All in all, the membrane potential is changed for about a hundred mV in just a few milliseconds (for example, from -70 to +30 mV and back to -70 mV).

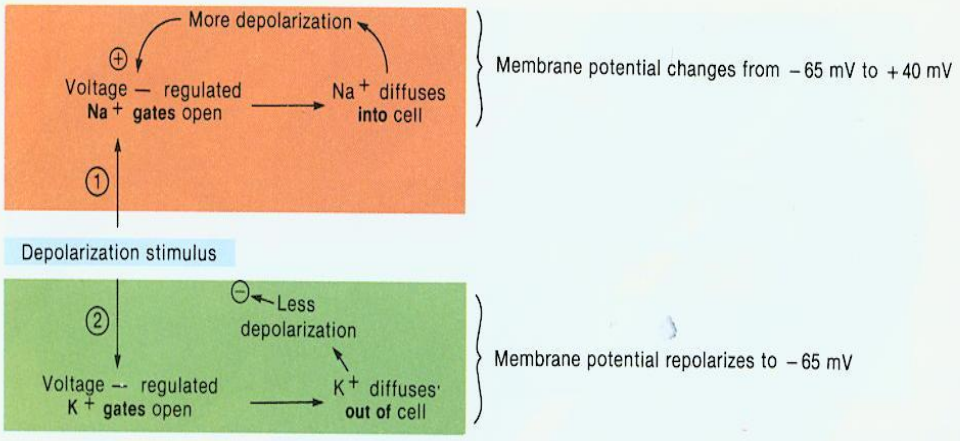



**Figure 14.11** Depolarization of an axon has two effects: (1) $Na^+$ gates open and $Na^+$ diffuses into the cell, and (2) after a brief period, $K^+$ gates open and $K^+$ diffuses out of the cell. An inward diffusion of $Na^+$ causes further depolarization—this causes further opening of $Na^+$ gates in a positive feedback + fashion. The opening of $K^+$ gates and outward diffusion of $K^+$ make the inside of the cell more negative and thus have a negative feedback effect − on the initial depolarization.

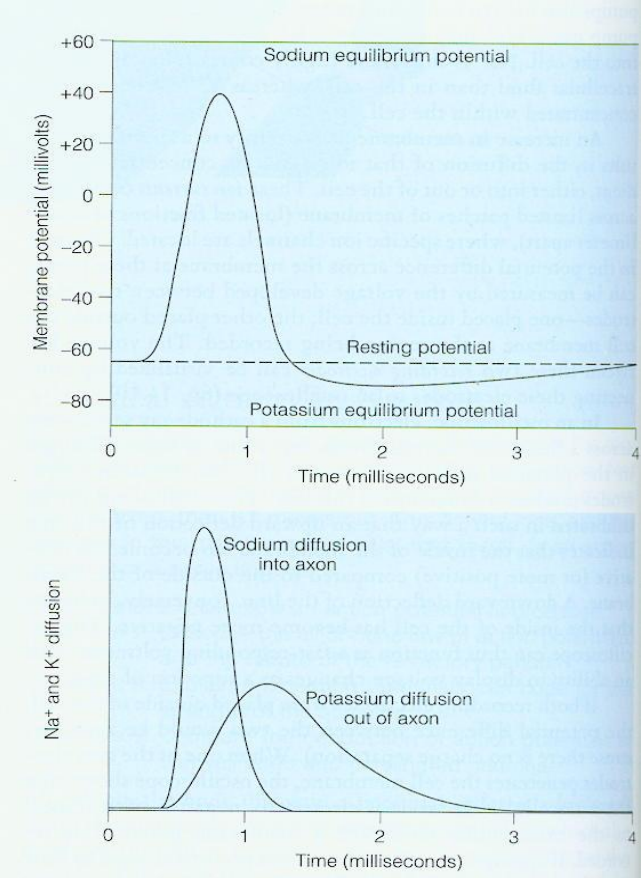
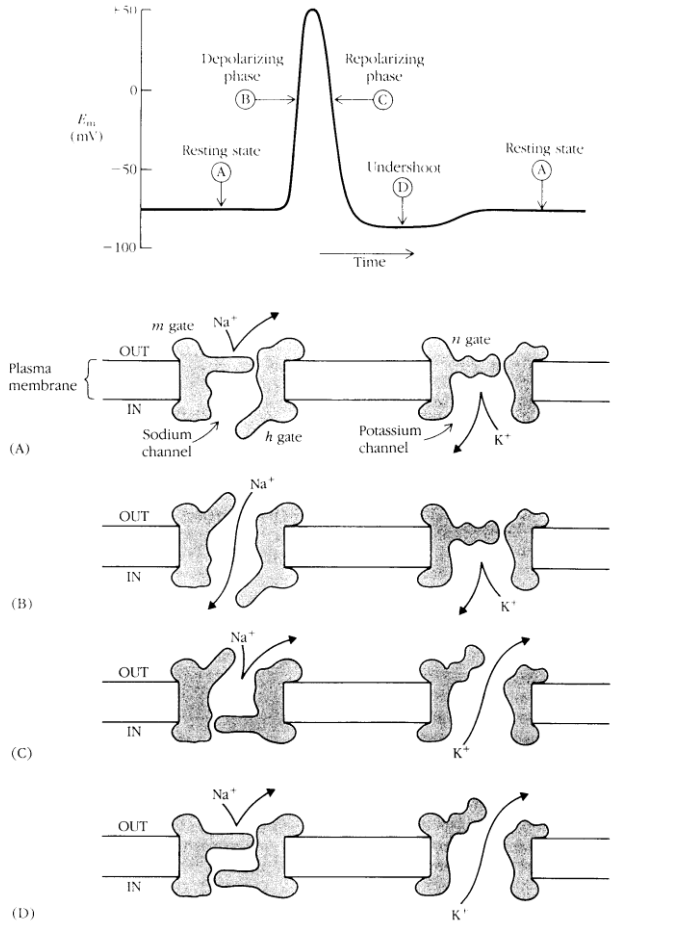



**Figure 14.12** An action potential (*top*) is produced by an increase in sodium diffusion followed, with a short time delay, by an increase in potassium diffusion (*bottom*). This drives the membrane potential first toward the sodium equilibrium potential and then toward the potassium equilibrium potential.

The cause of increased membrane permeability to ions: the very beginning of depolarization (immediately, before reaching the threshold trigger!) opens the gates of sodium (first), then potassium channels changing their protein conformation. These are **voltage-dependent** channels. Sodium ions start entering the cell immediately after the gate is opened. However, the explosive entrance of sodium ions is possible only after reaching the firing threshold. Prior to that, entry of sodium ions and the resulting depolarization push the potassium ions out of the cell (in part through the always open leak channels). Consequently, the entire process is quickly smothered without any major expansion outside of the stimulation point. However, if the threshold is reached, the explosive entrance of sodium can no longer be stopped and the shape of the voltage change no longer depends on the size of the initial depolarization (one match will ignite the same fire as two or three matches). This is explained by the **positive feedback** effect of entry of sodium ions into the cytoplasm. Specifically, the entrance of sodium, as a positive ion, increases the initial depolarization, which increases opening of sodium gates, because they are opened by the depolarization stimulus. This in turn further enhances the entry of new sodium ions and so on, until all sodium gates are opened. This process occurs regardless of the initial excitation, as long as it is above the firing threshold. Thus, the firing threshold is the spark that is sufficient to light a fire, which then burns itself down.




Specifics of the individual channels: sodium channels have two gates: *activation* and *inactivation* gate. The former are opened by depolarization, and latter, are closed by depolarization with a delay. Thus, depolarization opens sodium channel only briefly (until the inactivation gate responds). Therefore, for these reasons (and additionally because of opening of potassium channels), sodium does not reach its equilibrium potential of, for example +60 mV, but the amplitude of the action potential is lower, about +40 mV (these values are specific to individual neurons). Potassium channels have only one gate, which open as a result of depolarization at a time when sodium channels begin to close. Different are the mechanisms of passage through the channels: sodium ions must dehydratize and potassium ions do not. Details of the mechanism by which the depolarization stimulus changes the conformation of sodium and potassium channels are not known. Initially it was thought that the mediator is binding of some molecules, ligands, but this theory is rejected.

Repolarization phase, before returning to its original state, contains **hyperpolarization** phase, during which the cell is more negative, more polarized than in unstimulated condition. The cause is the continuation of potassium ion efflux even after reaching the equilibrium potential of unstimulated state, until all of potassium gates become completely closed, i.e., membrane permeability for potassium ions returns to unstimulated condition, when all the conditions of cable equation and Goldman-Katze equation are met.

An action potential will occur after only one, sufficiently strong, but short-term stimulation. Several successive excitations, or a continuous excitation of sufficient intensity will cause a **series of action potentials**. However, another action potential can not be generated as long as the sodium channels (you cannot ignite a new fire in the fire that is still burning). This is called the **absolute refractory period**. During this time the nerve fiber is absolutely non-excitable regarding the emergence of a new action potential.

Even after this time, since the sodium gates are closed, a new impulse will hardly take place, until the potassium gates are closed as well, that is, until an equilibrium state is achieved. The reason is that the entrance of sodium ions, following the new excitation stimulus, must be so strong to overcome the competing efflux of potassium ions. This is called the **relative refractory period**. During this time the nerve fiber can be excited only by the sufficiently strong stimuli.

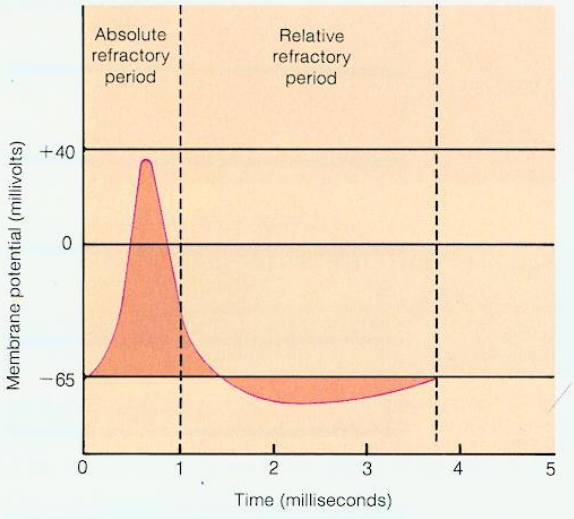



**Figure 14.14** The absolute and relative refractory periods. While a segment of axon is producing an action potential, the membrane is absolutely or relatively resistant (refractory) to further stimulation.

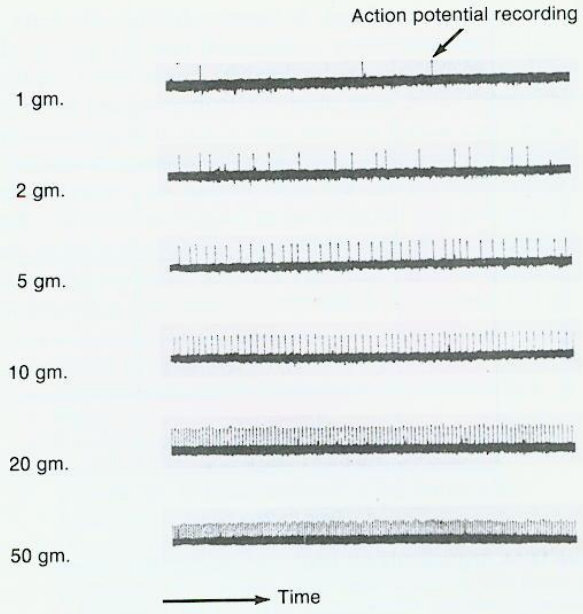



**Figure 14.13** Recordings from a single sensory fiber of the sciatic nerve of a frog stimulated by varying degrees of stretch of the gastrocnemius muscle. Note that increasing amounts of stretch (indicated by increasing weights attached to the muscle) result in an increased frequency of action potentials.

It is very important to note that increasing, continuous stimuli will induce increasingly earlier firing of the new action potentials, up until the time determined by the upper limit of absolute refractoriness. This way enables that stimulation intensity is coded by the frequency of action potentials.

Namely, the main function of the nervous impulse is transmission of information (from the sensory cell to the brain, or from central nervous system to the target organ), which encrypts: 1. the location of stimulation, and 2. the intensity of stimulation. The amplitude of action potentials in a certain nerve fiber cannot be changed. It is determined by the density of sodium and potassium channels and is characteristic of the nerve fiber. Any stimulus that is above the threshold causes action potential of the **same** amplitude, regardless of its intensity. Thus, the neurons cannot code the intensity of stimulus (degree of depolarization) by the amplitude of action potentials. However, during the progressively stronger stimuli, the series of action potentials will be generated with increasingly shorter intervals between each impulse, i.e. with increasing rate (frequency). Therefore, **the stimulus intensity is coded by the neuron as the rate (frequency) of the series of action potentials** that lasts as long as the stimulation is present. In doing so, the neuron

cannot distinguish between the continuous excitation and a series of pulses at intervals much shorter than action potential duration. In other words, short breaks in the excitation have no effect.

FOUR FEATURES OF THE ACTION POTENTIAL

1. **existence of a threshold**: if depolarization is insufficient, the influx of sodium ions through voltage-dependent channels is compensated for by the efflux of potassium ions (through leak channels and voltage-dependent potassium channels which first begin to open).

2. **constancy of the action potential amplitude**, i.e. its independence on the depolarization magnitude (above the threshold), is explained by the positive feedback effect, i.e., self-reinforcing effect of entry of sodium into the cell

1. and 2. mean that the action potential is "**all or nothing**" event

3. **propagating without reducing the amplitude** (self-renewal)

4. existence of **insensitivity to subsequent stimuli** until "things do not go back into order."

The transport of sodium and potassium ions during a single action potential significantly changes the membrane potential of a neuron, whereas the changes in the cellular content of these ions are negligible. However, after many action potentials are fired in a neuron, the Na-K pump reestablishes 'the order in house'.

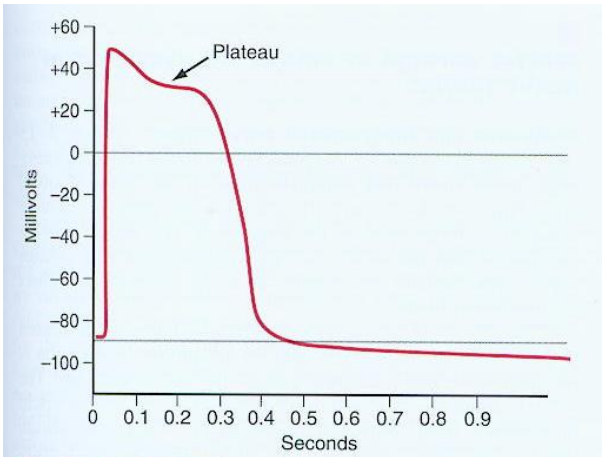



FIGURE 5-13

Action potential (in millivolts) from a Purkinje fiber of the heart, showing a "plateau."

The existence of a plateau in some neurons (cardiac muscle) is explained by the existence of slow calcium channels (through which sodium ions can also enter) and greater delays in potassium channel opening.

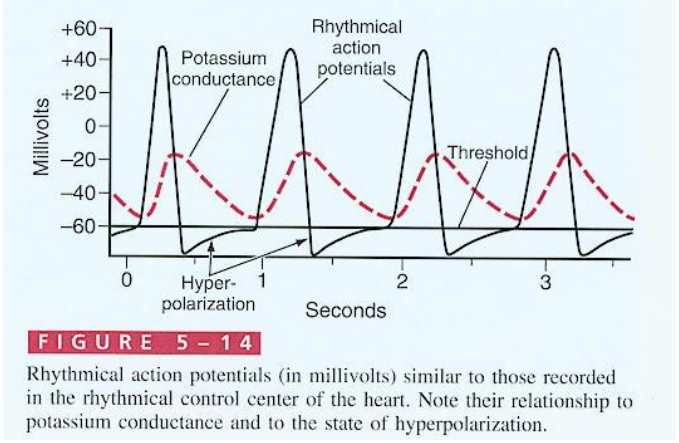



FIGURE 5-14

Rhythmical action potentials (in millivolts) similar to those recorded in the rhythmical control center of the heart. Note their relationship to potassium conductance and to the state of hyperpolarization.

Rhythmic self-firing of nerve signals in some tissues is explained by the relatively low polarization of neurons, just on the verge of opening of sodium channels.

## CONDUCTION OF ACTION POTENTIALS

Flux of ions that locally occurs at the site of an action potential formation is sufficient to cause depolarization stimulus in their environment. In this way, once generated form of voltage change propagates, without losses, along the whole axon in both directions from excitation site. Changes are propagating like a wave, i.e. an action potential does not travel, but is constantly restoring itself along the axon. In fact, by propagating the depolarization excitation, the changes of membrane permeability to sodium and potassium ions are also propagating. In doing so, ions travel in loops along the axon from the outside, through the membrane on its inner side. Propagation velocity of the nervous signal depends on the type of neurons and thickness of nerve fibers.

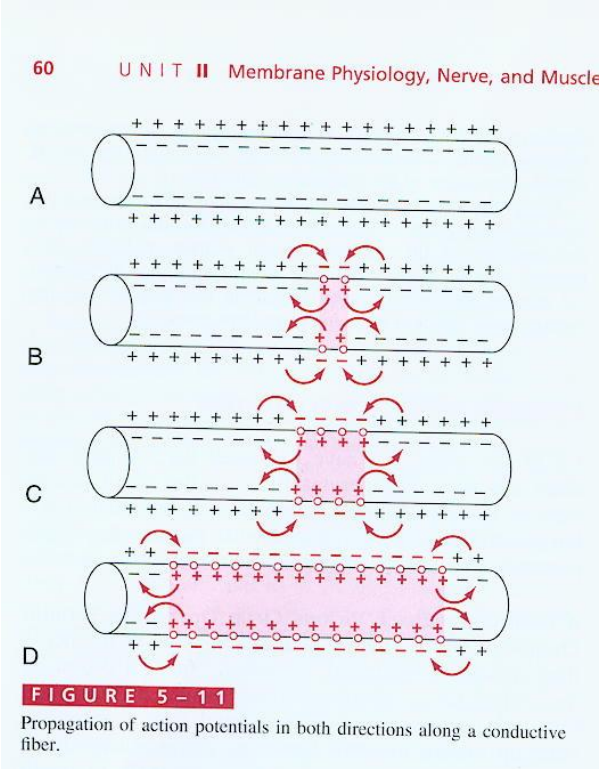
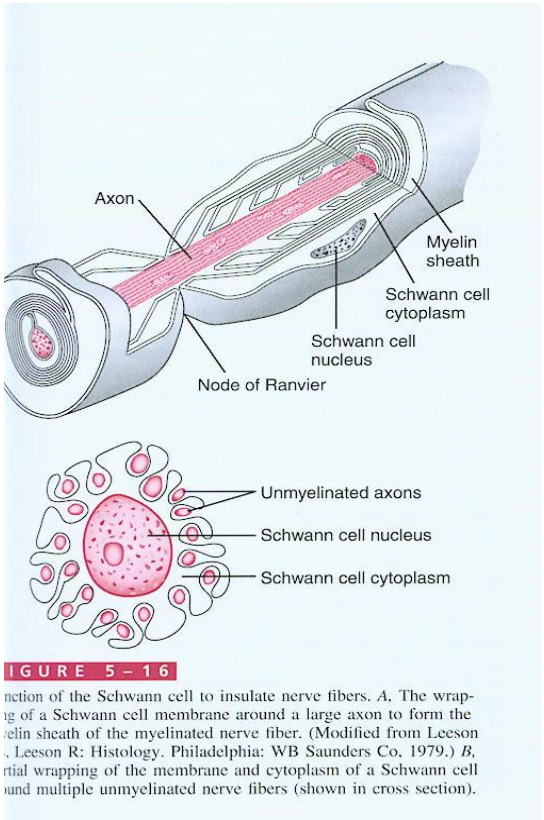


FIGURE 5–11
Propagation of action potentials in both directions along a conductive fiber.

Morphologically and functionally we distinguish two types of nerve fibers. **Myelinated nerve fibers** are wrapped with several layers of cells that act as insulators. These cells are Schwan cells in peripheral and oligodendrocytes in the central nervous system, respectively. **Nonmyelinated nerve fibers** do not have such an envelope. In the body the majority of fibers are myelinated, nonmyelinated are thin (<3 μm), for example postganglionic autonomic fibers. Myelinated nerve fibers conduct the nerve signal fast, but take up more space. So some nerves in our bodies often contain both types of fibers (sometimes even individual fibers are mixed or only partially myelinated). This is especially important in the visual and auditory nerve, so that the information is quick (through myelinated fiber), and but also spatially well-defined (by means of thin fibers nonmyelinated fibers).

Action potential in nonmyelinated fibers propagates **self-renewingly** and **continuously** along the fiber. In contrast, propagation of the signal in myelinated fibers is **stepwise** (**saltatory**), where the signal is regenerated only in small, non-insulated

parts of the axon, the so-called **Ranvier nodes**. This ensures a much higher (~50 times) speed of conduction, up to 120 m/s.



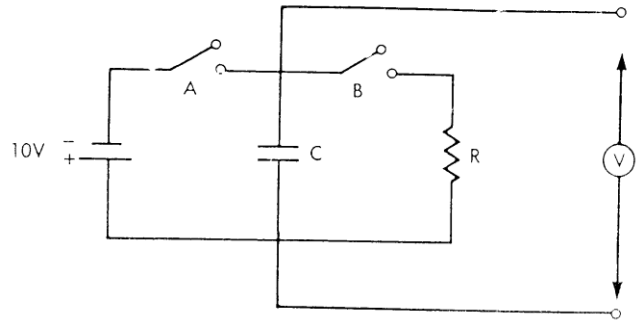

IGURE 5–16

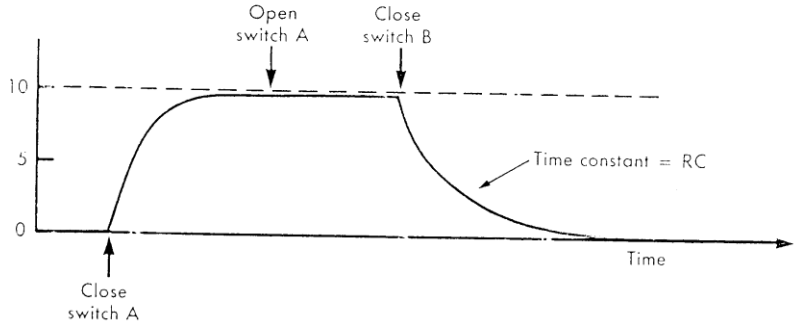
nction of the Schwann cell to insulate nerve fibers. *A*, The wrap-
g of a Schwann cell membrane around a large axon to form the
elin sheath of the myelinated nerve fiber. (Modified from Leeson
, Leeson R: Histology. Philadelphia: WB Saunders Co, 1979.) *B*,
rtial wrapping of the membrane and cytoplasm of a Schwann cell
und multiple unmyelinated nerve fibers (shown in cross section).

**Propagation velocity of action potentials in nonmyelinated fibers (axons)**

First let's note that the axon membrane acts as a capacitor or as two mutually isolated surfaces with stored opposite charges. However, ions can still transfer from one surface to another through ion channels. Therefore, each ion can be either stopped at the membrane, or pass through the ion channel. This is modeled by the **parallel connection** of capacitors and resistors. Furthermore, the existence of membrane potential is modeled by connecting the capacitor to an external power source, as shown in figure.

We are interested in what happens when a charged capacitor with capacitance **C** is allowed to discharge through the resistor **R**. This will tell us how fast the membrane potential is changed in certain location in the neuron, depending on the capacitance of the membrane and the resistance which ions need to overcome on their way. Given that the capacitance of the capacitor is the electric charge that can be stored for a unit change in voltage, we expect that the speed of the voltage decrease will be inverse to the size of capacitance and resistance , i.e. the fewer charges need to be transported and smaller the resistance to this transport, the faster the voltage change.




By mathematical analysis, we obtain that voltage of the capacitor connected in parallel with a resistor, is decreased in time from initial value **U** based on the relation:

$$\mathbf{U}(t) = \mathbf{U} \exp(-t/RC)$$

The **RC** product is called a **time constant**. The smaller it is, the faster the voltage drops in the capacitor (which discharges through the resistor). After the RC time, the voltage amplitude falls by a factor 1/e.

To be closer to reality, notice that in the loops in which ions are passing, extracellular ohmic resistance $\mathbf{R_e}$, intracellular ohmic resistance $\mathbf{R_i}$ and the membrane resistance **Rm** needs to be overcome. This is illustrated by the model as shown (the equivalent electrical model of electrotonic conduction).

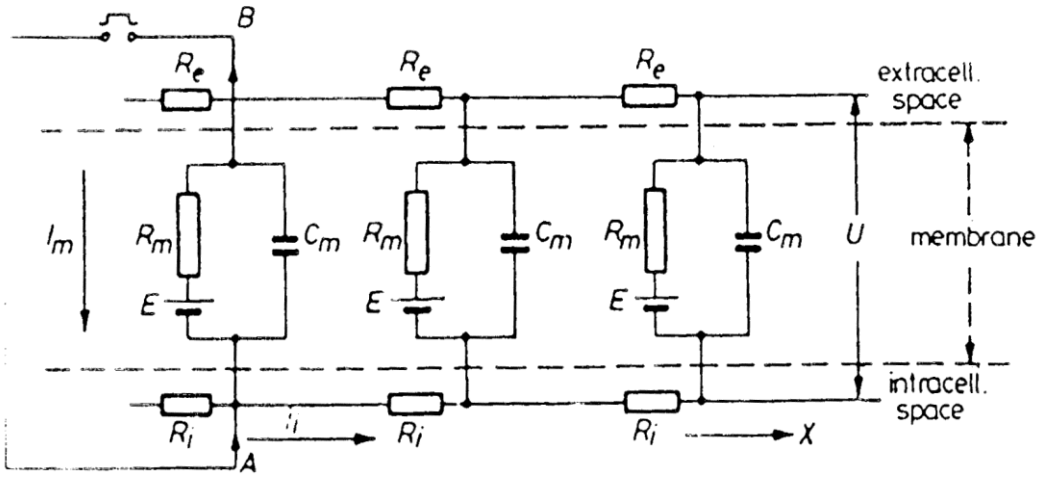



*Fig. 6.5.* Schematic circuit of the electric model of the cell for the interpretation of the effect of a pulse on the membrane

*A* and *B*: exciting electrodes

Compared with the passage of ions through the cytoplasm, where there are more macromolecules, the extracellular resistance is negligible. The analysis of the equivalent electrical model of nerve fibers we obtain that the total (effective) ohmic resistance is the geometric mean of $\mathbf{R_m}$ and $\mathbf{R_i}$:

$$\mathbf{R} = \sqrt{\mathbf{R_m R_i}}$$

Therefore, the time constant is $\mathbf{C}\sqrt{\mathbf{R_m R_i}}$ . It is a measure of the speed at which the membrane responds to stimuli. The smaller it is, the faster the membrane becomes polarized and action potential appears faster along the axon, i.e., the conduction velocity is greater.

Besides the time constant, another parameter determines propagation velocity of action potentials. The **spatial constant** describes the loss of ions during action potential propagation along the axon. Action potential will spread faster the smaller the transmembrane loss of ions, and this will be the larger the ratio of membrane and cytoplasmic resistance $\mathbf{R_m/R_i}$ is. The spatial constant of an axon is $\sqrt{\mathbf{R_m/R_i}}$ . From electric model it follows that it is a distance after which the initial amplitude of the stimulus is decreased by a factor 1/e. If the spatial constant is large, the stimulus that originated in one place of an axon extends far away. Because of this, the remote area of the axon will be more quickly excitated above the threshold, i.e., the action potential will travel more quickly.

*Thus, the smaller the neuron's time constant and the larger its spatial constant, the faster the velocity of action potential.*

The thicker nonmyelinated neurons conduct action potential faster than the thinner ones. This is because their time constant is smaller and spatial constant higher. Their time constant is lower, despite the fact that their capacitance is greater (it increases with the thickness of the axon, **D**). This is because the relatively greater reduction in the effective resistance to ion current prevails (it drops with $\mathbf{D}^{3/2}$). Overall, assuming that the conduction velocity (**v**) is inversely proportional to the axon time constant, the following appears:

$$\mathbf{v} \sim \sqrt{\mathbf{D}}$$

The outline of the derivation is shown in the next sketch:

Thicker axon has smaller RC and greater Rm/Ri

## Derivation:

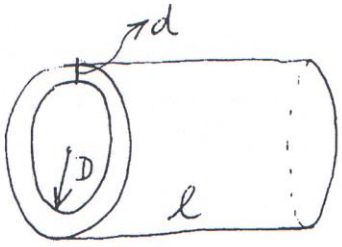
Physical background

- Ohm's law $\left(R = \rho \frac{\ell}{S}\right)$
- $C =$ dielectric constant $(\varepsilon) * \frac{\text{Area (S)}}{\text{Distance between plates (d)}}$
- Equivalent current model of electrotonic conduction

$$R = \sqrt{R_m \cdot R_i}$$

$R \cdot C_m$ — temp. constant

$\sqrt{R_m / R_i}$ — space constant




$d \ll D$

$$R_m = \rho_m \frac{d}{2\pi D \cdot \ell}$$

$$R_i = \rho_P \frac{\ell}{D^2 \pi}$$

$$C_m = \varepsilon \frac{2\pi D \cdot \ell}{d}$$

$$\Rightarrow R = \frac{1}{\pi\sqrt{2}} \sqrt{\rho_m \rho_P \cdot d} \cdot \frac{1}{D^{3/2}} \Rightarrow R \cdot C_m \sim \frac{1}{\sqrt{D}}$$

$$v \sim \frac{1}{R \cdot C_m} \sim \sqrt{D}$$

$v$: conduction velocity of action potential

$\equiv$ In unmyelinated nerve, the conduction velocity of action potential increases with the square root of the axonal thickness (despite increasing Cm, because it surpasses relatively larger decrease of R!!!

