Compendium of Medical Physics, Medical Technology and Biophysics

for students, physicians and researchers

By Nico A.M. Schellart

Dept. of Medical Physics,
Academic Medical Center
University of Amsterdam
Amsterdam

Second edition, version 3

Introduction

This compendium describes from the physical point of view in simple concepts the principle of working of some 100 methods, techniques, apparatus and equipment used in medical practice. These subjects, divided in various chapters, have been written for especially students in medicine, and for students in medical biology, medical information and other medical disciplines. However, the text is also readable for clinicians and other academic physicians, and moreover for paramedical scientists and technicians with a higher education.
This compendium gives for some subjects more explanation than one would expect in a typical compendium. For extensive explanation, especially when it respects derivation of equations the reader is directed to textbooks and other publications (journals, the internet).

The subject descriptions comprise three parts.
The first one, "**Principle**" gives the basic principles.
''**Applications**" is devoted to medical applications, but often also gives a very comprehensive description of applications in science, technology and daily live. In many cases "Application" does not literally consider a method, but the physical behavior of a biomaterial, organ or tissue governed by the described working principle.
A more detailed description of "Principle" can be found in "**More Info**". In general, this last part is also readable for students in the above mentioned disciplines, but in other cases it gives details and occasionally equations that require more knowledge of physics and mathematics.

The compendium has not yet been finished (various subject descriptions are still under construction) but in its present form launching can better not be delayed since this compendium covers the majority of the field. Underlined blue text fragments refer to chapters (occasionally in preparation). Their electronic links has not yet been implemented.

To write the compendium innumerous sources, especially the internet, papers in biomedical journals and textbooks are scrutinized. Important ones are given at the end of the subject descriptions. Many text fragments, figures and tables of the Wikipedia Free Encyclopedia were used, mostly modified but also literally adopted. I greatly acknowledge all the concerning institutions their authors for the free use of this material.

This compendium (a slightly shorter version with a different lay-out) can also be found at other websites:

http://onderwijs1.amc.nl/medfysica/compendium.htm (normal size)
http://onderwijs1.amc.nl/medfysica/compendium2.htm (PDA format)

I hoop that this electronic book meets a need, since comprehensive books in this field accessible for non-physicists are scarce.
I would very much like to obtain comments about obscurities, errors, etc. in order to improve this publication (n.a.schellart@amc.uva.nl, telefax +31-20-6917233, Dept. of Medical Physics, PO Box 22660, 1100 DD, Amsterdam).

Nico A.M. Schellart, June 2007.

Contents

SI units and derivatives, abbreviations, constants and variables

SI units and derivatives

In this manuscript the basic SI (Système Internationale) units (or their derivatives) are preferably used. The seven basic units are: m, kg, s, A, K, mol and cd.

J	Joule, unit of energy (/ 1 N·m)
K	Kelvin, absolute temperature; 273.15 K / 0 °C
kg	kilogram
m	meter
mol	1 mol gas at 1 bar and 0 °C comprises $6.02 \cdot 10^{24}$ particles and has a volume of 22.7 L
N	Newton, unit of force (/ 1 kg·m·s^{-2})
Pa	Pascal / N/m^2, 1 bar / 10^5 Pa
s	second
W	Watt, unit of power (/ 1 J/s)

Abbreviations, constants and variables (some very occasionally applied ones are not listed)

a	length (m)
a	acceleration (m·s^{-2}
A	age (year), area (m^2), absorption or extinction
A-mode	amplitude mode
atm	1 physical atmosphere = 101325 Pa; often 1 atm is rounded to 1 bar
"	radiation constant (≈ 2.821), heat convection coefficient " = 1.35 ( ) T/H)$^{1/4}$, thermal diffusivity = $\lambda/(\rho \cdot c_p)$, Womersley number
b	Wiens displacement constant ($2.898 \cdot 10^{-6}$ nm·K)
B	magnetic field strength (N·Amp^{-1}·m^{-1})
B	angle
bar	/ 10^5 Pa. The pressure of 10 m H_2O at 293 K is 0.98067 bar
BEE	basal energy expenditure (J/h)
BF	body fat (%)
BMI	body mass index, mostly W/H^2 (kg/m^2)
BOLD	blood oxygenation level dependent, fMRI signal
BTPS	body temperature pressure saturated, gas volume standardization
β	compressibility, thermal expansion coefficient (1/K)
c	specific heat coefficient (at 0 °C), speed of light and sound (m/s)
c	concentration (mol/L)
c_p	specific heat coefficient at constant pressure (J/(kg·K)
C	heat transport by convection (W), Celsius
C	compliance ≡ 1/*E* (1/Pa), (for hollow subjects L/Pa)
$\boldsymbol{C}_{L,dyn}$	dynamic lung compliance
$\mathbf{C}_V$	volume of dissolved gas at 1bar per volume of liquid (L_n/L)
Cal	calorie (= 4.185 J)
CCD	chrage-coupled device
CEU	contrast enhanced ultrasound
(	c_p/c_v ratio (ratio of specific coefficients of heat capacity with constant p and constant V), surface tension (N/m)
d	diameter (m)
dB	decibel
DC	direct current
dyn	dynamic (subscript)
D	diameter (m), diffusion constant (m^2·s^{-1}·bar^{-1}), diameter (m)
Δ	difference
Δp_{tp}	transpulmonal pressure difference (Pa)
) H_x	specific evaporation heat of fluid x (kJ/kg)
Δp	pressure difference over the airways system from mouth to alveoli (Pa)
Δp'	Δp plus pressure difference over mouth piece and demand valve (Pa)
E	energy, work (J), heat transport by evaporation (W)
E	Young's modulus or modulus of elasticity or stiffness (the reciprocal of compliance) (Pa), electric field strength (N/C or V/m)
E	elastance, the elasticity modulus for hollow subjects (Pa/L)
ECG	electrocardiogram
el	elastic (subscript)
EPR	electron paramagnetic resonance

ERV expiratory reserve volume (L)
ESR electron spin resonance
, (relative) dielectric constant, emittance factor of radiation, tensile strain (= relative extension), remaining error (statistics)
0 dynamic viscosity coefficient (Pa·s or Poise; 1 Poise = 0.1 Pa·s)
0_{air} dynamic viscosity coefficient: $17.1 \cdot 10^{-6}$ (Pa·s)
Θ angle
Φ_m total metabolic power (W)
F force (N)
FE specific fraction of expired air
FEF forced expiratory flow
FEV1 forced expiratory volume in 1 s (L/s)
FCIP fluorescent calcium indicator proteins
FI specific fraction of inspired air
fMRI functional MRI
FRC functional residual capacity (L)
g gravity constant (N/kg)
G shear modulus or modulus of rigidity (Pa)
GPF green fluorescent protein
Gr Grashof number
h Planck's constant ($6.626 \cdot 10^{-34}$ Js), height (m), hour
H height (m)
HBO hyperbaric oxygen
HPLC high performance liquid chromatography
HR heart rate (beat/min)
Hz frequency in Hertz (s^{-1})
i electric current (A), numerical index
I spectral radiance ($J \cdot s^{-1} \cdot m^{-2} \cdot sr^{-1} \cdot Hz^{-1}$)
I intensity of radiation
IR infra red
J sound intensity (W/m^2)
k constant, k is Boltzmann's constant (J/K)
k extinction coefficient, heat conductivity constant
K Kelvin, bulk modulus (Pa), constant
l, length, distance (m)
L length (m), lung (subscript)
LED light emitting diode
λ wavelength (m), heat conduction coefficient (W/(m·K)
m mean (arithmetic), molecular mass (relative mass with respect to the mass of one proton)
m mass (kg)
max maximal (subscript)
MI magnetic resonance imaging
min minute
MR magnetic resonance imaging
MRI magnetic resonance imaging
μ Poisson ratio ≡ $-\Delta d \cdot d^{-1} / \Delta L \cdot L^{-1}$
MVV maximal voluntary ventilation (L)
MVV_{30} MVV with 30 breaths/min (L)
n number, exponent, particle concentration (m^{-3})
n number of kmoles
N number of particles
N_A number of particles in 1 mole is number of Avogadro: $6.0225 \cdot 10^{23}$ mol^{-1}
ν kinematic viscosity= η/ρ (= dynamic viscosity/density) (m^2/s)
O area (m^2)
p pressure (bar, Pa)
p_{tp} transpulmonary pressure / $\Delta p + \mathbf{E} \cdot V_{tidal}$ (Pa, cmH_2O)
P pressure (bar, Pa), power (W)
P poise, a unit of the dynamic viscosity (= 0.1 Pa·s)
Pa Pascal, 1 Pa / 1 N/m^2
PET positron emission tomography
ppb parts per billion
ppm parts per million
Pr Prandtl number
P_{ST} pressure due to surface tension
q charge of particle (single charge is $1.60 \cdot 10^{-19}$ C), Bunsen's absorption coefficient (L_{gas}/L_{liquid} or g_{gas}/L_{liquid})
Q electric charge (C)

r	radius (m)
rad	radian
R	universal gas constant (= 8315 J/(kmol·K)), resistance
R	heat transport by radiation (W), reflection coefficient
R	resistance
Ra	Rayleigh number
R_{aw}	resistance of airways ($Pa \cdot L^{-1} \cdot s^{1}$)
Re	Reynolds number
RMV	respiratory minute volume (L/min)
RV	residual (lung) volume (L)
D	specific density at 0 °C (kg/m^3)
$\mathbf{D}_{air}$	specific density of air: 1.29 kg/m^3 at 0 °C
Q	electric charge (C), unit of flow (e.g. heat)
radian	$1/2\pi$
s	Laplace operator (=$i\omega$)
S	energy state of electron orbital
SaO_2	arterial Hb bounded with SaO_2
SPL	sound pressure level, 1 dB SPL = 20 μPa
stat	static (subscript)
STPD	standard temperature and pressure dry, gas volume standardization
F	constant of Stefan-Boltzmann; 5.7×10^{-8} $W/(K^4.m^2)$, tensile stress (Pa)
t	time, temperature in °C
T	temperature in Kelvin
τ	time constant (s)
TLC	total lung capacity (l)
Φ_m	total metabolic power (W)
UV	ultra violet
v	velocity (m/s)
vis	viscous (subscript)
ν_{max}	peak frequency (Hz) of radiation
V	volume (m^3 or l), voltage (V)
V_{tidal}	volume of inspiration (L)
$\dot{V}$	volume flow (m^3/s or L/s)
VC	vital capacity (L)
VEGF	vascular endothelial growth factor
VO_2	aerobic capacity (mL/kg or mL/(min.kg))
VOR	vestibule-ocular reflex
W	weight (kg), wall (subscript)
ω	angular frequency (radians/s)
x	distance (m)
Z	impedance, e.g. electric (V·s/Amp), acoustic (Pa·s/m)
ζ	resistance coefficient of flow through curvature or bend

Systems and basic concept

Autoregulation

Principle

In biomedicine, autoregulation is the process of changing the principal system parameter to cope with a change in conditions. It is most known for the circulatory systems, but macromolecular systems can also have autoregulation.

Application

The principle is applied in the cerebral, systemic, muscular and renal circulation.

Cerebral circulation
The cerebral flow of blood is autoregulated by altering cerebral vessel diameters. Proper cerebral perfusion pressure (CPP) is achieved by dilatation (widening) of arterioles that lowers pressure and creates more room for the blood, or constriction to raise pressure and lower cerebral blood volume. Thus, changes in the body's overall blood pressure (see also Blood pressure: measurement) do not normally alter cerebral perfusion pressure drastically, since they constrict when systemic blood pressure is raised and dilate when it is lowered. Arterioles also constrict and dilate in response to different chemical concentrations. For example, they dilate in response to higher levels of CO_2 in the blood.
Systemic circulation
The capillary bed of an organ usually carries no more than 25% of the maximal amount of blood it could contain, although this amount can be increased through autoregulation by inducing relaxation of smooth muscle. The capillaries do not possess this smooth muscle in their own walls, and so any change in their diameter is passive. Any signaling molecules they release (such as endothelin for constriction and NO for dilation) act on the smooth muscle cells in the walls of nearby, larger vessels, e.g. arterioles.
Muscular circulation
Myogenic autoregulation is a form of homeostasis (see Bioregulation, homeostasis). It is the way arteries and arterioles react to an increase or decrease of blood pressure to keep the blood pressure within the blood vessel constant. The smooth muscle of the blood vessels will react to the stretching of the muscle by opening ion channels which cause the muscle to depolarize leading to muscle contraction.
The effect of this is to reduce Blood flow through the blood vessel as the volume of blood able to pass through the lumen is significantly reduced. Alternatively, when the smooth muscle in the blood vessel relaxes due to a lack of stretch, the ion channels will close resulting in vasodilatation of the blood vessel. This increases the rate of flow through the lumen as its radius is greater.
Renal circulation
This system is especially significant in the kidneys, where the glomerular filtration rate is particularly sensitive to changes in blood pressure. However, with the aid of the myogenic mechanism it allows the glomerular filtration rate to remain very insensitive to changes in human blood pressure.

More Info

Cerebral autoregulation
When pressures are outside the range of 50 to 150 mmHg, the blood vessels' ability to autoregulate pressure is lost, and cerebral perfusion is determined by blood pressure alone, a situation called pressure-passive flow. Thus, hypotension (inadequate blood pressure) can result in severe cerebral ischemia in patients with conditions like brain injury, leading to damage (ischemic cascade).
Other factors that can cause loss of autoregulation include free radical damage, nervous stimulation, and alterations of the partial CO_2 and O_2 pressure. Even in the absence of autoregulation a high pCO_2 can dilate blood vessels up to 3.5 times their normal size, lowering CPP, while high levels of oxygen constrict them. Blood vessels also dilate in response to low pH. Thus, when activity in a given region of the brain is heightened, the increase in CO_2 and H^+ concentrations causes cerebral blood vessels to dilate and deliver more blood to the area to meet the increased demand. In addition, stimulation of the sympathetic nervous system raises blood pressure and blocking lowers pressure.

Colloid

Principle

Colloids are all types of, generally, liquid-like mixtures existing of a solvent and a (semi-)macro-molecular substance. More precisely, a colloidal solution or colloidal dispersion is a type of mixture intermediate between a *homogeneous mixture (*also called a solution*)* and a *heterogeneous mixture*. Also its properties are intermediate between the two.
Typical membranes restrict the passage of dispersed colloidal particles more than for ions or dissolved molecules. Many familiar substances, including butter, milk, cream, aerosols (fog, smog, and smoke), asphalt, inks, paints, glues, and sea foam are colloids. In a colloid, the size of dispersed phase particles range from 1 nm to 1µm. Dispersions where the particle size is in this range are referred to as colloidal aerosols, colloidal emulsions, colloidal foams, or colloidal suspensions or dispersions. Colloids may be colored or translucent because of the Tyndall effect (see Light: scattering), which is the scattering of light by particles in the colloid.
When the substance is a liquid or solid, the mixture is a Suspension. When both phases are liquid, than the mixture is an Emulsion.
Since there are 3 basic aggregation states one would expect 9 kinds of solids but since gas is always soluble in another gas, 8 types remain. Table 1 gives these combinations.

Table 1 Types of colloids

		Dispersed Medium		
		Gas	Liquid	Solid
Continuous Medium	Gas		**Liquid Aerosol** Examples: fog, mist	**Solid Aerosol** Examples: smoke, dust
	Liquid	**Foam** Examples: whipped cream, nose spray	**Emulsion** Examples: milk, hand cream, salve	**Sol** Examples: paint, pigmented ink, blood
	Solid	**Solid Foam** Examples: aerogel, styrofoam, pumice	**Gel** Examples: gelatin, jelly, cheese, opal	**Solid Sol** Examples: cranberry glass, ruby glass

Applications

They are numerous, as Table 1 indicates, also in medicine.

More info

Interaction between colloid particles
The following forces play an important role in the interaction of colloid particles:
- Excluded Volume Repulsion This refers to the impossibility of any volumetric “overlap” between hard particles.
- Electrostatic interaction Colloidal particles often carry an electrical charge and therefore attract or repel each other. The charge and the mobility of both the continuous and the dispersed phase are factors affecting this interaction.
- Van der Waals forces This is due to interaction between two dipoles (permanent or induced). The van der Waals force is always present, is short range and is attractive.
- Entropic forces: According to the second law of thermodynamics, a system progresses to a state in which Entropy (irreversible increase of disorder at microscopic scale) is maximized. This can result in effective forces even between hard spheres.
- Steric forces between polymer-covered surfaces or in solutions containing non-adsorbing polymer can modulate interparticle forces, producing an additional repulsive steric stabilization force or attractive depletion force between them.

Stabilization of colloid suspensions
Stabilization serves to prevent colloids from aggregating. Steric and electrostatic stabilization are the two main mechanisms for colloid stabilization. Electrostatic stabilization is based on the mutual repulsion of like electrical charges leading to very large charge double-layers of the continuous phase around the particles. In this way the specific density differences are so small that buoyancy or gravity forces are too little to overcome the electrostatic repulsion between charged layers of the dispersing phase.
The charge on the dispersed particles can be observed by applying an electric field: all particles migrate to the same electrode and therefore must all have the same sign charge.

Destabilizing a colloidal suspension

Unstable colloidal suspensions form flocks as the particles aggregate due to interparticle attractions. This can be accomplished by a number of different methods:

- Removal of the electrostatic barrier that prevents aggregation of the particles (by addition of salt to a suspension or changing the pH). This removes the repulsive forces that keep colloidal particles separate and allows for coagulation due to van der Waals forces.
- Addition of a charged polymer flocculant (bridging of individual colloidal particles by attractive electrostatic interactions).
- Addition of non-adsorbed polymers called depletants that cause aggregation due to entropic effects.
- Unstable colloidal suspensions of low volume fraction form clustered liquid suspensions wherein individual clusters of particles fall to the bottom or float to the top, since Brownian motion become too small to keep the particles in suspension. Colloidal suspensions of higher volume fraction can form colloidal gels with viscoelastic properties. These gels (e.g. toothpaste) flow like liquids under shear but maintain their shape when shear is removed. It is for this reason that toothpaste stays on the toothbrush after squeezing out.

Colloidal particles are large enough to be observed by Confocal microscopy.
Just as a solution, a colloid has an osmotic effect (see Osmosis).

Compliance (hollow organs)

Principle

Elastance is a measure of the tendency of a hollow organ to recoil toward its original dimensions upon removal of a distending or compressing force. It is the reciprocal of *compliance*.

Compliance of a hollow organ is calculated using the following equation:
$C = \Delta V/\Delta P$,
where ΔV is the change in volume, and ΔP is the change in pressure. In SI-units, its dimension is L/Pa. It should not be mixed up with the one-dimensional compliance, or better the reciprocal, i.e. modulus of elasticity as defined in the theory of strength of materials, see Elasticity and Hooke's law, Elasticity 1: elastic or Young's modulus and Tensile strength.

Application

Cardiovascular system
The terms elastance and compliance are of particular significance in cardiovascular physiology. Specifically, the tendency of the arteries and veins to stretch in response to pressure has a large effect on perfusion and blood pressure.
Veins have a much higher compliance than arteries (largely due to their thinner walls). Veins, which are abnormally compliant, can be associated with edema. Pressure stockings are sometimes used to externally reduce compliance, and thus keep blood from pooling in the legs. An extreme application is the use of pressure trousers or suites by astronauts.
Lungs
Compliance of the lungs is an important measurement in pulmology. Fibrosis is associated with a *decrease* in pulmonary compliance. Emphysema is associated with an *increase* in pulmonary compliance.

More Info

For the cardiovascular system, see Elasticity of the aorta, Blood pressure: models and Windkessel model.
For the lungs, see Lung gas transport 2, pressure, volume and flow and Lung gas transport 2, resistance and compliance.

Emulsion

Principle

Emulsions are part of a more general class of two-phase systems of matter, called *colloids* (see Colloid). Although the terms colloid and emulsion are sometimes used interchangeably, in emulsion *both the dispersed and the continuous phase are liquid*. Emulsions are also a subclass of Suspension.
An emulsion is a mixture of two immiscible (unblendable) substances. One substance (the dispersed phase) is dispersed in the other (the continuous phase).
Emulsions have a cloudy appearance, because the many phase interfaces (the boundary between two phases) scatter light (see Light: scattering).
Emulsions are unstable and thus do not form spontaneously. Energy input through stirring, etc., or spray processes are needed to form it. Over time, emulsions tend to revert to the stable state of for instance oil separated from water. Surface active substances (surfactants, see Surface tension) can increase the kinetic stability of emulsions greatly so that, once formed, the emulsion does not change significantly over years of storage.
Emulsification is the process by which emulsions are prepared.

Application

In medicine
Emulsions are frequently used in drugs.
In daily life
A large field of application is food and cosmetic industry. Examples of emulsions include oil in water and butter. In butter and margarine (are also suspensions), a continuous lipid phase surrounds droplets of water (water-in-oil emulsion).

More Info

There are three types of emulsion instabilities:
- flocculation, where the particles form clumps;
- creaming, where the particles concentrate towards the surface (by buoyancy or by e.g. centrifugation) of the mixture while staying separated;
- breaking, where the particles coalesce (recombination to form bigger ones) due to lack of shaking and form a layer of liquid.

Emulsifier
An emulsifier, also known as an emulgent or surfactant, is a substance which stabilizes an emulsion. An example of food emulsifiers is egg yolk (where the main emulsifying chemical is the phospholipid lecithin). Proteins and low-molecular weight emulsifiers are common as well.
Detergents, another class of surfactant, chemically interact with both oil and water, thus stabilizing the interface between oil or water droplets in suspension. This principle is exploited in soap to remove grease for the purpose of cleaning. A wide variety of emulsifiers are used to prepare emulsions such as creams and lotions.

Whether an emulsion turns into a water-in-oil emulsion or an oil-in-water emulsion depends on the volume fraction of both phases and on the type of emulsifier. Generally, the so-called Bancroft rule applies: emulsifiers and emulsifying particles tend to promote dispersion of the phase in which they do not dissolve very well. For example, proteins dissolve better in water than in oil and so tend to form oil-in-water emulsions (that is they promote the dispersion of oil droplets throughout a continuous phase of water).

Steradian

The steradian (symbol: **sr**) is the derived SI unit of describing the 2D angular span in 3D-space, analogous to the way in which the radian describes angles in a plane. Just as the radian, the steradian is dimensionless. More precisely, the steradian is defined as "the solid angle subtended at the center of a sphere of radius *r* by a portion of the surface of the sphere having an area r^2." If this r^2 area is a circle, the solid angle is a simple cone subtending a half apex angle θ of $\cos^{-1}(1–1/2\pi) \approx 0.572$ rad or 32.77° (or with apex angle 65.54°).
Since the surface area of this sphere is $4\pi r^2$, then the definition implies that a sphere measures 4π steradians. By the same argument, the maximum solid angle that can be subtended at any point is 4π sr. A steradian can also be called a *squared radian.*

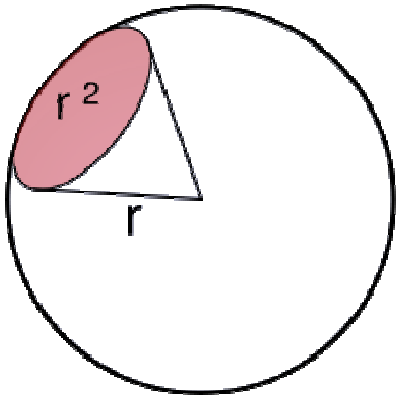


Fig. 1 A graphical representation of 1 steradian.

A steradian is also equal to the spherical area with an angle of 1 radian. This area is equal to $1/4\pi$ of a complete sphere, or $(180/\pi)^2$ or 3282.8 square degrees.
As an example, radiant intensity can be measured in watts per steradian ($W \cdot sr^{-1}$).

Suspension

Principle

A suspension is a colloidal (see Colloid) dispersion (mixture) in which a finely-divided substance is combined with another substance, with the former being so finely divided and mixed that it doesn't rapidly settle out. In everyday life, the most common suspensions are those of solids in liquid water. Basically, all Emulsion (immiscible fluid in fluid mixtures) are also suspensions.
A suspension of liquid droplets or fine solid particles in a gas is called an *aerosol* (see Colloid). In the atmosphere these consist of fine dust, sea salt, cloud droplets etc.

Application

Suspensions are widely applied in medicine, daily life and industry. Common examples are ice cream, a suspension of microscopic ice crystals in cream. Mud or muddy water, is an emulsion where solid particles (sand, clay, etc.) are suspended in water. Paint is an emulsion where the solid pigment particles are mixed in a water-like or organic liquid.

More Info

The amount of energy determines the maximum size of particle that can be suspended. In the absence of additional energy (agitation), all particles down to colloidal size will eventually settle out into a distinct phase. Suspensions separate over some period of time, solutions never because the intermolecular forces between the different types of molecules are of similar strength to the attraction between molecules of the same type.

Wheatstone bridge

Principle

A Wheatstone bridge is used to measure an unknown electrical resistance by balancing two legs of a bridge circuit, one leg of which includes the unknown component.

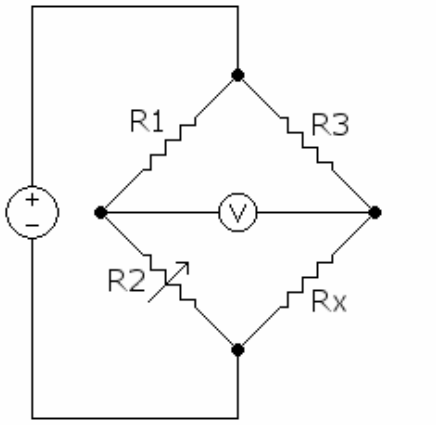


Fig. 1 Principle of the Wheatstone bridge

In Fig. 1, R_x is the unknown resistance to be measured; R_1, R_2 and R_3 are resistors of known resistance and the resistance of R_2 is adjustable (a potentiometer). If the ratio of the two resistances in the known leg (R_2 / R_1) is equal to the ratio of the two in the unknown leg (R_x / R_3), then the voltage between the two midpoints will be zero and no current will flow between the midpoints. The voltage is measured by a sensitive voltage meter, a sensitive galvanometer. R_2 is varied until this condition is reached. The current direction indicates if R_2 is too high or too low.
Detecting zero current can be done to extremely high accuracy. Therefore, if R_1, R_2 and R_3 are known to high precision, then R_x can be measured to high precision. Very small changes in R_x disrupt the balance and are readily detected.
At the balance point the value of R_x can be found from:
$R_x = R_2 \cdot R_3 / R_1$.
Variations on the Wheatstone bridge can be used to measure the capacitance of a condenser or the inductance of a coil.

Application

It is typically applied in biomedical apparatus, but mostly hidden for the user. In basic research, e.g. in the field of bio-electricity manually a minimum of some quantity can be adjusted.

Molecular structure of matter

Adhesion

Principle

Molecules in liquid state experience strong intermolecular attractive forces. When those forces are between unlike molecules, they are said to be adhesive forces. The adhesive forces between water molecules and the walls of a glass tube are stronger than the cohesive forces (the attractive forces between like molecules). This leads to an upward turning meniscus at the walls of the vessel (see Capillary action).
The attractive forces between molecules in a liquid can be seen as residual electrostatic forces and are called van der Waals forces or van der Waals bonds (see Cohesion).
More generally, and from a macroscopically point of view, adhesion is the molecular attraction exerted between bodies in contact.
Notice that *in medicine*, an adhesion has a completely other meaning. It is a fibrous band of scar tissue that binds together normally separate anatomical structures. It usually occurs as a result of surgery, infection, trauma or radiation.

Application

Adhesion is of particular interest to (medical) biologists to understand the workings of cells and to engineers who wish to stick objects together.

More Info

Five mechanisms have been proposed to explain why one material sticks to another.

- *Mechanical adhesion*
  Two materials may be mechanically interlocked. Sewing forms a large-scale mechanical bond, Velcro forms one on a medium scale, and some textile adhesives form one at a small scale.
- *Chemical adhesion*
  Two materials may form a compound at the join. The strongest joins are where atoms of the two materials swap (ionic bonding) or share (covalent bonding) outer electrons. A weaker bond is formed if oxygen, nitrogen of fluorine atoms of the two materials shares a hydrogen nucleus (hydrogen bonding).
- *Dispersive adhesion*
  This is also known as adsorption. Two materials may be held together by van der Waals forces.
- *Electrostatic adhesion*
  Some conducting materials may pass electrons to form a difference in electrical charge at the join. This results in a structure similar to a capacitor and creates an attractive electrostatic force between the materials.
- *Diffusive adhesion*
  Some materials may merge at the joint by Diffusion: general. This may occur when the molecules of both materials are mobile and soluble in each other. This would be particularly effective with polymer chains where one end of the molecule diffuses into the other material. It is also the mechanism involved in sintering. When metal or ceramic powders are pressed together and heated, atoms diffuse from one particle to the next. This joins the particles into one.

What makes an adhesive bond strong?
The strength of the adhesion between two materials depends on which of the above mechanisms occur between the two materials, and the surface area over which the two materials contact. Materials that wet against each other tend to have a larger contact area than those that do not. Wetting depends on the surface energy (the disruption of chemical bonds that occurs when a surface is created) of the materials.

Capillary action

Principle

Capillary action described in the Cohesion-tension theory is considered a mix of cohesion and adhesion. Capillary action or capillarity (also known as capillary motion) is the ability of a substance to draw a liquid upwards against the force of gravity. The standard reference is a tube in plants but capillarity can be seen readily with porous paper. It occurs when the adhesive intermolecular forces between the liquid and a solid are stronger than the cohesive intermolecular forces within the liquid. The effect causes a concave meniscus to form where the liquid is in contact with a vertical surface. The same effect is what causes porous materials to soak up liquids.
A common object used to demonstrate capillary action is the *capillary tube*. When the lower end of a vertical glass tube is placed in a liquid, such as water, a concave meniscus forms. Surface tension pulls the liquid column up until there is a sufficient weight of liquid for gravitational forces to come in equilibrium with intermolecular adhesive forces. The weight of the liquid column is proportional to the square of the tube's diameter, but the contact area between the liquid and the tube is proportional only to the diameter of the tube, so a narrow tube will draw a liquid column higher than a wide tube. For example, a glass capillary tube 0.5 mm in diameter will lift a theoretical 28 mm column of water. (Actual observations show shorter total distances.) In addition, the angle of incidence (contact angle) can be calculated exactly (see the textbooks of physics).
With some materials, such as mercury in a glass capillary, the interatomic forces within the liquid exceed those between the solid and the liquid, so a convex meniscus forms and capillary action works in reverse. Now the liquid level in the tube is lower.

Application

In medicine Capillary action is also essential for the drainage of constantly produced tear fluid from the eye. Two cannaliculi of tiny diameter are present in the inner corner of the eyelid; their openings can be visualized with the naked eye when the eyelids are everted.
In botany A plant makes use of capillary force to draw water into its system (although larger plants also require transpiration to move a sufficient quantity of water to where it is required).
In hydrology capillary action describes the attraction of water molecules to soil particles.
In chemistry Chromatography utilizes capillary action to move a solvent vertically up in a plate or paper. Dissolved solutes travel with the solvent at various speeds depending on their polarity. Paper sticks for urine and pH tests are also applications.
In daily life Towels (fluid transfer from a surface) and sponges (the small pores) absorb liquid through capillary action. Some modern sport and exercise fabrics use capillary action to "wick" sweat away from the skin.

More info

With notes on the dimension in SI units, the height h of a liquid column (m) is given by:

$$h = 2\gamma\cos\theta/\rho g r, \qquad (1)$$

where:
γ = surface tension (J/m² or N/m)
θ = contact angle, this is the angle between the meniscus (at the wall) and the wall
ρ = density of liquid (kg/m^3)
g = acceleration due to gravity (m/s²)
r = radius of tube (m)
For a water-filled glass tube in air at sea level,
γ = 0.0728 J/m² at 20 °C, θ = 20° (0.35 rad), ρ = 1000 kg/m^3 and g = 9.8 m/s². And so the height of the water column is given by:

$$h \approx 1.4\cdot10^{-5}/r \qquad (2)$$

Thus in a 2 m wide capillary tube, the water would rise an unnoticeable 0.014 mm. However, for a 1 mm wide tube, about the size of a hematocryte capillar, the water would rise 7 mm, and for a tube with radius 0.1 mm, the water would rise 14 cm.

Cohesion

Principle

Cohesion or cohesive attraction or cohesive force is the intermolecular attraction between (nearly) identical molecules. The cohesive forces between liquid molecules are responsible for phenomena such as Surface tension and capillary force.
Molecules in liquid state experience strong intermolecular attractive forces. When those forces are between like molecules, they are called cohesive forces. For example, cohesive forces hold the molecules of a water droplet together, and the strong cohesive forces constitute surface tension.
When the attractive forces are between unlike molecules, they are said to be adhesive forces (see Adhesion). When the adhesive forces between water molecules and the walls of a glass tube are stronger than the cohesive forces lead to an upward turning meniscus at the walls of the vessel. This is Capillary action. Mercury is an example of a liquid that has strong cohesive forces, as becomes clear from the very convex meniscus in the tube of a classical air pressure meter.

Application

There are many, in medicine, science and daily life. Often they are based on surface tension.
Clinical tests Normal urine has a surface tension of about 0.066 N/m, but if bile is present, (a test for jaundice) it drops to about 0.055. In the Hay test, powdered sulfur is sprinkled on the urine surface. It will float on normal urine, but sink if the bile lowers the surface tension.
Surface tension disinfectants Disinfectants are usually solutions of low surface tension. This allows them to spread out on the cell walls of bacteria and disrupt them.
Walking on water Small insects can walk on water. Their weight is not enough to penetrate the surface.
Floating a needle If carefully placed on the surface, a small needle can be made to float on the surface of water even though it is several times as dense as water.
Soaps and detergents They help the cleaning of clothes by lowering the surface tension of the water so that it more readily soaks into pores and soiled areas.
Washing The surface tension of hot water is lower and therefore it is a better "wetting agent" to get water into pores and fissures rather than bridging them with surface tension.

More Info

There are various phenomena, which are based on cohesion.

Surface Tension
Water is a polar molecule due to the high electronegativity of the oxygen atom, which is an uncommon molecular configuration whereby the oxygen atom has two lone pairs of electrons. When two water molecules approach one other they form a hydrogen bond. The negatively charged oxygen atom of one water molecule forms a hydrogen bond with a positively charged hydrogen atom in another water molecule. This attractive force has several manifestations. Firstly, it causes water to be liquid at room temperature, while other lightweight molecules would be in a gaseous phase. Secondly, it (along with other inter molecular forces) is one of the principal factors responsible for the occurrence of surface tension in liquid water.
Water at 20 °C has a surface tension of 0.073 N/m compared to 0.022 N/m for ethyl alcohol and 0.47 N/m for mercury. The latter high value is the reason why in a capillary filled with mercury the meniscus is very convex. The surface tension of water decreases significantly with temperature.
The surface tension arises from the polar nature of the water molecule. See for a further explanation of the *surface tension* the description of the subject itself.
Cohesion in crystals
In crystals (of molecular-, ionic-, valence- and metal-type) many types of forces play a role such as van der Waals forces and forces of chemical bonds. A van der Waals force is the attraction between two molecules with positively and negatively charged ends. This polarity may be a permanent property of a molecule (Keesom forces) or a temporarily property, which occurs universally in molecules, as the random movement of electrons within the molecules may result in a temporary concentration of electrons at one end (London forces).

Evaporation and perspiration; calculations

Principle

Evaporation

With evaporation, the opposite process of condensation, the atoms or molecules of the liquid gain sufficient energy to enter the gaseous state. It is exclusively a surface phenomenon that occurs at all temperatures and should not be confused with boiling. Even at cool temperatures, a liquid can still evaporate, but only a few particles would escape over a long period of time. Boiling occurs throughout a liquid (Fig. 1) and is characterized by the boiling point (e.g. for H_20 at 1 bar 100 ^oC). For a liquid to boil its vapor pressure must equal the ambient pressure and bubbles are generated in the liquid.
For particles of a liquid to evaporate, they must be located near the surface, and moving in the proper direction, and have sufficient kinetic energy to overcome the Surface tension. Only a very small proportion of the molecules meet these criteria, so the rate of evaporation is limited. Since the average kinetic energy of the molecules rises with temperature and a larger fraction of molecules reaches the requested velocity, evaporation proceeds more quickly at higher temperature. As the faster-moving molecules escape, the remaining molecules have lower average kinetic energy, and the temperature of the liquid thus decreases. This phenomenon is also called *evaporative cooling*.
For simplicity, from now the liquid is considered water and the gas air with some water vapor.

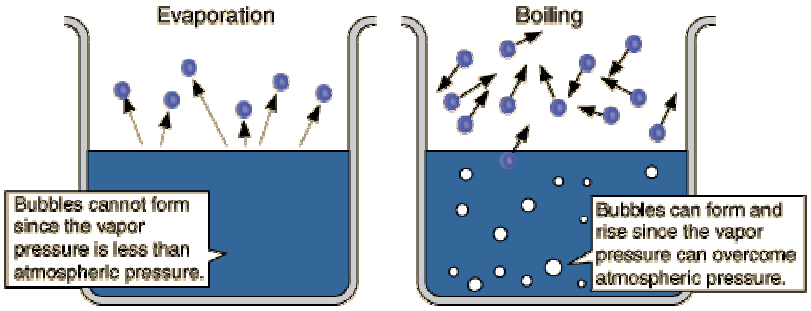


Fig. 1 Evaporation and boiling

Factors influencing rate of evaporation of water

- Evaporation is proportional with the difference in the partial water pressure (p_{H2O}) given by the temperature of the water and p_{H2O} in the air.
- Evaporation increases with the temperature of the water.
- The higher the speed of air compared to the water surface, the faster the evaporation, which is due to constantly lowering the vapor pressure in the boundary layer. Wind speeds < 0.15 m/s are neglected.
- Other substances, especially polar ones (e.g. salts) decrease evaporation (Raoult's law). When a (monomolecular layer of) surfactant (a lipid-polar substance, see Surface tension) or a non-polar liquid with a smaller density (oil etc.) covers the water surface, evaporation is very strongly reduced.

The vapor pressure of saturation increases about exponential with the temperature, see Fig. 2.

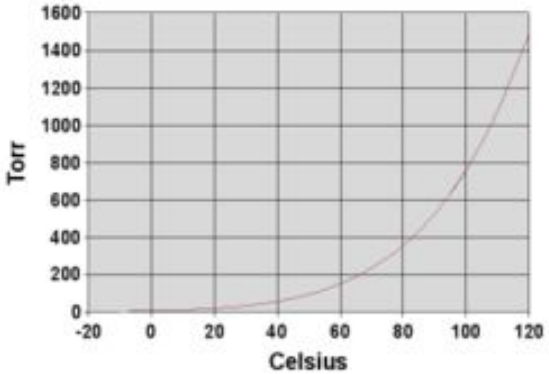


Fig. 2 Vapor pressure of equilibrium, p_{H2O}, as function of temperature. 760 Torr = 1atm = 1.013 bar. At 0 ^oC p_{H2O} is 6.2 mbar and at 100 ^oC 1001.3 mbar (= 1001.3 hPa).