Along the axon, at certain distance from excitation site, an action potential is triggered when a change of surface charge density causes depolarization above the threshold. This means that, except for the time and spatial constants, the conduction speed of nerve signal also depends on the firing threshold, and on how fast and intense the depolarization phase is. Specifically, adjacent area will be excited sooner, the more ions there are and the sooner they are moved.
In summary, the conduction velocity in non-myelinated nerve fiber increases with following:

- lower firing (triggering) threshold (higher density of $Na^+$ ion channels)
- lower membrane capacitance (smaller change of charge for a unit change in voltage)
- lower resistance to ion current
- larger ratio of transmembrane resistance to the resistances to ionic current along the membrane (less ions are lost through the membrane)
- locally, a shorter duration of action potential, and
- if the action potential amplitude is higher (more ions start at the same time)

**Conduction velocity in myelinated nerve fibers**
Every part of myelinated axon can be seen as composed of 2 **n** capacitors with capacitance **C** that are connected in series, where **n** is the number of coils (each coil contributes with two membranes), connected to the voltage **U**, which is parallel connected to the resistor 2 **n** times the size of individual membrane resistance. The voltage on each capacitor is **U**/2**n**, and the total capacitance **C**/2**n**.

Along the "insulated" part of myelinated axon, action potentials are not triggered because depolarization is divided into **2n** small depolarizations that are insufficient to reach the threshold, and, in addition, the ion channel density is large only at Ranvier nodes. This means that, along almost the entire length of myelinated axon, ions travel similar to electrons in an electric conductor cable. Although the transmembrane losses are reduced by multiple increases of the membrane resistance, in this way the signal cannot travel without being restored for more than a few millimeters. Because of that, the signal is restored at the densely spaced, very short intervals, where the axon is without the myelin sheath. It seems as if the signal jumps from one node to another. Saltatory conduction is much faster because it does not require a critical amount of charge for firing along insulated (almost the whole) axon. The other two reasons are: decreasing time and increasing spatial constants of the axon.

In myelinated fiber the conduction velocity of the nerve signal increases linearly with the increase in axon thickness. Empirical formula is as follows:

$$\textbf{velocity (m/s)} = 6 \times \textbf{thickness } (\mu\text{m})$$

The formula is valid for $\mathbf{D} > 3\ \mu$m. For very thin axons ($\mathbf{D} < 1\ \mu$m) myelination would not increase the speed of pulse propagation, because then the resistance to cytoplasmic ionic current greatly increases. Because of this, these fibers are not myelinated.

One of chronic complications of diabetes is neuropathy. It occurs because of hyperosmotically-mediated damage of Schwann sheaths of peripheral nerves. In multiple sclerosis, glia cells of central nervous system are damaged.

**SELF-ASSESSMENT**

1. What is the change in transmembrane potential of a cell if extracellular potassium concentration increases from 5.5 to10 mmol/l? Assume that in both conditions potassium ions establish the steady state. Explain why you can assume that intracellular potassium concentration is the same in both conditions. Is it necessary to know the value of intracellular potassium concentration to answer the above question?

2. Explain why large intake (or loss) of electrolytes is: 1. confined to extracellular compartment; 2. still may have major impact on function of some cells (which cells?)?

3. The factor that DOES NOT contribute to the resting transmembrane potential is:

a) fixed intracellular anions
b) larger membrane conductivity for potassium than for sodium ions
c) existence of voltage-gated $Na^+$ channels
d) capacitor characteristics of neuron membrane
e) action of Na-K pump

4. In resting neuron the flow of potassium ions through leakage channels is greater than flow of sodium ions (A) *because*- at rest the neuron membrane is more permeable to potassium than for sodium ions (B). The true statement is:

a) A correct, B correct; A and B associated
b) A correct, B correct; A and B unassociated
c) A correct, B incorrect
d) A incorrect, B correct
e) A and B incorrect

5. In resting neuron membrane, the conductivity of sodium ions is 50 times less than of potassium ions. This is valid if extracellular potassium concentration is normal (around 5.5 mmol/l). It is experimentally determined that in hyperkalemia the transmembrane potential of neuron approaches the equilibrium potential for potassium. What can one conclude from these facts regarding the effect of hyperkalemia on the ratio of membrane conductances for potassium and sodium?

6. The local response voltage maximum decreases to 1/e of the initial value at 3 mm from the point of excitation in axon 1, and at 6 mm in axon 2. Which is the ratio of space constants of two neurons? Assuming that both axons have the same inner resistance to ionic current (per unit length), answer which axon has the membrane of greater resistance to ionic current (per unit length).

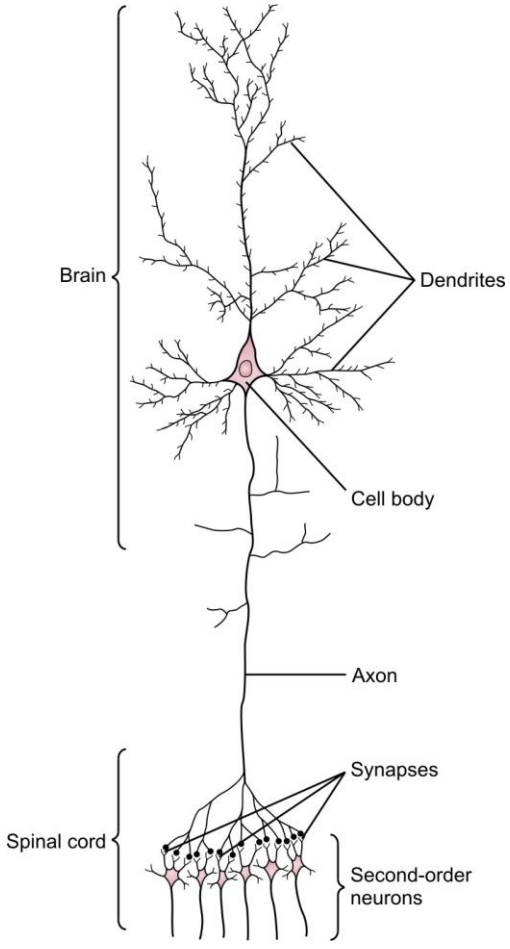
7. Two non-myelinated neuron fibers differ only in one being twice thicker. Which fiber transmits the action potential faster, which is the ratio of their space and time constants?

8. Myelinization increases the action potential amplitude (A) *because* Schwan cell wrapping decreases neuron RC constant (B)
a) A correct, B correct; A and B associated
b) A correct, B correct; A and B unassociated
c) A correct, B incorrect
d) A incorrect, B correct
e) A and B incorrect

# Chapter 4: BIOPHYSICS OF SENSORY FUNCTIONS

THE INTRODUCTION ON THE NERVOUS SYSTEM

**The central nervous system (CNS)** is primarily made of neurons in the brain and in the spinal cord. There are about 100 billion of those neurons. A typical motor cortex neuron (shown in the figure below) consists of a cell body (or **soma**) where many numerous branches (**dendrites**) are protruding from. There are between 200 and 200000 dendrites. Those dendrites are connected to other neurons which carry information to the brain, so called **afferent neurons** of the peripheral nervous system, or with other neurons in the CNS. Neurons do not touch each other directly, but are connected through a chemical communication throughout **synapses**. The outgoing information is passed on from the soma over the **axons** - the long neuronal branch of the nerve fiber. Only axonal part of the neuron has a cable function feature conducting action potential, since the other parts do not have sufficient density of sodium channels.



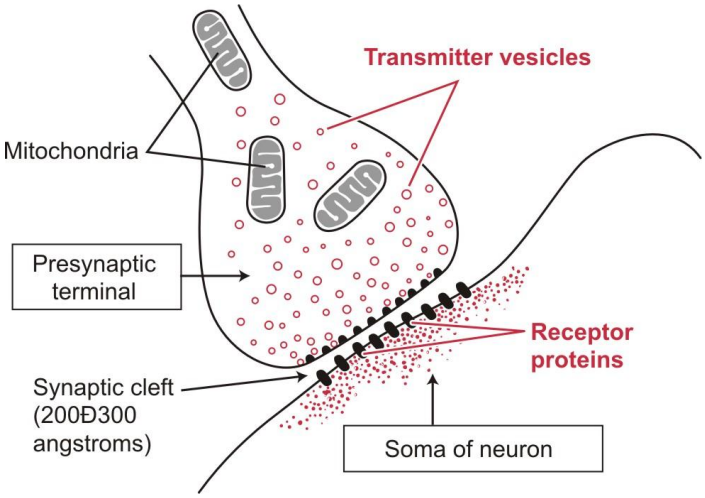

**synaptic potential.**

At the end of the nerve fiber, the axon branches into numerous spines of the dendrites making synaptic connections with other neurons in the CNS. The signals travel down the **peripheral efferent (descending) neurons** toward to the **'effectors' organs**: 1. skeletal muscles, 2. smooth muscles of internal organs, or 3. gland with the external or internal release of substances.

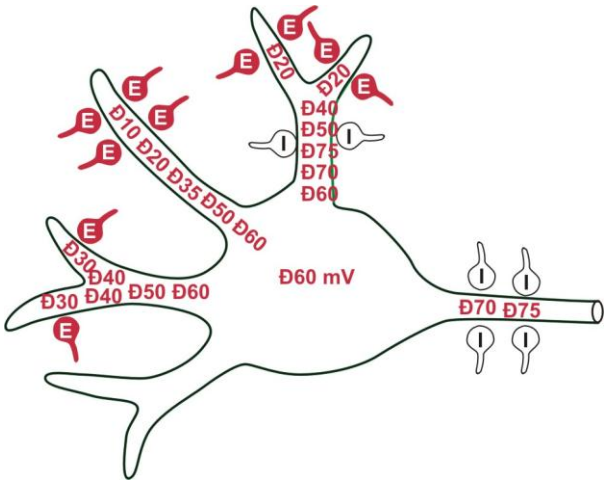
Most of the synapses in the CNS are of chemical origin. Such a synapse consists of axonal spines of **the pre-synaptic neuron**, **synaptic cleft** (200-300 angstroms in width) and dendritic spines of **the post-synaptic neurons**. Action potential at the end of axon spine causes the secretion of **neurotransmitters** in the synaptic cleft. These compounds diffuse to the membrane of the postsynaptic neuron, where they bind with protein receptors. The result of the binding can be opposite: for specific post-synaptic neurons, some neurotransmitters are facilitating, whereas some others are inhibitory. The difference is based on type of the binding: an opening of the certain ion channel causes local depolarization (**facilitation**) or hyper-polarization (**inhibition**), leading to either **facilitating** or **inhibitory post-**

Thus, the action potential signals are transmitted from one neuron to another only indirectly, through chemical substances. However, there is an exception - **electrical**

**synapses** – which are predominantly found in the smooth and cardiac muscle. The role of chemical synapses is much more complex and sophisticated.



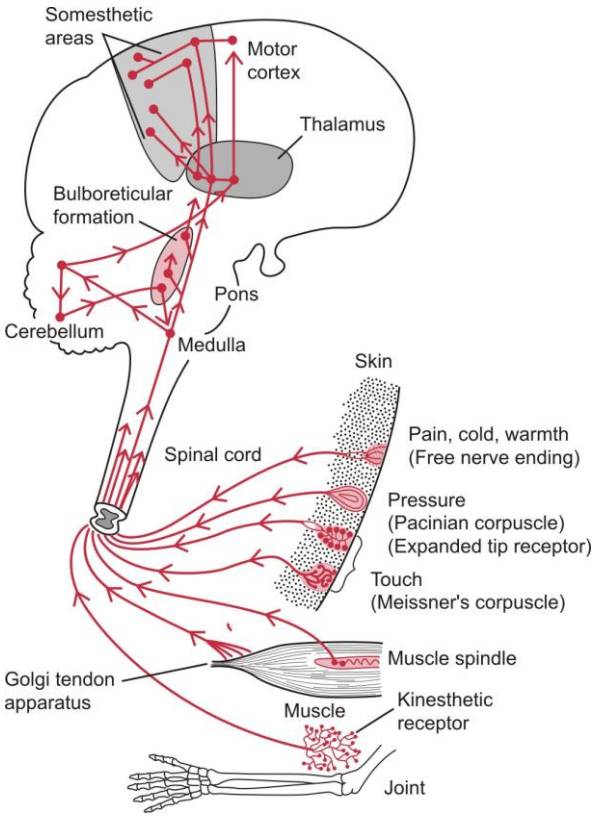

It is important to note that the axon branches of the pre-synaptic neurons are synaptically connected with the dendrites of the post-synaptic neurons. Both neurons have such connections with a lot more neurons in their neighborhood.This means that a single connection between two neurons is only one of the many contributions that elicits or inhibits the action potential in the post synaptic neurons. Individual contributions are either excitatory or inhibitory in the post-synaptic potentials. These potentials, seen as localized stimuli with decreasing intensity, are spreading out from dendrites to soma, adding up in time and space with many other signals from adjacent dendrites (and from soma as well) until summation signal reaches the axon - a place where the action potential is elicited.




The action signal will occur only if the depolarization of the axonal beginning is large enough. In this way, the synapses perform selection of action potentials, often blocking weak signals, while allowing strong ones to pass. In addition, they can redirect the incoming signal into several different directions.

THE SENSORY PART OF THE NERVOUS SYSTEM

As we are interested in the sensory system, the main principles and structures of registration, processing, transmission and interpretation of the external stimuli in the system will be considered: sensory cells, the primarily afferent neurons, neural networks and CNS. The sensory system transmits information from receptors located all over the skin and in some deeper structures, through the peripheral afferent nerves to all parts of the spinal cord and to certain parts of the brain. After that, **secondary signals** are transmitted from the brain to almost all parts of the nervous system in order to execute motor or mental response.




Sensory receptors selectively sample the small portions of energy from the environment, which are then used for the controlled creation of nerve signals. The appearance of these signals, together with the specific way in which information comes to the CNS makes the internal biological representation of specific components of the external world. **Sensation** is part of the complex processes of **perception**, which involves the integration of experience and comparison with other sensations, in order to evaluate/compare the quality, intensity and importance of sensations. In fact, one of the most important functions of the nervous system is to collect and interpret information in a manner that results in appropriate mental and motor responses. More than 99% of all sensory information is discarded by the brain as irrelevant.

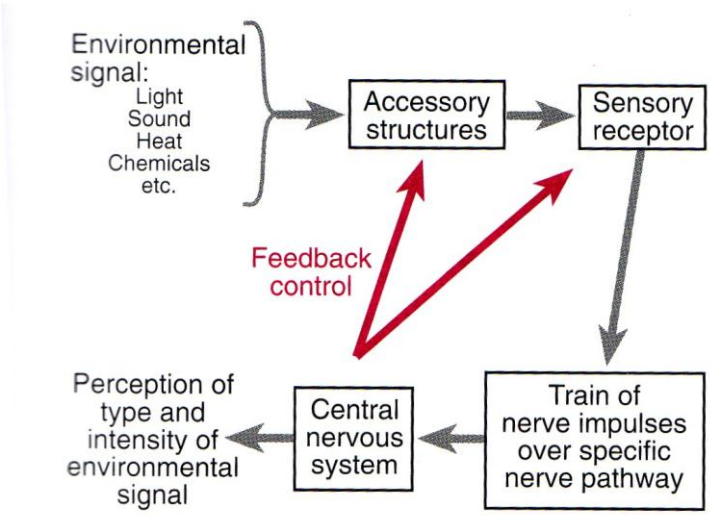
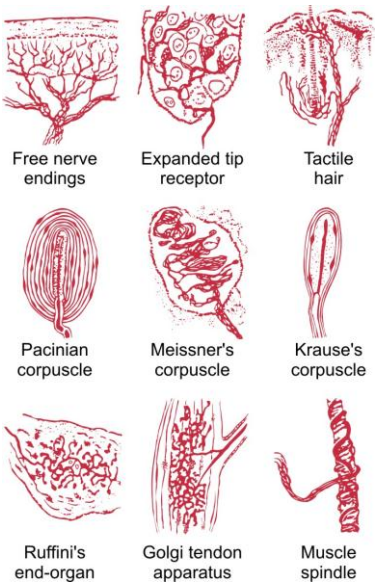



**Figure 4–1** ■ The steps involved in sensory transduction. While the details vary with each type of sensory process, the overall process is similar.

Excitation can be of the following types:

1) electromagnetic: heat, light;
2) mechanical: pressure, sound waves, vibrations;
3) chemical.

**Intensity** is a common dimension to all stimuli. It is a measure of energy (or the concentration of chemical excitation) which interacts with sensory receptors. Stimulus is based on **the sensory modality**, such as the traditional five senses: taste, smell, touch, sight and hearing. More complex stimuli (e.g. humidity) are a combination of primary ones (pressure and temperature).




**Photoreceptors** detect light, and they are the function of the senses of vision; **chemoreceptors** are in function of sensations of taste and smell, and control of arterial blood; **mechanoreceptors** sense physical deformity and participate in the sense of touch and hearing, as well as in the detection of stress in the muscles and tendons; **thermal receptors** detect thermal changes; **proprioceptors** are sensors that provide information about joint angle, muscle length, and tension position of the limb in space; **nociceptors** detect harmful substances, etc..Excitation is manifested at **the receptor cells.** These are either special cells (i.e. cells with hairs in the cochlea of the inner ear – hair cells) or non-myelinated endings of the afferent primary neurons (otherwise known as sensory or receptor neurons).

Each excitation causes the **receptor (generator, microphonic) potential**: a change in the resting potential of the neuron. This change is usually **depolarization**, but can be

**hyper-polarization**, or even both (for example, alternating in the case of hair cells in hearing).

The conversion of specific energy excitation in the receptors potential (i.e. mechanical, heat, light or chemical energy into electrical energy) is called **signal transduction**. It is based on **the change of membrane permeability**, either by mechanical deformation, binding of ligands, changing temperatures or absorption of the electromagnetic radiation.

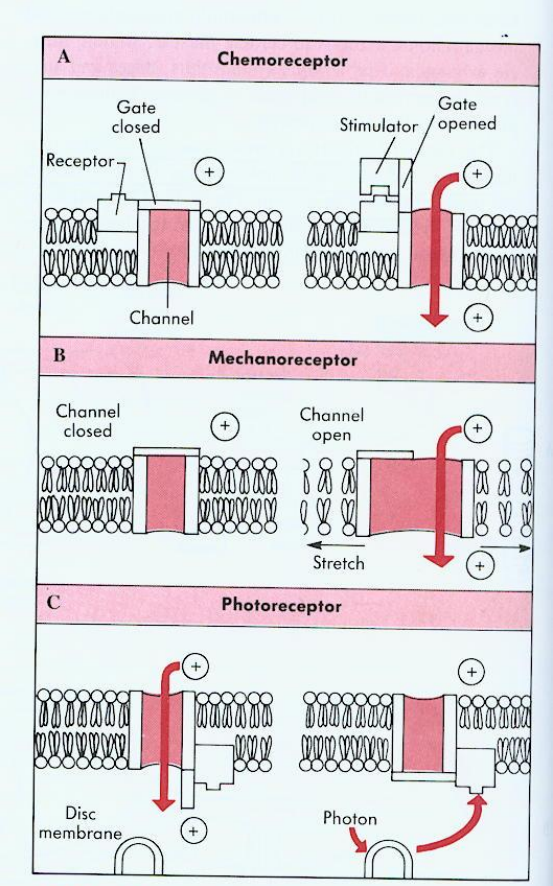



**Fig. 7-2** Conceptual models of transducer mechanisms in three types of receptors. **A,** Chemoreceptor; **B,** mechanoreceptor; **C,** vertebrate photoreceptor. (See text.)

If the receptor potential is large enough, the neuron triggers an action potential. If the receptor occupies only proximal part of the primary neurons, the receptor potential being depolarizing exceeds stimulus threshold leading to the action potential occurring in the first Ranvier's node of the axon. In the case when a receptor cell is a specific neuron, the receptor's potential will release neurotransmitters from the receptor cells into the synaptic cleft between two neurons. This release will lead to a depolarization of the post-synaptic neuron being above the threshold of excitability. This process of converting one form of electric energy into another is called **signal conversion**.

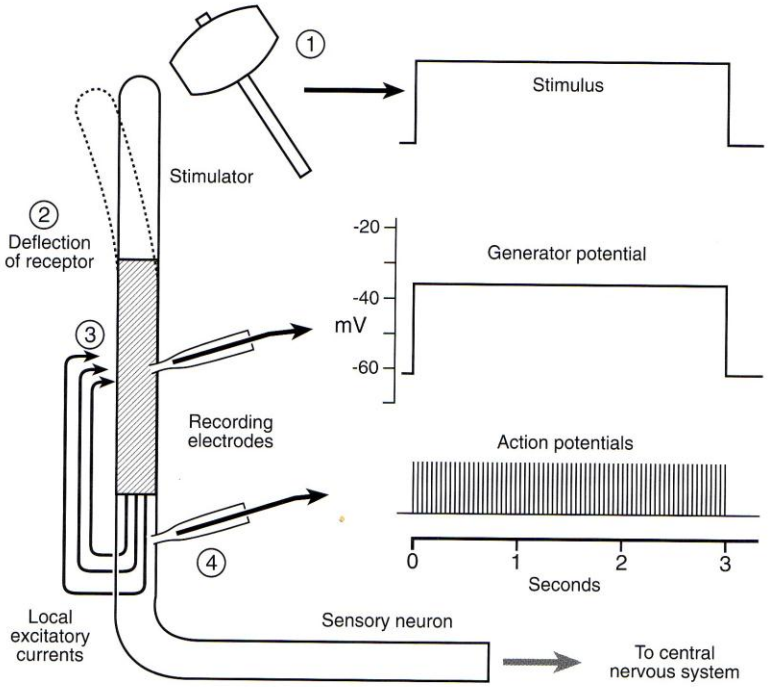



**Figure 4–2** ■ The relation between an applied stimulus and the production of sensory nerve action potentials. See text for details.

Regardless of the stimulation type, information starts to be transmitted as a series of action potentials in primary afferent neurons (which make synapses with the secondary, then those with tertiary, etc.). Thus, the stimulation type is recognized not by the type of response (which is always the same), but by **the specific place in the CNS** that is ultimately stimulated. We already learned that the amplitude of the action potential is independent of the intensity of the depolarizing stimulation, and that stronger stimulation induces more frequent action potentials (i.e. higher rate of action potentials). Both effects are related to the absolute and relative refractivity of neurons. In the end, **perception of the stimulation intensity** depends primarily on **the frequency of action potentials**, with a number of stimulated neurons playing additional role.

Therefore, the brain interprets the incoming signal as a sound if it is part of the auditory cortex, and its intensity is encoded by the frequency of incoming signals, i.e. the number of action potentials per unit of time. Sound frequency will be determined by the precise location of the signal stimulation in the inner ear. This fascinating mechanism will be described in more details later in the chapter *Physics of ear and hearing*.

The additional role of the number of stimulated neurons in the case of perception of pain is illustrated in the following second figure. This role will be described later in the case of hearing sensation.

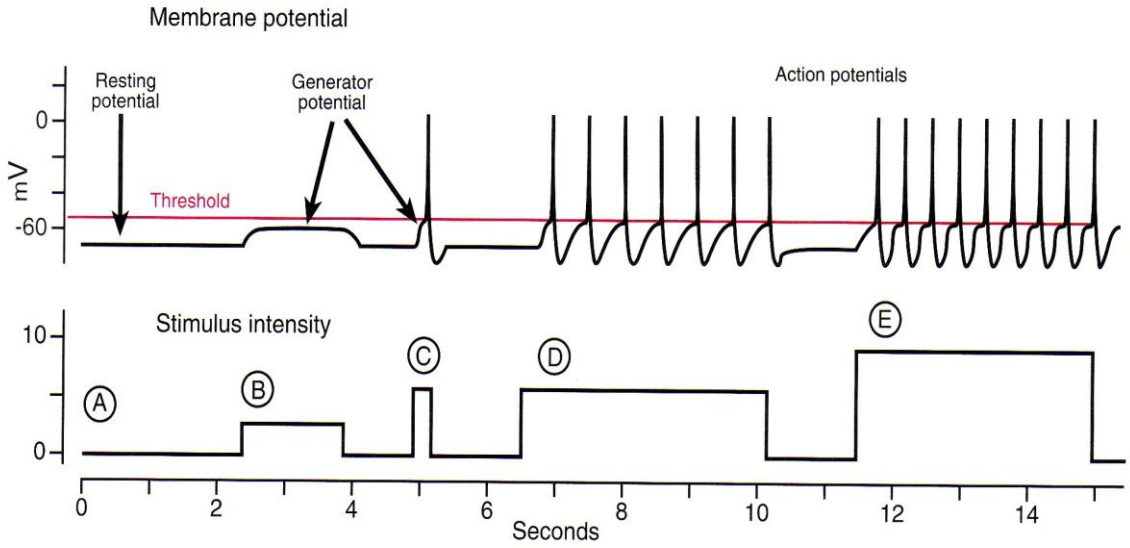



**Figure 4–3** ■ Sensory nerve activity with different stimulus intensities and durations. With no stimulus (A), the membrane is at rest. A subthreshold stimulus (B) produces a generator potential too small to cause membrane excitation. A brief but intense stimulus (C) can cause a single action potential. Maintaining this stimulus (D) leads to a train of action potentials. Increasing the stimulus intensity (E) leads to an increase in the action potential firing rate.

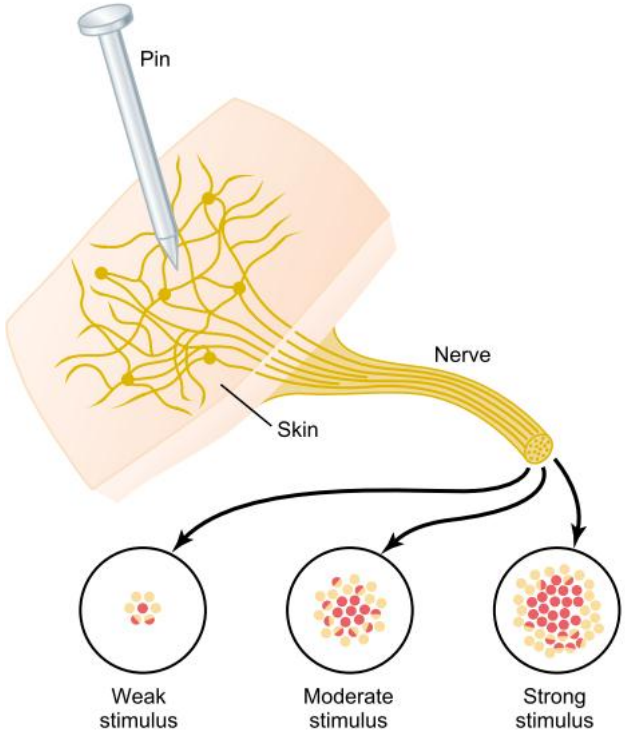



**Figure 46–7**

Pattern of stimulation of pain fibers in a nerve leading from an area of skin pricked by a pin. This is an example of *spatial summation*.

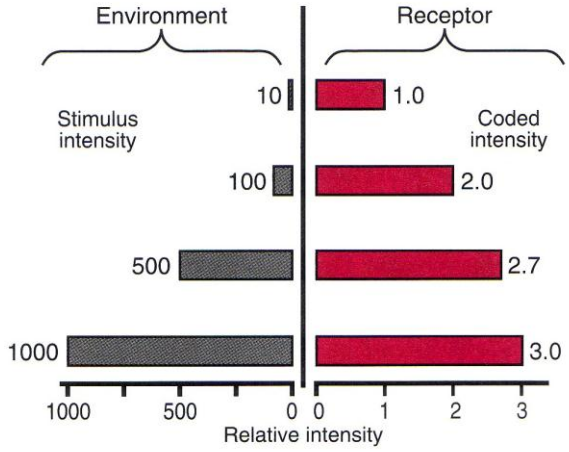



**Figure 4–5** ■ Compression in a sensory process. By a variety of means, a wide range of input intensities is coded into a much narrower range of biologic responses that can be represented by variations in action potential frequency.

The receptor potential grows with the stimulation intensity causing more frequent action potentials. However, the maximum of receptor potential is about 100mV. In addition, each axon is limited by maximal signal frequency that can be transmitted, given the absolute refractory period (a few hundred Hertz). Thus, increasing the intensity of the external stimulation initially will increase the strength of perception until saturation (when big differences in the intensity lead to relatively small differences in perception). In that case, the relationship between stimulation intensity and perception becomes sub-linear.

Limited magnitude of the receptor potential and limited triggering frequency for action potential signals in the axon are main reasons for the sub-linear relationship between stimulation intensity and the perceptual strength.

Thus, **signal compression** occurs at the level of transduction (very often) and signal conversion (always). This explains the enormous range of intensities in perception of stimulus features (1:10^12 in the auditory system, 1: 10^11 in the visual system). In nature, we are exposed to external stimuli at huge range of intensities. It was an evolutionary need that made us sensitive on even weak stimuli like the sound of sneaking predator or spotting prey in the dark. We had to develop a system that protects us from the strongest stimulation (younger generations will have problems in preserving hearing when regularly exposed to loud music). The price we pay is the inability to distinguish fine gradations of strong stimuli (which may be desirable, for example in the case of painful stimuli).

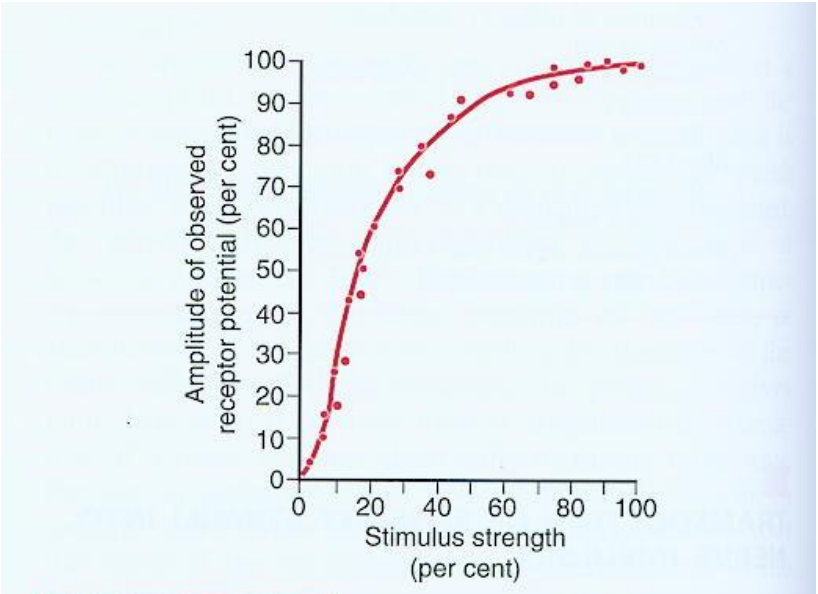
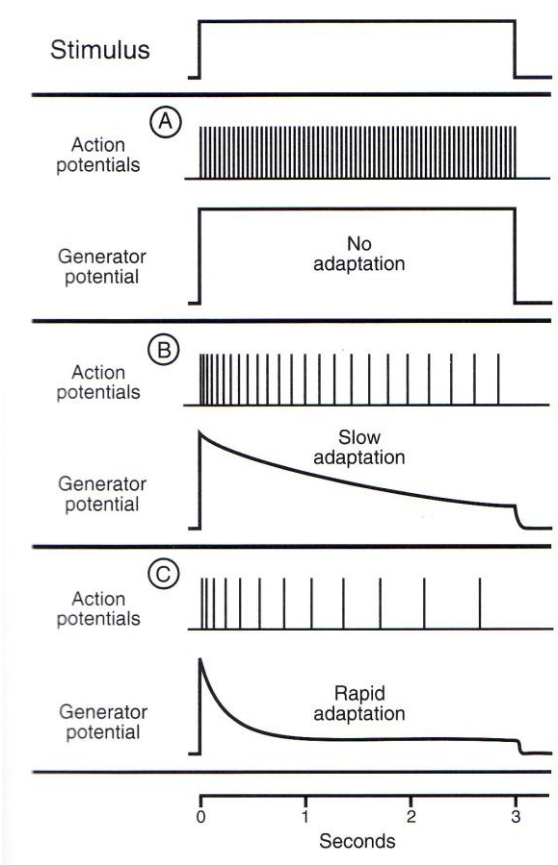



**FIGURE 46–4**

Relation of amplitude of receptor potential to strength of a mechanical stimulus applied to a pacinian corpuscle. (From Loëwenstein WR: Excitation and inactivation in a receptor membrane. Ann N Y Acad Sci 94:510, 1961.)




Generating potential of most of the sensory receptors exposed to the constant stimulus gradually declines over time leading to a reduction of the frequency of action potentials. Thus, the receptor is adjusting to continuous stimulus by reducing the sensitivity. This is called **adaptation** which can be fast or slow. The adaptation occurs at the level of transduction. It is usually caused by inactivation of sodium channels after extended period of depolarization, and can weaken fundamental capacity of receptors to produce the receptor potential (for example, light induced breakdown of photo-sensitive molecules in rods and cones of the eye), and it may influence the actions of supporting structures by the mechanism of negative feedback (for example, too much light closes the eye pupil). Adaptation can be fast or slow. Tonic receptors adapt just a little or not significantly at all.

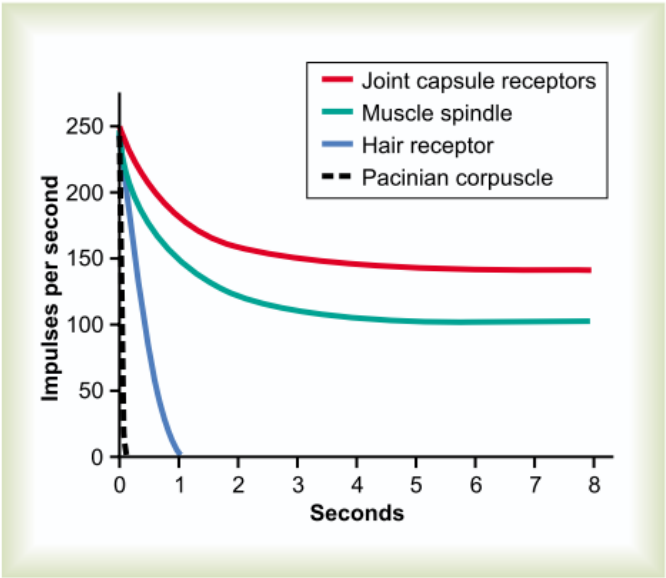



**Figure 46–5**

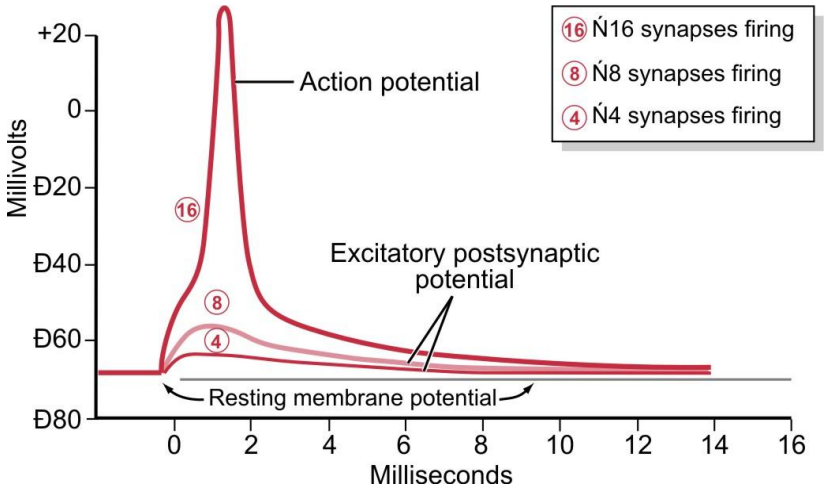
Adaptation of different types of receptors, showing rapid adaptation of some receptors and slow adaptation of others.

It is thought that adaptation occurs in order to: 1) reduce stimulus overload; 2) ignore continuous, but less important stimulation, and to have the ability of **perception of fast changes** (and not just the magnitude) of stimulation.

**Transmission and signal processing in neuronal networks**

We already know that the impulse that comes only from a single pre-synaptic terminal (axonal end of the pre-synaptic neuron) rarely causes action potential in the post-synaptic neuron. To be exact, the usual quantity of neurotransmitters released into the synaptic cleft causes only 0.5 to 1 mV of the local post-synaptic excitatory potential, while the 10-20 mV is needed to elicit above the threshold depolarization in remote, more sensitive part of the neuron (axonal beginning). However, two neurons are usually connected with several synapses, through which the impulse is often simultaneously transmitted. Although these terminals are spatially quite separated, its effect is multiplied by summation, especially due to the good ionic conductivity of some large neuronal soma. Depending on the number of excited synapses, post-synaptic neuron can be excited above or below the threshold.

The effect of addition of simultaneous post-synaptic potentials, achieved by excitation of mutually remote parts of the membranes of neurons is called **spatial summation**.




Stimulation from pre-synaptic terminal opening the membranes of the post-synaptic terminals lasts for only about 1 msec. However, duration of post-synaptic potentials is about fifteen times longer. Therefore, series of rapid stimulus repetition induce, by mutual superposition, significantly stronger post- synaptic potential compared to the isolated stimulation. So, repeatedly firing of one pre-synaptic terminal, if fast enough, can cause a summary response almost equal to the arithmetical sum of the components, which we call **temporal summation** of signal.

CNS consists of thousands of mutually connected neurons forming **neural networks**. Some of them are made of only a few neurons, while other networks can be made of

enormous number of neurons. The entire cortex, for example, can be considered as one large neural network. Each network has its own specific organizational characteristics, which are again part of a larger whole, in order to allow the entire system to jointly carry out various functions.

The picture shows a few neurons (left) which forward signals to neurons in another group (right) which are mutually connected. Individual neurons can provide hundreds of thousands of axon spines into a space that we call **axonal stimulatory field**. It has to be noted that neurons synapse mostly with their neighbors, but some may connect to remote neurons as well.

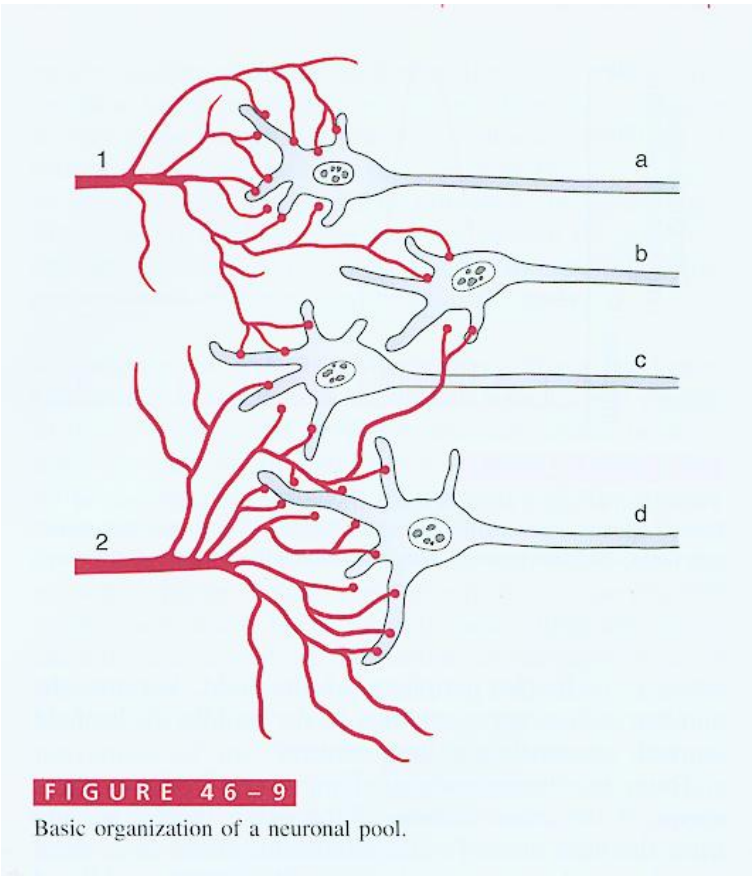



FIGURE 46 – 9
Basic organization of a neuronal pool.

Some pairs of neurons in the neural network can be connected in a way that the synaptic transmission of the signal (i) stimulate below the threshold (only a few synaptic connections), (ii) stimulate above the threshold (a lot of synapses), but also (iii) makes inhibition.

It is very often important that the signals in neural network stimulate far greater number of neurons than those who make the incoming signal. This phenomenon is called **divergence (branching)** of the signals. There are two types of divergence: (i) signal amplification in the same tract, and (ii) divergence in various tracts (directions). The effective control of the periphery by specific center can be achieved with signal divergence.

On the other hand, **convergence** allows signals from multiple neurons to flow (focus) into single neuron, stimulating it. In this way, spatial summation of weak signals can be achieved.

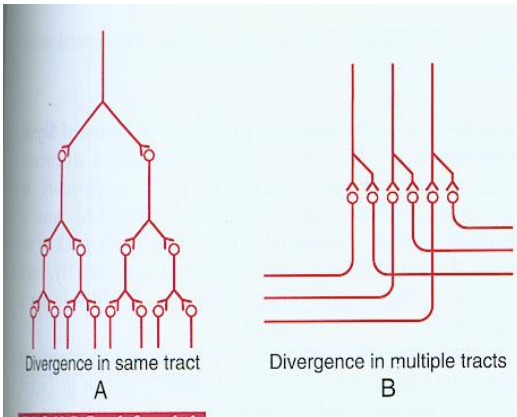



FIGURE 46-11

"Divergence" in neuronal pathways. *A*, Divergence within a pathway to cause "amplification" of the signal. *B*, Divergence into multiple tracts to transmit the signal to separate areas.

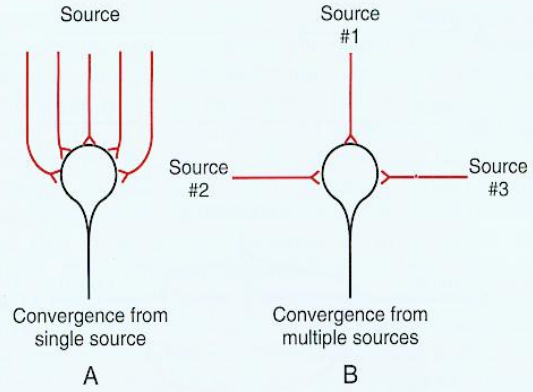



FIGURE 46-12

"Convergence" of multiple input fibers onto a single neuron. *A*, Multiple input fibers from a single source. *B*, Input fibers from multiple sources.

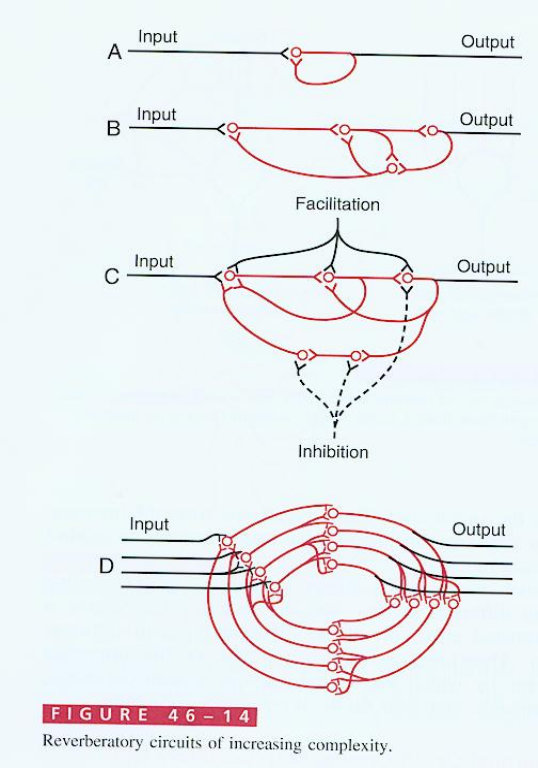



FIGURE 46-14

Reverberatory circuits of increasing complexity.

**Reverberation** is the principle of how to extend the duration of the input signal. It is achieved by repetitive self-stimulation of the neural network by a positive feedback effect. Similarly, the negative feedback loop ensures turning off.

Transduction and signal conversion have a meaning of **encoding** the signal into a series of action potentials, when the signal becomes compressed. In the final stage of processing of stimulation signal, the CNS performs **decoding** of the signal, which may produce complete or partial decompression.

**Decoding** is the conversion of a series of action potentials into step voltages, in one or more steps. But, there is little understanding of these processes.

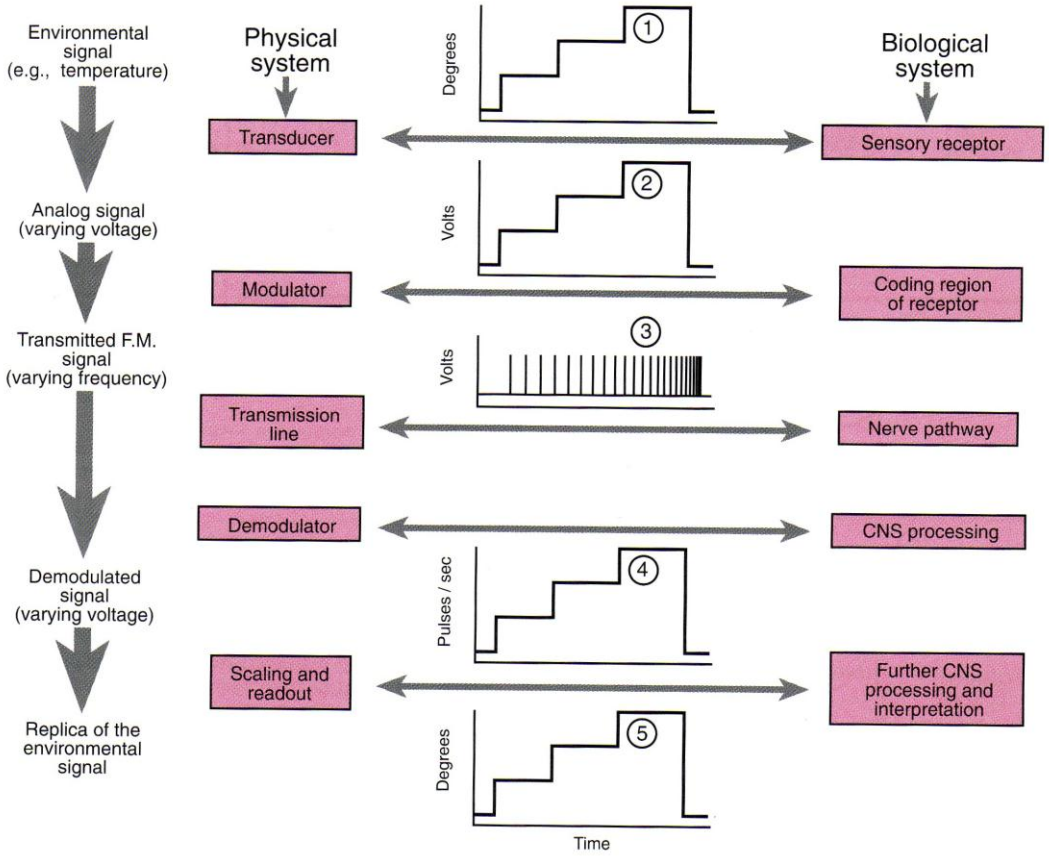
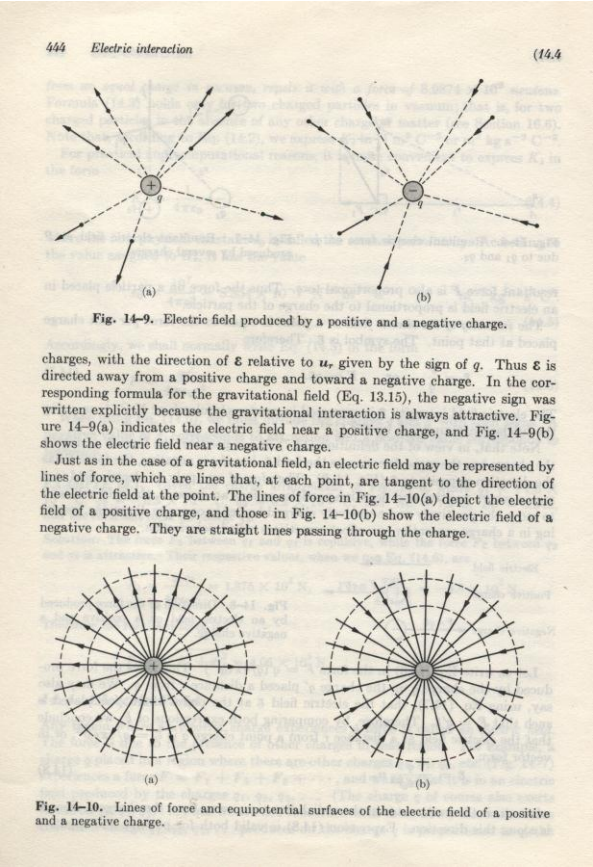



**Figure 4–6** ■ Transmission of sensory information. Because signals of varying amplitude cannot be transmitted along a nerve fiber, specific intensity information is transformed into a corresponding action potential frequency, and CNS processes decode the nerve activity into biologically useful information. The steps in the process are shown at the left, with the parts of a physical system that perform them. At the right are the analogous biologic steps involved in the same process.

# Chapter 5: POTENTIALS ON BODY SURFACE

A cell in a resting state is polarized, and has a strong electric field inside the membrane. In the cell, as well as outside, the electric field does not exist, i.e. all loci have the same potential (which is negative inside the cell, and zero outside of it). However, generation of action potentials due to action of cardiac and skeletal muscles, as well as brain activity, results with a formation of electric fields and the motion of ions throughout the extracellular space. Ions move because they are influenced by electrical force. This electrical force is caused by a multitude of dipoles, the separated charges of equal amount and opposite charge. The vectorial sum of these dipoles determines the total dipole momentum. The associated electric potential and electric field, modified by the presence of a conductive medium, can be detected on body surface.

ELECTRIC DIPOLE



444 *Electric interaction* (14.4

**Fig. 14–9.** Electric field produced by a positive and a negative charge.

charges, with the direction of $\mathcal{E}$ relative to $u_r$ given by the sign of $q$. Thus $\mathcal{E}$ is directed away from a positive charge and toward a negative charge. In the corresponding formula for the gravitational field (Eq. 13.15), the negative sign was written explicitly because the gravitational interaction is always attractive. Figure 14–9(a) indicates the electric field near a positive charge, and Fig. 14–9(b) shows the electric field near a negative charge.

Just as in the case of a gravitational field, an electric field may be represented by lines of force, which are lines that, at each point, are tangent to the direction of the electric field at the point. The lines of force in Fig. 14–10(a) depict the electric field of a positive charge, and those in Fig. 14–10(b) show the electric field of a negative charge. They are straight lines passing through the charge.

**Fig. 14–10.** Lines of force and equipotential surfaces of the electric field of a positive and a negative charge.

Let us recall first what the electric field and a corresponding electric potential of a single charge are. According to **Coulomb's law**, charges of the same sign repel each other, and the opposites attract. This force lies along the line that passes through charges; it is proportional to the product of charges ($q_1, q_2$) and inversely proportional to the square of their distance ($\sim 1/r^2$).

A **field** of the source of the electrical charge is described by its action on a single positive charge: the force - **field strength** ($\sim 1/r^2$), potential energy - **the electric potential** ($\sim 1 / r$).

Let us also recall that vector is a directed length, determined by magnitude and direction. The sum of two vectors is a diagonal, where the two vectors make the sides, and a scalar product of two vectors is a number (not a vector!), that is a product of the length of one vector and the projection of the other vector on its direction.

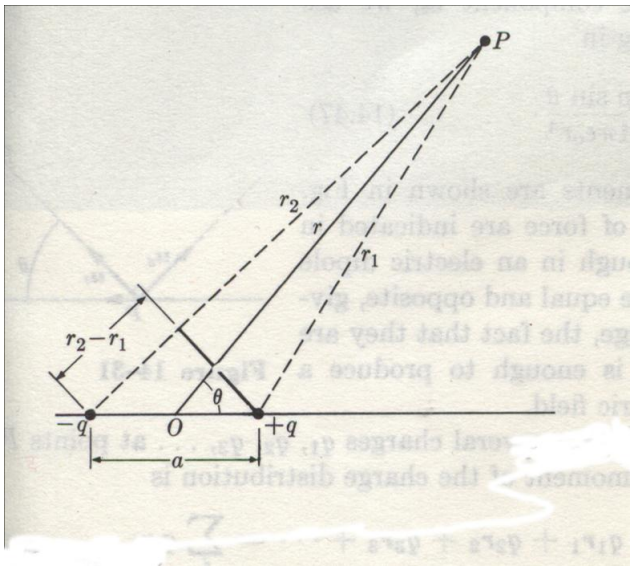
**Electric dipole** consists of two equal, mutually spaced opposite charges, **+q** and **-q**. It is described by an **electric dipole momentum**: a vector directed from the negative to the positive end, with the amount (**p**) that is equal to the product of charge **q** and the distance between them **a**:

$$\mathbf{p = q\,a}$$

Thus, an electric dipole is the stronger the greater the charges and the more spaced they are.

Electric potential of a dipole equals the algebraic sum of the potentials of both charges, while the electric field of a dipole is a vector sum of the fields of individual charges. A relatively simple trigonometric analysis shows that, at distances that are much larger than the dipole (**r>> a**), the electric potential of the dipole is proportional to a scalar product of electric dipole momentum and a vector of the position of the point of interest:

$$\mathbf{V}(\vec{\mathbf{r}}) \div \vec{\mathbf{p}}\,\vec{\mathbf{r}}$$

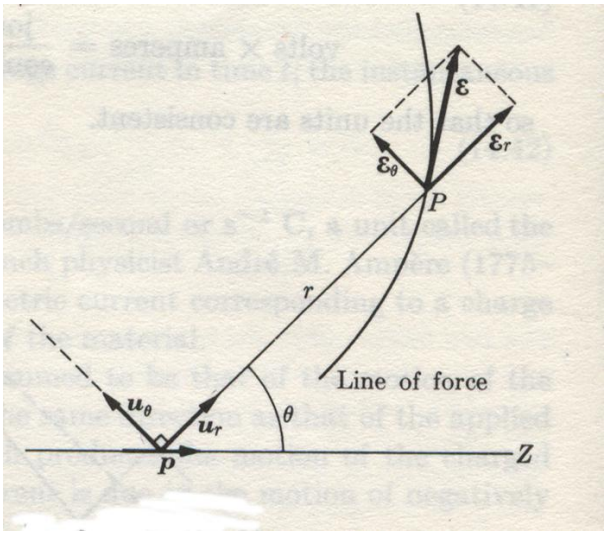



If **θ** marks the angle between the dipole vector and the position vector of the point of interest, and $\varepsilon_0$ the absolute permittivity of vacuum, a complete expression is:

$$\mathbf{V}(\vec{\mathbf{r}}) = \frac{\mathbf{p}\cos\theta}{4\pi\varepsilon_0\mathbf{r}^2}$$

Let us emphasize two important differences with respect to the potential of a single charge. Firstly, a dipole potential is **not spherically symmetric**, while the potential of a single charge is independent of the direction of observation, i.e. equal on the surface of a sphere whose charge lies in its center. Dipole potential is greatest in the direction of the dipole axis ($\cos\theta = 1$), and equals zero in the axis vertical to it ($\cos\theta = 0$). Another difference is that the dipole potential falls more rapidly to zero with increasing distance from it (as $1/r^2$), compared to the potential of a single charge (as $1/r$).

The vector of the strength of the electric field is determined by a direction and a magnitude of change of electric potential. Expressions for the components of the electric field (dipole force on a single positive charge) will not be described here. A take home message is that the dipole field strength falls to zero faster (as $1/r^3$) than the field strength of a single charge (as $1/r^2$).



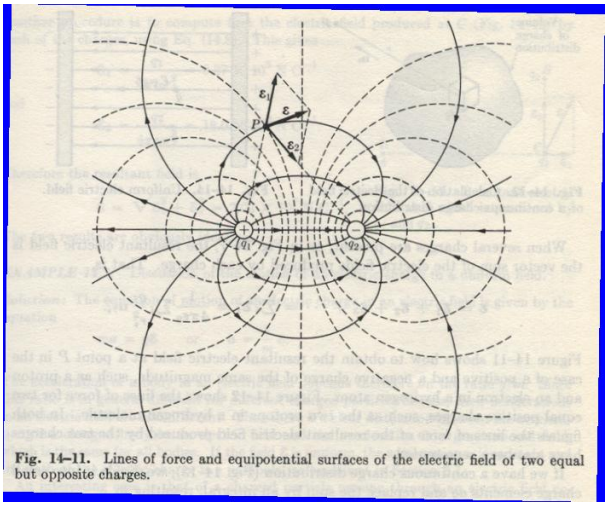



Fig. 14–11. Lines of force and equipotential surfaces of the electric field of two equal but opposite charges.

Upper diagrams show the force lines (curves whose tangent shows the direction of force) and the cross section of the equipotential surfaces (where the field strength and the potential have constant values).

This analysis of electric charges in a vacuum cannot be directly applied on a body; i.e. on the conductive agent. For that purpose, we need to consider the existence of a current source and a current sink at the points of the dipole. Applying Ohm's law, the above equation can be reproduced so that the charge is replaced with a current **I**, which originates and sinks in the dipole, and, instead of the constant of vacuum permittivity $\varepsilon_0$, conductivity of a medium $\lambda$ is used.

Thus, the expression for the electric dipole potential, from which the current loop of the intensity, **I**, emerges and sinks in, is:

$$\mathbf{V}(\vec{\mathbf{r}}) = \frac{\mathbf{I} \cdot \mathbf{a} \cdot \cos\theta}{4\pi\lambda\mathbf{r}^2}$$

In reality, it should be also kept in mind that the body is not homogeneous and that different tissues have different conductivities, and that there is, in addition to heat, the capacitive resistance of the cell membranes. All in all, a complicated calculation allows us just to approximate the real condition, often relying on empirical rules, which cannot always be accurately derived. It is important to remember that the presence of conductive medium (water in the extracellular fluid) modifies the dipole field (relative to vacuum), similar to the fact that there is an electric field in the electric cable, although we are far from the source of electrical energy.
As the potential difference between two points in an electrical circuit is determined by the resistance to the flow of electrons (voltage drop), so the potential difference between two points on the body surface (e.g. due to heart action) is determined by the resistance encountered by the ions moving from one point to another. If this resistance is small, the points are virtually on the same potential. This fact explains why rather

distant points on body surface, which are at various distances from the summary dipole resulting from the heart action, may have the same potential.

Let's start from the simplest case of one neuron. Empirical fact is that the electric field of an excited neuron resembles the dipole field in the direction of the negative to the positive part of the membrane, as indicated in the figure below.