Perspiration

Perspiration is the process of water evaporation from the human skin.

The process of evaporative cooling is the reason why evaporating sweat cools the human body. The cooling effect of flowing air of the (human) body is caused by convection but also by evaporation. Evaporation and so cooling happens also by expiration. At a body temperature of 37 ^{o}C, the alveoli, always, so irrespective the ambient pressure, have a p_{H2O} of 47 mm Hg (= 62.6 mbar = 6260 Pa). One square meter of a physiological salt solution evaporates about 6 L per day (see **More Info**). Assuming that the human skin is covered by a thin layer of physiological salt and that its surface is 1.9 m^2, then the perspiration is 1.9 m^2 x 6.0 $L \cdot m^{-2} \cdot day^{-1}$ = 11.4 L/day. Actually, the loss is only 0.75 L/day (see Body heat dissipation and related water loss). The reason is that the skin is generally not covered by a layer of physiological salt. Skin is not well permeable for water. Evaporation mainly occurs in the sweat glands, which cover only a small fraction of the skin. However, with extensive sweating (heavy endurance sport, sauna) the skin is covered with a thin layer of liquid and perspiration is some 0.5 L/hour. Under these conditions, total sweat production is higher, but much drips off.

More Info

Evaporation is a state change from liquid to gas, and as gas has less order than liquid matter, the Entropy of the system is increased, which always requires energy input. This means that the Enthalpy change for evaporation (ΔH_{liquid}) and the standard enthalpy change of vaporization or heat of evaporization (ΔH^0_{liquid}) is always positive, making it an endothermic process and subsequently, a cooling process.
If the evaporation takes place in a closed vessel, the escaping molecules accumulate as a vapor above the liquid. Many molecules return to the liquid and more molecules return as the density and pressure of the vapor increases. When the process of escape and return reaches equilibrium, the vapor is said to be "saturated," and no further change in either vapor pressure and density or liquid temperature will occur. For a system consisting of vapor and liquid of a pure substance, this equilibrium state is directly related to the vapor pressure of the substance, as given by the Clausius-Clapeyron relation:

$$\ln p^*_x = -\Delta\hat{H}_x/RT + B, \text{ where} \qquad (1)$$

p^*_x is the vapor pressure (bar)
$\Delta\hat{H}_x$ the heat of vaporization of liquid x (kJ/mole)
R is the gas constant (8.315 J/(mol·K)
T is the temperature (K)
B is a variable based on the substance and the system parameters.
$\Delta\hat{H}_{H2O}$ varies from 45.05 kJ/mol at 0 ^{o}C to 40.66 kJ/mol H_20 at 100 ^{o}C. This gives a strong dependency of p* in eq 1 from temperature.
Rewriting equation (1) for water, p^*_{H2O} is approximated by:

$$p^*_{H2O} = 1.27 \cdot 10^6 \, e^{-5219/T} \text{ (bar)}. \qquad (2)$$

The rate of evaporation in an open system is related to the vapor pressure found in a closed system. If a liquid is heated with the vapor pressure reaching the ambient pressure, the liquid will boil.

The underlying physics of (1) is not too hard, but calculating the amount of liquid mass that is evaporated is another thing. The idea is that the fastest water molecules escape from a monomolecular layer of water, which accounts for the surface tension. The solution is an approximation, actually only for the state of equilibrium but for an open system, not in equilibrium, it works rather well.
The number of evaporating particles per second per unit area (ref. 1) is equal to:

$$N_e = (1/A)(v/d)e^{-W/kT}, \text{ where} \qquad (3)$$

- A is the cross sectional area of the particle (m^2, water molecule $0.057 \cdot 10^{-18}$ m^2),
- d the thickness of a monomolecular layer of the particles (m, water molecule ca. $0.27 \cdot 10^{-9}$ m),
- k Boltzmann's constant ($1.3805 \cdot 10^{-23}$ J/K).
- v the root mean square velocity of the liquid particle (m/s), a function of $T^{0.5}$: It can be calculated from Einsteins equation (ref. 2) of the Brownian motion (random movement of particles suspended in a fluid): $l^2 = (2kT/(3\pi\eta d) \cdot t$ where η is the dynamic viscosity coefficient, for water η • 0.001 Pa·s, at 37 ^{o}C, l the free travel distance (estimated in water at 7 nm, nearly ten times less than the 66 nm in air) and t the time of the free path length. Rewriting gives $V = (2kT/(3\pi\eta d)/^{-1}$. This yields v =0.908 m/s).
- W is the energy needed to evaporate. $\Delta\hat{H}_{water}$ = 2428 kJ/kg at 37 ^{o}C, and a rough estimate of physiological salt is $\Delta\hat{H}_{physiological\ salt}$ = 2437 kJ/kg at 37 ^{o}C. Knowing Avogadro's number being $6.0225 \cdot 10^{23}$, $W_{water} = 7263 \cdot 10^{-23}$ J/molecule),
- The factor $e^{-W/kT}$ is the fraction of particles that have enough velocity to escape, so it presents also the probability (water at 310 K gives $42.6 \cdot 10^{-9}$, physiological salt $40.0 \cdot 10^{-9}$).

Finally, for water at 37 ^{o}C an evaporation of 7.48 $mg \cdot s^{-1} m^{-2}$ is found. This means lowering the surface with 6.4 mm per day or an evaporation of about 6.4 L/m^2 per day (physiological salt 6.0 L/m^2 per day).

At 14 ^{o}C, experimentally (no air convections, humidity low) a surface lowering of 1.04 mm/day was found whereas 1.07 mm/day is predicted.

References

1. Feynman R.P., Leighton R.B. and Sands M. The Feynman lectures on Physics. Addison-Wesley, Reading, Mass., 1963.
2. Kronig R. (ed.). Leerboek der Natuurkunde, Scheltema & Holkema NV, Amsterdam 1966.

Mass spectrography

Principle

A mass spectrograph is used to separate electrically charged particles, for instance isotopes, according to their masses. In the mass spectrograph, streams of charged particles (ions) travel in a vacuum, pass through deflecting magnetic and electric fields (produced by carefully designed magnetic pole pieces and electrodes) and are detected by photographic plates. The degree of bending depends on the masses and electric charges of the ions. The detector measures exactly how far each ion has been deflected, and from this measurement, the ion's 'mass to charge ratio' can be worked out. From this information it is possible to determine with a high level of certainty what the chemical composition of the original sample was.

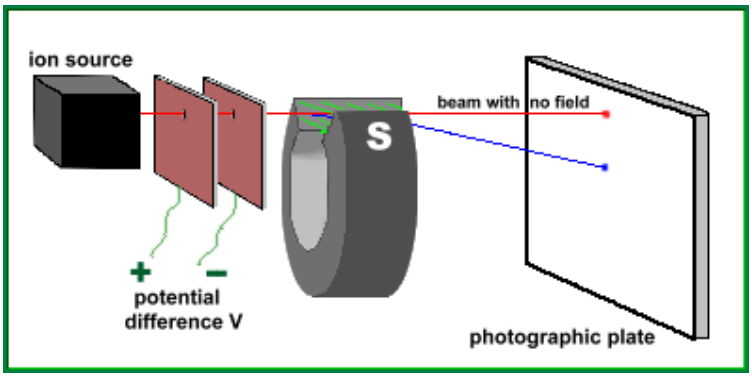


Fig. 1 A striped setup of a mass spectrograph for positive ions

The mass of an ion can be calculate directly:

$$m = \frac{B^2 q r^2}{2V}$$

where
B is the strength of the magnetic field provided by the magnet (in $N.Amp^{-1}.m^{-1}$),
r is the radius of curvature of the beam while in the field (m),
V is the voltage applied between the two plates (V),
q is the charge on an ion (1.60×10^{-19} coulomb if singly charged)
The equation shows that a smaller mass gives a smaller radius, due to the smaller momentum (mass·velocity) of the particle, provided that the ions enter the electric field with the same velocity.

Application

The mass spectrograph is very widely used in chemical analysis and in the detection of impurities in medicine, biology and (bio)chemistry. Applications include:

- identifying unknown compounds by their mass and/or fragments thereof;
- determining the isotopic composition of one or more elements in a compound;
- determining the structure of compounds by observing the fragmentation of the compound;
- quantifying the amount of a compound in a sample using carefully designed methods (mass spectrometry is not inherently quantitative);
- determining chemical or biological properties of compounds, proteins and peptides with a variety of other approaches, like age determination of graves (forensic medicine, medical archeology)

It is often combined with High performance liquid chromatography (HPLC, see Chromatography).

More info

Mass analyzers separate the ions according to their mass per charge. There are many types of mass analyzers. For these types see for instance Wikipedia. Usually they are categorized based on the principles of operation.

The common process comprises the following steps:
- The analyte is brought into some vacuum;
- it is ionized;
- the ions are accelerated in an electric field according to the Lorentz force acting on the charged particles.
Also in the two next steps the Lorentz force is the ruling physical law;
- by diaphragms and a velocity selector (a combination of electric and magnetic fields) the ions of interest are collected in a narrow beam;
- the beam enters a high vacuum chamber with two perpendicularly arranged magnetic fields. These bends the ions into semicircular paths ending at the photographic plate. The radius of this path depends upon the mass of the particles (all other factors, such as velocity and charge, being equal).
-, the position of the blackened spots on the plate makes possible a calculation of the isotope masses of the elements of the analyte. Nowadays an electron amplifier is used.
Depending on the selection process various versions of analyses exist.

A modification is a mass spectrometer, often used to measure the masses of isotopes. The quantity to measure is the mass-to-charge ratios of molecules and atoms. A mass spectrometer does not measure the kinetic energy of particles - all particles have the same known kinetic energy (or an integer multiple thereof, depending on the charge) - so it is disputable whether mass spectrometry strictly is a type of spectroscopy (see Spectroscopy).

Reference
http://www.worsleyschool.net/science/files/mass/spectrograph.html

Paramagnetism, diamagnetism and magnetophoresis

Paramagnetism

Paramagnetism is the tendency of atomic magnetic dipoles to align within an external magnetic field. Paramagnetic materials attract and repel like normal magnets when subjected to a magnetic field. This effect occurs due to quantum-mechanical spin as well as electron orbital angular moment (the vector cross product of radius and mass times velocity). This alignment of the atomic dipoles with the magnetic field tends to strengthen it and is described by a relative magnetic permeability μ slightly greater than unity (or, equivalently, a small magnetization or positive magnetic susceptibility X_v).

Application

In medical Magnetic resonance imaging (MRI) the paramagnetism of contrast substances as barium sulfate is applied to enhance contrast.
Hb is diamagnetic when oxygenated but paramagnetic when deoxygenated. Therefore, an important application is the paramagnetism of deoxy-Hb. The magnetic resonance (MR) signal of blood is therefore variable depending on the level of oxygenation. These differential signals can be detected using an appropriate MR pulse sequence as Blood Oxygenation Level Dependent (BOLD) contrast (see Functional MRI (fMRI)).
The existence of four unpaired electrons in deoxy-Hb can give rise to a process called magnetophoresis (see below).

More info

Paramagnetism requires that the atoms individually have permanent dipole moments even without an applied field, which typically implies partially filled electron shells. In pure paramagnetism, these atomic dipoles do not interact with one another and are randomly oriented in the absence of an external field, resulting in zero net magnetic moment. If they *do* interact, they can spontaneously align or anti-align, resulting in ferromagnetism (permanent magnets) or antiferromagnetism (opposed electron spin) respectively.
In atoms with no permanent dipole moment, e.g. for filled electron shells, a weak dipole moment can be *induced* in a direction *anti*-parallel to an applied field, an effect called diamagnetism. Paramagnetic materials also exhibit diamagnetism, but the latter effect is typically orders of magnitude weaker. Paramagnetic materials in magnetic fields will act like magnets but when the field is removed, the magnetic alignment is quickly disrupted. In general, paramagnetic effects are small. They are, as ferrimagnetism (see below) temperature dependent.

In order to be paramagnetic, i.e. to be attracted, the electrons align themselves in one direction in a magnetic field. The following example may clarify this. Carbon has an electron configuration with the six orbitals: 1s^2, 2s^2, 2p^2. The way that this would look is 1s^2: in this orbital we have one arrow (the momentum) going up and one arrow going down. 2s^2: in this orbital there is one arrow going up and one arrow going down. 2p^2: this has three orbitals with one arrow going up in the first orbital, one arrow going up in the second orbital and nothing (no electron) in the third orbital. Since in 2p^2 both electrons have the same spin, a permanent dipole moment results and there is a strong paramagnetic behavior in an magnetic field in which all the dipoles are lined up. In a similar way it can be understood that the alkaline-earth metals, and e.g. Pt are paramagnetic. For ions, basically the same concept holds to explain paramagnetism.

Although most components of the human body are weakly diamagnetic, many organisms have been shown to contain small amounts of strongly magnetic materials, usually magnetite (Fe_3O_4; ferrimagnetism). The most extreme case is that of magnetotactic bacteria. With their magnetite particles they orient themselves in the earth's magnetic field. Magnetite crystals have also been found in pigeons, honeybees, many mammals, and even in the human brain, but in proportionately much smaller amounts than in the bacteria.
It seems very unlikely that there is enough magnetite within the human body to provide a possible mechanism to explain the highly speculative effects of magnetic therapy.

Diamagnetism

Diamagnetism is the exact opposite as paramagnetism. A diamagnetic substance is repelled, very weakly, by a magnet and is only exhibited in the presence of an externally applied magnetic field as the

result of changes in the orbital motion of electrons. Consequently, diamagnetic substances have a relative magnetic permeability smaller than unity (or negative magnetic susceptibility).
The magnetic field creates a magnetic force on a moving electron, being F=qv x **B**. (see Lorentz force). This force changes the centripetal force on the electron, causing it to either speed up or slow down in its orbital motion. This changed electron speed modifies the magnetic moment of the orbital in a direction against the external field.
It should be noted that dia- and paramagnetism can change when the aggregation state changes and also when an element becomes an ion.

Application
Medical MRI generally utilizes the diamagnetism of water, which relies on the spin numbers of excited hydrogen nuclei in water of the tissue under study. Some nuclei with non-zero spin numbers are (to some extent) parallel or anti-parallel aligned in the strong magnetic field of the MR apparatus (some 2-7 Tesla). The vast quantity of aligned nuclei in a small volume sum to produce a small detectable change in field.
Similar as paramagnetism of deoxy-Hb, the diamagnetism of oxy-Hb is applied in Functional MRI (fMRI) This technique is often applied to establish the change (relative) of oxygen consumption in particular brain structures when e.g. performing some intellectual or sensory-motor task.
A particularly fascinating phenomenon involving diamagnets is that they may be levitated in stable equilibrium in a very strong magnetic field, with no power consumption. The magnetic moment is proportional to the applied field **B**. This means that the magnetic energy of diamagnets is proportional to $\mathbf{B}^2$, the intensity of the magnetic field. In very strong fields, a thin slice of pyrolytic graphite (strongly diamagnetically), a live frog, and water amongst other things have been levitated with success.

More Info

Consider two electron orbitals; one rotating clockwise and the other counterclockwise, say in the plane of this page. An external magnetic field into the page will make the centripetal force on an electron rotating clockwise increase, which increases its moment out of the page. That field would make the centripetal force on an electron rotating counterclockwise decrease, decreasing its moment into the page. Both changes oppose a magnetic field into the page. The induced net magnetic moment is very small in most everyday materials. An example is He. As helium has two orbitals, one with spin-arrow facing up and one with a spin-arrow facing down. Together with the angular moments of both orbitals, hardly any magnetic momentum remains in the atom. Consequently, as other noble gases, Ne is diamagnetic.
All materials show a diamagnetic response in an applied magnetic field; however for materials which show some other form of magnetism, the diamagnetism is completely swamped. Substances which only, or mostly, display diamagnetic behavior are termed diamagnetic materials, or diamagnets. Materials that are said to be diamagnetic are those which are usually considered by non-physicists as "non magnetic", and include water and DNA, most organic compounds such as oil and plastic, and many metals. Superconductors may be considered to be perfect diamagnets.

Magnetophoresis

The existence of unpaired electrons in the four heme groups of deoxy-Hb gives this molecule paramagnetic properties as contrasted to the diamagnetic character of oxy-Hb. Based on the measured magnetic moments of Hb and its compounds, and on the relatively high Hb concentration of human erythrocytes, differential migration of these cells is basically possible if exposed to a high magnetic field. In this way it is possible to measure the migration velocity of deoxy-Hb-containing erythrocytes, exposed to a mean magnetic field of 1.40 T (Tesla) and a mean gradient of 0.131 T/mm, a process called magnetophoresis (e.g. *Biophys J.* 2003;84:2638-45).

Surface tension

Principle

In physics, surface tension is an effect within the surface layer of a liquid (gas-liquid interface) that causes the layer to behave as an elastic sheet. It is the effect that allows insects to walk on water, and that causes Capillary force.

Fig. 1 Surface tension prevents this flower from submerging.

Flat surfaces
Surface tension is caused by the attraction between the liquid molecules, due to various intermolecular forces. In the bulk of the liquid, each molecule is pulled equally in all directions by neighboring liquid molecules, resulting in a net force of zero. At the surface of the liquid, the molecules are pulled inwards by other molecules deeper into the liquid, but there are no liquid molecules on the outside to balance these forces. So, the surface molecules are subject to an inward force of molecular attraction which is balanced by the resistance of the liquid to compression. There may also be a small outward attraction caused by air molecules, but as air is much less dense than the liquid, this force is negligible.
Surface tension is measured in Newton/meter ($N \cdot m^{-1}$) and represented by the symbol γ, σ, or *T* and is defined as the force along a line of unit length parallel to the surface or work done per unit area. In other words, the surface tension ($N \cdot m^{-1}$) is equivalent to energy per square meter ($J \cdot m^{-2}$). This means that surface tension can also be considered as free surface energy. If a surface with surface tension γ is expanded by a unit area, then the increase in the surface's stored energy is also equal to γ.

Drops and gas bubbles
With a completely flat water surface there is no force, which tries to pull a liquid molecule outside the liquid, provided that the gas above the surface is saturated with the molecules of the liquid. However, with a gas bubble in a liquid, the interface attempts to have an as small as possible surface, since doubling the bubble surface doubles the increase in the surface's stored energy. Consequently the surface tension is directed toward the centre of the bubble. Without a counteracting force the bubble will vanish. This can only be prevented when the bubble pressure P_{bubble} balances the surface pressure P_{ST} due to the surface tension plus the ambient pressure P_{amb} in the liquid being the sum of hydrostatic and atmospheric pressure:

$$P_{bubble} = P_{ST} + P_{amb}, \text{ and} \qquad (1)$$

$$P_{ST} = 2\gamma/r$$

where r the radius. P_{bubble} is the sum of the partial pressures of the composing types of gas in the bubble ($\Sigma P_{bubble\ gases}$).

The above equation holds for a pure liquid like water. However, in biological systems there are generally molecules with a hydrophobic and hydrophilic part, like fatty acids. These molecules form a monomolecular membrane at the interface with the hydrophobic end inside the gas and the hydrophilic end in the water.

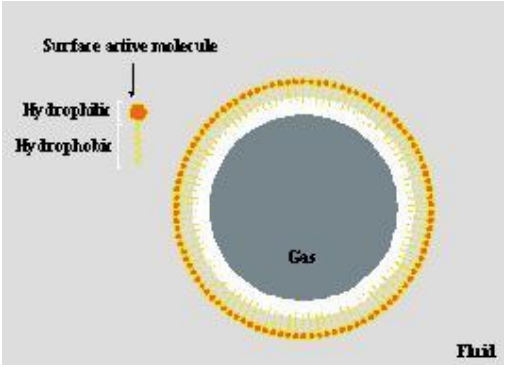


Fig. 2 Skin of surface active molecules surrounding a gas bubble.

Just as the water molecules at the interface "pull" towards each other in surface tension, the surface active molecules, the surfactant, "push" against each other. So, this force is directed outward. This counteracts the effect of surface tension, and therefore eliminates the loss of gas in the bubble by outward diffusion. This membrane reduces the motion of gas molecules from the bubble to the liquid. $P_{surfactant}$ is:

$$P_{surfactant} = 2\gamma_c/r$$

γ_c accounts for the springy "push back" effect of the surfactants. Equation (1) becomes:

$$\Sigma P_{bubble\ gases} + 2\gamma_c/r = P_{amb} + 2\gamma/r. \qquad (2)$$

Due to the surface tension drops, e.g. the small water droplets in fog or floating oil drops in liquid are spherical.

Application

The surfactant is of crucial importance to maintain the shape of the alveoli. Without surfactant they will shrink immediately. The surfactant decreases the effect of γ by a factor of ten. Lung surfactant and surfactant disorders are subject of extensive research.
Venous gas embolism (VGE) and decompression sickness (DCS) of divers is due to the occurrence of pathological nitrogen bubbles in tissues and blood. Generally, VGE occurs during the ascent and DCS after the ascent. The occurrence of these bubbles relies on surface tension of the bubble skin and pressure differences.

More info

A soap bubble comprises a double (or more) molecular membrane of fatty acids (the 'skin') covered by a thin water film at both sides. The inside directed force yielded by the surface tension of both water films is counteracted by the outward directed force provided by the skin. The net effect of both surface tension evokes a pressure within the bubble. Since there is an outer and an inner interface the constant 2 in eq. (2) should be doubled to 4.
Eq. (2) suggests equilibrium, but this cannot be reached. The outer water film will vaporize due to the low p_{H2O} of the outside air. This makes the bubble slightly growing. Also N_2 and O_2 enter the bubble due to partial pressure differences whereas H_2O cannot leave easily the bubble since the skin is a barrier for this polar molecule. This also makes the bubble growing. Also soap molecules and water collect at the bottom due to gravity. These processes destabilize the bubble leading to rupture from top to bottom.

A quantity related to surface tension is the energy of Cohesion, which is the energy released when two bodies of the same liquid become joined after removal of the boundary of separation. Since this process involves the removal of surface from each of the two bodies of liquid, the energy of cohesion is equal to twice the surface energy. A similar concept, the energy of Adhesion applies to two bodies of different liquids.

Waves and wave phenomena

Doppler principle

Principle

The Doppler effect, discovered by Christian Doppler in 1842, is the apparent *change in frequency* (and wavelength) of a wave that is perceived by an observer moving relative to the source of the waves. The waves can be electromagnetic (visible light, X-ray, radio-waves, gamma radiation, etc.), sound waves, gravity waves, surface waves at a liquid (water) etc. For sound waves, the velocity of the observer and the source are reckoned relative to the transmitting medium.

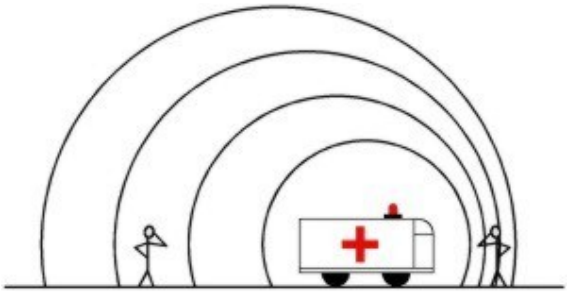


Fig. 1 Sound waves emanating from an ambulance moving to the right.

The total Doppler effect results from either motion of the source or motion of the observer. Each of these effects is analyzed separately. For waves which do not require a medium (such as light) only the relative difference in velocity between the observer and the source needs to be considered. Fig. 1 visualizes how the sound of an ambulance are compressed (perceived frequency increase) in front of the ambulance and 'diluted' (frequency decrease) behind it.
Remind that the frequency of the sounds that the source *emits* does not actually change.
The following *analogy* helps *to understand the Doppler principle*. Someone throws one ball every second in your direction. Assume that balls travel with constant velocity. If the thrower is stationary, you will receive one ball every second. However, if he is moving towards you, you will receive balls more frequently than that because there will be less spacing between the balls. The converse is true if the person is moving away from you. So it is actually the *wavelength*, which is affected; as a consequence, the perceived frequency is also affected.
If the moving source is emitting waves through a medium with an actual frequency f_0, then an observer stationary relative to the medium detects waves with a frequency f given by:

$$f = f_0\left(\frac{c}{c + v_s}\right) \qquad (1)$$

where c is the speed of the waves in the medium and v_s is the speed of the source with respect to the medium (negative if moving towards the observer, positive if moving away), with the observer on the pathway of the source (radial to the observer). With v_s is<<c and Δf, the frequency shift, being $f - f_0$, and applying (1) Δf is:

$$\Delta f = f_0 \frac{v_s}{c} \qquad (2)$$

A similar analysis for a moving observer and a stationary source yields the observed frequency (the observer's velocity being represented as v_o):

$$f = f_0\left(1 + \frac{v_0}{c}\right) \qquad (3)$$

A stationary observer perceives the moving ambulance siren at different pitches depending on its relative direction. The siren will start out higher than its stationary pitch, slide down as it passes, and continue lower than its stationary pitch as it recedes from the observer. The reason the siren slides is because at the moment of passing there is some distance between the ambulance and you. If the siren approached you directly, the pitch would remain constant (as v_s is only the radial component) until the vehicle hit you, and then immediately jump to a new lower pitch. The difference between the higher pitch and rest pitch (when $v_0 = 0$) would be the same as the lower pitch and rest pitch. Because the vehicle passes by you, the radial velocity does not remain constant, but instead varies as a function of the angle between your line of sight and the siren's velocity:

$$f = f_0\left(\frac{c}{c + v_s \cos\theta}\right) \qquad (4)$$

where θ is the angle between the object's forward velocity and the line of sight from the object to the observer.

Applications

The Doppler effect is broadly applied to measure the flow velocity of blood in vessels and the heart with ultrasound. A limitation is that the Ultrasound beam should be as parallel to the blood flow as possible. Other limitations are absorption at interfaces and scatter (e.g. on blood cells). Δf, the Doppler shift is 2 times that of 2) since the emitted ultrasound beams impinges *and* reflects on the blood cells. Δf is generally some hundreds of Hz and can directly made audible by a microphone.

Contrast enhanced ultrasound (CEU) using gas-filled microbubble contrast media can be used to improve velocity or other flow-related medical measurements.
Velocity measurements of blood flow are also used in other fields of Echography (obstetric, neurological).
Instruments such as the *Laser Doppler velocimeter* (LDV), and *Acoustic velocimeter* (ADV) have been developed to measure velocities in a fluid flow.

Measurement of the amount of gas bubbles in the venous system is performed in diving medicine. They are generally measured in the pulmonary artery by a Doppler probe at the 3rd intercostal space. Gas bubbles in the circulation may result in decompression sickness.

"Doppler" has become synonymous with "velocity measurement" in medical imaging. But in many cases it is not the frequency shift (Doppler shift) of the received signal that is measured, but the phase shift (*when* the received signal arrives).

The Doppler effect is also a basic 'tool' in astronomy (red shift, temperature measurements by line broadening) and daily life radar (navigation, speed control).

More info

The LDV (also known as laser Doppler anemometry, or LDA) is a technique for measuring the direction and speed of fluids like air and water. In its simplest form, LDV crosses two beams of collimated, monochromatic light in the flow of the fluid being measured. A microscopic pattern of bright and dark stripes forms in the intersection volume. Small particles in the flow pass through this pattern and reflect light towards a detector, with a characteristic frequency indicating the velocity of the particle passing through the probe volume. LDV may be unreliable near solid surfaces, where stray reflections corrupt the signal. The ADV emits an acoustic beam, and measure the Doppler shift in wavelengths of reflections from particles moving with the flow. This technique allows non-intrusive flow measurements, at high precision and high frequency.

The echo-Doppler technique The description of velocity measurement holds for a continuous emitted ultrasound. With pulsed Doppler, short periods of emitting and receiving are alternated. By adjusting the period length the depth of reflection can be selected. In this way, motion of deeper layers is not disturbing the analysis.
The Doppler method and echography are combined in one apparatus, the *echo-Doppler* and yields a duplex image: in the black and white echo-image, direction and velocity are depicted in red for approaching the probe and blue for leaving. The more blue or red the higher the speed. The most common application is Doppler echocardiography.

Lissajous figure

Principle

In mathematics, a Lissajous figure is the graph of a 2D-motion which occurs when two sinusoidal oscillations with the same frequency or different frequencies (a and b), and each with its own amplitude (A and B) and with a phase difference (φ), are added as vectors with one sinus along the horizontal and one along the vertical axis.
In equations:

$x = A\sin(at+\varphi)$ and $y = B\sin(bt)$,

where t is time.
Depending on the ratio a/b, the values of A, B and φ, the figure is simple or complicated. A=B, a/b=1 and φ=0 yields a line, changing φ to φ=π/2 (radians) a circle is obtained and making A≠B an ellipse. Other ratios of a/b produce more complicated curves (Fig. 1), which are closed only if a/b is a rational fraction (a and b integers). The perception of these curves is often a seemingly 3D-trajectory of motion, an example of a visual illusion.

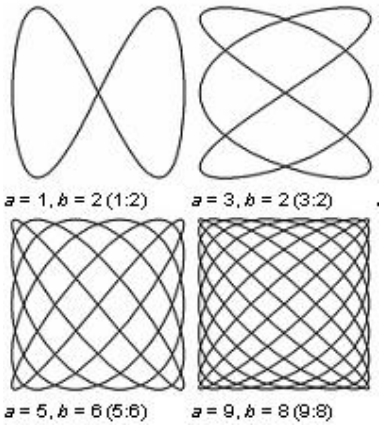
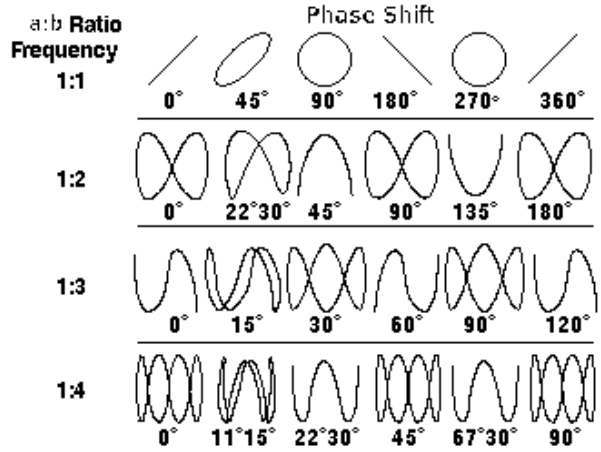


Fig. 1 Examples of Lissajous figures with φ = π/2, a odd, b even, |a − *b*| = 1.

More info

Also parabolas and more complicated figures can be obtained, as depicted in Fig. 2, which shows the relationship of frequency ratio and phase shift φ.



Even though they look similar, *spirographs* are different as they are generally enclosed by a circular boundary where a Lissajous curve is bounded by a square (±1, ±1).

Biomechanics and tissue elasticity

Ballistocardiography

Principle

Ballistocardiography is a noninvasive technique for the assessment of cardiac function by detecting and measure the recoil (the reaction) of the human body due to the blood that the heart is currently pumping (the action). It is the derivative of the momentum (mass x velocity) and consequently has the dimension mass x length/time.
To be more precise a ballistocardiogram (BCG) measures the impact of blood colliding with the aortic arch, which causes the body to have an upward thrust (reaction force) and then the downward thrust of the blood descending. The ballistocardiogram is in the 1-20 Hz frequency range.
One example of the use of a BCG is a ballistocardiographic scale, which measures the recoil of the person's body that is on the scale. A BCG scale is able to show a persons heart rate as well as his weight.
Sensors are often hidden in the upholstery of the chair and the electronics is also hidden. In this way the subject is not aware of recording.

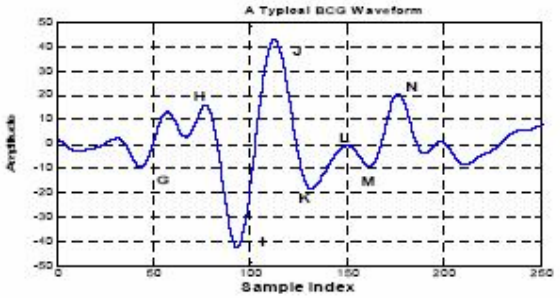


Fig. 1 A BCG signal with spikes and wave complexes (from ref. 1).

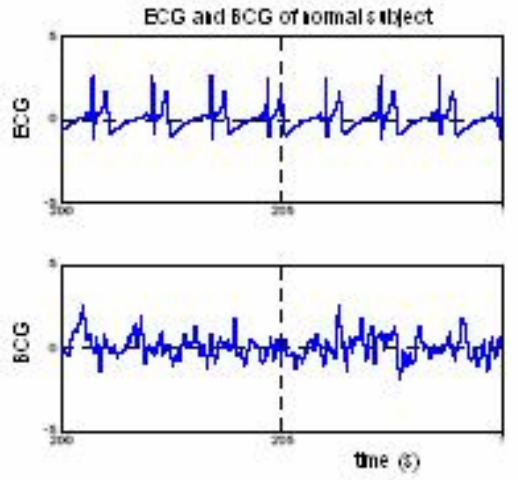


Fig. 2 ECG and BCG records of a normal subject using 1-45 Hz band pass filter for ECG and 1-10 HZ for BCG. (Motion artifacts in BCG signal are not removed).

Application

The charm of the method is that no electrodes are needed to be attached to the body during measurements. This provides a potential application to assess a patient's heart condition at home. Recordings can be transmitted real time to the clinic.
The BCG can show main heart malfunctions by observing and analyzing the BCG signal.

BCG is used in heamodynamic modeling for calculating pulsatile fluid flows with large Womersley number in large arteries around the heart and valves.
A 3D version of BCG has been used in spaceflight during free-floating microgravity.

More info

A BCG can be recorded from the surface of body with accelerometers, specific piezoelectric foil sensors or charge-capacitive sensors (e.g. EMFi sensor). The measured signal is amplified by a charge amplifier.
Often the BCG is recorded together with the ECG (one lead by strips of copper on the arm rests to measure R-peaks for heart rate), respiratory rate, respiratory amplitude, and body movements, all wireless. This integrated technique is also called the static charge-sensitive-bed (SCSB) method. It is for instance used for recording body movements after exercise and during sleep.
As holds for many biological signals (EEG, ECG etc.) analysis is nowadays often performed by wavelet analysis or an or another component analysis. The analysis may be proceeded by applying an artificial neural networks (ANNs).

References

1. www.arehna.di.uoa.gr/Eusipco2005/defevent/papers/cr1069.pdf -
2. Akhbardeh A., Junnila S., Koivistoinen T., Värri A. Applying Novel Supervised Fuzzy Adaptive Resonance Theory (SFART) Neural Network and Biorthogonal Wavelets for Ballistocardiogram Diagnosis. Proceedings of the 2006 IEEE, International Symposium on Intelligent Control, Munich, Germany, October 4-6, 2006.
3. Xinsheng Yu; Dent, D.; Osborn, C. Classification of ballistocardiography using wavelet transform and neural networks. Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE. 1996, 3,:937 - 938

Biomechanics

Principle

Biomechanics is the research and analysis of the mechanics of living organisms on multiple levels, from the level of molecules to tissue, organ and organism.
Some simple examples of biomechanics research include the investigation of the forces that act on limbs, static as well as dynamic, the aerodynamics of bird and insect flight, hydrodynamics of swimming in fish, and locomotion in general across all forms of life, from individual cells to whole organisms. The biomechanics of human motion is called kinesiology.

Applied mechanics, and mechanical engineering disciplines such as fluid mechanics and solid mechanics, play prominent roles in the study of biomechanics. By applying the laws and concepts of physics, biomechanical mechanisms and structures can be simulated and studied.
Loads and deformations can affect the properties of living tissue. For example, the effects of elevated Blood pressure on the mechanics of the arterial wall and bone growth in response to exercise can remodel the vessel or bone anatomy.
Biomechanics in sports science attempts to improve performance in athletic events through modeling, simulation, and measurement.

Application

The following fields of research and application can be considered as Biomechanics.

Fluid dynamics
applies to gases and liquids (see Poiseuille's Law, Reynolds Number, Entrance effect and entrance length, Flow in a bended tube, Flow in bifurcations, Flow through a stenosis, Navier-Stokes equations) of circulation (see Blood flow Bernoulli's and Pascal's Law, Womersley number) and the respiratory system (Lung gas transport 1, basic principles, Lung gas transport 2, pressure, volume and flow, Lung gas transport 3, resistance and compliance).

Biomechanics of bones and soft tissues
Bones are anisotropic (different properties in the different directions) and so bones are strongest along one particular axis and less strong along the two other axes, with the latter two mostly of the same strength.
Soft tissues such as skin, tendon, ligament and cartilage are combinations of matrix proteins and fluid. The function of *tendons* is to connect muscle with bone and is subjected to tensile loads. Tendons must be strong to facilitate movement of the body while at the same time remaining compliant to prevent damage to the muscle tissues. *Ligaments* connect bone to bone and therefore are stiffer than tendons but are relatively close in their Tensile strength. Cartilage, on the other hand, is primarily loaded in compression and acts as a cushion in the joints to distribute loads between bones.
The stress-strain relations can be modeled using Hooke's Law, in which they are related by linear constants known as the Young's modulus or the elastic modulus, the shear modulus and Poisson's ratio. See for detailed descriptions Elasticity and Hooke's law, Elasticity 1: elastic or Young's modulus, Elasticity 2: shear modulus, Elasticity 3: compressibility and bulk modulus, Stiffness and Tensile strength.

Biomechanics of the muscle
The biomechanics of muscle, being skeletal muscle (striated), cardiac muscle (striated) or smooth muscle (not striated) is highly dynamic and often not linear.

More Info

Biomechanics of soft tissue
The theory rather well applies to bone but much les to soft tissues due to non-linear behavior and evident visco-elasticity properties. With visco-elasticity there is energy dissipation, or Hysteresis, between the loading and unloading of the tissue during mechanical tests. Some soft tissues can be preconditioned by repetitive cyclic loading to the extent where the stress-strain curves (see Tensile strength) for the loading and unloading portions of the tests nearly overlap.

Elasticity and Hooke's law

Principle

Hooke's law of elasticity is an approximation which states that the amount by which a material body is deformed (the *strain*) is *linearly related to* the force (the *stress*) causing the deformation. Materials for which Hooke's law is a useful approximation are known as linear-elastic or "Hookean" materials.
As a simple example, if a spring is elongated by some distance ΔL, the restoring force exerted by the spring F, is proportional to ΔL by a constant factor k, the spring constant. Basically, the extension produced is proportional to the load. That is,

$$F = - k \Delta L. \qquad (1a)$$

The negative sign indicates that the force exerted by the spring is in direct opposition to the direction of displacement. It is called a "restoring force", as it tends to restore the system to equilibrium. The potential energy stored in a spring is given by:

$$U = 0.5\, k\, \Delta L^2. \qquad (1b)$$

This comes from adding up the energy it takes to incrementally compress the spring. That is, the integral of force over distance. This potential can be visualized as a parabola on the U- ΔL plane. As the spring is stretched in the positive L-direction, the potential energy increases (the same thing happens as the spring is compressed). The corresponding point on the potential energy curve is higher than that corresponding to the equilibrium position (ΔL =0). Therefore, the tendency for the spring is to decrease its potential energy by returning to its equilibrium (un-stretched) position, just as a ball rolls downhill to decrease its gravitational potential energy. If a mass is attached to the end of such a spring and the system is bumped, it will oscillate with a natural frequency (or resonate with a particular frequency, see Resonance).