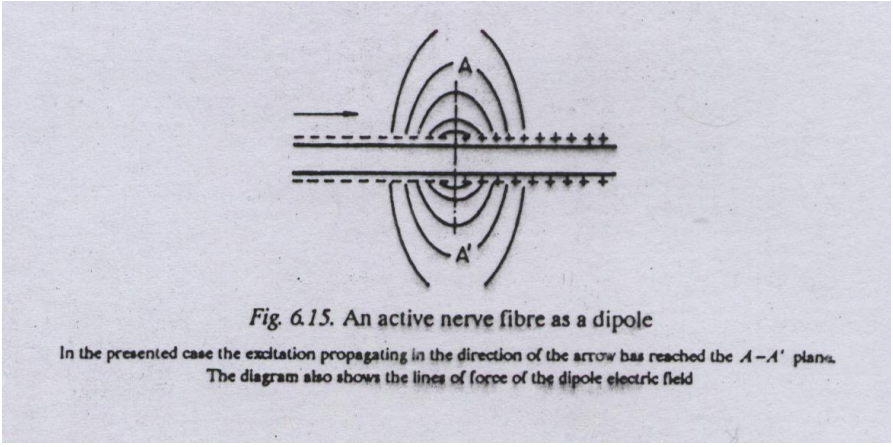



*Fig. 6.15.* An active nerve fibre as a dipole

In the presented case the excitation propagating in the direction of the arrow has reached the $A-A'$ plane. The diagram also shows the lines of force of the dipole electric field

Therefore, excited neuron can be modeled by a single effective dipole. Dipole is oriented along the membrane, toward the positive part, and its intensity changes over time, depending on the polarized state of the membrane. The strength of the dipole (its dipole momentum) increases from the moment of excitation, when it is zero (and when the entire membrane is polarized); it is largest at the moment when exactly half of the membrane is depolarized (shown on the diagram above); it again falls back the zero at the end of depolarization, and since then, during repolarization, increases again but in the opposite direction, reaching a maximum when half of the membrane is repolarized, and again, upon completion of repolarization, is zero. For accurate understanding of this process, one should take into account that the velocity of action potential propagation ensures that the excitation process reaches the distant end of axon long before the action potential is finished at the excitation site. Thus, the potential of the neuron dipole shows **biphasic** changes over time.

Similarly, the result of measuring the **potential difference** between two points on neuron surface is a biphasic curve, as shown below. The curve has a negative and positive phase, its shape during spreading of a single impulse depends on the shape of the action potential (i.e. as the impulse changes over time in one point) and its propagation velocity.

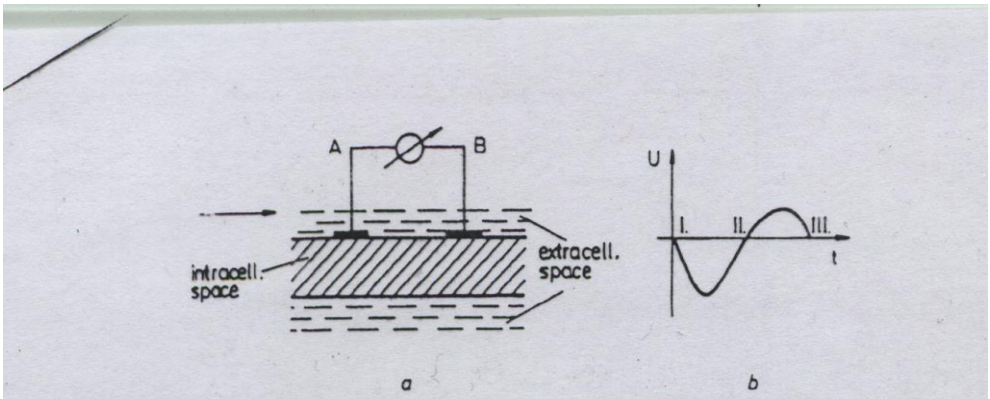



*Fig. 6.13.* Diagram relating to recording of the biphasic action potential (*a*) and the recorded action potential (*b*) in case of excitation propagating in the direction of the arrow

If we analyze multiple neurons, individual dipole vectors are added, giving the total electric dipole moment of the group of neurons. This value changes over time, and the resulting potentials can be measured on the body surface.

Most commonly, we measure a potential difference between two points on the body surface, i.e. on the skin. These potentials which arise due to the work of the heart are called an **electrocardiogram (ECG)**, ones due to brain activity **electroencephalogram (EEG)**, and those which results from skeletal muscle activity electromyogram (EMG).

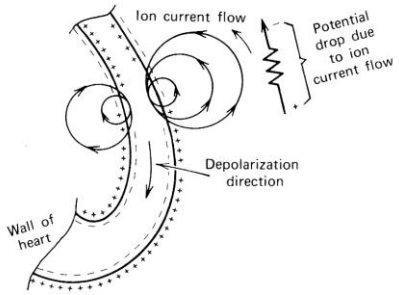



**Figure 9.15. Schematic of an action potential moving down the wall of the heart. Some of the ion current, indicated by the circles, passes through the torso, indicated by the resistor. The potential on the chest wall is due to current flow through the resistance of the torso.**

Among them, the ECG has the greatest diagnostic value. It is not difficult to explain the shape of an ECG curve in a healthy subject, nor the most common abnormalities. One needs to be aware that the heart muscle is inervated by neural fibers which originate in a group of cells in the right atrium, the so called **sinoatrial (SA) node**. These cells can be stimulated to a greater or lesser activity, but they rhythmically self-trigger even without the external stimulus. This rhythmical self-activation of cells in the SA node is explained by the relatively low polarization of these cells, just close to the threshold for opening of the sodium channels. The excitation, therefore, starts in the right atrium, from there it expands first on the left atrium, and then on both ventricles. Between the atria and ventricles, a connective tissue with low conductance is placed, so that the signals are transmitted exclusively through a narrow conductive area, so called **atrioventricular (AV) node**. Its impulse propagation velocity is low, so the signal

slows down. This effect has its purpose, because it is necessary to temporarily separate the contraction of atria from the contractions of ventricles. Within the ventricles, the action potentials are propragated via fast **Purkinje fibers**. The action potential is transmitted from the cells of Purkinje fibers to the cardiac  myocites, where it causes a change in cytosolic calcium concentration, thus initiating their contraction.

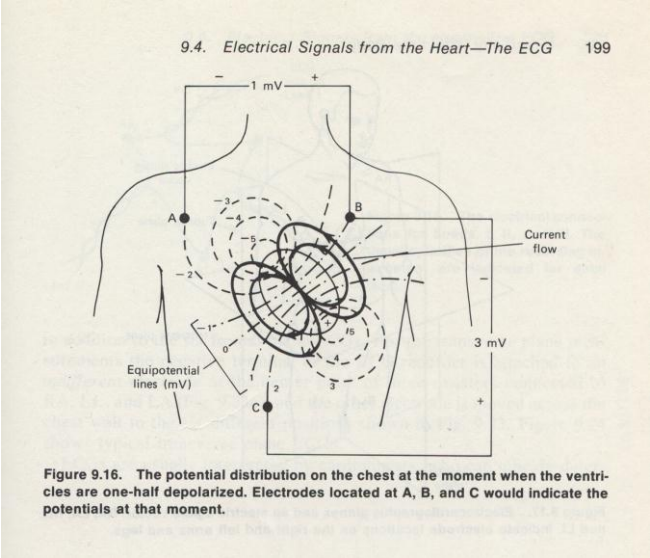
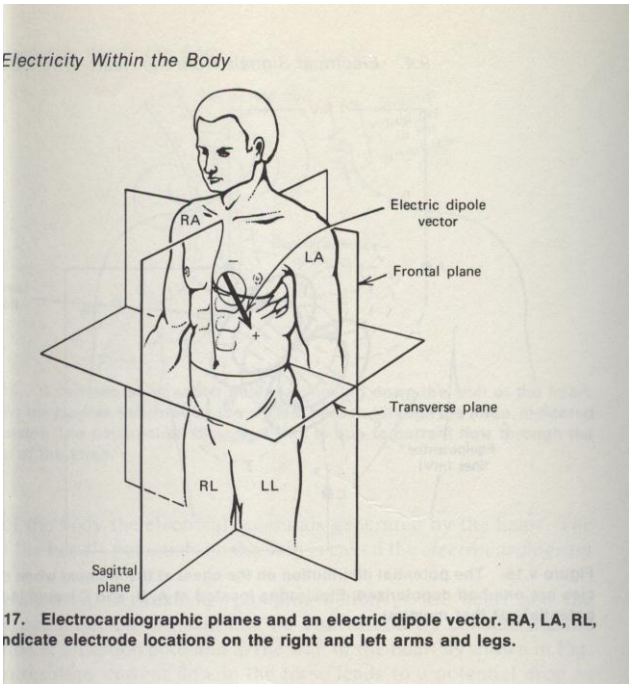



**Figure 9.16. The potential distribution on the chest at the moment when the ventricles are one-half depolarized. Electrodes located at A, B, and C would indicate the potentials at that moment.**

This means that depolarization signal spreads from the cardiac base to apex. It can be presented with a summary vector of a dipole moment, which is greatest when half of the heart muscle is depolarized, and it is directed, same as the cardiac axis, down and left, and forward relative to the frontal plane.




**17. Electrocardiographic planes and an electric dipole vector. RA, LA, RL, ndicate electrode locations on the right and left arms and legs.**

Measurements are performed by placing the electrodes on body surface. Depending on the position of electrodes, we detect different shapes of ECG curves. Thus, in order to facilitate the interpretation, it was necessary to standardize them. For that purpose, Einthoven suggested placing three electrodes on the front of the chest wall, at the tips of an equilateral triangle (**Einthoven triangle**), where the base connects the points located below both shoulders and the top is in a median line, at the lowest point of the chest wall, just above the diaphragm. However, since the limbs are very good conductors (much better than the chest), for practical reasons, the electrodes are placed on the left and right wrists and on the left ankle. The right leg is farthest from the heart, so it is not used directly, but only as a ground, in order to reduce interference noise from the power source and the device itself. The potential measured at the left (right) wrist is practically equal to the potential at the left (right) point on the base of the Einthoven triangle, while the potential of the left ankle corresponds to the potential measured at the top of the triangle. Therefore, we can proceed with further analyses as if the electrodes were placed at the tips of Einthoven triangle.

We measure the potential difference (voltage) between left and right arm (**I ECG lead**), left leg and right arm (**II ECG lead**), and left leg and left arm (**III EKG lead**). In doing so, the first potential is subtracted from the second. These voltages are called **standard ECG leads**, and are presented as three separate curves on a device printer (display). Since these voltages are relatively low (up to 3 mV), they should be first amplified. Amplifier input jacks are placed on the electrodes, two on each (one for each of the two leads that the electrode is part of), as illustrated below. According to the standard, positive amplifying input jack deflects the printer needle up and the negative down.

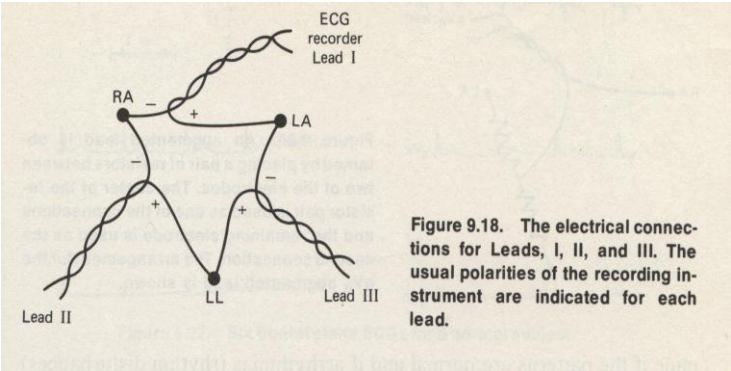



**Figure 9.18. The electrical connections for Leads, I, II, and III. The usual polarities of the recording instrument are indicated for each lead.**

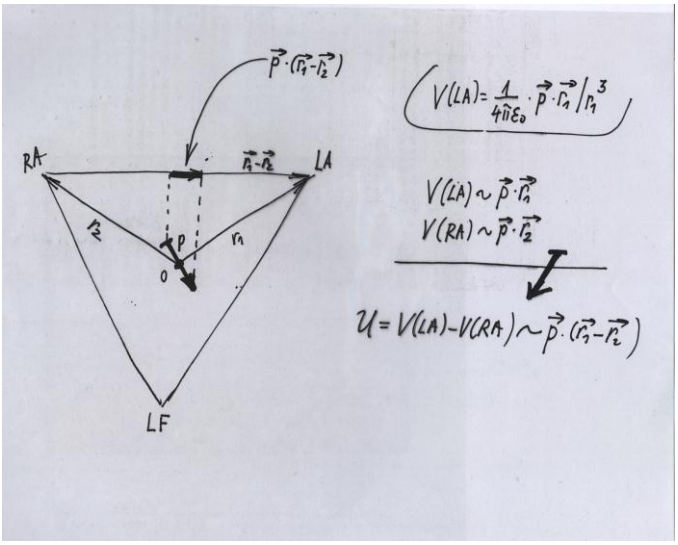
This ensures that the recorder shows the voltages according to the agreed standard.

From the introduction, we know that the potentials on body surface are proportional to the projections of the total electric dipole moment of an organ to the vector of the position observed.

From this it follows that the potential difference between two points on the body surface, resulting from the organ activity, is proportional to the projection of its total electric dipole momentum on the line that passes through the observed points.

Presented is a sketch of derivation of the above rule for standard ECG leads (Lead I is highlighted). Analogous considerations apply to other ECG leads, as well as any other

voltages present on bodily surface. Thus, each ECG lead is proportional to the projection of a total cardiac dipole momentum, and changes along with it in time, forming a distinctive curve.




The agreement on  polarity of the standard leads' electrodes becomes now clear. This way, at the time when the cardiac dipole reaches the maximum value (at the point when half of the cardiac muscle is depolarized), signal in all three leads is positive (the projection of the dipole is directed from the negative to the positive pole), as shown below.

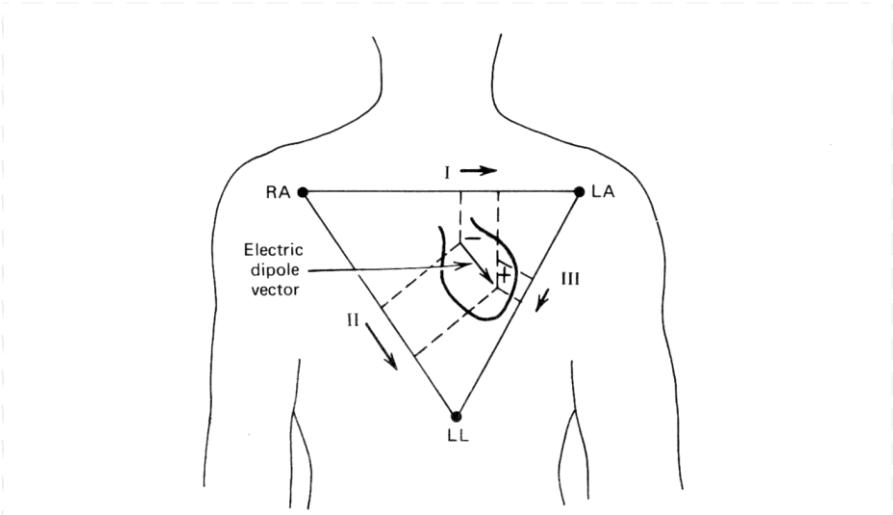
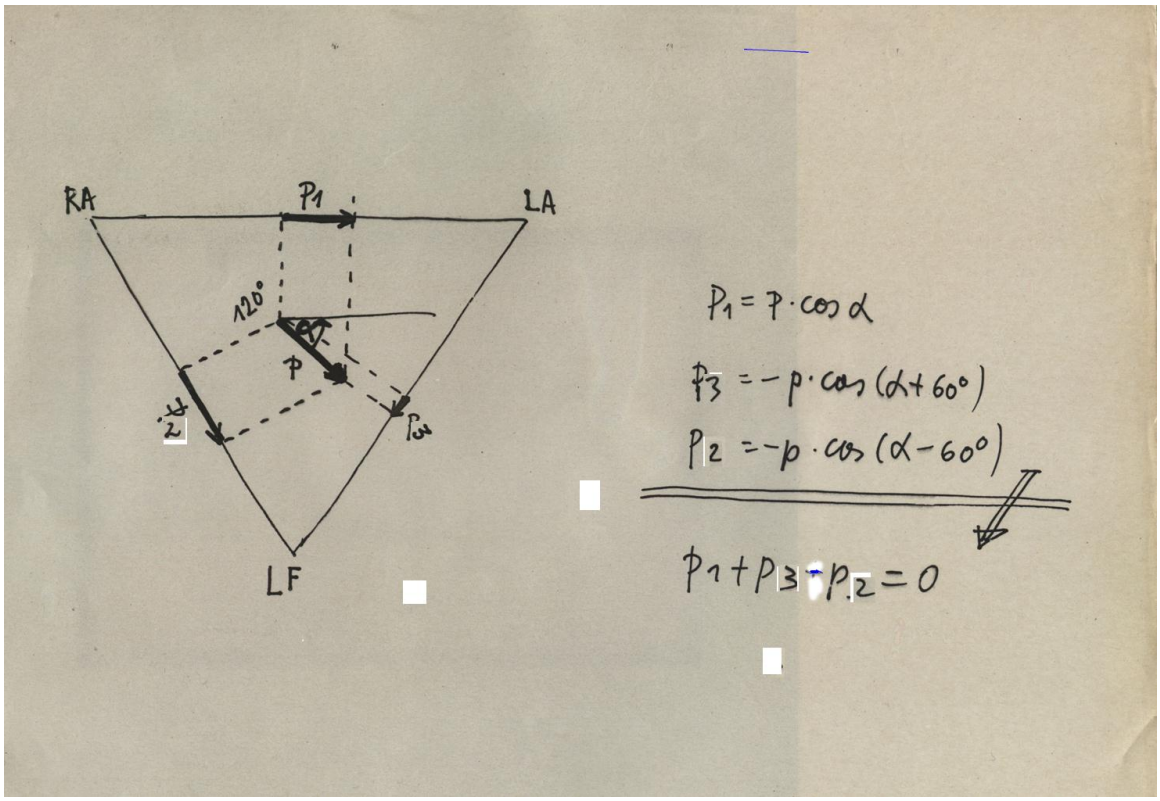



**Figure 9.19. Schematic of the electric dipole of the heart projected on the frontal plane. For electrical purposes the three electrodes (RA, LA, and LL) can be thought of as the points of a triangle, the Einthoven triangle. The potential in Lead I at any moment is proportional to the projection of the dipole vector on the line RA-LA; the potentials in Leads II and III are proportional to the projections on the other sides of the triangle.**

An interesting fact is that three standard leads do not represent independent information. Specifically, at any time point the magnitude of one of them is equal to the sum of

magnitudes of the remaining two. This is called **Einthoven's rule**. Sketch of derivation s attached.




On the figure above it can be seen that, at a time when cardiac dipole is greatest, the amplitude recorded in the II lead is greater than in the other two. At that time, in healthy subjects, projection of the cardiac dipole in the frontal plane is almost parallel to the axis of the II lead. This is the reason why this lead is most often analyzed. Figures bellow show the recording from lead II in a healthy subject.

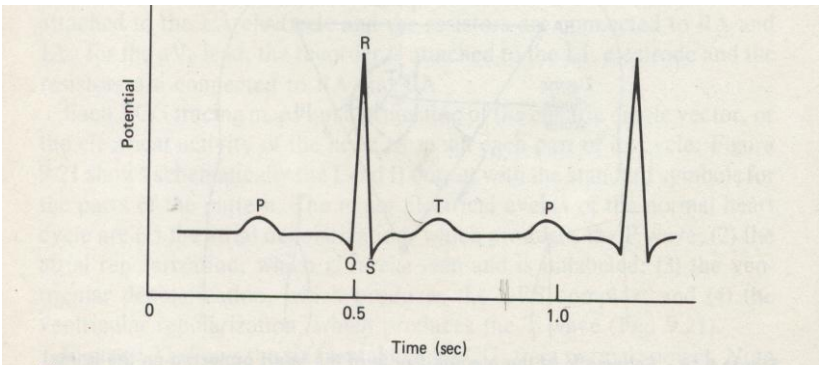
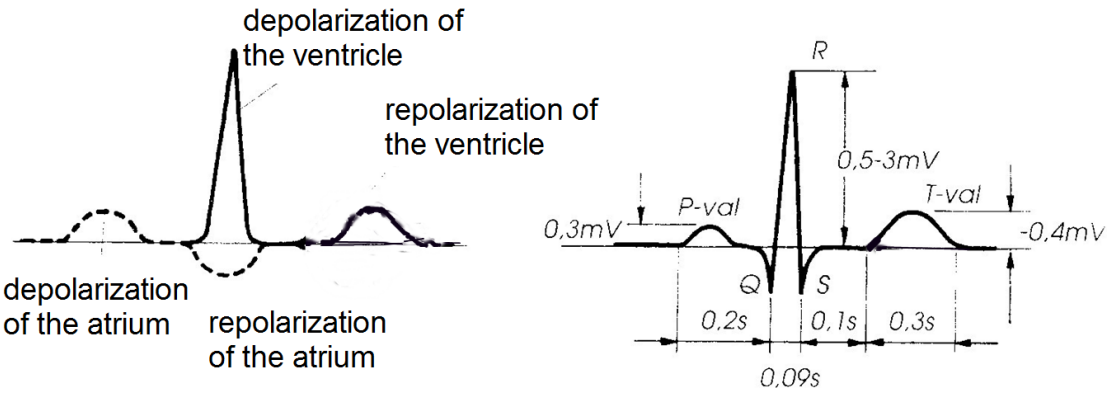



**Figure 9.21. Typical ECG from Lead II position. P represents the atrial depolarization and contraction, the QRS complex indicates the ventricular depolarization, the ventricular contraction occurs between S and T, and T represents the ventricular repolarization.**




## Heart voltage

It is important to note that each ECG signal consists of two components, smaller one is generated by the signal coming from atria (dotted curve in the upper left figure), and the larger one belongs to the ventricles (solid line). In ECG, the summary signal is recorded, and it consists of depolarization and repolarization of both, ventricles and atria. The first deflection from zero voltage (isoelectric line) is the result of atrial depolarization and is called the **p-wave**. This is followed by almost simultaneous depolarization of ventricles and repolarization of atria. Atrial repolarization causes direction of the summary dipole opposite of their depolarization, and this wave is negative. However, in the summed ECG signal, this wave is hidden because it is overcome (masked) by the positive signal coming from the depolarization of greater mass, the ventricles. Yet, it causes lowering of the base of ventricular depolarization wave, the so-called **QRS complex**: the first and last notches of the complex (Q and S) are lowered below the isoelectric line. S wave is followed by a rest, represented with zero voltage, so called the **S-T segment**, and then the **T wave,** which represents repolarization of ventricles. One might expect the T wave to be negative, but repolarization of ventricles, due to the high pressure inside them, cannot start at the same location where the depolarization had started (at the base), but rather where it had ended (apex). This is the reason for the corresponding summary dipole to be in opposite direction.

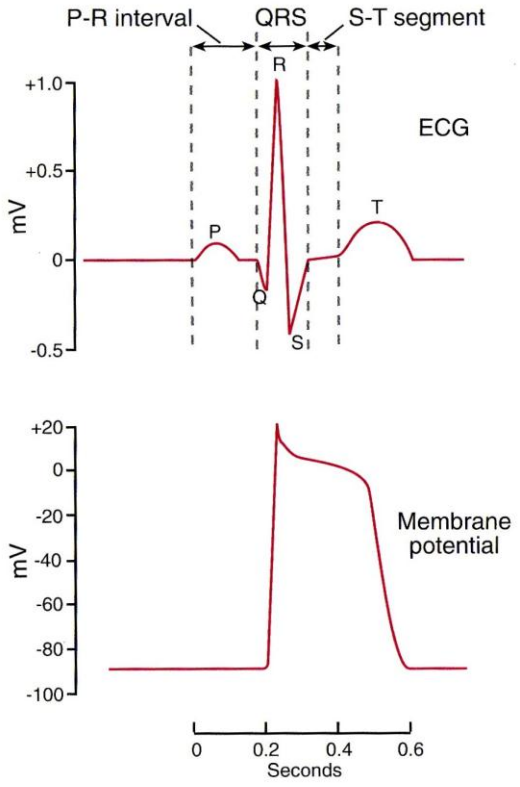



**Figure 13–12** ■ Timing of ventricular membrane potential and ECG. Note that the ST segment occurs during the plateau of the action potential.

The time alignment of one ECG cycle and a single action potential, taking place in a ventricular myocyte, is shown below. It is important to notice that the S-T segment coincides with action potential plateau. A single impulse arising in SA node leads to a single contraction of cardiac cavities; atria and ventricles. Contraction of an individual cardiac myocyte starts immediately after propagation of the action potential through its cytosol (seen as the completion of a QRS complex in the ECG). Both filling and emptying of cardiac cavities requires a certain time period, muscle synchronicity and matching of atrial and ventricular action. Therefore, the eventual premature impulse would impair the otherwise optimal sequence of events.

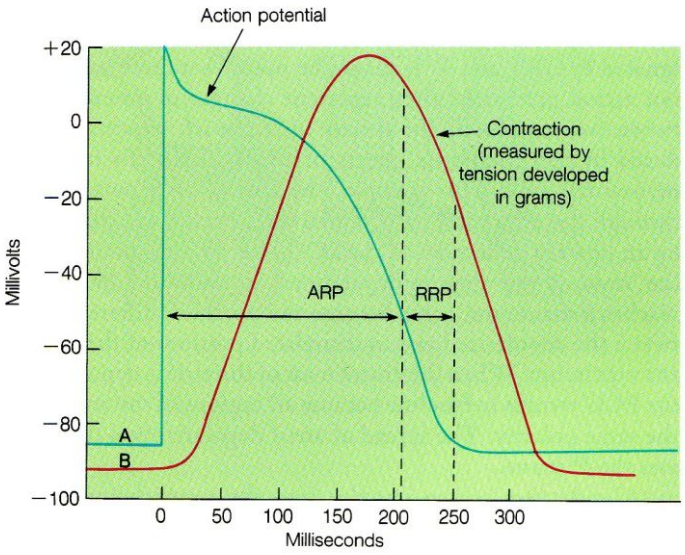


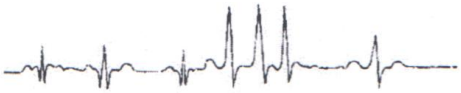

**Figure 21.10** The time course for the myocardial action potential (*A*) is compared with the duration of contraction (*B*). Notice that the long action potential results in a correspondingly long absolute refractory period (ARP) and relative refractory period (RRP). These refractory periods last almost as long as the contraction, so that the myocardial cells cannot be stimulated a second time until they have finished their contraction from the first stimulus.

The existence of a plateau prolongs duration of cardiac action potential, and long refractory period prevents the initiation of a new contraction until the previous cycle is completed. Consequently, unlike skeletal muscle, cardiac muscle cannot, by frequent, long-lasting excitation, retain the state of prolonged contraction. The relatively long time of a single contraction enables the functionally synchronous contraction of different parts of myocardium that are not simultaneously reached by an excitation impulse, because a little delay in the initiation of contraction is not important.

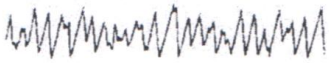
Analyses of shapes of ECG curves in different leads enable us to detect many irregularities of cardiac function and the underlying causes: rhythm disorders or **arrhythmias**, the existence of the necrotic parts of heart muscle as a result of previous **myocardial infarction**, the presence of an inadequate tissue perfusion (**ischemia**), thickening of the heart muscle (**hypertrophy**) or cavity increase (**dilatation**) due to increased resistance to blood flow through the heart and other parts of circulatory system.

Filling of the cardiac cavities is called **diastole**, and emptying (contraction) a **systole**. At rest, a healthy heart receives blood at about 70 times per minute (left side of the heart from the pulmonary vasculature, and right side from the rest of the body) and the same amount is ejected to large blood vessels (the left ventricle into the aorta and right ventricle into pulmonary artery). It is said that the **heart rate** (or pulse) is 70 beats (cycles)/min. Slower cardiac rhythm is called **bradycardia**, and accelerated rhythm **tachycardia**. Normally, ECG is a periodic curve, which is called **sinus rhythm**. The pulse is normally accelerated during physical activity, excitement, and a lack of strength of contraction, while slowing of the pulse is normally found in athletes, in whom the heart ejects a large blood volume in a single beat (they have large **stroke volume**), so frequent contractions are not necessary. If the signal is still periodic, we are talking about **sinus tachycardia** or **sinus bradycardia**. Within the periodic signal we can

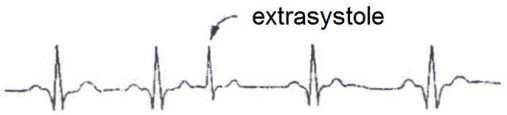
encounter a lack of ventricular contraction, as evidenced by the absence of QRS complexes in ECG. The most common cause is the so called **A-V block**, due to ischemia, inflammation or compression of the AV-node. In contrast, in ECG signal a premature ventricular systole (**extrasystole**) might also be observed, as evidenced by premature QRS complex. This systole is usually functionally inefficient, as ventricles empty before they are sufficiently filled with blood. This is manifested by the absence of peripherally measured pulse. The most common cause for this event is the existence of so called **ectopic sites** in atria or ventricles, which initiate uncoordinated cardiac contractions. They arise either in a modified ischemic tissue or due to toxic irritation of the AV node, Purkinje fibers or cardiac muscle (**myocardium**) with alcohol, nicotine, caffeine or drugs. If extrasystoles are frequent and consecutive (one after another), we are talking about **paroxysmal tachycardia**. The most severe are cardiac arrhythmias, called fibrillation (flickering), especially in the ventricles. During a fibrillation, the SA node is no longer in control and there is a quick series of completely ineffective contractions, which, if the ventricles are affected, means the **cessation** of blood flow.