For a Hookean material it also holds that ΔL is reciprocally proportional to its cross sectional area A, so $\Delta L \sim A^{-1}$ and ΔL ~ L. When this all holds, we say that the spring is a linear spring. So, Hooke's law, i.e. equation (1a) holds. Generally ΔL is small compared to L.
For many applications, a rod with length *L* and cross sectional area A, can be treated as a linear spring. Its relative extension, the *strain* is denoted by ε and the tension in the material per unit area, the *tensile stress*, by σ. Tensile stress or tension is the stress state leading to expansion; that is, the length of a material tends to increase in the tensile direction.
In formula:

$$\varepsilon = \Delta L/L, \qquad \text{(dimensionless)} \qquad (2a)$$
$$\sigma = E\varepsilon = E\Delta L/L = F/A, \qquad (\mathrm{N/m^2} \equiv \mathrm{Pa}) \qquad (2b)$$

where ΔL is the extension and *E* the modulus of elasticity, also called Young's modulus. Notice that a large *E* yields a small ΔL. *E* is a measure of the stiffness (not the stiffness to prevent bending; see Stiffness) and reciprocal to the mechanical compliance. As (2b) says, it is the ratio σ/ε and so the slope of the stress/strain (σ/ε) curve, see Fig. 1.

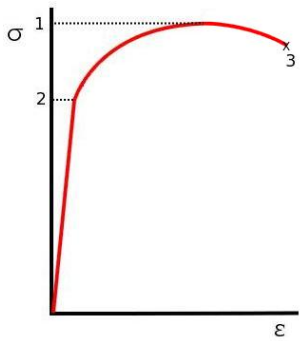


Fig. 1 Stress-strain curve. The slope of the linear part is by definition *E*. 1. ultimate strength, 2 limit of proportional stress. 3 rupture. Notice that ε is the *fraction* of extension (dimensionless) and σ the tensile *force/area*.

The modulus of elasticity *E* is a measure of the stiffness of a given material to prevent linear extension. Hooke's law is only valid for the linear part, the elastic range, of the stress-strain curve (see Fig. 1). When the deforming force increases more and more, the behavior becomes non-linear, i.e. the stress-strain curve deviates from a straight line. In the non-linear part *E* is defined as the rate of change of stress and strain, given by the slope of the curve of Fig. 1, obtained by a tensile test. Outside the linear range for many materials like metals stress generally gives permanent deformation.
With further increasing stress the slope diminishes more and more with increasing strain and materials like metals liquidize. Finally it ruptures. See Tensile strength for more info how materials with different mechanical properties, such as bone and tendon, behave before rupturing.

Materials such as rubber, for which Hooke's law is never valid, are known as "non-hookean". The stiffness of rubber is not only stress dependent, but is also very sensitive to temperature and loading rate.

Applications
Strength calculations of skeleton and cartilage structures, of tendons (especially the archilus and patella tendon, also aging studies) and the elastic behavior of muscles, blood vessels and alveoli. Further aging of bone, thrombolysis (compression modulus, see **More Info**).There are also applications in the hearing system: calculations of the mechanical behavior of the tympanum, middle ear ossicles, auditory windows and the cochlear membranes.
Failure strains (points of rupture) are important to know, especially for biomaterials. For tendons and muscle (unloaded) they go up to 10%, sometimes until 30%. Ligaments have occasionally a failure strain up to 50%. Spider fiber reaches 30% and rubber strands more than 100%. However, these high values all include plastic deformation.
Rubber is a (bio)material with one of the most exotic properties. Its stress-strain behavior exhibits the so called Mullins effect (a kind of memory effect) and Payne effect (a kind of strain dependency) and is often modeled as hyperelastic (strain energy density dependency). Its applications, also in medicine, are numerous.

More Info

Application of the law as described above can become more complicated.

Linear vs non-linear This has already been discussed. There exist all kind of non-linearity's, depending on the material (steel-like, aluminum-like, brittle-like as bone, elastic-like as tendons).
Isotropic and anisotropic. Most metals and ceramics, along with many other materials, are isotropic: their mechanical properties are the same in all directions. A small group of materials, as carbon fiber and composites have a different *E* in different directions. They are anisotropic. Many biomaterials, like muscle, bone, tendon, wood, are also anisotropic. Therefore, *E* is not the same in all three directions. Generally in the length of the structure it is different than in both par-axial directions. Now, σ and ε comprise each 3x3 terms. This gives the tensor expression of Hooke's Law and complicates calculations for biomaterials considerably. Often there is circular symmetry (muscle, tendon) which brings the dimensionality down to 2D (2D elasticity)
Inhomogenity Sometimes, a (biological) material is not homogeneous in some direction, so *E* changes along some axis. This happens in trabecular bone when it is denser at the surface.

Literature
http://www.vki.ac.be/research/themes/annualsurvey/2002/biological_fluid_ea1603v1.pdf

Elasticity 1: elastic or Young's modulus

Principle

There are three primary elastic moduli, each describing a different kind of deformation. These are the elasticity or Young's modulus *E*, shear modulus G (see Elasticity 2: Shear Strength) and the bulk modulus K (see Elasticity 3: compressibility and bulk modulus):
Because all elastic moduli can be derived from Young's modulus, the latter is often referred to simply as the *elastic modulus*.
The modulus of elasticity *E* is a measure of the stiffness of a given material or its resistance to be stretched when a force is applied to it. The elastic modulus is defined as the slope of the curve of tensile stress and tensile strain:

$$E = \Delta\sigma / \Delta\varepsilon,$$

where relative extension (strain) is denoted by ε and the tension in the material per unit area, the tensile stress, by σ. Because strain ε is a unit-less ratio, σ and *E* are measured in Pa. For small values of ε and σ, *E* is constant for some materials (the stress strain curve is linear). But often, especially with biomaterials, generally the curve is nowhere a straight line.
For the linear behavior (straight line) we have:

$$E = \frac{\sigma}{\varepsilon} = \frac{F/A}{\Delta L / L}$$

Table 1 gives some values of *E* for some materials applied in prostheses and biomaterials (room temperature):

Table 1 Young's modulus for some (bio)materials

material	*E* (GPa)	G (GPa)
rubber	0.001	
ZYLON PBO	138-167	
steel	200	75.8
titanium		41.4
vessel wall	0.0001-0.001	
muscle	0.00001	0.00006
tendon	0.1	
bone	0.1-10	4.5 – 8.0*

* human pipe bones

For solids G is about halve K. See for a further description Elasticity and Hooke's law.
Aging makes many biomaterials more brittle with lowering ultimate strength. For bone this is 2%/decade with a 2% increase of *E*.

Application

They are found in the field of the biomechanics of bone and tendon structures in basic medical science. The modulus of elasticity (and other material properties) is of importance for all kind of implants, such as stents, tubes, wires, cardial valves, artificial hearts, etc. and for dental and orthopedic prostheses etc.

More Info

When an object has an isotropic molecular crystalline structure, *E* as well as K are larger than G. In other words, the crystal resist stronger against strain and compression than against shear. Strain gives a distance decrease of some particle with the neighboring particles in the direction of the applied force and in the two perpendicular directions, but to a smaller amount. The latter depends on the Poisson ratio (see Elasticity 3: compressibility and bulk modulus). In the unstressed state this potential energy, the result of solely internal forces, is minimal. Whit strain, external force is added and the potential energy rises.
Compression gives a distance decrease of some particle with all neighboring particles that gives the largest change in potential energy. With shear there is only a distance change in one dimension. Shear and strain makes the crystal anisotropic. This can change various material properties.
Young's modulus is a physical consequence of the Pauli exclusion principle.

Elasticity 2: shear modulus

Principle

The shear modulus or modulus of rigidity (G or sometimes S) is a type of elastic modulus which describes an object's tendency to shear (the deformation of shape at constant volume) when acted upon by opposing forces; it is defined as shear stress over shear strain.

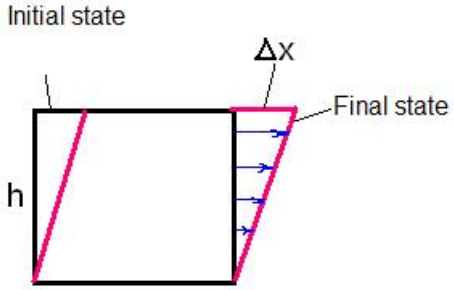


Fig. 1 Deformation by shear stress.

$$G = \frac{F/A}{\Delta x/h} = \frac{Fh}{\Delta xA}, \qquad 1)$$

where F/A is shear stress (force F over cross area A) and *Δx/h* is shear strain (distance of deformation *Δx* over height h).
The shear modulus is one of several quantities, such as compressibility, stiffness, hardness, and toughness (see Tensile strength) for measuring the strength of materials. All of them arise in the generalized Hooke's law (see Elasticity and Hooke's law).
Anisotropic materials such as wood and bone are poorly described by elastic moduli.
G and *E* are related by the Poisson ratio μ (see Elasticity 3: compressibility and bulk modulus):

$$G = E/(2+\mu) \qquad 2)$$

Shear modulus is usually measured in GPa. See Elasticity 1: elastic or Young's modulus for the values of G of some materials.

Application

G is smaller than *E* as (2) shows. Therefore, bone fracture mostly occurs when bended or twisted which gives a high shear. In contrast, tendons rupture by tensile stress.

More Info

In solids, there are two kinds of sound waves, pressure waves and shear waves. The speed of sound for shear waves is controlled by the shear modulus.
The shear modulus is part of the derivation of viscosity.

Elasticity 3: compressibility and bulk modulus

Principle

The bulk modulus *(K)* describes volumetric elasticity, or the tendency of an object's volume to deform when under pressure. It is defined as volumetric compressive stress over volumetric compression, and is the inverse of compressibility.
Compressive strength is the capacity of a material to withstand axially directed pushing forces. When the limit of compressive strength is reached, materials are crushed (bone, concrete). An object under high pressure may undergo processes that will affect the stress-strain curve (see Tensile Strength), due to chemical reactions.

Application

Material choice for the design of prostheses and strength calculations of biomaterials like bone. When a mammalian or avian tubular bone with thin walls is bended, in the outer curvature there is tensile stress and in the inner curvature compressive strength. The shear stress (see Elasticity 2: Shear Modulus) is small but becomes more important the thicker the wall.
Bone can withstand greater 1D compressive stress than tensile stress. K of bone, ca. 50 GPa, is some 2-4 times that of concrete.

More info

The Poisson ratio μ
Here, for simplicity it is assumed that in the three dimensions material properties are the same (a so called isotropic material). When a beam is stretched by tensile force (see Tensile Strength), its diameter (d) changes. The change in diameter Δd, generally a decrease, can be calculated:

$$\Delta d/d = -\mu \Delta L/L, \qquad (1a)$$

where L is the length of the beam, and ΔL the length increase by the tensil force F (see also Tensil Strength). The constant of proportionality μ is the Poisson ratio. Since ΔL/L. = F/*E*A with A is the cross sectional area it holds that:

$$\Delta d/d = \mu F/(EA). \qquad (1b)$$

The range of μ is from 0 to 0.5, and mostly $0.2 < \mu < 0.4$. For metals $\mu \approx 1/3$ and for rubber $\mu \approx 0.5$, which means that rubber hardly changes volume. Some materials behave atypically: increase of d when stretched. This happens with polymer foam.
Since generally ΔL << L, it can easily been proven that the change in volume ΔV (generally increase) is:

$$\Delta V = d^2\Delta L + 2dL\,\Delta d. \qquad (2a)$$

Substitution in (1a) and dividing by $V = d^2L$ gives the relative volume change of the beam:

$$\Delta V/V = (1-2\mu)\Delta L/L = (1-2\mu)F/EA \qquad (2b)$$

With the material constants *E* and μ the changes in shape of an isotropic elastic body can be described when an object is stretched or compressed along one dimension. Objects can be subjected to a compressive force in all directions. For this case a new material constant, the bulk modulus K, comprising *E* and μ, is introduced. The bulk modulus can be seen as an extension of Young's modulus *E* (see Elasticity 1: elastic or Young's modulus) to three dimensions.

Derivation of K
When an elastic body, here as an example a cube (edge length b), is compressed from all sides by a pressure increase of ΔP, then the change of the edge length Δb is:

$$\Delta b = -\frac{1-2\mu}{E} b \cdot \Delta P \qquad (3a)$$

The volume change is by approximation:

$$\Delta V \approx 3b^2\,\Delta b. \qquad (3b)$$

After substitution of Δb from (3b) in (3a) it follows that:

$$\frac{\Delta V}{V} = -\frac{3(1-2\mu)}{E}\Delta P \qquad (4a)$$

The bulk modulus K is defined as:

$$K = \frac{E}{3(1-2\mu)} \qquad (4b)$$

The value of μ determines whether K is smaller (brittle materials), about equal (many metals) or larger (highly elastic materials like rubber) than *E*. Finally, It follows that

$$\frac{\Delta V}{V} = -\frac{1}{K}\Delta P \qquad (4c)$$

This equation says that the higher the compressive strength, the smaller the volume changes.
The material constant K (in Pa) is known for many materials.

Elasticity of the aorta

Nico A.M. Schellart, Dept. of Med. Physics, AMC

Principle

When the heart contracts, blood is pumped into the aorta. Then, pressure P_a, flow and volume V of the aorta increase. We consider now firstly how under assumed validity of the law at of Hooke and Laplace the P-V relation will be. Firstly, we calculate using the law of Laplace how the flow of the aorta relates to P_a.

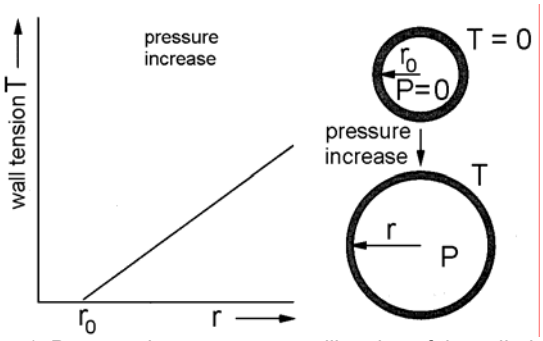


Fig. 1 Pressure increase causes dilatation of the cylinder

Since the force in the aorta wall is $F_w = 2P_a r_i l$, where r_i the aorta inner radius and l its length. Without overpressure, the half inner circumference of the halve of the aorta cylinder is $w = \pi r_{i,0}$ with $r_{i,0}$, the inner radius before the blood ejection of the left ventricle (Fig. 1, right upper part). Then the material cross section is $A = l \cdot d_0$. After ejection the inner radius is increased to $r_{i,t}$. and the change of w, the strain of the elastic aorta wall, is $\Delta w = \pi(r_{i,t} - r_{i,0})$. The wall stress goes from (assumed) zero (equilibrium before ejection) to a value given by Hooke's law:

$$\sigma = E\varepsilon = E\Delta L/L = F/A, \qquad (1)$$

(see Elasticity and Hooke's law), and after ejection, we obtain:

$$\sigma = F_w/A = E\Delta w/w = E(r_{i,t} - r_{i,0})/r_{i,0}. \qquad (2)$$

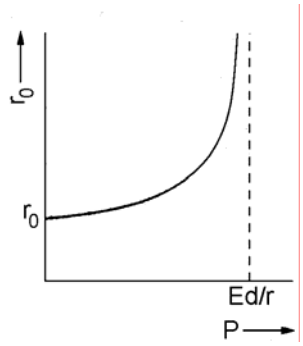


Fig. 2 The relation between the radius r of a cylinder and the pressure in the cylinder

This gives to how the wall tension increases at a cylinder, if by pressure increase the radius of the cylinder increases. In literature one finds except wall stress σ (in N/m^2) also wall tension T= Pr, generally given in N/m. Substitution of $T = P_a r_i$ in (2) provides (after conversion):

$$r_{i,t} = r_{i,0}(1 - P_a r_{i,0}/Ed)^{-1}. \qquad (3)$$

This function is illustrated in Fig. 2.
When the pressure in the aorta increases, also the volume V increases. Using $V = \pi r_{i,0}^2 l$ it follows that:

$$V(P_a) = V_0(1 - P_a r_{i,0}/Ed)^{-2}. \qquad (4)$$

All values must be expressed in preferably the SI system (otherwise convert, e.g. 100 mmHg ~ 13600 N/m^2 = 13.6 kPa). At the age of 50 years, realistic values for the aorta are $R_{i,0}$ = 5.6 mm, E = 5,105 N/m^2, d = 2 mm, V_o=40 mL and *l* = 40 cm.
Fig. 3 depicts equation (4). Above 70 mmHg, this P-V relation ceases to be quadratic. V increases much less and finally the relation becomes sigmoid. The model, derived from the linear approach of elasticity (law of Hooke), only holds well at low values of P_a. At higher values of P_a, collagen in the vessel wall will dominate the properties of the elastine.

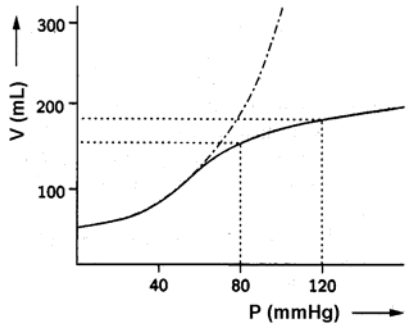


Fig. 3 Aorta volume V as a function of the pressure of the aorta

More Info

As a measure for the ductility of the aorta (compliance) C is used, either in its dynamic (C_{dyn}) or static (C_{stat}) form. C_{dyn} is the pressure derivative of the volume/pressure curve (drawn line in Fig. 3): $C_{dyn} = dV/dP_a$. Fig. 4 shows the impact of the compliance of the aorta. The drawn line represents the pressure in the left ventricle and the dotted line that in the aorta. An approach of the dynamic compliance between 80 and 120 mmHg (the regular values) is obtained by taking dV/dP_a at 100 mmHg, which can be approximated by $C_{dyn,100}$ = (185-160)/(120- 80) = 0.625 mL/mmHg). The static compliance C_{stat} is V/P_a. For P_a = 100 mmHg it follows that $C_{stat,100} \approx 170/100 = 1.7$ mL/(mmHg). When the volume-pressure relation of the aorta is approached by a straight line between the origin and the point with pressure P_a (the ' work point '), the slope of this line represents static compliance. This definition of C_{stat} holds for the aorta as a whole, here with an assumed length of 40 cm. In the literature, C_{stat} is sometimes expressed per unit length of the vessel. The compliance of the aorta results in a rather regular flow, in spite of the pulse-wise time course due to the contractions of the heart.

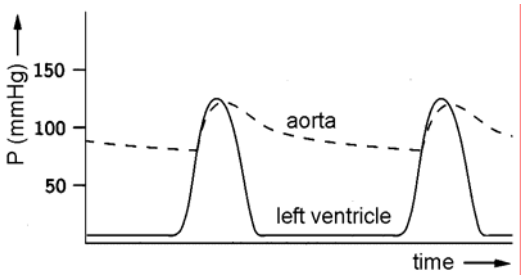


Fig. 4 Effect of the compliance on the aorta pressure

In Fig. 4 shows that after closing the aorta valve P_a (and volume) decreases about exponentially. The total pulsation of P_a is much smaller than that of the ventricle due to the high aorta compliance. If the aorta would be a rigid tube, the blood flow would be more strongly pulsating. This will result in a larger mechanical load of the rest of the circulatory system as follows from the Windkessel model (see Blood pressure: Windkessel model) of the circulatory system. Aging diminish the 'empty' volume V_0 of the aorta. This hardly affects C_{stat} but it seriously decreases C_{dyn}: a decrease of approximately 0.6 mL/mmHg at the age of 50 to approximately 0.2 mL/mmHg at 80 years.