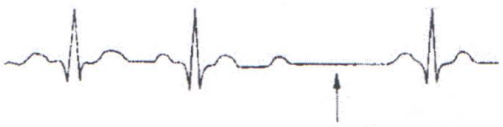
paroxysmal tachycardia



fibrillation in the ventricles



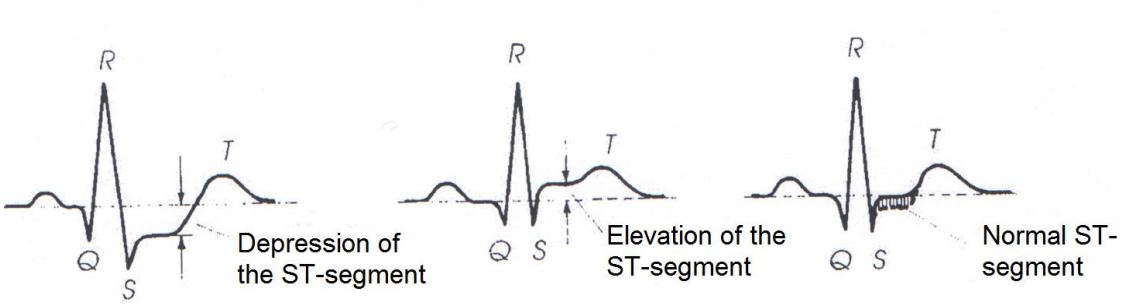

extrasystole



A-V block R-peak missing

If the ventricular fibrillation persists, the outcome can be lethal. It is caused by a permanent cycling of action potential in the cardiac myocytes with reduced refractory period and the resulting constant, uncoordinated and unsynchronized contraction of parts of the myocardium, without any efficient cardiac pumping action. Life can be saved using a **defibrillator**. Two large electrodes are placed on both sides of the heart, and a current of a very large intensity is released through the body (e.g. capacitor is charged to 10 000 V and discharged in a few ms). This current makes the entire heart muscle temporarily unexcitable, after which a normal rhythm is typically established.

Except in diagnosis of arrhythmias, ECG is also used to discover necrotic and poorly perfused parts of myocardium. Detection is based on the appearance of the so called **injury current**. Specifically, ischemic, injured tissue may lose the ability of repolarization, and remain partially or completely depolarized. In this case, the ion current continuously circulates between this depolarized and normally polarized surrounding tissue, producing a permanent dipole. This further causes the presence of constant voltage on body surface, which is seen in the ECG as an ST segment deflection (elevation or depression) from the isoelectric line of zero potential. If patient's ECG shows the injury current at rest, it is most likely due to a scar tissue after myocardial infarction. If no such signal is present at rest, but the ST segment *depression* appears at certain exercise intensity, it indicates a compromised myocardial blood perfusion during physical activity, when the cardiac muscle needs greater blood supply. This phenomenon is called a stress-induced ischemia. A diagnostic procedure of ECG recording at rest and during various exercise loads is called **stress ergometry.**



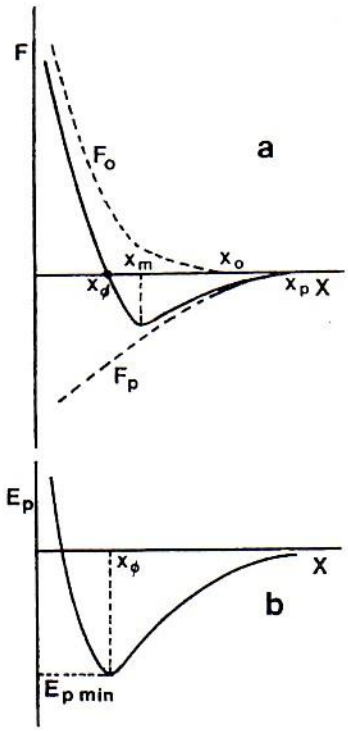

In addition to these disorders, ECG is very useful in detecting myocardial hypertrophy and cavity dilatation. These pathological conditions cause shift in cardiac axis, which is evidenced by a shift in dipole vector and characteristic alterations in ECG.

# Chapter 6: Mechanical tissue characteristics

INTERMOLECULAR FORCES

Atoms and molecules of a substance attract or repel each other, depending on the distribution of positive and negative charges. Only **quantum mechanics** provides the adequate insight, while intelligible classical physics is insufficient.

The figure below illustrates how force (upper panel) and potential energy (lower panel) between two atoms (in crystal or molecule) depend on their separation. The total force has attractive and repulsive components; the former, by definition, bears the positive sign. Both components increase to infinity as atoms approach each other, and decrease to zero as they move away. However, the attractive component decreases slower with increasing distance, and the repulsive component increases faster with decreasing distance between the atoms. In consequence, at larger distances the attractive component prevails. At distances when atoms' electron clouds begin to overlap, the repulsive force rapidly increases, preventing further approaching. In absence of external forces the equilibrium position exists where both components are equal, resulting in zero total force and maximal negativity of potential energy. Note that, as atoms separate, by moving away from equilibrium, the total force is at first an increasing attraction; then, only after position of maximal attraction, the attraction decays to zero.




The external force (of neighboring atoms and molecules) changes the equilibrium position, depending on its direction. Anyhow, in the new position of equilibrium, the external force matches in magnitude and has opposite direction from the inter-atomic force **F**. We should observe that the relation between change in inter-atomic distance ($\Delta$**x**) and magnitude of external force (**-F**), is approximately linear for relatively small shifts. In other words, since, around equilibrium, the resultant curve may be approximated by straight line (panel **a**), it holds:

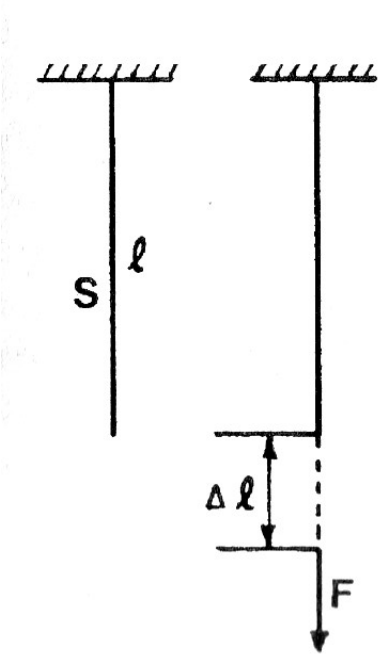
$$-\mathbf{F} = \mathbf{k}\,\Delta\mathbf{x}, \text{ where } \mathbf{k} \text{ is constant}$$

In solid state atoms oscillate around their equilibrium position, forming crystals. In liquid state, atoms or molecules are more separated, in the region where the total force is attractive, but cannot overcome chaotic thermal movement.

HOOKE'S LAW

Intermolecular forces determine mechanical properties of a substance. These are counter forces which oppose the external force; an object deforms until equilibrium is reached. **Deformation** is thus caused by external force. Under the same force different objects deform in wide range, from extensively (rubber) to unnoticeable to bare eyes (steel stick).

After external force ceases, the object may retain its original form, which is an **elastic deformation**, or stay partly deformed, which is **plastic deformation**.




Consider the simplest case of a narrow, straight and flat body, exposed to a pair of either tension or compression forces. For example, let us concentrate on a vertically hung wire, loaded with weight on the free end, as on the figure. To relate this macroscopic situation with previous case of two atoms, it is first necessary to normalize the external force **F** to unit cross-sectional area of the wire, **S**. Evidently, it is the harder to extend the wire the thicker it is. It depends on the wire cross-sectional area how many parallel units will oppose the tension; each unit opposes independently, and the individual contributions add together. Therefore one defines the **tension stress** as **p** = **F/S**. Secondly, it is necessary to express the elongation **Δx** in percents of the initial, unstressed length **l**, i.e. as **d** = **Δx/l**. It is clear since the larger initial length the more serially connected units (atoms, molecules) extend, each of them suffering smaller deformation. That is why it is easier to extend larger wire (or coil). In going with elementary law of (approximate) proportionality between external force and a change in inter-atomic distance (**F=kΔx**), for small, elastic deformations, it holds:

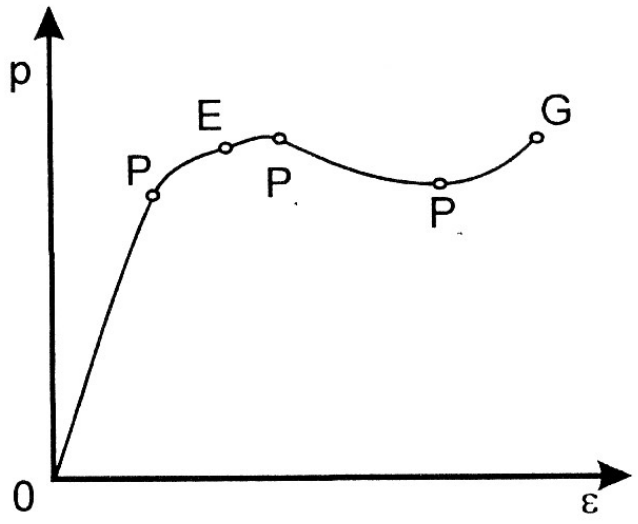
$$\mathbf{p = E\,d}$$

where the constant **E** is **Young's modulus of elasticity** (specifying the material, no matter the dimensions!), and the law is **Hooke's law**.

The greater module of elasticity the harder material extends. Similar holds for compression stress, inducing **contraction**. Materials resist to stresses of compression and tension very differently, as in the table below.

**Table 3.2. Strengths of Bone and Other Common Materials**

| Material | Compressive Breaking Stress ($N/mm^2$) | Tensile Breaking Stress ($N/mm^2$) | Young's Modulus of Elasticity ($\times 10^2$ $N/mm^2$) |
|---|---|---|---|
| Hard steel | 552 | 827 | 2070 |
| Rubber | — | 2.1 | 0.010 |
| Granite | 145 | 4.8 | 517 |
| Concrete | 21 | 2.1 | 165 |
| Oak | 59 | 117 | 110 |
| Porcelain | 552 | 55 | — |
| Compact bone | 170 | 120 | 179 |
| Trabecular bone | 2.2 | — | 0.76 |

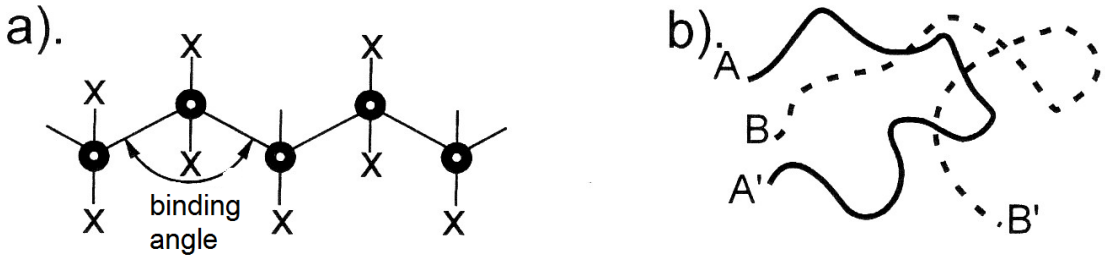


Deformation curve

Large external forces may deform the body beyond the limit of linear relationship stress-deformation. After that point, most materials deform more easily. The figure below shows the curve of deformation of a material. The **area of elasticity** (0-E) contains the segment of **linear elasticity** (Hook's law valid) and the segment of **non-linear elasticity** (P-E). After that, upon releasing stress, body does not retain its original form, which is the onset of the **area of plasticity**. This area terminates by breaking of material (point G). The area of plasticity sometimes includes the **relaxation segment** $P_1$-$P_2$ where stress in a material lowers. **Stiff** body has large area of elasticity and large modulus of elasticity (e.g. steel). **Fragile** body has breaking point close to elasticity area, or even has no plasticity area at all (e.g. glass). **Plastic** body has small area of elasticity, and large area of plasticity (e.g. lead). Materials that have relaxation segment are **flexible** (e.g. clay and plasticine). **Viscoelastic** materials do not have area of linear elasticity, which will be explained in the next section. Some of them have breaking point under low stress, but the corresponding deformation is large. That is how some muscles and connective tissues (with lot of elastin) can extend to ten times the unstressed length before breaking.

HOOK'S LAW AND BIOLOGICAL MATERIALS

Tissues differ from inorganic solid materials. Tissues are comprised from cells and interstitium, with various complex macromolecules surrounded by water. Deformation of a tissue includes several possibilities, depending on type and function. The following figure is a simple model of large polymer molecule, having twisted, interwoven chains.




Polymer molecules: a) Structure; b) Curved shape of twisted molecules AA' & BB'

Note that the extension of a single chain can be realized by:

- chain straightening, which is easy, involving breaking of a few weak bonds between opposite chain segments (tertiary structure),

- rotation over the primary (straightened) chain (i.e. in a way that each second elementary unit rotates for 180°; this breaks more common bonds between neighboring units, which are, however, also weak bonds (secondary structure), and by

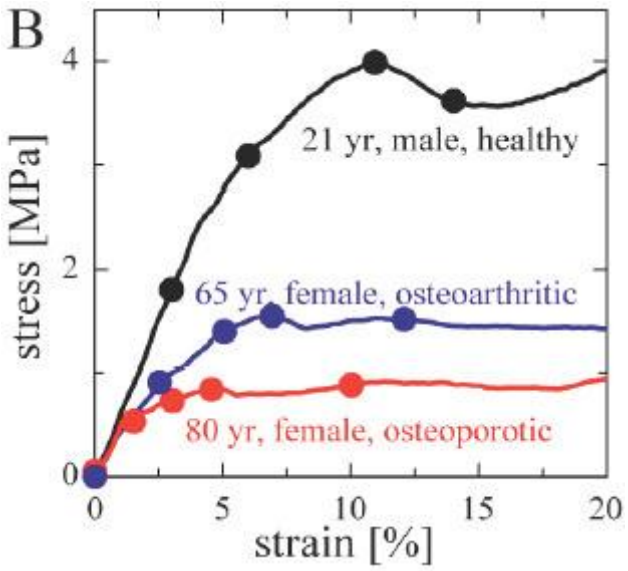
(c) opening of the covalent bond angle, which requires large forces and does not occur before secondary structure breaks?

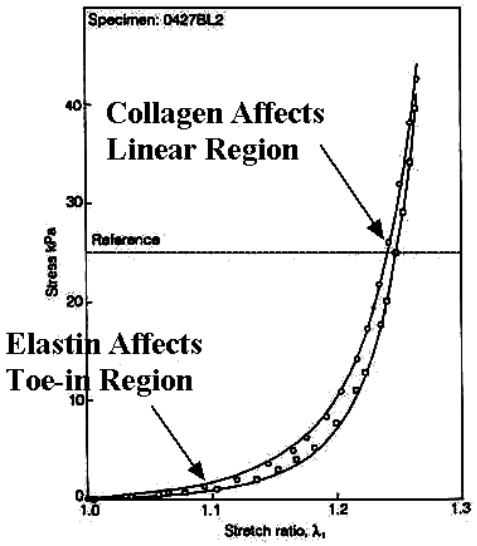
In case of relatively small stress, the chain straightening and rotation over primary chain produce elastic deformation in which the work of external force stores as a tissue potential energy, without heat losses. However, in reality, molecules slide over each other, which includes thermal interaction and heat loses, analogous to viscous resistance of fluid flow. One says that tissues have both elastic properties of solid structures and viscous properties of fluids; i.e. that tissues are **viscoelastic**.

One should observe that viscous resistance of a medium increases with velocity of an object moving through. For example, the air resistance to raindrops increases with square of raindrop velocity. In this way the resistance rapidly equalizes with gravity and raindrops fall at the earth surface with moderate velocity (otherwise they would be real water projectiles). Analogously, in case of large stresses, the viscous resistance of a medium opposes large rates of deformation.

Viscoelasticity ensures that doubling the brief stress does not result in double, but relatively lower deformation. Thus, large stress does not necessarily induce large damage, provided its action lasts short.

After an area of elasticity, some tissues oppose stress more vigorously (like blood vessel), other offer less resistance (like bones), as presented in the figure below.




DEFORMATION TYPES

Several deformation types can develop, depending on the direction and point of action of external forces.

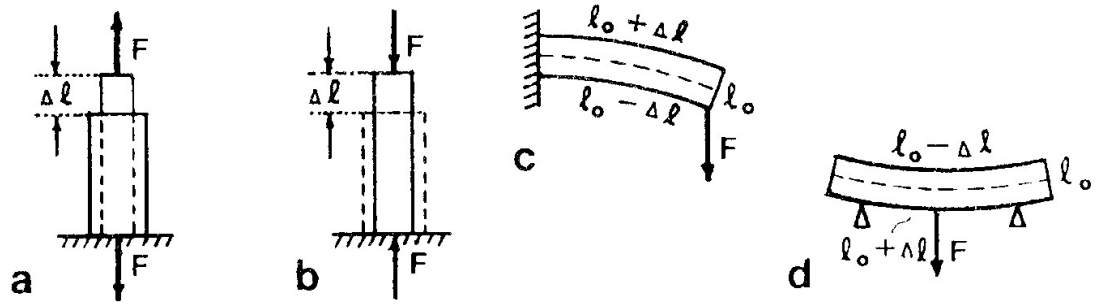
**Extension** (**dilatation**) occurs due to **tension stress**, when a pair of anti-parallel forces, acting on opposite ends, *increase* the distance between body's particles in the direction of their action. Therefore, extension is always accompanied by body's cross-sectional thinning (lower figure; panel a).

**Contraction** occurs due to **compression stress**, when a pair of anti-parallel forces; acting on opposite ends, *decrease* the distance between body's particles in the direction of their action. Therefore, extension is always accompanied by body's cross-sectional thickening (lower figure; panel b).

**Bending** occurs in two ways:

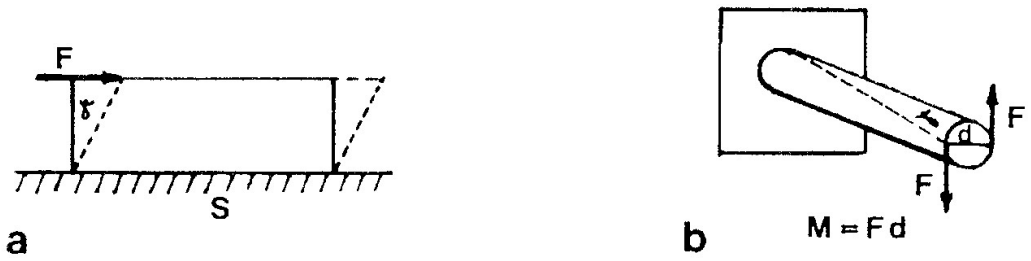
(a) The force acts at right angle to axis of an elongated body, at its free end, with the opposite end fixed. In consequence the body bends. The upper layers distend, while the lower layers contract. Medial layers keep the length constant, experiencing least stress (lower figure; panel c).

(b) An elongated body, supported at both ends, experiences lateral stress in the middle. The upper layers contract, and the middle layers keep their length constant (lower figure, panel d).




**Sliding** occurs when force acts in the plain of a body's surface, with opposite surface fixed. The layers move relatively to each other, in a way that the layer in the plain of force experiences maximal displacement; other parallel layers displace progressively lesser. The angle of bending of the side surface is proportional to **shear stress**, i.e. the external force normalized to unit surface of a body (lower figure, panel a).

**Torsion or twisting** occurs when a pair of opposite forces acts on a free end of an elongated body, fixed at the opposite site (lower figure, panel b).




WHERE AND WHY THE BONES BREAK

Bones are the most durable part of a body, lasting for hundreds, even millions of years. Bone structure depends on function (support, locomotion, protection, storage, food crushing, sound transmission). The basic structural elements (except water) are protein compound- **collagen**, and inorganic component, the **bone mineral-calcium hidroxyapatite** ($Ca_{10}(PO_4)_6(OH)_2$). Mineral matrix builds in a complex protein structure, realizing the ideal combination of flexibility (due to collagen) and stiffness (due to bone mineral). The long bone resembles the hollow pipe. This is an ideal form regarding the material economy and function, as well as an ability to respond to stresses coming from various directions. As seen previously, the lateral force causes combined contraction of layers at the site of impact and elongation at opposite site (figure below). The material in the middle deforms least. That is why the hollow pipe is the best structure to withstand lateral stresses (this explains the design of a brick).

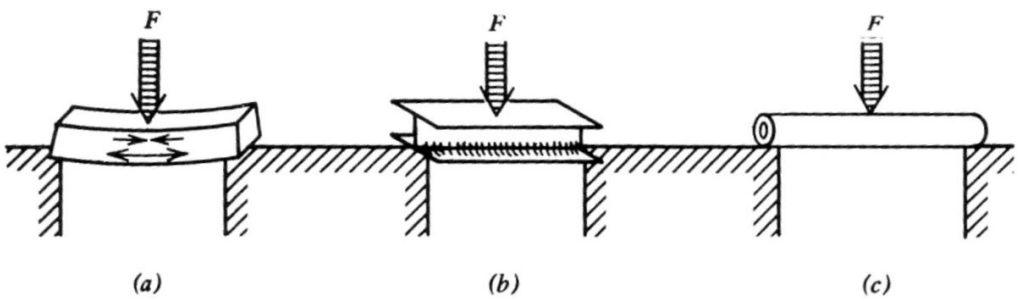
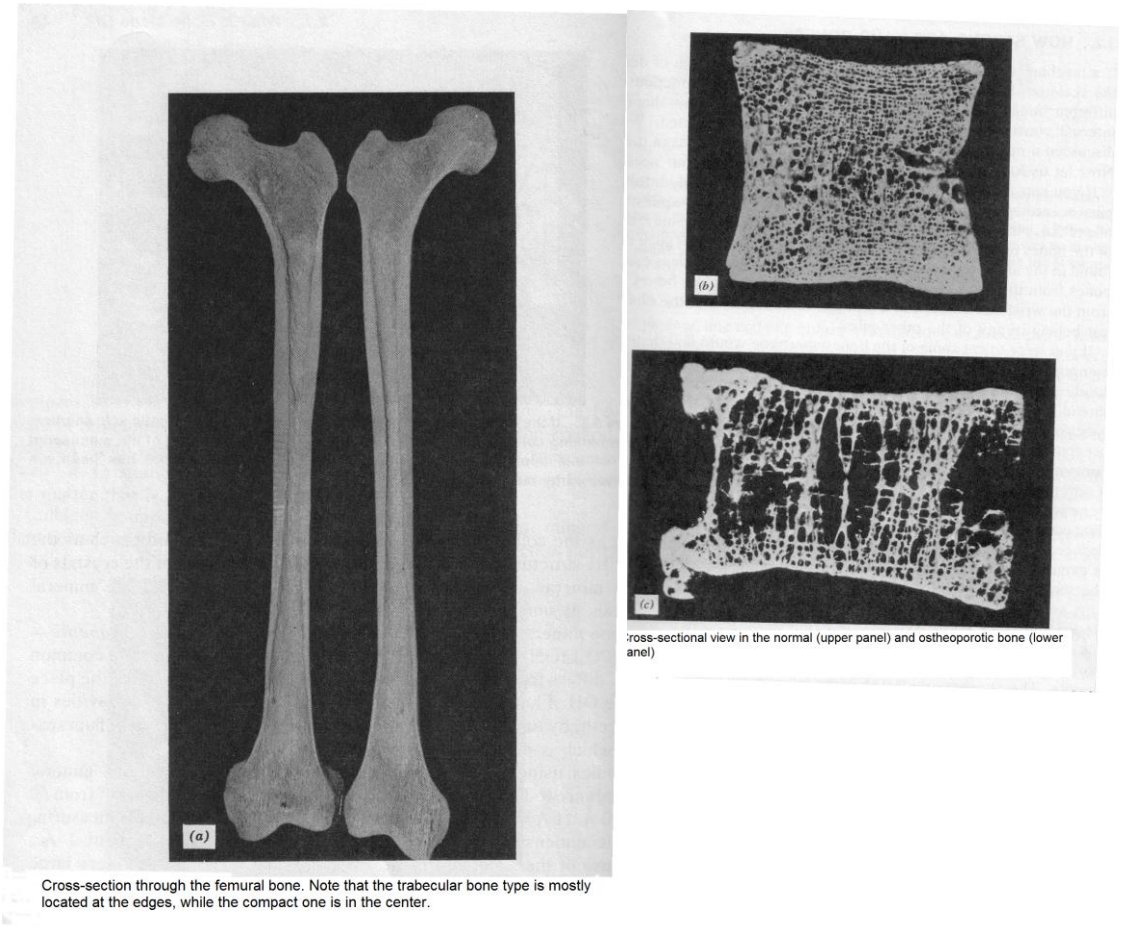



**Figure 3.5. Various types of beams subjected to a force *F*. (*a*) In a simple rectangular beam the greatest stresses are near the top and bottom. There is little stress in the middle of the beam. (*b*) Because the stress in the middle is small, a beam that has less material there—an I beam—can be used. (*c*) A tubular beam can be thought of as a rotated I beam with the center web removed. It is used when the force may come from any direction.**

There are two types of bones: **compact bone** and spongy or **trabecular bone**. In a certain bone, one type dominates, but mostly both types are present. On the lower figure, we see the longitudinal cross section of a thighbone- femur (a) and transversal slices through normal (b) and osteoporotic vertebra (c).

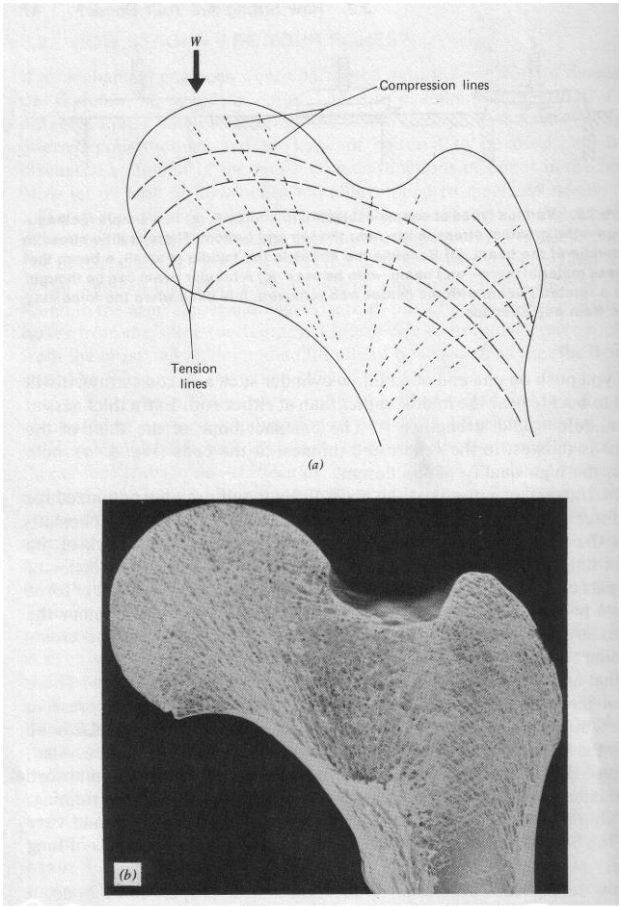



ross-sectional view in the normal (upper panel) and ostheoporotic bone (lower anel)

Cross-section through the femural bone. Note that the trabecular bone type is mostly located at the edges, while the compact one is in the center.

Trabecular bone dominates the ends of a long bone, in articulations with large surface areas. The percent of compact bone increases towards middle section, comprising

mostly superficial layers. Vertebrae are almost exclusively made of trabecular bone, with exception of a tiny compact superficial layer. Trabecular bone is substantially weaker than the compact bone, due to less bone mass in a given volume, which is especially emphasized in osteoporosis. However, on microscopic level, a piece of trabecular bone tissue does not differ from the piece of compact bone, both in case of normal and osteoporotic bone. For this reason, it is more correct to state that osteoporosis denotes the condition of diminished bone mass, not density. Namely, the osteoporotic bone has normal density; it only contains more tissues other than bone in a given volume.



Cross-section through the femur's head and neck.

Thighbone is an example of a perfect natural design. On widened ends, the thighbone receives the compressive stresses on relatively large areas and then dissipates the forces to various directions over the larger volume than the compact bone of the same mass. For this reason, as well due to larger flexibility of trabecules, the spongy bone is more resistant to compressive forces, which occur in articulations during walking, running and jumping, than the compact bone of the same mass (!). On the other hand, spongy bone is less resistant to lateral stresses which cause bending. This is why there is increasingly more compact bone towards thighbone middle section.