Literature

Van Oosterom, A and Oostendorp, T.F. Medische Fysica, 2nd edition, Elsevier gezondheidszorg, Maarssen, 2001.
N. Westerhof, Noble M.I.M and Stergiopulos N. Snapshots of hemodynamics: an aid for clinical research and graduate education, 2004, Springer Verlag.

Laplace's law

Principle

The law of Laplace holds for fluid or gas filled hollow objects (e.g. blood vessels, heart, alveoli). It gives the relation between transmural pressure P, and wall tension, T. With a soap bubble as an example, the classical approach, this tension is directly related to the Surface tension and consequently has the dimension N/m.
For a circular cylinder with length *l* and an inner radius r_i., as model of a blood vessel, P acts to push the two halves apart with a force equal to P times the area of the horizontal projection of the cylinder, being $2l \cdot r_i$ (Fig. 1). It acts on both halves, so the force F_P is:

$$F_P = 2Pl r_i. \qquad (1a)$$

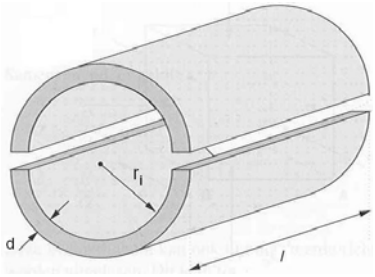


Fig. 1

The two halves are kept together by wall stress, σ_w, acting in the wall only. The related force F_w, visualized in Fig. 2, is thus:

$$F_W = 2\sigma_w l \cdot d, \qquad (1b)$$

where d the wall thickness. 2d*l* is two times the longitudinal cross area of the wall itself, at which this force is acting (Fig. 2). This force is in equilibrium with F_P and thus: $2Pl r_i = 2\sigma_w l \cdot d$. This results in the form of the law of Laplace as mostly used in hemodynamics. It presents the relation between P within the lumen and σ:

$$\sigma_w = P r_i d^{-1}. \qquad (2)$$

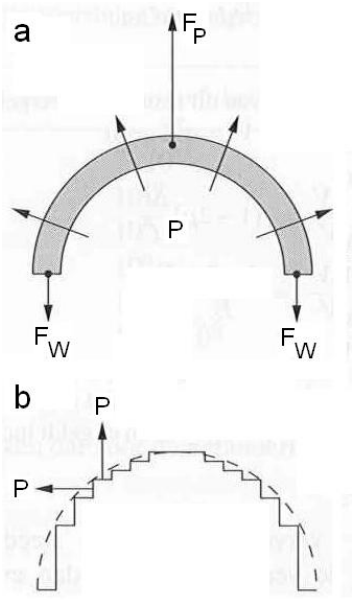


Fig. 2 (a) Visualization of the forces acting on and in the wall. (b) Cancellation of the sideward directed forces due to P. The horizontal projection of the inside of the cylinder and P directly gives F_P (eq. 1a).

Stress has the dimension N/m^2, the same as pressure. In other words it says that pressure and wall stress are related by the ratio of radius over wall thickness. With constant d and P, but increasing r_i, σ_w increases proportionally with r_i, At a certain r_i, ultimate strength is reached (see Tensile strength) and rupture occurs (along the length axis of the cylinder).

Another measure is the wall tension T, the force per unit length of the cylinder:

$$T_W = F_W(2l)^{-1} = Pr_i. \qquad (3a)$$

For a sphere, a similar derivation holds and the result is $\sigma_w = ½Pr_i/d$ and T_W is $= ½Pr_i$.

.

$$T_W \text{ is } = ½Pr_i. \qquad (3b)$$

Application

The law is of importance in (experimental) vascular medicine and cardiology, and experimental pulmology.
Assuming a simple shape such as a sphere, or circular cylinder, the law may be applied to the ventricular wall in diastole and systole, as well as to the vessel wall. The law of Laplace can also be used in the contracting heart, where the force is generated in the wall and the rise in pressure is the result of the contracting muscle.
Since the wall cross sectional area of a blood vessel is constant d decreases inversely with r_i, and consequently $\sigma_w \sim r_i^2$. Therefore, with hypertension, rupture of a vein or aneurism may occur with a small increase of r_i.

More Info

A more general form of the Laplace equation holding for a hollow organ of any shape is $T_W = Pr_c$, where r_c, is the (local) curvature radius. The “mean” curvature radius of any 3D shape is defined as:

$$1/r_c = 1/r_1 + 1/r_2, \qquad (4)$$

where r_1 is the radius of the largest tangent circle and r_2, that of the largest one perpendicular on the plane of the former. For a sphere $r_1 = r_2$, and for a cylinder $r_1 = \infty$ and r_2, $= r_i$. After calculating r_c for the cylinder and sphere respectively (3a) and (3b) directly follow.

The law of Laplace applies to any geometry, irrespective of whether the material is linear or nonlinear elastic or if the wall is thin or thick. The only limitation of Laplace's law is that it yields the average wall stress and thus it cannot give any information on the stress distribution across the wall. For cylindrical geometries, and assuming linearly elastic (Hookean, see Elasticity and Hooke's law) material the distribution of circumferential stress or hoop stress across the wall thickness can be approximated by:

$$\sigma_{w,r} = Pr_i^2 (1+ r_o^2/r^2)/ (r_o^2- r_i^2), \qquad (5)$$

where r_o and r_i are the external and internal radius, respectively, and r is the position within the wall for which local stress is calculated.
The relevant force related to (local) wall stress of the heart muscle of the left ventricle is often calculated for the ventricular “equatorial” plane. It is $F = P{\cdot}A_e$, where A_e is equatorial cavity cross-sectional area and P luminal pressure. The wall stress σ_w is given by F/A_w, with A_w the equatorial cross-sectional area of the muscle ring. Thus $\sigma_w = P{\cdot}A_e/A_w$.
Many other relations between wall force or stress and ventricular pressure have been reported, but since measurement of wall force is still not possible, it is difficult to decide which relation is best.

Relation to the Young’s modulus
For a relatively thin arterial wall (d« r_i and incompressible) one can use Laplace's Law to derive the following expression for the incremental elastic modulus:

$$E = (r_i^2/d)\Delta P/\Delta r_i. \qquad (6)$$

For thick walls, as is often the case in arteries, the Young’s modulus (see Elasticity 1: elastic or Young’s modulus) is best derived from the measurement of pressure and radius:

$$E = 3\, r_i^2\, r_o(\Delta P/2\Delta r_o)/(r_o^2- r_i^2) \qquad (7)$$

Literature

Van Oosterom, A and Oostendorp, T.F. Medische Fysica, 2nd edition, Elsevier gezondheidszorg, Maarssen, 2001.
N. Westerhof, Noble M.I.M and Stergiopulos N. Snapshots of hemodynamics: an aid for clinical research and graduate education, 2004, Springer Verlag.

Stiffness

Principle

Stiffness is the resistance of an elastic body to deflection or bending by an applied force.
In solid mechanics, a material behaves elastically if it changes shape due to an applied load, and that when the load is removed, recovers its original shape. According to solid mechanics theory, *every* material will change shape when loads are applied to it (even very small loads). Furthermore, every material will undergo elastic deformation as long as the loads are kept under a certain limit. This limit is known as the yield strength (see Tensile strength and other subjects about elasticity) of the material, and is one way of defining its strength.
The elasticity of a solid is inversely proportional to its stiffness. Stiffness, when corrected for the dimensions of the solid, becomes modulus of elasticity *E*, an intrinsic material property (see Elasticity 1: elastic or Young's modulus). Considering *E*, stiffness should be interpreted as the resistance against linear strain, not as stiffness against bending. However, the latter is directly related to *E* and of principle importance in daily engineering, but also in bio-engineering and bio-mechanics.

The bending-stiffness *k* (N/m) of a body that deflects a distance δ under an applied force F (Fig. 1) is:

$$k = F/\delta$$

The elastic deflection δ (m) and angle of deflection φ (radians) in the example of Fig. 1, a cantilever (single sided clamped) beam with width B and height H is:

$$\delta = F \cdot L^3 / (3 \cdot E \cdot I) \qquad (1)$$

$$\varphi = F \cdot L^2 / (2 \cdot E \cdot I) \qquad (2)$$

where
F = force acting on the tip of the beam
L = length of the beam (span)
E = modulus of elasticity

$$I = \text{area of momentum} = B \cdot H^3/12 \qquad (3)$$

where H is height and B width.

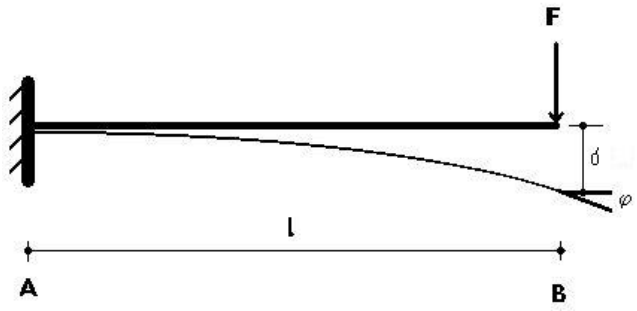


Fig. 1 Cantilever beam with load at the tip.

From (1) it follows that the ratio L/H is the most determining factor. Doubling L with the same H and B increases deflection 8 fold. For given L, F and *E*, the extend of bending is only dependent on ***I***, that is given by the cross sectional size and shape. With the momentums of a circular rod versus a square beam and a pipe versus a square box girder, with the conditions that the outer cross sectional area's are equal (on the basis of anatomical space limitations) and the cross sectional material area's are also the same (equal ''investment" of grow and "maintenance"), it can be proved that irrespective wall thickness the ratio $\delta_{pipe}/\delta_{box\ girder} = 12/4\pi = 0.955$. This small draw back of the tubular shape of bones is probably compensated by a more efficient development of round structures versus structures with sharp edges.

Application

In the field of mechanics of biomaterials, stiffness is mainly of importance for bones.
Equation (1) implies that a high *E* is sought when deflections are undesirable (orthopedic prostheses, like bars, plates, joints), while a relatively low *E* is required when (some) flexibility is needed (flexible orthopedic prostheses, stents).
Tubular bones can generally be considered as round tubes when the bone tissue in the medulla is ignored. (Actually, the mesh of small bone-beams of the medulla considerably increases stiffness of the tubular bone). The middle section (tubular structure) of the human femur and tibia have a medulla/total diameter ratio of ca. 0.55. That of the humerus is ca. 0.6.
Using the same amount bone material, tubular bending stiffness increases with a factor $(2k^2-1)$ where k is the factor of increase of the outer diameter with respect to the diameter of the massive rod. For the femur and tibia this means an increase in stiffness of some 67% compared to the rod design (k=1.15). Wall thickness can also be expressed in k: $0.5k-0.5(k^2-1)^{0.5}$. However, the increase in outer diameter isn't unlimited. Other mechanical (and anatomical) properties and constraints determine an optimal diameter.
The diaphysis of the human femur has a practically circular outer diameter, more than any other human tubular bone. However, the medulla is less cylindrical than the outer diameter with a larger width in the antero-posterior direction (59% of diameter) than in the medio-lateral direction (46%). The reversed may be expected since in the anterior-posterior direction a femur should be stiffer than in lateral direction. Forces act mainly in the median plane, not in the medio-lateral (coronal) plane. However, in the medial plane limb stiffness (with the knee as a hinge-joint) is also provided by tendons and muscles, whereas in the coronal plane this contribution is much less. The cross section of the humerus is more concentric, yielding more equal stiffness in both directions. The forces acting on the humerus are probably also more equal in the different planes than for the femur.

More Info

The above example of the femur can be approximated by a square box girder with a rectangular inner shape as depicted in Fig. 2. Its area moment is $(BH^3 - bh^3)/12$. Starting from a square inner shape material is taking away from the anterior and posterior walls and used for an increase of wall thickness of the lateral sides, such that the total material spent is the same. Normalizing the moment of the square profile at 1, and calculating the moment for the dashed inner shape, the area moment for the anterior/posterior and lateral direction is (1-0.81*0.36) and (1-0.16*0.36) respectively. The latter is clearly larger (33%), consequently bending in the medial plane is larger (see (1)).
In addition to stiffness, bending rupture and yield strength are of importance. Taking also the bone density of the medulla into account, making calculations more realistic, the femur seems to be more designed for high yield strength (see Tensile strength) and to withstand bending fracture than stiffness. The above makes clear that even for parts of extremities with static conditions biomechanics is already complicated. To understand static mechanics better we need the *concept of the neutral line*, which is beyond the scope of this compendium.

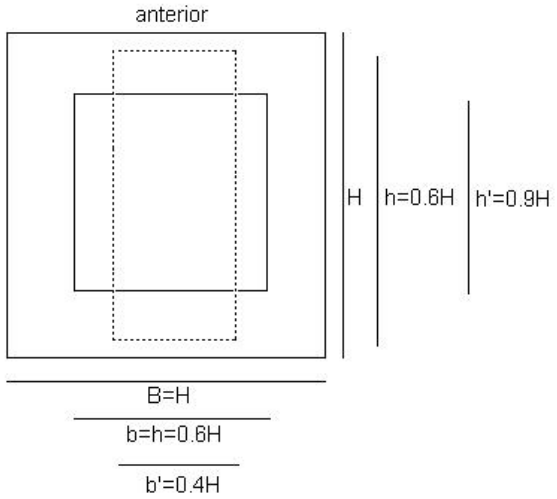


Fig. 2 Basic approximation of a non-concentric tubular bone cross section. The box girder with dashed inner circumference has the same material cross section as the square one.

As both the F and ð are vectors, in general their relationship is characterized by a stiffness matrix for the 3 dimensions. The displacement can, in general, refer to a point distinct from that where the force is

applied and a complicated structure will not deflect purely in the same direction as an applied force. The stiffness matrix enables such systems to be characterized in straightforward terms. For a mathematical treatise see http://www.efunda.com/formulae/solid_mechanics/mat_mechanics/hooke_orthotropic.cfm. Modern engineering design software can calculate stiffness, bending strength etc. for complicated 3-D (biological) structures.

References

Évinger S, Suhai B, Bernáth B, Gerics B, Ildikó Pap I, and Gábor Horváth G. How does the relative wall thickness of human femora follow the biomechanical optima? An experimental study on mummies. J Exp Biol 208, 899-905 (2005)

Drapeau MSM and Streete MA. Modeling and remodeling responses to normal loading in the human lower limb, Am J Physic Anthropol 129:403–409 (2006)

Elizabeth Weiss, Humeral Cross-Sectional Morphology From 18th Century Quebec Prisoners of War: Limits to Activity Reconstruction, Am J Physic Anthropol, 126:311–317 (2005).

Tensile strength

Principle

Tensile stress or tension is the stress state leading to expansion; that is, the length of a material tends to increase in the tensile direction. The volume of the material stays more or less constant. Therefore in a uni-axial material the length increases in the tensile stress direction and the other two directions will (with most materials) decrease in size. The par-axial decrease is given by the Poisson ratio μ. Tensile stress is the opposite of compressive stress. Objects designed to resist tensile stress are ropes, nails, bolts etc. Beams or rod subjected to bending have tensile stress as well as compressive stress (see Elasticity 3: compressibility and bulk modulus) and/or shear stress (Elasticity 2: shear modulus). Tensile stress may be increased until the reach of tensile strength, namely the *limit state* of stress. This is the maximum amount of tensile stress that it can be subjected to before failure. The definition of failure can vary according to material type and design methodology. There are three typical definitions of tensile strength:

- Yield Strength - The stress a material can withstand without permanent deformation. This is the linear part of the stress-strain curve.
- Ultimate Strength - The maximum stress a material can withstand. This is generally higher than the breaking strength.
- Breaking Strength - The stress value on the stress-strain curve at the point of rupture.

The various definitions of tensile strength are shown in Fig. 1 and Fig. 1 of Elasticity and Hooke's law.

Many metals have a linear stress-strain relationship up to a sharply defined yield point. For stresses below this yield strength all deformation is recoverable, and the material will relax into its initial shape when the load is removed. For stresses above the yield point, a portion of the deformation is not recoverable, and the material will not relax into its initial shape. This unrecoverable deformation is known as plastic deformation. Sometimes, these plastic deformations can be large without fracture (in metals such as being drawn into a wire): the material is ductile. For many technical applications plastic deformation is unacceptable, and the yield strength is used as the design limitation. Brittle materials such as concrete, carbon fiber and bone (more or less) do not have a yield point at all. They behave very different from ductile materials. A stress-strain curve for a typical brittle material is shown in Fig. 1.

The table gives the some yield strengths and ultimate strength of some materials used in prostheses and some biomaterials:

Material	Yield strength (MPa)	Ultimate strength (Mpa)	Density (g/cm^3)
Polypropylene	12-43	19.7-80	0.89-0.93
Stainless steel AISI 302	520	860	7.9
Titanium Alloy (6% Al, 4% V)	830	900	4.51
Nylon, type 6/6	45	75	1.14
Kevlar		3600	1.4
ZYLON (PBO)		5800	1.56
Rubber		15	1.1-1.2
Spider silk	1150 (?)	1200	1.3
Pine Wood (parallel to grain)	11	40	0.61
Bone (limb)	17	130	1.8
Patella tendon		60	0.8 – 2.7
Dentine	17		2.1-3
Enamal	67		

Applications

Material properties of artificial materials of human prostheses (orthopedic, vascular, artificial heart valves, wires etc.) are of ultimate importance since they determine the risk of a defect and the interval of displacement. For cardiovascular applications they are crucial. Much research is done to improve their mechanics.

The stiffness of compact bone tissue to prevent extension depends on the bone from which it is taken. Fibular bone has a Young's modulus *E* (see Elasticity 1: elastic or Young's modulus) about 18% greater,

and tibial bone about 7% greater, than that of femoral bone. The differences are associated with differences in the histology of the bone tissue. In biomaterials, tensile strength is different for the different directions. Bone is elastically anisotropic, i.e. its properties depend on direction. Such behavior is unlike that of steel, aluminum and most plastics, but is similar to that of wood. For example for the human femur, in longitudinal direction *E* is 135 MPa, but transversally only 53 MPa.
Severe tendon injuries nearly always result in chronic lengthening. The failure strain of tendons is about 10% of their rest length.

More info

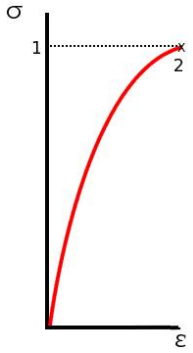


Fig. 1 Stress-strain curve typical of a brittle material (e.g. bone). 1. ultimate strength, 2. rupture.

Tensile strength is measured in Pa (≡ N/m²), The breaking strength of a rope or tendon is specified in Newton without specifying the cross-sectional area of the rope (e.g. supraspinal tendon 900 N). This is often loosely called tensile strength, but this not a strictly correct use of the term.
In brittle materials such as rock, concrete, cast iron, tensile strength is negligible compared to the compressive strength and it is assumed zero for most engineering applications.
Tensile strength can be measured for liquids as well as solids. For example, when a tree draws water from its roots to its upper leaves by transpiration the column of water is pulled upwards from the top by capillary action and this force is transmitted down the column by its tensile strength.

Single-walled carbon nanotubes made in academic labs have the highest tensile strength of any material yet measured, with labs producing nanotubes with a tensile strength of 63 GPa (63,000 MPa) well below its theoretical tensile strength of 300 GPa (300,000 MPa). As of 2004, however, no macroscopic object constructed using a nanotube-based material has had a tensile strength remotely approaching this figure, or substantially exceeding that of high-strength materials like Kevlar.

A parameter derived from the stress-strain curve is toughness, the resistance to fracture of a material when stressed. It is defined as the amount of energy that a material can absorb before rupturing, and can be found by finding the area (i.e., by taking the integral) underneath the stress-strain curve. It is the same as $-0.5k\Delta L^2$ where k is the spring constant and ΔL the strain (see Elasticity and Hooke's law). Toughness, often expressed as the *Modulus of Toughness*, is measured in J/m^3.

Literature
http://silver.neep.wisc.edu/~lakes/BoneAniso.html

Torsion

Principle

Torsion occurs when an object is twisted or screwed around its axis. Torsion can be the result of an applied torque. It is a kind of shear stress (see Elasticity 2: shear modulus). For circular sections, the shearing stress at a point on a transverse plane is always perpendicular to the radius to the point.
The torsion coefficient is a property of torsion springs. It is the torque required to twist a body through an angle of one radian (1 rad=360/π) and is usually denoted by *K*. Therefore it is given as
$K = \tau/\theta$ (Nm/rad)
where τ is the torque and θ is the angle in radians. This equation is analogous to the spring constant given by Hooke's law (see Elasticity and Hooke's law) only used for linear strain along an axis.
Torque can be thought of informally as "rotational force". The force applied to a lever, multiplied by its distance from the lever's fulcrum (pivot point), is the torque. Torque, τ, also called moment or couple, is commonly measured in Nm. More generally, one may define torque as the *vector product*:
$\tau = r \times F$
where F is the force vector and r is the vector from the axis of rotation to the point on which the force is acting. This assumes the force is in a direction at right angles to the straight lever.
Since τ is a vector, it cannot be expressed in Joule, a scalar unit. The rotational analogues of force, mass, and acceleration are torque, moment of inertia and angular acceleration respectively. For example, a force of 3 N applied 2 m from the fulcrum exerts the same torque as 1 N applied 6 m from the fulcrum.

The joule is also defined as 1 Nm, but this unit is not used for torque. Since energy can be thought of as the result of "force dot distance", energy is always a scalar whereas torque is "force cross distance" and so a vector-valued quantity. A torque τ of 1 Nm applied through a full revolution will require an energy E of exactly 2πJ (from $E = \tau \cdot \theta = 1 \times 2\pi$ J).

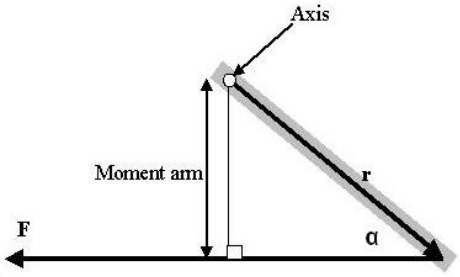


Fig. 1 Moment-arm diagram

If a force of magnitude F is at an angle α from the displacement arm of length *r* (and within the plane perpendicular to the rotation axis, see Fig. 1), then from the definition of cross product, the magnitude of the torque arising is:

$$\tau = r\sin(\alpha) \times F.$$

Application

Torque is applied in the models about axial rotation in joint and about the saccadic eye movements.
A simplified model of eye rotation is that of Westheimer, which comprises inertia J due to the mass of the eye ball, friction B due to the ball/orbita friction, and stiffness K due to the muscles and tendons. Together they yield a second order linear system (see **More Info** and Linear systems). The system is visualized in Fig. 2.

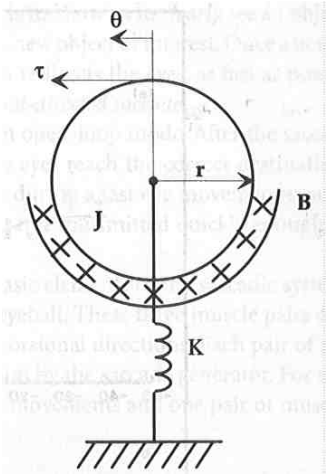


Fig. 2 The eyeball position is given by angle Θ. The stiffness is visualized by the symbol of a spring.

Other important applications are the torque calculations in a hip and shoulder joint, specifically in the field orthopedic revalidation, orthopedics of elderly, and sports medicine.

More info

Static equilibrium
For an object to be in static equilibrium, not only must the sum of the forces be zero, but also the sum of the torques (moments) about any point. So, for the forces F, the sum of the forces requirement is $\Sigma F = 0$ and, and for the torque $\Sigma \tau = 0$. That is, to solve statically determinate equilibrium problems in 3-D we need 2 times 3 (dimensions) is 6 equations.

Torque as a function of time
Torque is the time derivative of angular momentum, just as force is the time derivative of mass times velocity. For multiple torques acting simultaneously:

$$\Sigma\tau = d\mathbf{L}/dt \qquad !1)$$

where **L** is angular momentum:

$$\mathbf{L} = \omega \cdot \boldsymbol{I} = 2\pi f m r^2,$$

where ω is angular velocity, ***I*** moment of inertia, f rotation frequency (rotations/s) and m mass. So, if ***I*** is constant,

$$T = \boldsymbol{I} \cdot d\omega/dt = \boldsymbol{I}\boldsymbol{\alpha} \qquad (2)$$

where **α** is angular acceleration, a quantity usually measured in rad/s^2.

Relationship between torque and power
If a force is allowed to act through a distance, it is doing mechanical work. Similarly, if torque is allowed to act through a rotational distance, it is doing work. Power is the work per unit time. However, time and rotational distance are related by the angular speed where each revolution results in the circumference of the circle being traveled by the force that is generating the torque. This means that torque that is causing the angular speed v_{ang} is doing work and the generated power may be calculated as:
$P = \tau \cdot v_{ang}$. (3)
Mathematically, the equation may be rearranged to compute torque for a given power output. However, in practice there is no direct way to measure power whereas torque and angular speed can be measured directly.
Consistent units must be used. For metric SI units torque is in Nm and v_{ang} in rad/s (not revolutions per second). For different units of power, torque, or angular speed, a conversion factor must be inserted into the equation. For example, if v_{ang} in is measured in revolutions instead of radians, a conversion factor of 2π must be added because there are 2π radians in a revolution:

$$P = \tau \cdot 2\pi \cdot n_{rev}, \qquad (4)$$

where n_{rev}, the rotational speed, is in revolutions per unit time.

For a rotating object, the *linear distance* covered at the circumference in a radian of rotation is the product of the radius with the angular speed. That is: linear speed = radius x angular speed. By definition, linear distance = $v \cdot t = r \cdot v_{ang} \cdot t$.
By substituting $F = r/\tau$ (from the definition of torque) into the definition of power, being P = F times linear distance/t , the power can be calculated: $P = (\tau/r)\ (\ r \cdot n_{ang}) \cdot t)/t = \tau \cdot n_{ang}$. (5)
If the rotational speed is measured in revolutions per unit of time, the linear speed and distance are increased proportionately by 2π in the above derivation to give (4).

The model of rotatiuon of the eye ball
This model can be written as:

$$\tau(t) = Jd\theta/dt^2 + Bd\theta/dt^2 + kd\theta/dt^2 \qquad (6).$$

The solution in Laplace notation is:

$$H(s) = \frac{\omega_n^2 / K}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \qquad (7)$$

where ω_n = 120 rad/s and ζ = 0.7. Consequently, the system is well damped.

Reference
Enderle J.D. The fast eye movement control system. In: The biomedical engineering handbook, Bronzino (ed), CRC Press, Boca Raton, pp 2473-24

Transport

Bernoulli's and Pascal's Law

Principle

The hydrostatic pressure in a tube is visualized in Fig. 1 and is given by Pascal's Law:

$$P_1 + \rho g h_1 = P_2 + \rho g h_2 = \text{constant}, \qquad (1)$$

where
ρ = fluid density (kg/m^3);
g = acceleration due to gravity on Earth (m/s^2);
h = height from an arbitrary point in the direction of gravity (m).
P = pressure somewhere imposed in the tube (N/m^2, Pa).
The second term is due to gravity. Fig. 1 illustrates the hydrostatic condition in the tube.

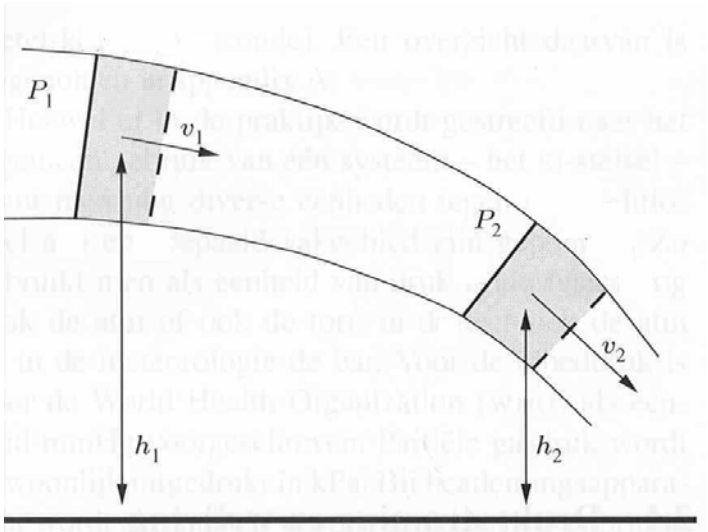


Fig. 1 Visualizing of Pascal's law and Bernoulli's equation in a tube.

By adding an inertia term, ½ρv^2, the kinetic energy per unit mass of the moving liquid with velocity v, Pascal's Law evolves to Bernoulli's equation:

$$\tfrac{1}{2}\rho v^2 + \rho g h + P = \text{constant} \qquad (2)$$

Resuming, in fluid dynamics, Bernoulli's equation, describes the behavior of a fluid (gases included) moving in the tube. More precisely, the flow is along a streamline (a line which is everywhere perpendicular to the local velocity flow). There are typically two different formulations of the equations; one applies to incompressible fluids and (2) holds. The other applies to compressible fluids (see **More Info**).

These assumptions must be met for the equation to apply:
- inviscid flow viscosity (internal friction) = 0;
- steady flow;
- incompressible flow, *ρ* is constant along a streamline. However, *ρ* may vary from streamline to streamline;

For constant-density steady flow, it applies throughout the entire flow field. Otherwise, the equation applies along a streamline.
A decrease in pressure occurs simultaneously with an increase in velocity, as predicted by the equation, is often called Bernoulli's principle.
Bernoulli's equation (2) under the above conditions and considering one dimension is a special, simple case of the Navier-Stokes equations.

Applications

In devices and instruments to calculate pressure in gas and liquid flows. In science, the Bunsen burner, the water-jet pump (venturi-principle), the Pitot tube, In the (path)physiology of the vascular and airways system to calculate and measure pressure, for instance before and after a stenosis (see Flow through a stenosis). For flow in the vascular and airways system many refinements are needed as those of **More Info**, and even more than that. Two approaches are feasible, a semi-analytic approach with applying practical 'rules' of technical liquid dynamics or a numerical approach with mathematical tools like the finite element approach.

More info

A second, more general form of Bernoulli's equation may be written for compressible fluids, in which case, following a streamline, we have:

$$v^2/2 + \varphi + w = \text{constant} \qquad (3)$$

Here, φ is the gravitational potential energy per unit mass, which is just φ = gh in the case of a uniform gravitational field, and w is the fluid enthalpy per unit mass, which is also often written as h. Further it holds that:

$$w = \varepsilon + P/\rho. \qquad (4)$$

ε is the fluid thermodynamic energy per unit mass, also known as the specific internal energy or "sie". With constant P, the term on the right hand side is often called the Bernoulli constant and denoted *b*. For steady inviscid flow (see Gas Laws) with no additional sources or sinks of energy, *b* is constant along any given streamline. Even more generally when *b* may vary along streamlines, it still proves a useful parameter, related to the "head" of the fluid.

Classically, the pressure drop over a tube (volume flow Q, length L and diameter d) with incompressible and with compressible flow is:

$$\Delta P = k_1 Q + k_2 Q^2 \qquad (5)$$

The first term applies for the laminar, the viscous component of the resistance and the 2nd for the turbulent part. Both are not implied in (2) and (3). The laminar part comprises the friction within the tube, and extra friction introduced by irregularities like non-smoothness of the tube wall, constrictions (see Flow through a stenosis), bends (see Flow in a bended tube) and bifurcations (see Flow in bifurcations). The same irregularities can also contribute to the second term when the flow through the irregularities is turbulent on basis of the Reynolds number. For transitional flow both terms contribute.
The influence of non-smoothness of the tube wall is expressed in a resistance factor λ. For Re<2300 it is irrelevant (always laminar). For 4000<Re<10^5 the pressure drop ΔP is given by the Darcy-Weisbach equation:

$$\Delta P = 0.5\lambda\cdot\rho\cdot L\cdot v^2/d = 8\cdot\pi^{-2}\cdot\lambda\cdot\rho\cdot L\cdot d^{-5}\cdot Q^2, \qquad (6)$$

where
$\lambda = 0.316/Re^{0.25}$ (according to the Blasius boundary layer).
For gas flow in a smooth tube there is an expression covering the various types of flow (ref. 1):

$$\Delta P_{tube} = 8\cdot\pi^{-7/4}\cdot\eta^{1/4}\cdot\rho^{3/4}\cdot L\cdot d^{-19/4}\cdot Q^{7/4}, \qquad (7)$$

where η is the dynamic gas viscosity (Pa·s).

Literature

1. Clarke J.R. and Flook V. Respiratory function at depth, in: The Lung at Depth, Lundgren C.E.G. and Miller J.N. (Ed), Marcel Dekker Inc., New York, Basel, pp 1-71,1999.
2. Van Oosterom, A and Oostendorp, T.F. Medische Fysica, 2nd edition, Elsevier gezondheidszorg, Maarssen, 2001.
3. Wikipedia

Blood flow

Principle

Blood flow in arteries can be characterized by the combination of three key phenomena:
- internal laminar (or sometimes turbulent) flow, with and without flow irregularities due to a stenosis, a curvature or a bifurcation;
- pulsatile flow, diminishing from aorta to capillaries.
- compliant arterial wall boundaries;

Each of them has a dramatic effect on the flow pattern.

More Info

For aorta flow and geometry the entrance effect due to the aortic valve (see Entrance effect and entrance Length), characterized by the entry length, is at least 150 cm, which is far greater than the length of the aorta. The flow in the aorta thus cannot be characterized as fully developed, i.e. laminar Poiseuille flow (see Poiseuille's Law). In addition to the entrance effect the proximal aorta is bended. The bend causes extra resistance and therefore pressure drop (see Flow in a bended tube).
In the bend, there is strong asymmetry due to the centripetal forces at the outer curvature. Branching causes also entry phenomena such as asymmetries in the velocity patterns (see Flow in bifurcations). Also complicated secondary flows perpendicular at the axial flow direction may occur, and even flow separation, all of which are far more difficult to analyze than simple steady-state fully developed Poiseuille flow. Secondary flows in curvatures and bifurcations are characterized by a swirling, helical component superimposed on the main streamwise velocity along the tube axis. In fact, all the larger arteries of the circulatory system, including the epicardial coronary vessels are subject to entrance effects.
In large vessels the inertia character overwhelms the viscid character. The inertance with the dimension of mass can be expressed as $\mathbf{L} = \rho \cdot l/A$, where ρ is the density, l the tube length and A the wetted tube area. Then the pressure drop ΔP over distance l is equal to $\Delta P = \mathbf{L} \cdot \dot{V}$ with $\dot{V}$ the volume flow (L/s).
Furthermore, flow in the larger arteries is, in general, not Poiseuille flow due to the pulsatile character, especially in the aorta. In the ascending aorta of large mammals viscous effects of the entrance region are confined to a thin-walled boundary layer. The core is characterized as *largely inviscid*, caused by the heavy pulsatile character. These factors, together with specific density and viscosity are comprised in the Womersley number, which can be considered as the pulsatile version of the Reynolds number. With high numbers inertia dominates, yielding a rather well flat flow front. With low numbers viscosity dominates, yielding parabolic-like flows, however skewed towards the outer wall. An example for this is the flow in the left common coronary artery. In other coronary arteries the Reynolds numbers are much lower, the viscous effects are more dominant and flow is laminar. The velocity profile in many regions will be more like a parabolic Poiseuille flow, except that there will be skewing of this profile due to vessel curvature and branching. Also, significant entrance effects may result in the blunting of the velocity profiles.

Although there have been numerous fluid-dynamic studies of secondary flow phenomena, instrumentation limitations have prevented in vivo observations.
An additional complication introduced by the geometry of the arterial system is flow separation from and reattachment to the wall, causing recirculation zones. This phenomena in pulsatile flows is an extremely complex.

Literature

Author N. Westerhof, Noble MIM and Stergiopulos N. Snapshots of hemodynamics: an aid for clinical research and graduate education, 2004, Springer Verlag.
http://mss02.isunet.edu/Students/Balint/bloodflow.html.

Blood pressure

See for a description, the methods of measurement, models and special issues of blood pressure the chapters:

Blood pressure: (Central) venous
Blood pressure: description and measurement
Blood pressure: models
Blood pressure: pulse pressure
Blood pressure: Windkessel model

Blood pressure: models

Principle

Blood pressure (BP) is the pressure exerted by the blood at right angles to the walls of the blood vessels P_i minus the environmental or ambient pressure P_a, so:

$$BP = P_i - P_a.$$

Unless indicated otherwise, BP refers to systemic arterial BP, i.e. the pressure in the large arteries delivering blood to body parts other than the lungs, such as the brachial artery (in the arm). The pressure of the blood in other vessels is lower than the arterial pressure. BP values are generally stated in mmHg, but can be converted to an SI-unit, i.e. in Pascals. The conversion is:

$P = \rho_{Hg}gh$, (Pa)

where ρ_{Hg} is the specific density of Hg (kg/m^3), g the gravitational acceleration (m/s^2) and h the value of BP, but now expressed in mHg (not mm).

Mean arterial pressure (MAP)

The mean arterial pressure (MAP) is defined as the average arterial pressure during a single cardiac cycle. It is a result of the heart pumping blood from the veins back into the arteries. The up and down fluctuation of the arterial BP results from the pulsatile nature of the cardiac output (see Blood pressure: pulse pressure). The pulse pressure is determined by the interaction of the stroke volume versus the volume and elasticity of the major arteries.

MAP can be calculated by:

MAP = (stroke volume x systemic resistance) + CVP,

where CVP is central venous pressure (see Blood pressure: (Central) venous). Mostly, CVP can be neglected. The first term at the right is the heamodynamic analog of Ohm's law for an electric cirquit (V=iR). Cardiac output represents the efficiency with which the heart circulates blood throughout the body. Since MAP stroke volume and systemic resistance are not easy to easure, a simple equation of approximation has been developed:

$$MAP = P_{dias} + (P_{sys} - P_{dias})/3.$$

It shows that MAP is nearer the level of diastolic than systolic pressure.

Factors influencing BP

The physics of the circulatory system, as of any fluid system, are very complex. Many physical and physiological factors influence BP. Some physical factors are:

Heart rate The higher the heart rate, the higher BP (assuming no change in stroke volume).

Blood volume The higher the blood volume, the higher the cardiac output.

Resistance The higher the resistance, the higher the BP. Resistance is related to size (the larger the blood vessel, the lower the resistance), as well as the smoothness of the blood vessel walls. Smoothness is reduced by the buildup of fatty deposits on the arterial walls. Deposits can affect the laminar character of the flow (see Poiseuille's law and Reynolds number). Various substances (vasoconstrictors and vasodilators) change vessel diameter, thereby changing BP.

Viscosity or thickness of the fluid Increase of viscosity results in increase of resistance and so of BP. Anemia reduces viscosity, whereas hyperemia increases viscosity. Viscosity also increases with blood sugar concentration.