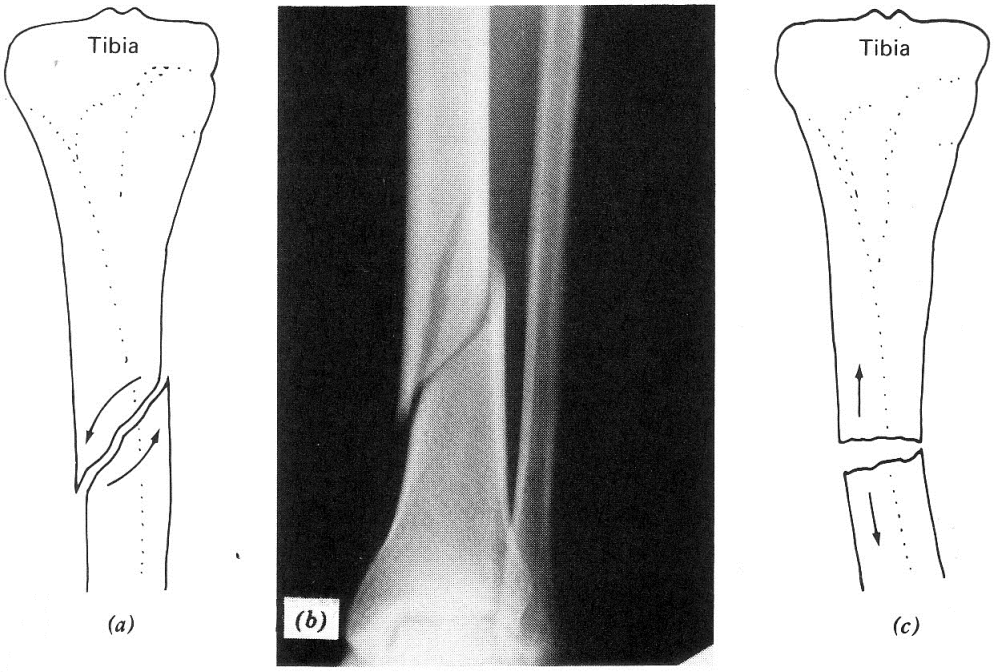
In conclusion, compared to the same mass of a compact bone, the spongy bone is superior in withstanding compression, but inferior in case of tension.

The bones rarely break due to excessive compression. This requires extreme conditions, i.e. falling from high object, when, most commonly, the vertebrae trunk and plateau of tibia.

Tensile stresses are even rarer. The may occur during epileptic seizure, due to strong, unwilling muscular contractions, when fibula or humerus may fracture.

The long bones most commonly fracture in middle section, after suffering lateral impact, which cause excessive bending and extension at contralateral site, which breaks first (figure below, left panel).


Torsion stress is a common cause of fractures, of a typically spiral appearance (figure above, right panel).




Tibia fractures: schema of spiral fracture due to torsion (a), its corresponding X-ray image (b) and fracture schema of tibia due to stretching (c)

Normal torsion stress is a typical cause of fracture of the neck of a femur in older people, especially with osteoporosis. Due to shorter neck compared to whole bone length, larger forces develop on this side of a lever.

## SELF ASSESSMENT

1. All elementary forces decrease with increasing distance between two objects (masses, charges, nucleons). In this light, explain apparently paradoxical fact that the larger the deformation, the more an object opposes the external stress.

2. Explain the phenomenon of viscoelasticity.

3. Why is a hollow brick better construction material than the compact one?

4. Explain why an excessive stress makes the blood vessel more rigid, which is not the case for bone.

5. A soccer player experiences a kick in his tibia. Explain why his tibia does not commonly break at the place of impact.

# Chapter 7: PHYSICS OF HEART AND CIRCULATION

## Liquids at rest

In liquid state the attractive forces between molecules do not suffice to keep them fixed in space (in contrast to solid state), but do suffice to prevent them to spread away due to thermal motion (in contrast to gaseous state). Like solids the liquids are relatively dense and keep constant their volume (**incompressibility, indistensibility**), but not their shape, taking the shape of a container. Like in gases, due to molecular flexibility, the pressure in liquids in transferred in all directions. This common feature of gases and liquids (**fluids**) is called the **Pascal's law**. However the freedom of movement of molecules in liquids is much less than in gases, so that pressure due to molecular random thermal movement in liquids is negligible when compared to pressures either due to weight of a liquid (**hydrostatic pressure**) or to external forces (**hydraulic pressure**). In contrast, the particles dissolved in liquid act as gas particles, exerting the pressure in proportion to their molar concentration and temperature (**osmotic pressure**).

Since the hydrostatic pressure in a liquid increases with depth and acts in all directions, the liquid exerts the force on the submerged objects in direction opposite to gravity and equal to the weight of the same volume of a liquid as is the volume of an object. This force is termed **buoyancy** and the phenomena the **Archimedes's** law. It stands for both liquids and gases. All in all the liquid state shares some features with solid state (large density, constant volume) and other features with gaseous state (spreading the pressure in all directions, the phenomena of buoyancy).

## Liquids in motion

When the liquid flows through the rigid pipe, the same volume that entered a section of a pipe in a given time must leave it at its distal end. This is the consequence of liquid incompressibility.

In these conditions and taking into account that the flow rate equals the product of the flow velocity (**v**) and the cross sectional area (**S**), if a caliber of a pipe changes, the flow velocity changes inversely:

$$\mathbf{v_1/v_2=S_2/S_1}$$

The above relation is called the **continuity equation**. The condition of incompressibility stands for the blood flow. However, the blood vessels are relatively rigid only in some parts of circulation. The compliant large arteries and veins pulsate rhythmically, as they receive the pulsate outputs of the heart. In these parts of circulation the continuity equation applies only when cycle-averaged values are considered.

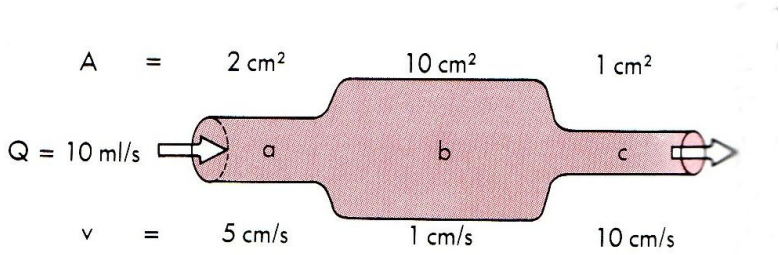
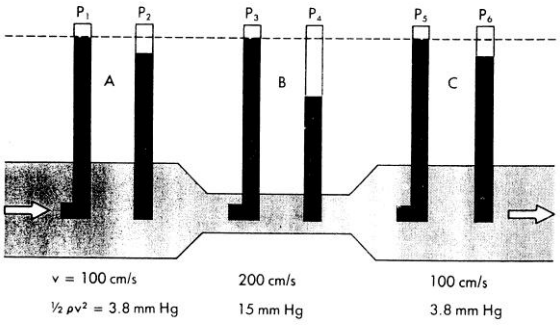



**Fig. 26-1** As fluid flows through a tube of variable cross-sectional area, *A*, the linear velocity, *v*, varies inversely as the cross-sectional area.

The pressure in a liquid can be viewed as the energy of its unit volume. At rest, it is a potential energy, comprised of two components: hydrostatic and hydraulic pressure. Hydrostatic pressure due to weight of a liquid of a density **ρ** at the depth **h** from the referent plane equals **ρgh**. In blood circulation this pressure changes with body posture; the referent plane crosses transversally the right atrium. The hydraulic pressure is due to the external force acting on a surface of a liquid; in blood circulation, the origin of this force is the heart. The sum of hydrostatic and hydraulic pressure is the **static pressure** (**$p_s$**). The static pressure acts in *all directions.* When liquid flows, a **dynamic pressure**, (**$p_d$**) is produced in *the flow direction,* which is the kinetic energy of a liquid unit volume- **$\rho v^2/2$**. The total pressure in a liquid is the sum of static and dynamic pressures and equals the total energy (potential and kinetic) of its unit volume (aside from the gravitational potential energy).

If one neglects the flow-related frictional heat loses (which is the model of an **ideal liquid**), the total energy is conserved; thus the total pressure in a horizontal circulatory segment does not change, which is the statement of the **Bernoulli's equation**:

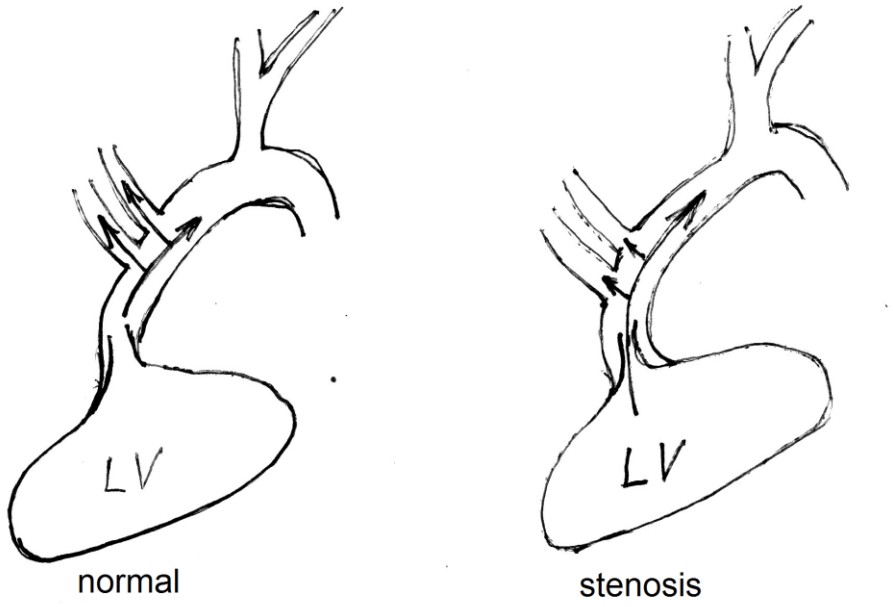
$$p_s + \rho v^2/2 = \text{constant}$$




Observe on the figure that the static pressure, which acts in all directions, is measured at the right angle to flow direction (straight manometers). In this way one avoids the dynamic pressure, which acts only in flow direction. In flow direction one measures the constant total pressure (L-bottom manometers). Observe that at narrow sections the static pressure reduces on account of increased flow velocity, which increases dynamic pressure.

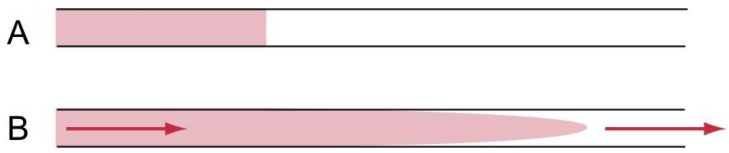
Bernoulli's equation can be applied in blood circulation in large blood vessels where the resistance to flow and the related energy frictional loses are negligible. In human circulation, the static pressure is usually much larger than the dynamic pressure. The reason behind is that the blood flow is relatively slow.

However, when the blood flows through the narrow heart valves, the blood velocity increases enough (especially in case of pathologic narrowing) that the dynamic pressure (which increases with the square of velocity) becomes the significant component of the total pressure; the static pressure is reduced in consequence. In **aortic stenosis** (the disease due to narrowed opening of **aortic valve**) the flow through the aortic inlets of the coronary arteries (**coronary or Valsalve sinuses**) is reduced. These sinuses are at right angle to the direction of blood flow, i.e. in the direction where only the reduced static pressure acts. The decrease in the coronary blood flow occurs in hard conditions when the heart work and energy consumption are increased, which acts to produce the vicious circle, requiring urgent intervention.




The small vessels offer significant resistance to flow and the Bernoulli's equation has to be completed with the term that describes the frictional energy (pressure) loses. The friction is between the fluid and the vessel wall, which decelerates the flow in the near-by fluid layers and induces further friction between fluid layers flowing with different velocities. One says that **real liquid** is **viscous**.

When the velocity is slow the liquid flow is **laminar**. The liquid layer just by the vessel wall does not move at all (there the friction is static), while the subsequent layers move progressively faster; the maximum is at vessel's axis.



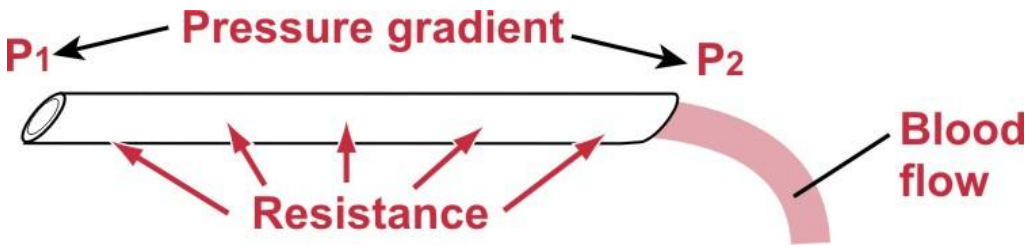

Thus, there is a **radial velocity gradient** from periphery towards axis, **Δv/Δx**, where **x** is the distance to vessel axis. The frictional force between two neighboring liquid layers is **shear stress**. It tends to accelerate the slower layer and decelerate the faster one. According to **Newton's law of viscosity**, the shear stress is proportional to the velocity gradient; the constant of proportionality is the **viscosity coefficient, η**, which is a characteristic of a given liquid:

$$\text{shear stress} = \eta\ \Delta v/\Delta x$$

The greater the viscosity coefficient (in short-viscosity) is the harder the fluid floes. Viscosity coefficient has little in common to density of a liquid. For example oil is more viscous than water although it has lower density (oil floats on water). Viscosity of gases increases with temperature, contrary to liquids (by cooling the liquids approach the solid state of infinite viscosity). That is why at wintertime the blood viscosity increases in areas not protected by clothing (head, fist, and feet). This may cause serious circulatory problems; an increase in incidence of cerebrovascular stroke in London at wintertime was observed in homeless people not using the head covers.

Due to viscous loses the total pressure in a liquid (unit total energy) decreases in direction of flow, which is prominent in case of small blood vessels. In other words, in order to sustain the constant flow of a liquid in a vessel one needs to establish the difference in pressure (**pressure gradient**) between the vessel's inlet and outlet. The required pressure gradient **ΔP** is proportional to the product of liquid flow **θ** and vessel's resistance to flow **R**.

$$\Delta P = \Phi \times R$$

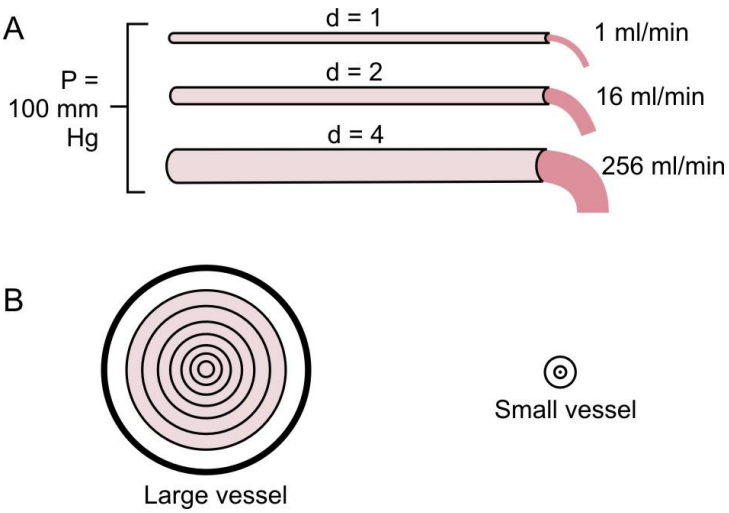



The resistance to flow of a liquid with viscosity **η** through the horizontal circular vessel with length **l** and radius **r** can be expressed as follows. If flow is constant, the total force acting on a cylinder of a fluid within the vessel must be zero. Since the shear forces between the fluid layers cancel each other (law of action and reaction), the two remaining forces that act on a cylinder must also cancel. These forces are shear stress

between the outer liquid layer and vessel wall times superficial area of a cylinder (force opposing flow) and pressure gradient times the cross-sectional cylinder area (force facilitating flow). Taking into account the Newton's law of viscosity and solving the resulting equation in **R**:

$$\mathbf{R} = \frac{8\eta}{\pi}\frac{\mathbf{l}}{\mathbf{r}^4}$$

The flow resistance **R** depends on a liquid and on dimensions of a vessel. It is proportional to vessel's length and rapidly increases as its radius decreases. It is very important to observe the later inverse $\mathbf{r}^4$ dependence. The strong dependence of a flow resistance on a vessel's lumen is a mighty mechanism of local regulation of blood flow. In case when metabolic activity increases, the blood flow becomes insufficient. The locally produced metabolites act vasodilatory; i.e. they act to decrease the tension (**tonus**) of smooth muscles of small blood vessels (arterioles and capillaries). The $\mathbf{r}^4$ dependence means that a small dilatation suffices to lower the flow resistance significantly. That is how the blood flow redistributes from passive to active areas, for example to skeletal muscles and heart during exercise. In contrast, the constriction of circulatory smooth muscle of arterioles drastically increases its flow resistance.




Combining the above two equations:

$$\mathbf{\Phi} = \frac{\pi}{8\eta}\Delta\mathbf{P}\frac{\mathbf{r}^4}{\mathbf{l}}$$

The above equation is known as the **Poiseuielle's law**: flow of liquid through horizontal vessel is proportional to the product of pressure gradient and the 4-th power of vessel's radius and inversely proportional to liquid viscosity and length of a vessel.

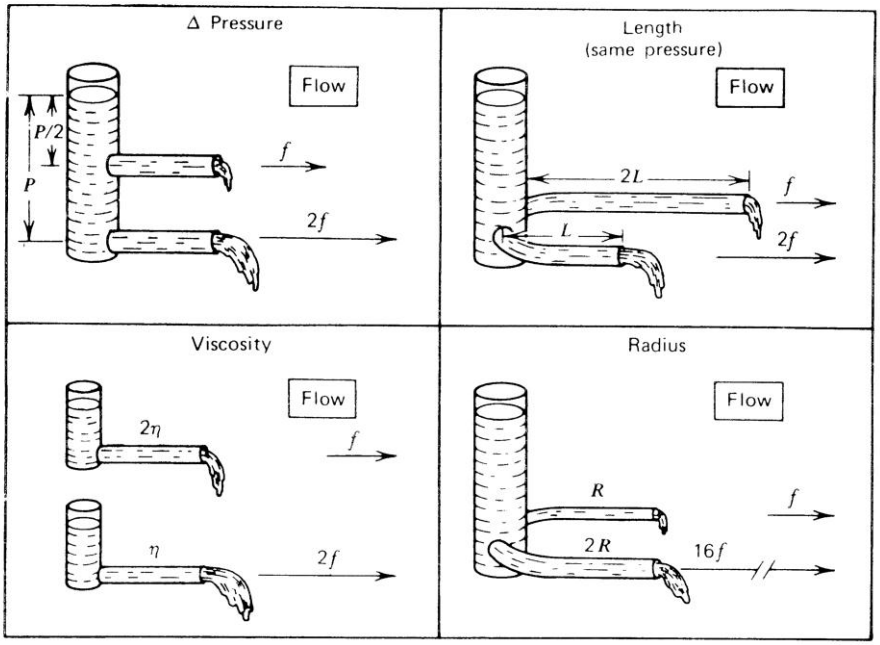



**Figure 8.13. Poiseuille's findings. The flow rate through a tube depends on the pressure difference from one end of the tube to the other, the length of the tube, the viscosity of the fluid, and the radius. The radius has the largest influence on flow rate.**

Laminar fluid flow is sustained as long as the flow velocity does not reach the critical value, when the layers begin to mix, establishing irregular, partly unpredictable fluid motion, known as **turbulences**.

In straight, smooth vessel the turbulences will occur the sooner (or the critical velocity will be the smaller) the denser the fluid, the wider the vessel and the less viscous the fluid is. *Osborne Reynold* empirically described the likelihood of turbulences with the dimensionless number; in his honor called the **Reynolds's number** ($\mathbf{R_e}$):

$$\mathbf{R_e} = \frac{\rho \, v \, D}{\eta}$$

The greater the Reynold's number, the greater the likelihood of turbulences. Fluid inertia (product of fluid density and flow velocity) promote cessation of laminar flow and induction of irregular motion, while fluid viscosity opposes mixing of fluid layers.

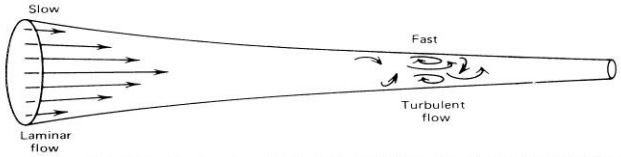



**Figure 8.15. If fluid is flowing in a long tapering tube, the velocity will gradually increase to the point where it exceeds the critical velocity $v_C$, producing turbulent flow.**

In straight, smooth blood vessels, the turbulences occur when the Reynolds number exceeds 2000. The fatty layers (**plaques**) on vessel's wall and presence of bifurcations promote turbulences even when the Reynolds's number is much less than 2000.

In turbulent motion, a part of fluid energy is lost as heat and another part as caustic energy of vibration. These energy loses are behind both the advantage and disadvantage of turbulences. The disadvantage of turbulences is the energy loss (decrease in total fluid pressure), or the additional resistance to flow. However, the turbulences can be heard, providing valuable diagnostic information.

At rest the turbulences in human circulation are rare, they occur regularly only in aorta when the heart rapidly ejects the blood through relatively narrow aortic valve. In exercise, the flow velocities multiply and the turbulences are more common.

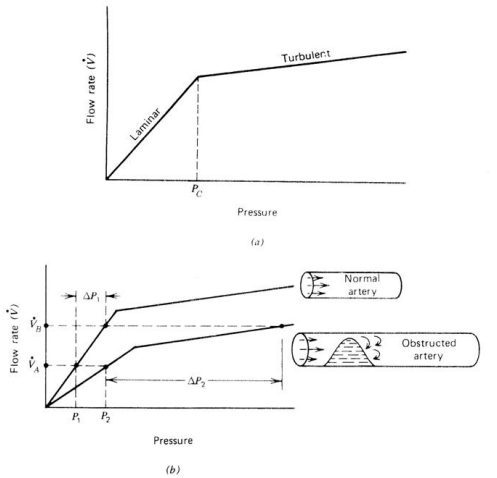
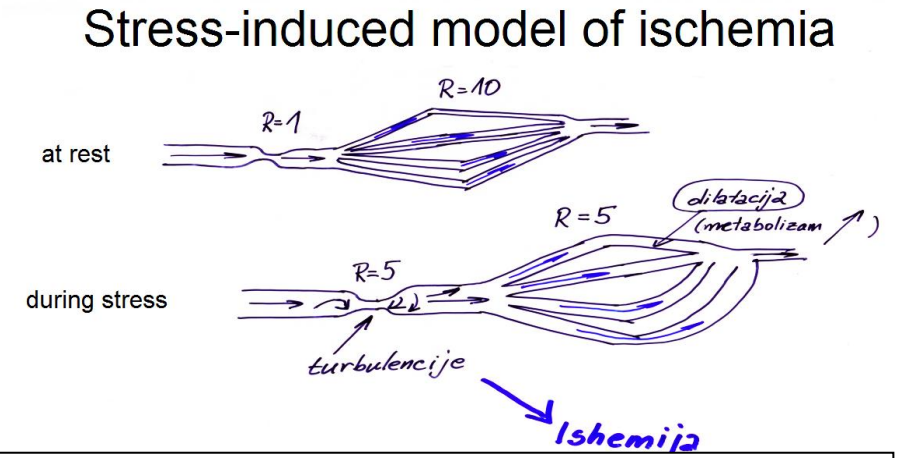



**Figure 8.16.** (*a*) When the flow in a tube becomes turbulent (at pressure $P_C$) the slope of the flow rate versus pressure decreases so that compared to laminar flow a greater increase of pressure is necessary to obtain a given increase in flow rate. (*b*) In an obstructed artery the pressure needed to produce a given flow rate is greater than in a normal artery of the same size. In addition, if the heart is called upon to increase the flow rate from $V_A$ to $V_B$ the turbulence produced in the obstructed artery requires a much larger pressure increase ($\Delta P_2$ vs $\Delta P_1$) and thus greater effort from the heart. (Adapted from I.W. Richardson and E.B. Neergaard, *Physics for Medicine and Biology*, Wiley-Interscience, New York, 1972, pp. 46–47.)

If blood vessel is pathologically narrowed (**stenotic**), the induction of turbulent flow can increase its resistance enough to reduce the blood flow significantly. Depending on a degree of stenosis, this may not happen at rest, but only during exercise, as shown in the following model.




At rest, the stenosis (narrowing of lumen) of coronary artery has no hemodynamic consequence as long as it is under 50%. Reason: the majority of resistance is in small vessels. In stress, due to 1. **turbulences** induced at narrowings and 2. **metabolic vazodilatation** of resistance blood vessels, the contribution of stenotic artery to total resistance becomes significant. In consequence the coronary blood flow is insufficient, leadind to depressed oxygen consumption in myocardium, i.e. the anaerobic metabolism (ischemia).

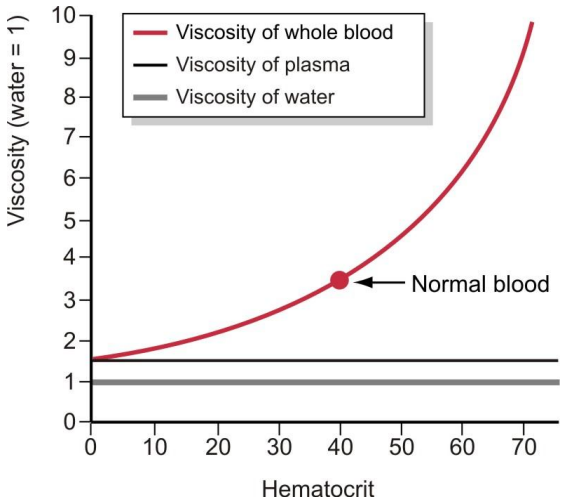
## How blood flows: hemorheology

The blood flow cannot be exactly analyzed. Blood is a **complex liquid** comprised of dilution of protein and lipoprotein complexes- **blood plasma**, with suspended blood cells, **erythrocytes** in great majority. Suspension denotes the presence of large particles in a liquid which at rest sediment at its bottom. The percentage of total volume of blood occupied by cells is **hematocrit** (**Ht**).

When the complex fluid flows, the suspended particles (erythrocytes) mix with solvent (plasma), in a way that depends on a flow velocity and dimensions of a vessel. This is one of reasons why the flow of a complex fluid is hard to deal with exactly. In addition, the blood vessels are not rigid pipes; their caliber can change in two ways: passively, by elastic deformation, under the influence of blood pressure (all blood vessels) and by active contraction of their smooth muscle (primarily small blood vessels). Even more, the heart is not a continuous pump, ejecting blood periodically, only during the part of a cycle. Nevertheless, the tissues receive blood continuously, in a way that a part of blood ejected in a single contraction stores in elastic arteries; the subsequent relaxation of these arteries provides the constant tissue blood supply and fills the heart, preparing it for the new contraction. During these events, the blood constantly accelerates and decelerates, so that **inertial effects** must also be taken into account. On top of all, when the rate of flow exceeds the critical value the laminar blood flow becomes irregular, turbulent one (this also occurs in case of simple liquids).

Despite all these complexities, the main features of blood flow- **hemorheology** can be described fairly well.

Although the blood plasma is a thin solution of protein and lipoprotein complexes, having density similar to water, its viscosity exceeds the viscosity of water for about 30% on body temperature. The reason is presence of large molecules. The presence of cells causes that blood viscosity is much larger than viscosity of water. The average viscosity of blood is about 3.5 times the viscosity of water and exponentially increases with hematocrit.




We are addressing the average values because in circulation the viscosity of blood changes according to changes in hematocrit and conditions that influence on how easy erythrocytes change shape (**deformability**) and how many of erythrocytes glue together

(**aggregability**), making doublets, triplets, etc. Due to varying viscosity, we say that blood is **non-Newtonian liquid**.

The reason behind changes in hematocrit (and thus blood viscosity) is the **Fahreus-Lindquist effect**: erythrocytes in blood vessel tend to accumulate around the vessel axis, where the velocity of flow is greatest; only at central axis the lateral pressures on erythrocyte are equal (recall Bernoulli's equation!). In consequence, the layer near-by the vessel wall is erythrocyte-free. The width of this layer corresponds to erythrocyte size. In small vessels, the volume of this layer is significant. That is why the hematocrit in small vessels is lower than in large blood vessels. This does not explain the capillary hematocrit since those vessels are so tiny that erythrocytes must deform in order to go through (which slowers the passage of erythrocytes, providing more time for exchange of gases).

For a given **Ht** the viscosity of blood increases with increasing aggregability (aggregates behave as rigid ellipsoids) and decreases with increases in erythrocyte deformability (by deforming, erythrocytes adapt to conditions of flow). The model is:

$$\eta_{krv} = \eta_{plazma}\ \mathbf{e}^{\mathbf{a\,Ht}}$$

where parameter **a** increases with aggregability and decreases with erythrocyte deformability.

We believe that in microcirculation **Ht** drops to about 2/3 of its value in large vessels, while in slow flow conditions erythrocytes make aggregates more easily. The effect of hematocrit prevails and the viscosity in microcirculation is probably lower than in macro circulation. We cannot take this statement for granted since in microcirculation nobody has yet measured viscosity directly.

Increased plasma cholesterol concentration increases both plasma viscosity and erythrocyte aggregability. Increased cholesterol also affects adversely the vessel wall, making it more rigid, less distensible. This impairs the ability of a vessel to vasodilate in response to increased flow demands (flow autoregulation). Slower flow promotes erythrocyte aggregation, which, owing to increased cholesterol plasma concentration, is *per se* more prone to making aggregates. This makes a vicious circle, which can end-up in dying out of a tissue supplied by impaired blood vessels.

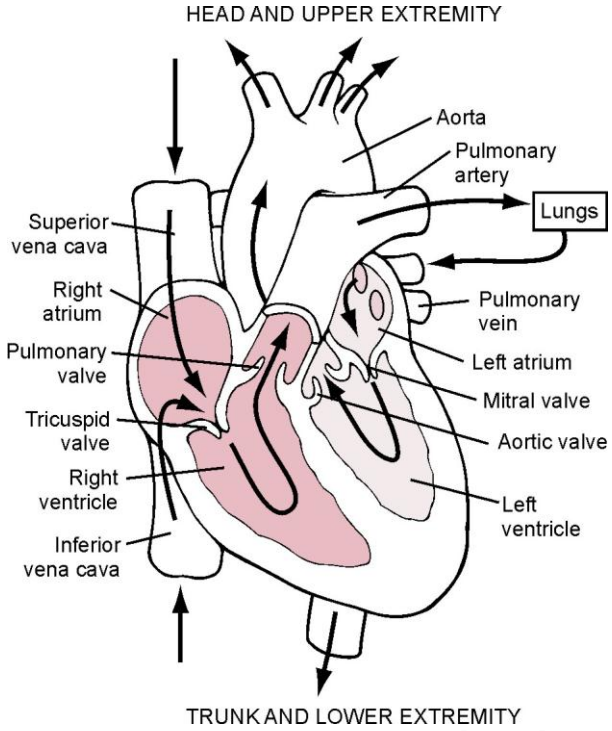
## Heart as a pump

In absence of viscous resistance, the heart should not work at all. It is the work of heart that provides sufficient pressure on its outlets (aorta on the left, pulmonary artery on the right site), which gradually spends as blood flows through circulation (systemic or peripheral and pulmonary or central); primarily in small blood vessels (recall the $1/r^4$ law).

In essence, heart is a double pump. The left heart, a stronger pump, accepts the blood from pulmonary veins and pushes it through peripheral circulation towards the weaker pump, the right heart. The right heart pushes the same amount of blood through pulmonary circulation, which offers much lower resistance. In pulmonary capillaries, the diffusion of respiratory gases occurs until initially low oxygen and high carbon

dioxide partial pressures in blood equate with high oxygen and low carbon dioxide partial pressures in alveolar air.

In healthy person, the two circulatory systems are in series: the blood output of the left heart comes to the right heart, then in lungs and back to the left heart again. Sometimes a child is born with heart defect, which disturbs normal blood flow. In most such cases, there is a disruption in integrity of heart septum, enabling some of blood to by-pass the normal circulatory route. In short, there is an **intracardial shunt**. The septal defect can be on atrial level (**ASD**) or ventricular level (**VSD**). Owing to larger pressures on the left site (at least in early phase of a disease), the shunting is from left to right heart.




Let us concentrate on a single cardiac cycle. It has two phases: the contraction phase- **systole** and relaxation phase- **diastole**. When there is no reference to specific muscle part, the phases apply to state of the heart ventricles. However, the atria also contract and relax, and thus have systole and diastole. More important functionally are systole and diastole of the ventricles. The disturbance in function of atria does not affect much the heart function. Atria are only widenings of pulmonary veins which facilitate ventricular filling. At rest, the ventricular filling is adequate even without contribution of atria. The role of atria is important in exercise and in diseases that affect adversely ventricular filling.

The two sites of heart work synchronously, in the same phase. After electrical stimulation from SA node, atria are first to contract, following by ventricles, some 0.1 to 0.2 sec later. During the period when both atria and ventricles relax, atria are filled with blood from great veins of peripheral and central circulation. The rise in atrial pressure opens **atrio-ventricular (AV) valves (mitral valve** on the left **and tricuspid valve** on the right site), which, if functioning properly, allow the passage of blood from atria to ventricles, but not in opposite direction. This initiates the filling of ventricles,

which, when about 80% finished, is facilitated by atrial contraction. The volume of blood in ventricles at the end of filling is **end-diastolic volume (EDV)**, normally around 120 ml. The contraction of ventricles first closes the A-V valves, then, when pressure in ventricles exceeds the pressure in output arteries, the output valves open (**pulmonary valve** on the right site and **aortic valve** on the left site; the common term is **semilunar valves**) and ventricles eject their **stroke volume (SV)**. Stroke volume is about 2/3 of EDV (70-80 ml). In other words the **ejection fraction** of a ventricle (**EF**), which is **SV/EDV** ratio, is normally around 65%. The rest is **end-systolic volume** (**ESV**) or **residual volume**. What follows is the next cycle, starting with relaxation of ventricles. At rest the average number of cycles in one minute **(heart rate, pulse)** is around 75, each cycle lasts 0.8 sec; 0.5 sec is diastole and 0.3 sec systole.

The pressures within ventricle and in output vessels produced by the left heart are around six times greater compared to the right site. The reason behind is that pulmonary circulation offers around six times less resistance compared to systemic circulation (small pulmonary blood vessels are more spread and large vessels are more compliant). Focusing on the left site, the several phases of heart action should be singled out:

1. The beginning of contraction of a ventricle closes the mitral valve. After that, as long as the pressure in a ventricle does not exceed the aortic pressure (around 80 mm Hg), the volume of a ventricle does not change (both input and output is zero); this is the phase of **isovolumic contraction**.

2. After the pressure in ventricle exceeds the aortic pressure, the aortic valve opens, the blood leaves the ventricle, the pressure in the ventricle and aorta still rises for some time, exhibiting maximum of around 120 mm Hg. This maximal aortic pressure is **systolic pressure**.

3. After that the pressure in ventricle falls, and, when it becomes lower than the pressure in aorta, the backpressure closes the aortic valve. The pressure in a ventricle falls towards zero value, and in aorta at around 80 mm Hg, which is the **diastolic pressure**.

4. During **isovolumic relaxation** mitral and aortic valve are closed**,** the pressure in ventricle decreases, but its volume does not change.

5. When pressure in ventricle falls below atrial pressure, **the phase of rapid filling** commences, when the blood from the pulmonary veins passes through atrium and fills the ventricle.

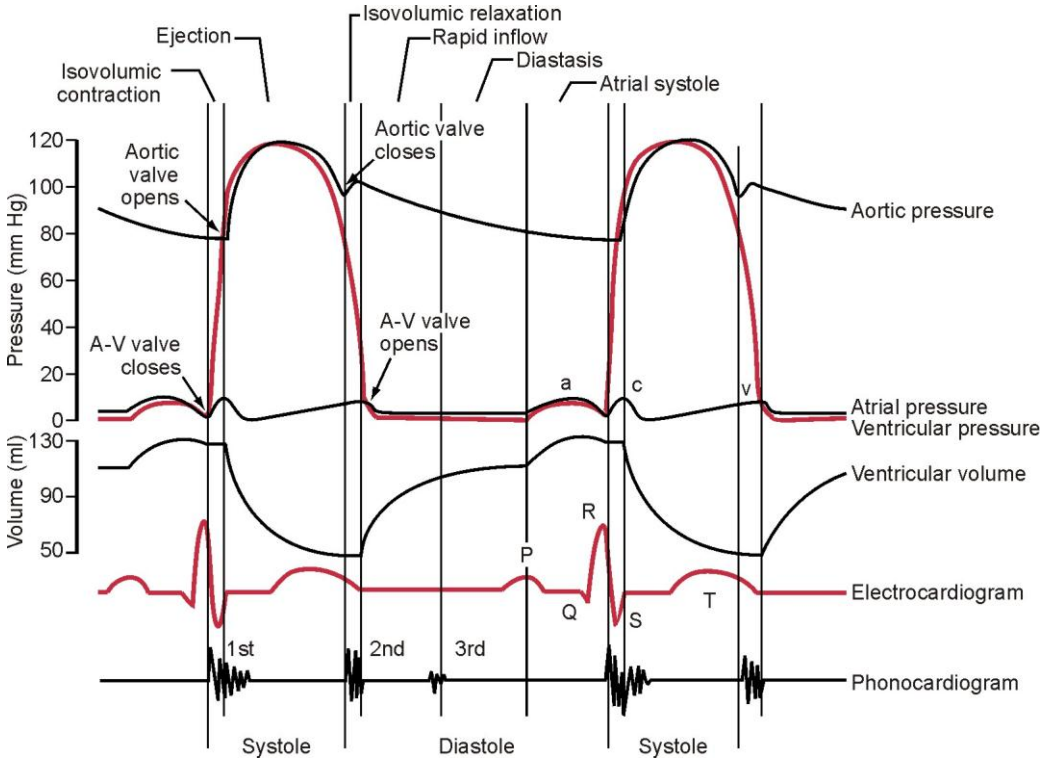
6. After ventricular filling is almost complete, the pressure in atrium equates with ventricular pressure and, during this period of **diastasis**, no blood enters the ventricle.

7. Atrial contraction (atrial systole) finishes the ventricular filling, which ends before the onset of the next **isovolumic contraction**.

Ventricular systole consists of phases 1. and 2. and the part of phase 3. until closing of semilunar valve, the rest is diastole. Notice that neither systole (contraction) nor diastole (relaxation) correspond exclusively to phases of blood ejection or filling, but contain also the phases during which ventricular volume does not change.

During blood passage through heart cavities and ejection to large arteries a part of energy is released in form of acoustic vibration. These sounds can be heard on the surface of chest cavity, especially with aid of **stethoscope.** The opening of heart valves

normally do not produce sounds, because these movements are relatively slow. On the contrary, during closing of valves the pressures and velocities are larger, causing the part of valves and neighboring fluid to vibrate in audible range. One first hears the closing of A-V valves- relatively long, low pitch sound, which is the **first heart sound**. At the end of systole the semilunar valves close, producing the **second heart sound**, which is shorter, higher pitch sound. Various heart abnormalities produce different sounds, commonly termed **murmurs**. The following figure illustrates the various events during the heart cycle.




The blood pressure in peripheral arteries can be measured because these vessels are near the surface of body and one can apply the external pressure (by encircling the limb by inflatable cuff, for example) to stop the flow. One measures this **occlusion** pressure (by Mercury or other manometer) and slowly lowers it. At the same time one listens to sounds produced by turbulent flow through narrowed artery. In this way, one can identify systolic pressure (before the onset of sounds produced by turbulences due to periodical openings of artery) and diastolic pressure (at time of cessation of sounds, when blood flows freely through unobstructed artery at all times during the heart cycle). The arteries in pulmonary circulation are deeper in body, so that one can only insert a catheter and measure the pressure directly, which is an invasive procedure.

## Basic hemodynamic equations

The heart blood output in a single minute is the **cardiac output (CO)**. It is the product of heart rate, **f**, and stroke volume:

$$\mathbf{CO} = \mathbf{SV} \times \mathbf{f}$$

Let's focus again on peripheral circulation. The pressure produced by left heart in aorta normally varies between 80 and 120 mm Hg. This pressure is almost totally wasted in peripheral circulation, so that right atrial pressure is only a couple of mm Hg. Thus, approximately the average pressure gradient in peripheral circulation equals the average aortic pressure ($\mathbf{P_a}$). This pressure is necessary to compensate for viscous loses in peripheral circulation. Applying the basic relation of fluid flow:

$$\text{Pressure gradient} = \text{Flow} \times \text{Resistance}$$

to peripheral circulation with total flow resistance **TPR**, one obtains:

$$\mathbf{P_a} = \mathbf{CO} \times \mathbf{TPR}$$

$$\mathbf{CO} = \frac{\mathbf{P_a}}{\mathbf{TPR}}$$

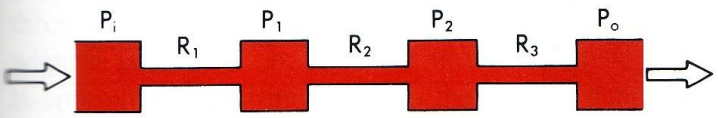
The above two equations (the variants of a single relation) should always be kept in mind when considering the control of blood flow (**hemodynamics**). Although, in the process of acquiring the steady state, each of the three variables may change in a different way, they are always bound by the same simple relation.

The common noninvasive manometers can not measure the blood pressure continuously, but only the extreme values, the systolic pressure ($\mathbf{P_s}$) and diastolic pressure ($\mathbf{P_d}$). Continuous noninvasive measurement of arterial pressure is possible by the method of **photopletismography**. Since the method is not widely accessible in practice the mean arterial pressure is commonly assessed as the weighted average between systolic and diastolic pressure. The weight factors are the parts due to systole or diastole in total duration of cardiac cycle. We can assume that at rest systole lasts 1/3, and diastole the remaining 2/3 of cardiac cycle (In ECG recording- time between neighboring R nods, known as **R-R interval**). With these assumptions:

$$\mathbf{P_a} = (2/3)\,\mathbf{P_d} + (1/3)\,\mathbf{P_s}$$

Observe a part of circulation, comprised by many blood vessels. Poiseuille's law gives the flow resistance of each vessel. Their total resistance depends on their calibers ($r^4$ dependence), but also on the way they are organized- in series or in parallel.

The same law of combining the resistances of conductors to current of electrons applies to flow resistance of blood vessels (**hydraulic resistance**). Serially organized resistors add together, whereas, in parallel design, one adds the inverse of resistances, or **conductances**.

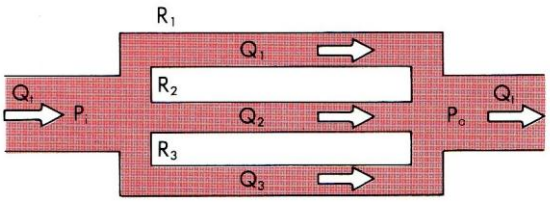



(a) $P_i - P_o = (P_i - P_1) + (P_1 - P_2) + (P_2 - P_o)$

(b) $\frac{P_i - P_o}{Q} = \frac{(P_i - P_1)}{Q} + \frac{(P_1 - P_2)}{Q} + \frac{(P_2 - P_o)}{Q}$

(c) $R_t = R_1 + R_2 + R_3$

**Fig. 26-9** For resistances ($R_1$, $R_2$, and $R_3$) arranged in series, the total resistance $R_t$, equals the sum of the individual resistances.




(1) $Q_t = Q_1 + Q_2 + Q_3$

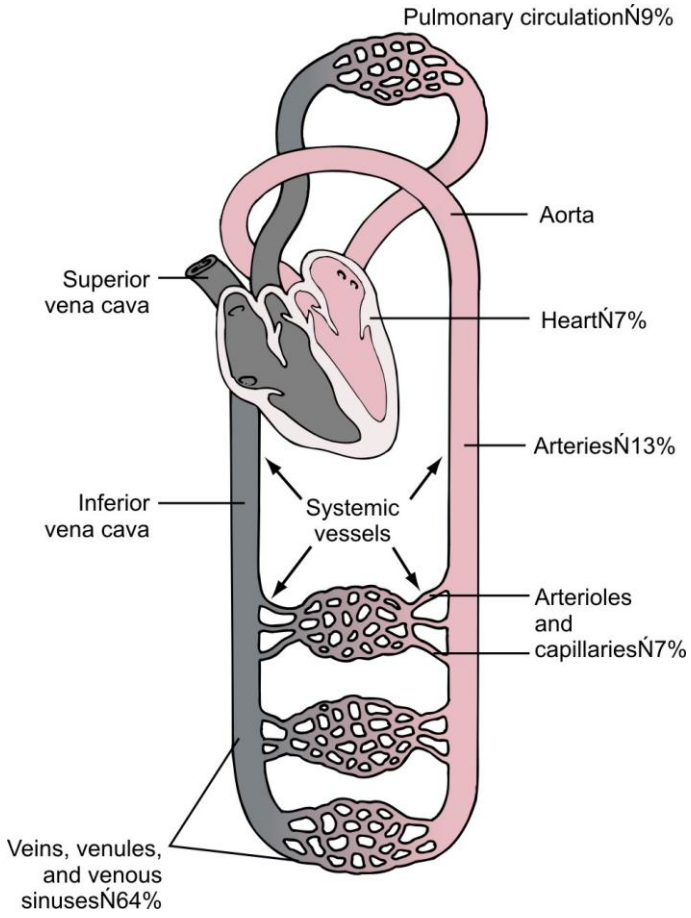
(2) $\frac{Q_t}{P_i - P_o} = \frac{Q_1}{(P_i - P_o)} + \frac{Q_2}{(P_1 - P_o)} + \frac{Q_3}{(P_i - P_o)}$

(3) $\frac{1}{R_t} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$

**Fig. 26-10** For resistances ($R_1$, $R_2$, and $R_3$) arranged in parallel, the reciprocal of the total resistance, $R_t$, equals the sum of the reciprocals of the individual resistances.

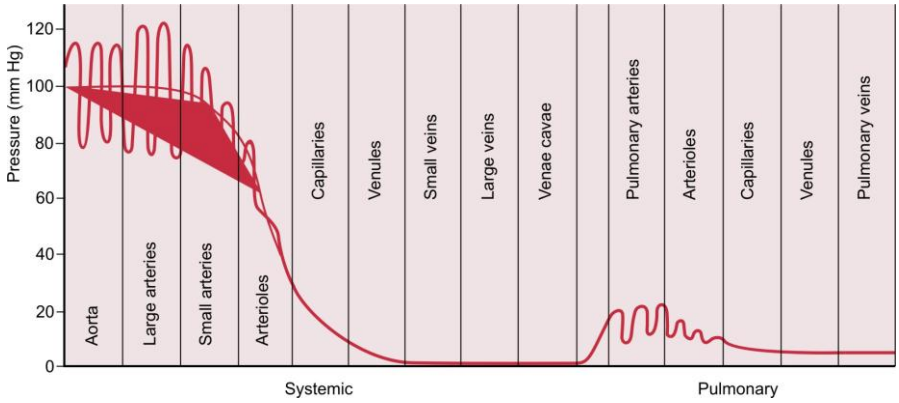
The circulations of organs are generally arranged in parallel, so that TPR is less than individual flow resistance of each organ. In the same way, the individual resistances of parallely combined circulatory segments of an organ are larger than total flow resistance of the organ. For example, the hydraulic resistance of the vessels of the left heart is greater than the hydraulic resistance of the whole heart. Even in case when one circulatory segment closes (infinite resistance), the heart will be able to perfuse the rest of parallel segments without significant extra burden. On the other hand, if the resistance of one of parallel routes declines significantly (zero resistance), the total

resistance declines much too (all blood redistributes to the low resistance route). In this way, TPR during heavy exercise may drop to only 20% of its rest value and blood flow through exercising muscles can increase up to 50 times!




On the contrary, if the blood flow is obstructed in a part serially connected with the rest of a system, the whole system is shutout from circulation. Closing of aorta or pulmonary artery stops the blood flow everywhere.

The circulation inlet of an organ is its main artery, which spreads in a system of smaller arteries, each of which spreads further in the network of arterioles and capillaries, supplying every tiny region of an organ. The capillaries further empty into venues, which join in the progressively larger body of veins. Each of these circulatory systems has its own flow resistance (depending on dimensions of vessels and degree of spreading). In that sense, we may say that the main artery is arranged in series with the system of smaller arteries, following by the system of arterioles, etc. Depending on the flow resistances of individual serial segments, the blood flow falls from the level in aorta to a tiny pressure in the right atrium. The largest drop in pressure occurs in arterioles, following by capillaries; therefore, the hydraulic resistances of those circulatory segments are largest. This is true despite the fact that total cross-section of arterioles is greater than the cross section of arteries, because the effect of their small lumen prevails ($r^4$ dependence!).



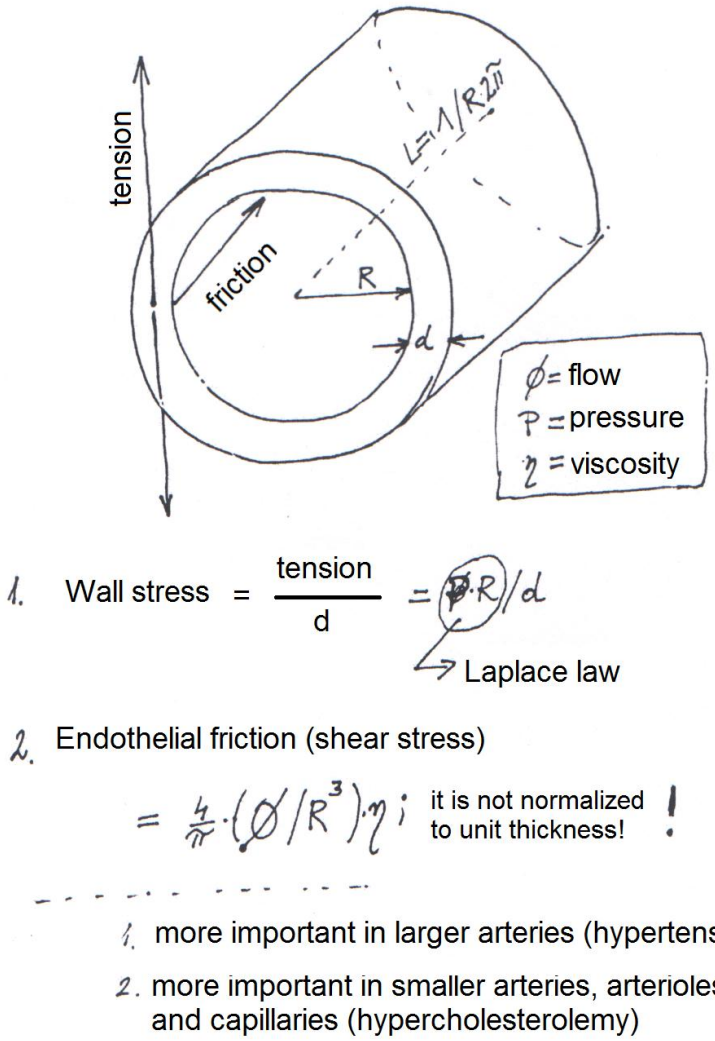

The blood vessel can change its resistance greatly by only small change of its lumen. If we consider the blood vessel as an elastic pipe, its resistance is determined by **transmural pressure** (difference in extramural and intraluminal pressure). This is a **passive** characteristic of blood vessel. However, blood vessel, especially arterioles, can change its lumen **actively**, by contraction or relaxation of its smooth muscle. This **autoregulation** of blood flow is independent or even against the change in transmural pressure.

It may look odd that the vascular smooth muscle is never relaxed, having so called **basal tonus.** It appears that the heart would have an easier job if all body arterioles were relaxed. Then, for the same cardiac output, in presence of reduced hydraulic resistance, the heart would have to produce much lower pressures. However, the blood flow demands of some organs may change very much. For example, during heavy exercise the majority of heart output redirects to working skeletal muscles, after meal the perfusion of stomach and intestines increases much, an increased mental activity requires an increase in brain perfusion, the part of a body affected by noxious substances needs an increased leukocyte delivery, etc. The redistribution of cardiac output would not be possible if some arterioles do not relax and other constrict. In conclusion, the basal tone of arterioles exists in order to decrease, in response to augmented perfusion demands of other body areas, which then take the larger portion of cardiac output.

## Stress in blood vessel

The flow of blood through blood vessels is both an essential need and major obstacle to longevity. Malignant diseases and diseases due to impaired endothelium of blood vessels cause virtually all fatal and serious diseases. Even what we call a natural death is mediated by general **atherosclerosis,** the occurrence of fatty layers in the wall of blood vessels. We used to think that the main cause of atherosclerosis is an increased concentration of cholesterol in blood plasma. However, it came out that endothelial damage enables the plaque constituents to protrude the blood vessel wall. The endothelial damage is due to stresses induced by flow of blood.

There are two types of stresses in blood vessel. One is caused by blood pressure, the other by blood flow friction.




The blood pressure creates a tension (reactive force per unit length) in a vessel's wall that opposes its distension. If a limit is exceeded a vessel tears apart. Consider the unit length segment of a cylindrical vessel intersected by a plane through its axis. The tension is the force opposing the separation of the two halves of this segment. The force on each half of a cylinder's superficial areas is the product of pressure **P** and radius of a cylinder base **R** (because, due to sphericity, only elements of a superficial area parallel with the cutting plane are taken into account; the exact derivation requires the use of integral calculus):

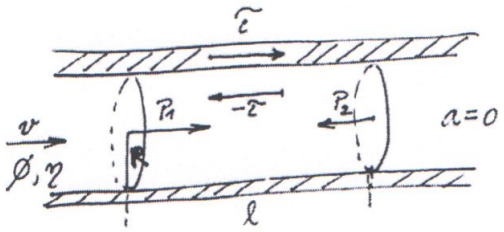
$$tension = -\mathbf{P} \times \mathbf{R}$$

The above relation is one of the forms of the **Laplace law**. The thicker the vessel's wall the harder the blood pressure deforms the vessel. It is thus natural to normalize the tension to a thickness of a vessel's wall. This quantity is called the **wall stress**. When we compare the risk of bursting between large and small blood vessels the wall stress is an appropriate measure. The blood vessel first distends easily (elastin distends, collagen

unrolls) until the excessive distension is opposed by unrolled, hard to distend collagen fibers.

Friction between blood and endothelial cells causes **the shear stress** on the surface of a vessel's wall. It is the reaction to the scraping force, primarily static friction with water molecules. We define it as the force per unit wall area. Since only the surface of a wall suffers this stress, the vessel's thickness is irrelevant (unlike in case of wall stress; both quantities, however, have the same dimension- force/surface area!). According to Newton's law of viscosity (the outline of derivation is below), this stress is proportional to blood flow **Φ** through the vessel and to blood viscosity η, but falls with third power of its lumen **R**:

$$\text{shear stress} = (4/\pi)\eta\, \mathbf{\Phi}/\, \mathbf{R}^3$$




$$\frac{dv}{dt} = a = 0 \Rightarrow (P_1 - P_2)\, R^2\pi = 2R\pi\, l \cdot \tau$$

shear stress of blood vessel wall [N/m²]

- Newton's law of viscosity

$$\tau = \eta \cdot (dv/dx)_{\text{wall}}$$

$\Rightarrow$ (a bit of math)

$$\tau = \frac{4}{\pi} \cdot \eta \cdot \Phi / R^3$$

Comparing these two stresses:

1. by dilating the blood vessel the wall stress increases and shear stress decreases and vice versa;

2. The wall stress may reach critical values sooner in large than in small blood vessels (especially in aorta in hypertensive persons), while shear stress causes more damage to arterioles and capillaries (especially in
persons with hypercholesterolemia).

## Work of heart and energy expenditure

Consider the work output and energy expenditure of a heart muscle during a single cycle. In heart, there are no antagonistic muscles for active dilatation of the ventricles to create the negative pressure and thus suck in the blood. The filling of ventricles is *passive*, on account of greater pressures in elastic great veins and atria. The energy needed to fill the ventricles is stored as an elastic deformation during heart contraction in the previous cycle. In short, contraction of heart muscle provides for filling of the ventricles in the subsequent cycle. Therefore, the heart works only during systole, during diastole the work is performed on the heart.

During contraction of heart muscle, about 85% of energy released by splitting of adenosine triphosphate appears as heat. The rest accounts for:

1. frictional heat, released during pushing the stroke volume through semilunar valves;
2. work of transferring and accelerating the stroke volume from lower pressure areas in great veins to higher pressure areas in large arteries.

Normally, at rest, the energy lost in traversing the semilunar valves is negligible component of the work of heart.

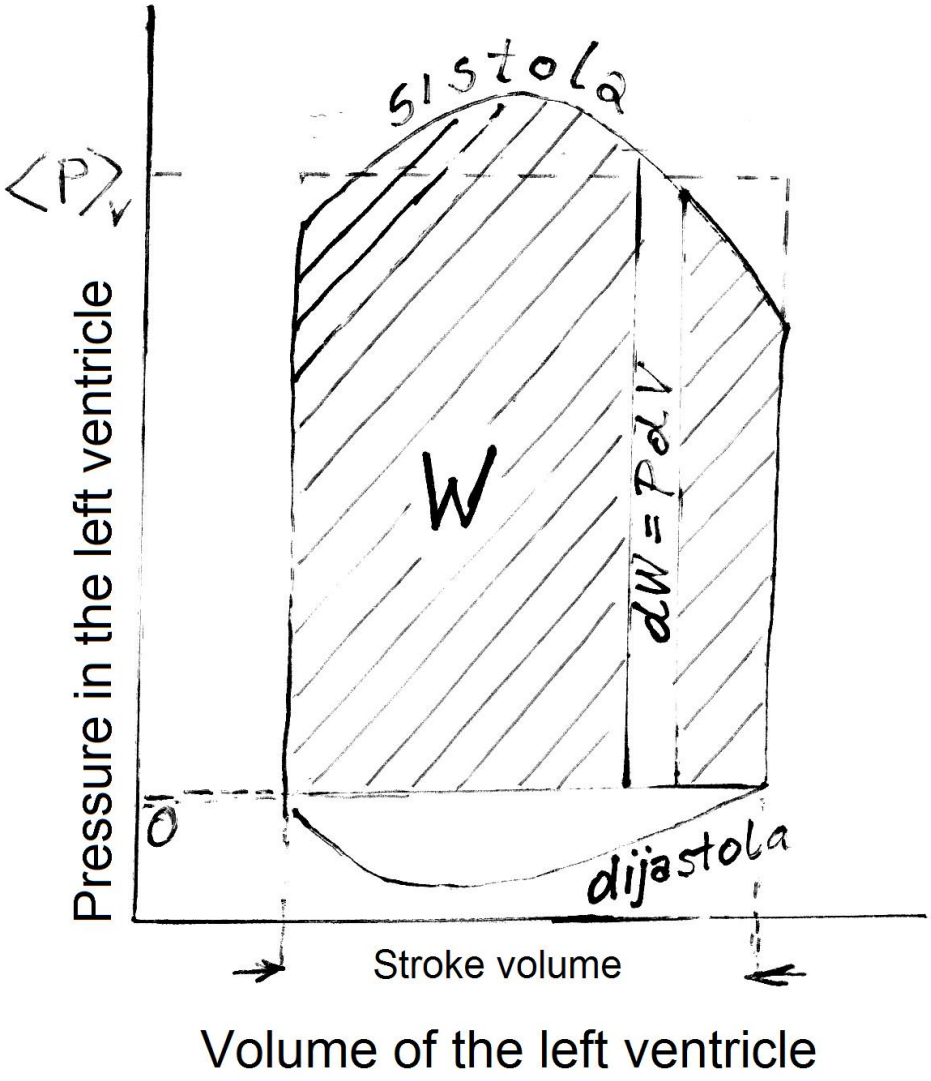
The work and energy expenditure of the left heart is around six times greater compared to the right heart. Let us concentrate on the left heart. The pressure in the left ventricle can be directly measured by inserting a catheter with the manometer. If pressure during systole were constant, the work of the left heart would simply be the product of stroke volume and pressure (the work is force x distance = pressure x volume). However, the pressure changes and **systolic work** of a ventricle can be equated to the sum (integral) of tiny contributions **dW**; the products of a pressure **P** at volume **V** and tiny volume **dV**, which is the area under ventricular P-V curve:

$$\mathbf{W} = \int_{\text{sstola}} \mathrm{P}\, \mathrm{dV}$$

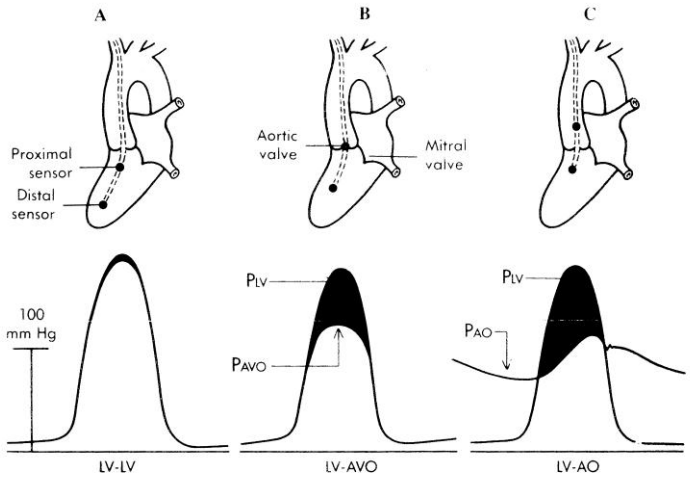
However, we do not consider the total area under P-V curve, but only its part over the line of the pressure at the end of diastole (since this pressure already exists before the onset of systole), as shown on the figure bellow.

In other words, the work of the left ventricle during a single systole can be equated to the product of volume-averaged pressure in the ventricle $\mathbf{<P>_V}$ and its stroke volume, **SV**:

$$\mathbf{W = <P>_V\ SV}$$




This quantity, aside from work needed to transfer the stroke volume to aorta, includes the work wasted as heat on traversing the aortic valve. That is why, during systole, the pressure in ventricle exceeds the pressure in aorta, especially during the fast ejection phase in the beginning. It is often wrongly quoted that the work of heart, as defined above, does not account for the gain in kinetic energy when blood accelerates in aortic valve. However, no external energy is needed for this acceleration, since a liquid itself provides for it. In aortic valve, the increase in dynamic pressure (kinetic pressure) accompanies the equivalent decrease in static pressure (potential energy). The reverse happens when blood decelerates after leaving the valve (recall Bernoulli's equation). Thus, aortic valve causes irreversible decrease in blood pressure, due to frictional heat loses, and only reversible decrease in static pressure, which accounts for transient acceleration of blood. Incidentally, these two pressure drops are approximately equal, which is used in diagnostic assessments of the degree of valvular stenosis (ultrasonically one can assess the velocities, not pressures.




■ **Fig. 26-4** Pressures *(P)* recorded by two transducers in a patient with aortic stenosis. **A,** Both transducers were in the left ventricle *LV-LV*. **B,** One transducer was in the left ventricle, and the other was in the aortic valve orifice *(LV-AVO)*. **C,** One transducer was in the left ventricle, and the other was in the ascending aorta *(LV-AO)*. (Redrawn from Pasipoularides A et al: *Am J Physiol* 246:H542, 1984.)

Measurement of pressure in the left ventricle is an invasive procedure, seldom used in humans. On the contrary, the pressures in aorta can be simply and noninvasively assessed, especially the peak values, the systolic and diastolic pressure. Further, the pressure in the left ventricle is around or slightly above the aortic pressure during the time of systole. For simplicity, let us approximate the volume-averaged systolic pressure in the left-ventricle by the systolic arterial pressure, **Ps**. Under this assumption, the systolic work of the left ventricle is simply assessed ad the product of systolic arterial pressure and stroke volume:

$$\mathbf{W = P_s\ SV}$$

This approximation is not valid in aortic stenosis, when the ventricular systolic pressure, due to increased resistance of the valve, must exceed the aortic pressure, and the work of left heart is greater than predicted by the equation above. To a lesser degree, the same holds in intensive aerobic exercise. Then the cardiac output is much increased, the flow velocities are fast and significant heat loses occur in normal semilunar valves also.

**The systolic work of heart** is the sum of systolic works of left and right ventricles. In normal conditions, the stroke volumes of the ventricles are equal, while, compared with the left ventricle, the systolic pressure and work of the right ventricle is about six times less.

**The power of heart contraction** equals the systolic work of heart divided by duration of systole (not with total duration of a cycle). The **minute work of heart**

equals the product of heart rate and systolic work of heart. Since the cardiac output is the product of heart rate and stroke volume:

$$\text{minute work of left ventricle} = \mathbf{P_s\ CO}$$

Notice that the above work, aside from obvious dependence on cardiac output, is proportional to systolic arterial pressure. This means that it depends not only on peripheral hydraulic resistance (defining the mean arterial pressure), but also on aortic distensibility. The more distensible aorta is, the lesser rise in diastolic pressure is required to accommodate the given stroke volume, lowering the systolic pressure and work of heart.

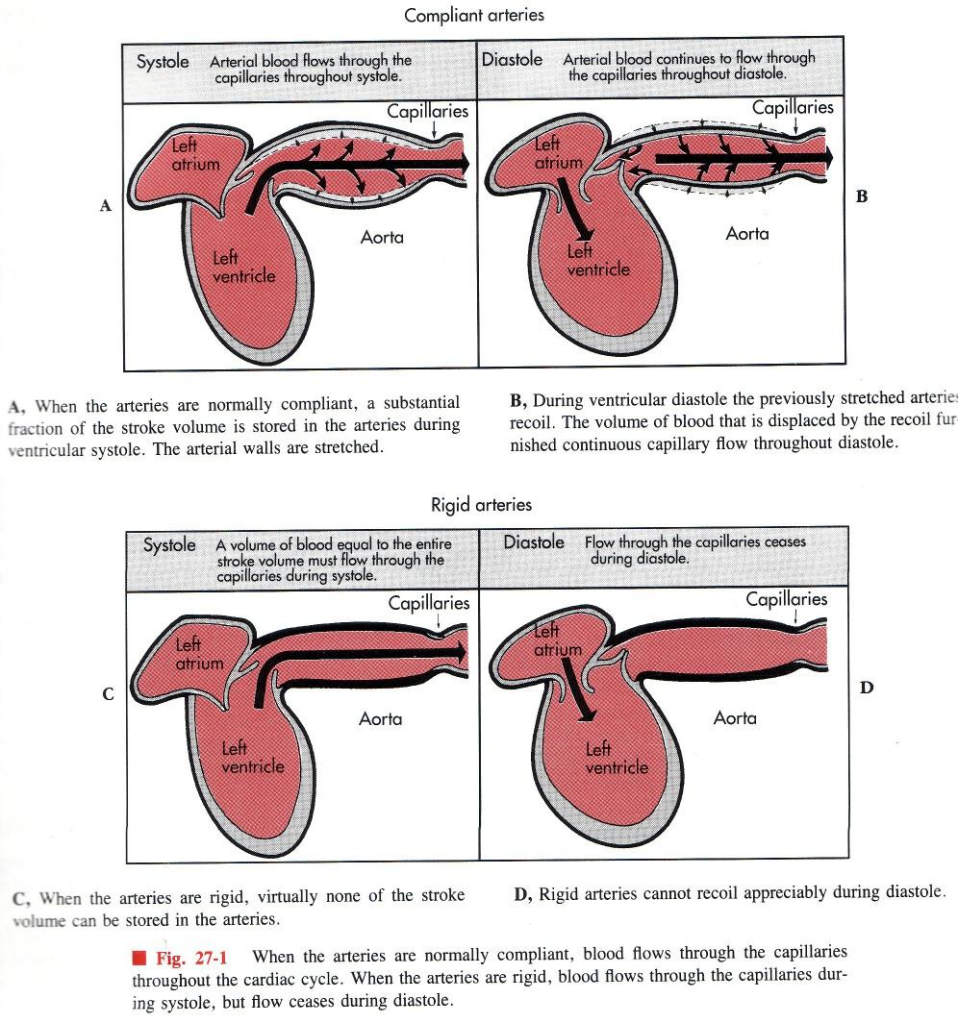
Accordingly, the work of heart per unit cardiac output increases both in **hypertension** (condition characterized by increased resistance of arterioles and capillaries), and when aorta becomes stiffer. The first condition increases the mean arterial pressure (approximately mean between diastolic and systolic pressure), the later augments the difference between these extremes; i.e. the systolic blood pressure.

We should now consider the central circulation also. The distinguishing features of pulmonary circulation are (i) greater spreading of vessels, causing about six times lesser hydraulic resistance and thus pulmonary mean arterial pressure and (ii) greater distensibility of pulmonary artery and thus lesser difference between systolic and diastolic pulmonary arterial blood pressures (25/15 mm Hg). However, the right myocardium is much thinner and weaker than the left one. In consequence, the pathologic changes on pulmonary blood vessels, or other causes of increased hydraulic resistance of central circulation have in general more serious consequences than the corresponding burdens to the left heart.

The amount of oxygen consumed by the heart is proportional to myocardial oxygen consumption. It follows that an increase in work of heart requires increased myocardial perfusion and thus oxygen delivery to the muscle. The energy expenditure in myocardium exceeds many times the work of heart. The **efficiency of the work of heart** in normal conditions is only between 15 and 20%. These figures become even smaller if heart works against increased arterial pressures. Thus, although the work of heart remains the same if one halves the stroke volume and doubles the systolic arterial pressure, the energy expenditure in this hypothetical case increases. One says that **the volume work of heart** is more efficient than **the pressure work of heart**. This contributes to vicious circle in aortic stenosis. In this condition, the left ventricle has to overcome an increased resistance of the narrowed aortic valve opening by producing larger pressures and thus performing more work. The additional burden is that the pressure work increases, demanding even larger energy expenditure. However, the substantially increased perfusion demands are confronted by problems due to accelerated blood flow through the valve and related decrease in lateral pressure at Valsalva sinuses (recall Bernoulli's equation), which are the inlets of heart circulation.

## Arterial system

The primary task of pulmonary and peripheral arterial systems is blood supply to tissue capillaries. Their roles of large and small blood vessels are different. The arterioles account for majority of hydraulic resistance, while large arteries and veins contain almost all of the blood. By complex, partly known mechanisms, the tension of smooth muscle of arterioles changes, adapting the regional hydraulic resistance, mainly in response to metabolic needs. In contrast, the large arteries passively distend in response to blood pressure changes. This serial arrangement of highly distensible and highly resistive systems is a **hydraulic filter,** an analog to the electrical filter, composed of capacitor and thermal resistor. The function of hydraulic filter is to convert the intermittent heart output into steady flow in tissue capillaries. The heart is an intermittent pump. Only about 1/3 of stroke volume flows through circulation during systole, supplying blood to tissues during this period of a cycle. The rest of stroke volume is stored in elastic arteries, mainly in the first third of systole, i.e. during only 1/9 of a cycle. The elastic recoil of arteries provides the flow of blood during diastole (longer period of a cycle), when heart stops pumping and arterial pressure declines. This provides virtually constant perfusion of tissues.




**A,** When the arteries are normally compliant, a substantial fraction of the stroke volume is stored in the arteries during ventricular systole. The arterial walls are stretched.

**B,** During ventricular diastole the previously stretched arteries recoil. The volume of blood that is displaced by the recoil furnished continuous capillary flow throughout diastole.

**C,** When the arteries are rigid, virtually none of the stroke volume can be stored in the arteries.

**D,** Rigid arteries cannot recoil appreciably during diastole.

■ **Fig. 27-1** When the arteries are normally compliant, blood flows through the capillaries throughout the cardiac cycle. When the arteries are rigid, blood flows through the capillaries during systole, but flow ceases during diastole.

If arteries were rigid pipes, the flow through capillaries would cease during diastole, when the arterial pressure would drop to zero. The heart would have to push all stroke

volume through circulation during systole, which is only 1/3 of a time it normally does. Due to increased viscous losses, such intermittent flow would require larger arterial pressure; i.e. larger work of heart and energy consumption. We have arrived at the same conclusion in the preceding section.

The aortic pressure-volume curve on the figure below was obtained during autopsy of men of various ages. One ties off all aortal branches and adds volumes of a liquid in that closed system, observing the corresponding pressures. The steeper the curve is the easier aorta distends. Observe the sigmoid shape of a curve in the youngest group; the curve is linear, except at small and large volumes, when aorta opposes distension more than at moderate volumes. A child inflating the balloon has a similar experience- the hardest effort is at the beginning and just before the balloon bursts. The older groups characterize curves of progressively lesser steepness. At each point of a curve, one defines the **aortic compliance** as the ratio of a differential change in volume (**ΔV**) and pressure (**ΔP**):

$$\text{aortic compliance} = \frac{\Delta \mathbf{V}}{\Delta \mathbf{P}}$$

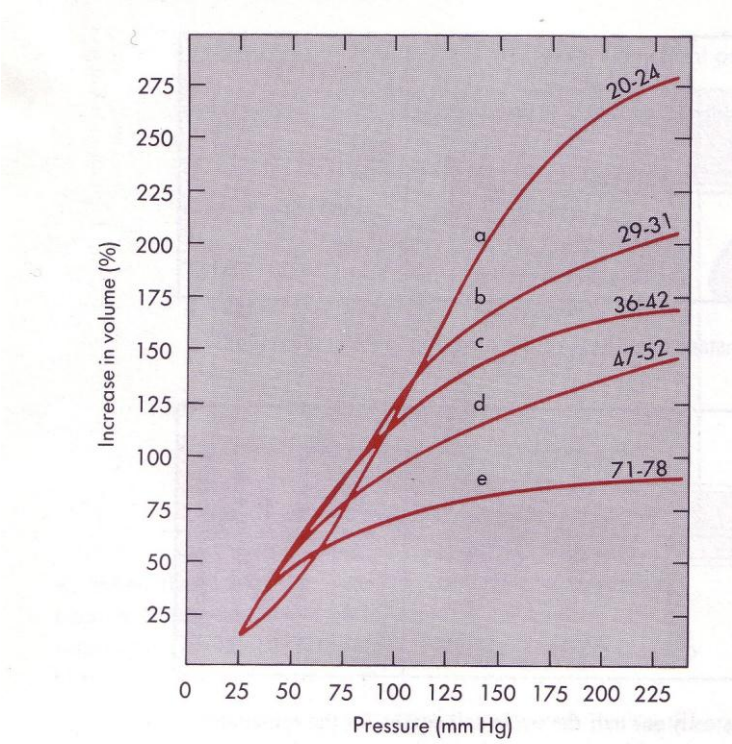



**Fig. 27-3** Pressure-volume relationships for aortas obtained at autopsy from humans in different age groups *(denoted by the numbers at the right end of each of the curves)*. (Redrawn from Hallock P, Benson IC: *J Clin Invest* 16:595, 1937.)

One should observe that aortic compliance depends on its volume; only for young persons, we may speak about constant slope of P-V curve at moderate volumes of aorta. In addition, we cannot compare aortic compliance between persons of various body sizes; the larger person will usually have larger aorta, which can accommodate larger increase in volume, despite eventually less distensible tissue of aortic wall. To obtain

the anthropometrically independent parameter, one defines the **aortic distensibility** as aortic compliance normalized to its relaxed volume, $V_0$:

$$\text{aortic distensibility} = \frac{\Delta \mathbf{V}}{\Delta \mathbf{P}} \frac{\mathbf{1}}{\mathbf{V_0}}$$

Neither aortic compliance nor distensibility can be measured in a living human. The parameter that can be assessed in clinical practice, by combining ultrasonically measured dimensions of blood vessels and noninvasive determination of arterial pressure is **arterial module of elasticity**:

$$\text{arterial module of elasticity } (\mathbf{E_p}) = \frac{\Delta P}{\Delta D/D}$$

where **ΔP** is the difference between systolic and diastolic arterial pressure- **the pulse pressure**, **D** the average diameter of aorta during cycle and **ΔD** the largest change in its diameter. Notice that module of elasticity, as distensibility, is anthropometrically independent quantity; neither pulse pressure nor the relative changes in aortic volume should depend on body size. The greater module of elasticity the lesser aortic distensibility is.

We have already seen that a decrease in aortic compliance raises the systolic arterial pressure. However, diminishing of aortic compliance (usually with aging) does not necessarily affect the mean aortic pressure. For the given cardiac output the mean arterial pressure depends only on peripheral resistance, which is determined by status of small blood vessels, regardless on aortic dimension or elasticity. This has an important implication: an increase in systolic pressure, in response to diminished aortic compliance, must be followed by a decrease in diastolic pressure, if peripheral resistance did not change. In this way the mean arterial pressure (index of peripheral resistance!) is kept constant.

In fact, a decrease in aortic compliance increases pulse pressure. Let us inspect this in some details. The left ventricle ejects the majority of stroke volume in the first third of systole, during the fast ejection phase, when the aortic pressure increases from diastolic to systolic value. In that time the aortic volume increases to accommodate the stroke volume minus the volume pushed through capillaries during that time (about 1/9 of SV). This volume increment is called the **systolic increment in aortic volume**. Thus:

$$\text{pulse pressure} = \frac{\text{systolic increment in aortic volume}}{\text{aortic compliance}}$$

Given the stroke volume, the systolic increment in aortic volume is also given. In that condition, a decrease in aortic compliance will augment the pulse pressure and vice versa. However, a change in hemodynamic conditions (e.g. during exercise) will also affect the stroke volume and pulse pressure.

Let us emphasize again the different roles of arterial large and small blood vessels. For the given stroke volume the pulse pressure (deference between systolic and diastolic pressure) is determined by aortic compliance, while arterial tonus determines the mean arterial pressure.

As an example, consider the normal arterial pressures of 80/120 mm Hg. The **arterial hypertension** (in short- hypertension) denotes the state of increased peripheral resistance, in most cases of unknown causes. The conservation of cardiac output requires an increase in mean arterial pressure (according to relation $P_a$ = CO x TPR). If aortic compliance is not affected, as expected in younger people, the pulse pressure will remain normal (40 mm Hg), and typical values of arterial pressure could be something like 110/150 mm Hg. However, hypertension is more common in older people, which usually also have lowered aortic compliance, resulting in aortic pressures like 110/180 mm Hg, with increases in both the mean arterial and pulse pressure.

## Gravitational effects

The circulatory pressures are not only due to heart contractions, but also to gravity, and thus depend on body posture. Thus far we have neglected hydrostatic pressure, due to blood weight, taking into account only the hydraulic pressure, produced by the pump. Gravity does not affect much when a person is lying, but has profound effects in standing posture. In standing posture, the referent plane is a horizontal through the right atrium, which intersects the heart outlets. Heart produces the initial pressures at the arterial outlets, which increase (down) and decrease (up) with vertical distance, $h$, from the referent plane for $\rho gh$, where $\rho$ is the blood density. This is why the peripheral arterial pressures in lowest body parts of a standing man are about three times larger than in highest parts (in smaller arterial vessels and in venous circulation one should also consider the viscous pressure loses).

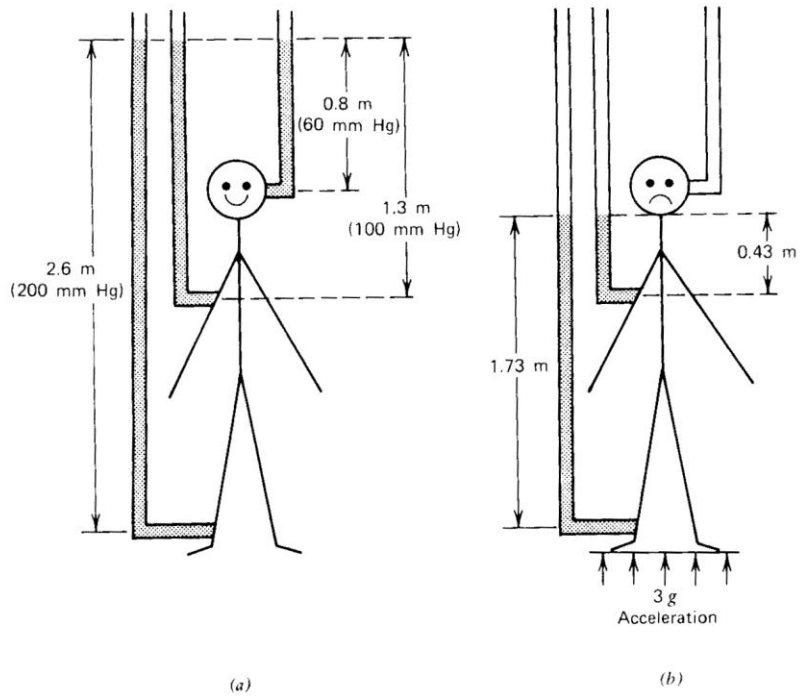


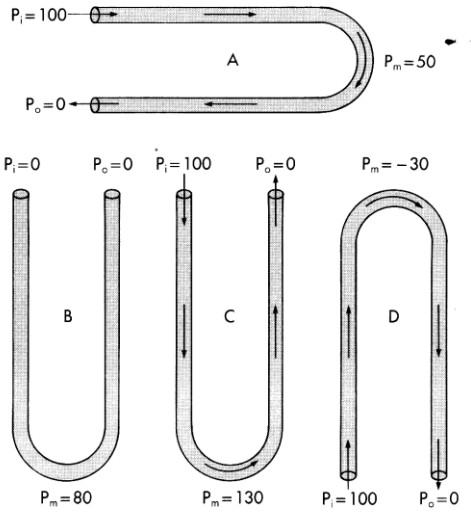

**Figure 8.8. (*a*) If glass capillaries were connected to the arteries at different locations the blood would rise to about the same level. (*b*) If the body were accelerated upward at 3 *g*, the blood would not reach the brain and black-out would result. If the body were horizontal the blood pressure would be about the same at the three points instead of differing by a factor of over three as shown here.**

Body posture also influences on blood flow. It is known that long still standing can result in loss of conciseness, secondary to decreased cardiac output. This fact is commonly misinterpreted. The explanation that standing impairs the venous return of blood neglects the compensatory effect on the arterial circulation, a part of the same loop (in pulmonary circulation these effects are only reversed). In this way nothing is lost or gained. To see it in more detail, consider the peripheral circulation as a U-tube. If blood vessels were rigid, the flow through U-tube would not depend whether the tube lies in horizontal or vertical plane or whether the openings face up or down. This is clear from the figure below, keeping in mind that in basic hemodynamic relation:

$$\text{flow} = \text{pressure gradient}/\text{resistance}$$

the pressure includes both the hydraulic component (produced by heart pumping) and the hydrostatic one. So, the above relation translates to:

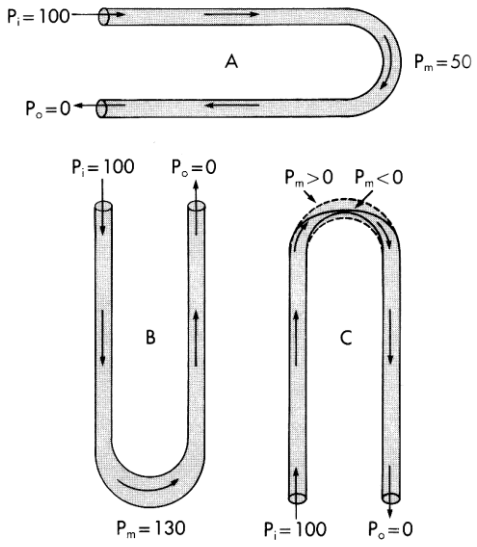
$$\text{flow} = (\text{hydraulic pressure gradient} - \rho g \Delta h)/\text{resistance}$$




■ **Fig. 30-15** Pressure distributions in *rigid* U tubes with constant internal diameters, all with the same dimensions. For a given inflow pressure ($P_i$ = 100) and outflow pressure ($P_o$ = 0) the pressure at the midpoint ($P_m$) depends on the orientation of the U tube, but the flow through the tube is independent of the orientation.

Thus, although pressures within U-tube depend on its orientation, this does not affect the flows, as long as the system is rigid.

The blood vessels are distensible, and their flow resistance can be modeled by the distensible U-tube, as on the figure below. The lower parts of a tube, owing to larger hydrostatic pressure would stretch more, which induces changes in hydraulic resistance, which influences on pressures and flows.




■ **Fig. 30-16** In U tubes with a *distensible section* at the bend, even when inflow ($P_i$) and outflow ($P_o$) pressures are the same, the resistance to flow and the fluid volume contained within each tube vary with the orientation of the tube. $P_m$, pressure at the midpoint of the tube.

This model suffices to explain the transient perfusion interruptions of apical (upper) parts of lungs in a standing person. The pressures produced by the right ventricle (25/15 mm Hg) suffice to raise the blood in the upper parts of lungs during systole, but not necessarily during diastole. The blood density approximately equals the density of water, so that the 15 mm Hg pressure corresponds to the blood column 15 x 13, 6 mm ≈ 20 cm in height. In tall person the vertical distance between pulmonary valve and top of lungs may exceed 20 cm. In that case the blood flow through highest lung areas temporarily stops and collapsible vessels collapse. The flow reestablishes when the local pressure exceeds the opening pressure of these vessels.

However, this model of distensible U-pipe does not explain the decrease in cardiac output when one gets-up or stands for a long time. The important fact not considered is that the blood volume is constant, so that an increase in blood volume at one site must result from blood that left the remainder. In this way, after getting up the blood accumulates in compliant veins of the legs, on account of blood above, mainly the central circulation (heart and lungs). Diminished volumes of heart cavities, especially at end-diastole, cause a decrease in stroke volume (if heart rate does not increase). This is so-called **Frank-Starling principle**; the fact that relates to better relationship of ventricular actin and myozin fibers in stretched muscle. Even this is not the whole story, since lowered cardiac output invokes an increase in activity of sympathetic autonomous system, leading to increase in myocardial activity, heart rate and vasoconstriction of leg veins. Only when this compensation is not fast enough or persistent a decrease in arterial pressure and blood flow may occur after changing posture from lying to standing or after standing still for a long time, respectively.

## SELF ASSESSMENT

1. At rest the mean aortic flow is 100 ml/s, and its mean diameter 2cm. The cross-sectional area of all peripheral capillaries is 8000 $cm^2$ and, at rest, 1/8 of them are open. Which is the mean blood velocity in capillaries at rest?

2. The total aortic pressure in the moment during systole is 13.3 kPa (100 mm Hg) and the velocity of blood near Valsalva sinuses 200 cm/s. Which is the pressure pushing the blood through Valsalva sinuses at that moment? Assume that those sinuses are at right angle to proximal aorta axis, neglect the radial changes in blood flow velocity.

3. During anesthesia the following patient data are obtained:

mean systemic blood pressure = 13.3 kPa (100 mm Hg)
mean pulmonary artery pressure = 6 kPa (45 mm Hg)
mean blood flow through peripheral circulation = 5 l/min, and
normal (or slightly elevated) left-atrial pressure

Which 4 of the following 8 statements are true beyond doubt and why?

A. The patient has pulmonary hypertension (elevated blood pressure in pulmonary artery).
B. The patient has primary pulmonary hypertension (pulmonary hypertension due to increased hydraulic resistance of pulmonary blood vessels).
C. The patient has secondary pulmonary hypertension (pulmonary hypertension due to left heart failure).
D. The patient has hyperdynamic circulation (increased blood flows through both systemic and pulmonary circulation).
E. The patient has left-to-right shunt (pathological communication between left and right heart, causing increased pulmonary blood flow).
F. The patient either has primary pulmonary hypertension or left-to-right shunt.
G. To arrive at right diagnosis one parameter is missing.
H. To arrive at right diagnosis one parameter is redundant.

4. After loss of blood due to bleeding, the person's hematocrit decreases from 0.5 to 0.35. Which is the relative change in blood viscosity of a person? Assume that, before loss of blood, the person's blood viscosity was 3 times the plasma viscosity.

5. During exercise the heart rate increases from 70 to 110 beats/min, the cardiac output from 5 to 10 l/min, and systolic aortic pressure from 13 to 15 kPa. Assess the relative changes in work of heart and power during one cycle and during 1 min. Approximate the volume-averaged left ventricular systolic pressure by aortic systolic pressure; assume that systole occupy 1/3 and ½ of heart cycle at rest and in exercise, respectively.

6. The blood vessel branches in two equal smaller vessels. Which is maximal relative decrease in lumen of these vessels which does not increase the blood shear stress after branching? What, in this borderline case, happens with wall stress and unit-length hydraulic resistance?

7. After large meal the cholesterol concentration in plasma increases for 20%. By this, the possibility of turbulences (only one statement is true):

a) remains the same

b) data on change of blood density is missing
c) data on change of thrombocyte aggregation is missing
d) increases
e) decreases

8. At rest the arterial blood pressures of a person are 80/120 mm Hg. During the steady state of exercising on bicycle ergometer, the person's cardiac output doubles, total peripheral resistance declines to 20%, and heart rate increases for 60%. Possible values of his arterial blood pressure during ergometry are (only one statement is true):

a) 90/140 mm Hg
b) 100/140 mm Hg
c) 130/190 mm Hg
d) 140/180 mm Hg
e) 120/140 mm Hg