More Info

Usually, the systolic pressure P_{sys} amounts to 120 mmHg, or about 16 kPa. At this P_{sys} and an air pressure P_a of 1 atm, about 100 kPa, the total, absolute pressure in the blood vessel P_i is:

$$P_i = P_{sys} + P_a = 116 \text{ kPa}.$$

16 kPa is equivalent to a the pressure of a column of water with a height of 1.63 m.

The flow speed of blood in the body is 10 up to 100 cm/s (respectively during diastole and systole) in the aorta, approximately 10 cm/s in the arteries and approximately 0.1 cm/s in capillares. According to the law of Bernoulli (see Bernoulli's and Pascal's Law) it holds that:

$$P_i + \rho gh + \tfrac{1}{2}\rho v^2 = c,$$

where c is a constant. This means that: $P_{sys} + \rho gh + \frac{1}{2}\rho v^2 = c - P_a$. Since the ambient pressure is the same in all tissues, it holds that:

$$P_{sys} + \rho gh + \frac{1}{2}\rho v^2 = \text{constant}.$$

At a flow speed v = 50 cm/s (the largest flow speeds in the blood vessels are of this order of size) it holds that $\frac{1}{2}\rho v^2 = 0.13$ kPa. Since P_{sys} is 16 kPa, the term $\frac{1}{2}\rho v^2$ in the law of Bernoulli, which we may call the flow-pressure, is negligible with respect to the blood pressure P_{sys}. However, the term ρgh is not negligible as is explained in Blood pressure: body posture.

See also:
Blood pressure: (Central) venous
Blood pressure: description and measurement
Blood pressure: body posture
Blood pressure: pulse pressure
Blood pressure: Windkessel model

Blood pressure: (central) venous

Venous pressure
Venous pressure is the blood pressure (BP) in a vein or in the atria of the heart. It is much less than arterial BP, with common values of 5 mmHg in the right atrium and 8 mmHg in the left atrium. Measurement of pressures in the venous system and the pulmonary vessels plays an important role in intensive care medicine but requires invasive techniques.

Central venous pressure
Central venous pressure (CVP) describes the pressure of blood in the thoracic vena cava, near the right atrium. CVP reflects the amount of blood returning to the heart and the ability of the heart to pump the blood into the arterial system. It is a good approximation of right atrial pressure, which is a major determinant of right ventricular end diastolic volume (right ventricular preload).
CVP can be measured by connecting the patient's central venous catheter to a special infusion set which is connected to a small diameter water column. If the water column is calibrated properly the height of the column indicates the CVP.

Blood pressure: description and measurement

Principle

Blood pressure (BP) is the pressure exerted by the blood at right angles to the walls of the blood vessels P_i minus the environmental pressure P_e, so:

$$BP = P_i - P_e.$$

Unless indicated otherwise, BP refers to systemic arterial BP, i.e. the pressure in the large arteries delivering blood to body parts other than the lungs, such as the brachial artery (in the arm). The pressure of the blood in other vessels is lower than the arterial pressure. BP values are generally stated in mmHg, but can be converted to an SI-unit, i.e. in Pascals. The conversion is:

$$P = \rho_{Hg}gh, \text{ (Pa)}$$

where ρ_{Hg} is the specific density of Hg (kg/m^3), g the gravitational acceleration (m/s^2) and h the height of the Hg column (m).
Systolic and diastolic adult BP of the brachial artery are typically 120 (16 kPa) and 80 mmHg (10 kPa) respectively.
The *mean arterial pressure* (MAP), see **More Info**, and *pulse pressure* (see Blood pressure: pulse pressure) are other important quantities.
BP varies from one heartbeat to another and throughout the day (in a circadian rhythm); they also change in response to stress (exercise etc.), nutritional factors, drugs, or disease.

Measurement
Invasive measurement
Arterial BP is most accurately measured *invasively* by placing a cannula into a blood vessel and connecting it to an electronic pressure transducer. This is done in human and veterinary intensive care medicine, anesthesiology, and for research purposes.
Non-invasive measurement
The non-invasive auscultatory and oscillometric measurements are simpler and quicker, have no complications, and are less unpleasant and painful, at the cost of somewhat lower accuracy and small systematic differences in numerical results. These methods actually measure the pressure of an inflated cuff at the points where it just occludes blood flow (systolic BP) or just permits unrestricted flow (diastolic BP).
The classical auscultatory method uses a stethoscope (for listening to the so-called Korotkoff sounds), a sphygmomanometer (upper arm cuff attached to a mercury or aneroid manometer).
Basic digital BP monitors are relatively inexpensive, making it easy for patients to monitor their own BP. Their accuracy can vary greatly; most have not been certified for accuracy by an approved authority. Upper arm, rather than wrist, monitors usually give readings closer to auscultatory. Some meters are automatic, with pumps to inflate the cuff without squeezing a bulb.

Fig. 1 Auscultatory method An aneroid sphygmomanometer with stethoscope

Oscillometric methods are used in the long-term measurement. The equipment is the same as for the auscultatory method, but with an electronic pressure sensor (transducer) fitted in the electronically operating cuff. The manometer is an electronic device with a numerical readout and checked periodically.
The cuff is inflated to a pressure initially in excess of the systolic BP ($BP_{systolic}$), and than reduced below $BP_{diastolic}$ over a period of about 30 s. When blood flow is nil (pressure > $BP_{systolic}$) exceeding systolic pressure) or unimpeded (cuff pressure <$PB_{diastolic}$), cuff pressure will be essentially constant. When blood flow is present, but restricted, the cuff pressure will vary periodically in synchrony with the cyclic

expansion and contraction of the brachial artery, i.e., it will oscillate. The values of $PB_{systolic}$ and $PB_{diastolic}$ are computed from the raw measurements and displayed.
Oscillometric monitors do not give entirely meaningful readings in certain “special conditions” such as arterial sclerosis, arrhythmia, preeclampsia, pulsus alternans, and pulsus paradoxus.
In practice the different methods do not give identical results; an algorithm and experimentally obtained coefficients are used to adjust the oscillometric results to give readings which match the auscultatory as well as possible. Some equipment uses computer-aided analysis of the instantaneous BP waveform to determine the systolic, mean, and diastolic points.
The term NIBP, for Non-Invasive BP, is often used to describe oscillometric monitoring equipment.

More Info

Regulation of BP
The endogenous regulation comprises the baroreceptor reflex, the renin-angiotensin system (RAS) and aldosterone release. This steroid hormone stimulates Na-retention and K-excretion by the kidneys. Since Na^+ is the main ion that determines the amount of fluid in the blood vessels by osmosis, aldosterone will increase fluid retention, and indirectly, BP.

The physics of the circulatory system, as of any fluid system, are very complex (see e.g. Elasticity of the aorta, Navier-Stokes equations, Windkessel model and Blood pressure: models).
Many physical and physiological factors influence BP. Cardiac output is heart rate times stroke volume. It represents the efficiency with which the heart circulates blood throughout the body.

For *factors influencing BP* see Blood pressure: models.

For *Mean arterial pressure (MAP)* see Blood pressure: models.

For *Orthostatic hypotension s*ee Blood pressure: influence of posture.

Blood pressure: pulse pressure

Principle

Formally it is systolic minus diastolic blood pressure. It can be calculated by:
Pulse pressure = stroke volume/compliance (Pa or mmHg).
Compliance is 1/elastance for a hollow organ, see Compliance (hollow organs).
Usually, the resting pulse pressure in healthy adults, sitting position, is about 40 mmHg. The pulse pressure increases with exercise due to increased stroke volume and reduced total peripheral resistance, up to pulse pressures of about 100 mmHg while diastolic pressure remains about the same or even drops (very aerobically athletic individuals). The latter effect further increases stroke volume and cardiac output at a lower mean arterial pressure. The diastolic drop reflects a much greater fall in total peripheral resistance of the muscle arterioles in response to the exercise (a greater proportion of red versus white muscle tissue).

Application

Low values
If resting pulse pressure < 40 mmHg, the most common reason is an error of measurement. If it is genuinely low, e.g. 25 mmHg or less, the cause may be low stroke volume, as in congestive heart failure and/or shock. This interpretation is reinforced if the resting heart rate is relatively rapid, e.g. 100-120 (in normal sinus rhythm), reflecting increased sympathetic nervous system activity.

High values
If the usual resting pulse pressure is consistently greater than 40 mmHg, e.g. 60 or 80 mmHg, the most likely basis is stiffness of the major arteries, aortic regurgitation (a leak in the aortic valve), an extra path for the blood to travel from the arteries to the veins, hyperthyroidism or some combination. (A chronically increased stroke volume is also a technical possibility, but very rare in practice.) Some drugs for hypertension have the side effect of increasing resting pulse pressure irreversibly. A high resting pulse pressure is harmful and tends to accelerate the normal ageing of body organs, particularly the heart, the brain and kidneys.
A high pulse pressure is an important risk factor (20% increase) for heart disease. A 10 mm Hg increase in pulse pressure increases the risk of major cardiovascular complications and mortality.

Blood Pressure: Windkessel model

Principle

The Windkessel model consists of four anatomical components: left ventricle, aortic valve, arterial vascular compartment, and peripheral flow pathway (see Fig. 1).

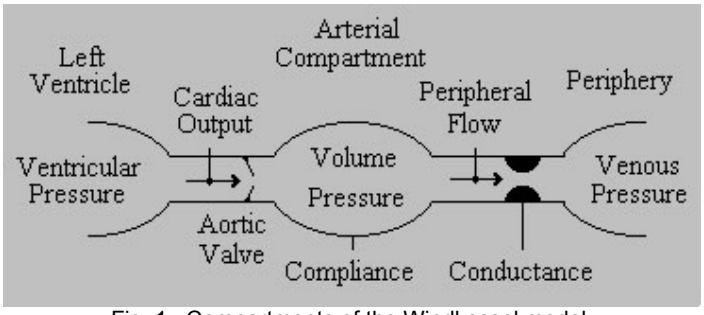


Fig. 1 Compartments of the Windkessel model.

The basic model is a closed hydraulic circuit comprised of a water pump connected to a chamber. The circuit is filled with water except for a pocket of air in the chamber. (*Windkessel* is the German word for air-chamber.) As water is pumped into the chamber, the water both compresses the air in the pocket and pushes water out of the chamber, back to the pump. The compressibility of the air in the pocket simulates the elasticity and extensibility of the major artery, as blood is pumped into it by the heart ventricle. This effect is commonly referred to as *arterial compliance*. The resistance, which the water encounters while leaving the Windkessel and flowing back to the pump, simulates the resistance to flow encountered by the blood as it flows through the arterial tree from the major arteries, to minor arteries, to arterioles, and to capillaries, due to decreasing vessel diameter. This resistance to flow is commonly referred to as *peripheral resistance*.
In terms of system analysis, the variable $I(t)$ ($mL \cdot s^{-1}$) is the flow of blood from the heart to the aorta (or pulmonary artery). It is assumed that the stroke volume of the heart is independent of the resistance and so $I(t)$ is the independent variable. The dependent variable P(t) (mmHg) is the blood pressure in the aorta (or pulmonary artery). Further, the system parameter C ($mL \cdot mmHg^{-1}$) is the compliance, this is the constant ratio of air volume to air pressure (see Compliance (hollow organs))
R ($mmHg \cdot min \cdot mL^{-1}$) is the peripheral resistance or flow-pressure proportionality constant (I/P=R as in Ohms law) in the systemic (or pulmonary) arterial system.

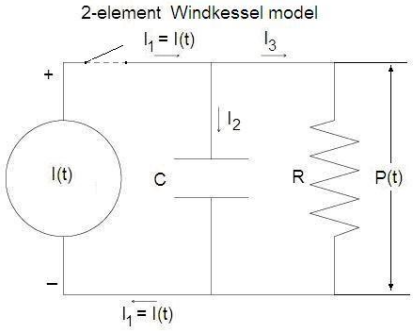


Fig. 2 Windkessel model with 2 elements.

Now, $I(t)$ can be expressed as a function of P(t) with R and C as parameters. Assuming that C is constant and flow I through the pipes connecting the air chamber to the pump follows Poiseuille's law and is proportional to the fluid pressure, the system behaves as a first order low pass system (or filter, see Linear systems: first order). It can be described with a simple differential equation. This equation describes the relationship between the imposed current, $I(t)$, and the resulting time-varying electrical potential, P(t), in the electrical analog depicted in Fig. 2. At the start of the diastole, at t = 0, C is loaded and has the starting pressure P_0. (In the aorta this is the systolic pressure.) During the diastole there is no blood flow from the heart, so $I(t) = 0$. Therefore, the source of the pressure, the heart, is disconnected (closed aorta valves), which action is presented by opening the switch in the upper left branch. The equation can be solved exactly for P(t):

$P(t) = P_0 e^{(t-tdias)/(RC)}$, (1)
where t_{dias} is the time at the start of diastole (for simplicity, one can take (t_{dias} = 0), and P_0 is the blood pressure in the aorta (or pulmonary artery) at the start of diastole. RC is the time constant (see Half-time and time constant), often denoted by τ (tau), which characterizes the speed of decay of P(t). After the time of 1τ, P has decreases to a fraction 1/e=0.368 of its original value (here P_0).
Equation (1) holds for a single beat. After some 6τ P(t)•0. Since the interval between te beats is much smaller than 6τ, some pressure at the end of cardiac cycles remains. This is the diastolic pressure.

A more advanced model of is the 3-element Windkessel model, which adds a resistive element between the pump and the air-chamber to simulate resistance to blood flow due to the aortic or pulmonary valve. It has been used to model blood pressure and flow in mammalian and avian animal models. This resistance depends of the instant in the cardiac cycle. It is very large (infinite) during the diastole and has a varying magnitude during the systole. It can be modeled by a fixed resistor combined with a switch.
Here is a schematic of the electrical circuit corresponding to the 3-element Windkessel model:

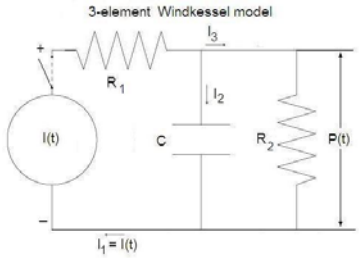


Fig. 3 Windkessel model with 3 elements.

Using the same circuit analysis technique as for the 2-element Windkessel model circuit, the differential equation for the 3-element Windkessel model becomes a little more complex, but it still comprises only a first order time derivative, belonging to a first order differential equation. However, when we measure the aortic pressure, that is behind R_1, then we obtain the same solution is in the 2-element model, with the switch opened and R substituted by R_1.
Until now, the hydrodynamic inertia of the blood flow has been ignored. Its electric analog is an inductance L, represented by a coil. The drop in electrical potential across an inductor with inductance, L, is L(d I(t)/dtL.

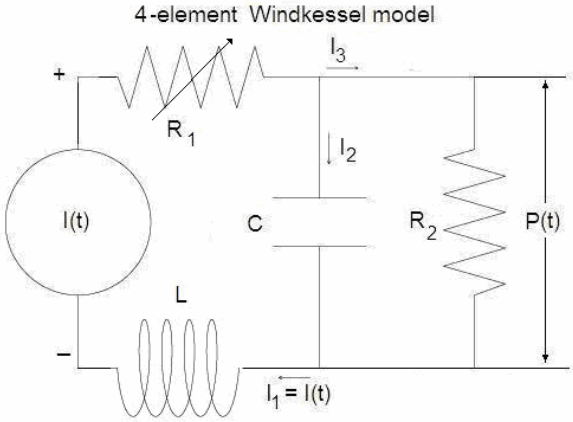


Fig. 4 Windkessel model with 4 elements.

Following the theory of circuit analysis, one finds a 2nd order differential equation (since the 2nd derivative also occurs) for this 4-element Windkessel model. R_1 is presented as a variable resistance (indicated by the arrow) to model the whole cardiac cycle.

L has units of mass per length to the fourth power. Note that for L = 0, this 4-element Windkessel equation reduces to the 3-element Windkessel equation. Also, during diastole when I(t) and its derivatives vanish, we can solve P(t) and get the same exponentially decreasing pressure function with decay time constant R_2C as in the 3-element Windkessel model.

Until know, the period of the systole was ignored. When we introduce the blood ejection during the systole, L also plays some role. For the systolic ejection one should choice a realistic time function as done in the upper panel of Fig. 5. After numerical parameterization Fig. 5 gives a result of a simulation.

Application

Windkessel models are frequently used to describe the load faced by the heart in pumping blood through the pulmonary or systemic arterial system, and the relation between blood pressure and blood flow in the aorta or the pulmonary artery. Characterizing the pulmonary or systemic arterial load on the heart in terms of the parameters that arise in Windkessel models, such as arterial compliance and peripheral resistance, is important, for example, in quantifying the effects of vasodilator or vasoconstrictor drugs. Also, a mathematical model of the relationship between blood pressure and blood flow in the aorta and pulmonary artery is useful, for example, in the development and operation of mechanical heart and heart-lung machines. If blood is not supplied by these devices within acceptable ranges of pressure and flow, the patient will not survive. The model is also applied in research with isolated hearts.

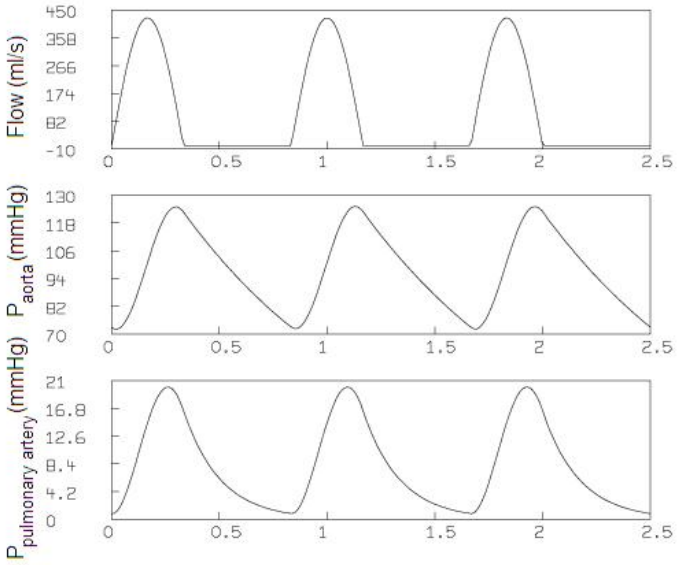


Fig. 5. Results of a simulation with the 4-element Windkessel model.

More Info

Often more complex modifications of the Windkessel models have been applied in experimental research. For instance the resistance R_1 is often replaced by an impedance (so a compliance is added), which is time dependent.

Reference

N. Westerhof, Noble M.I.M and Stergiopulos N. Snapshots of hemodynamics: an aid for clinical research and graduate education, 2004, Springer Verlag.

Body heat conduction and Newton's Law of cooling

Principles

Heat conduction is the transmission of heat across matter. Heat transfer is always directed from a higher to a lower temperature. The donor is refrigerating and the acceptor is warming. Denser substances are usually better conductors; metals are excellent conductors.

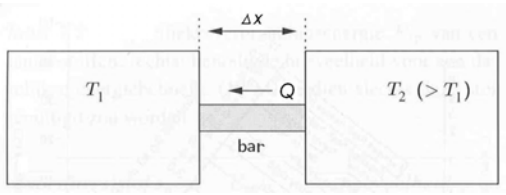


Fig. 1 Heat transfer in a bar, clamped between two solid objects with a homogeneous, constant temperature.

The law of heat conduction, also know as *Fourier's law*, states that the time rate of heat flow Q through a slab of a perfectly insulated bar, as shown in Fig.1, is proportional to the gradient of temperature difference:

$Q = k \cdot A \cdot \Delta T / \Delta x$, or more formally: (1a)
$dQ/dt = k \cdot A \cdot dT/dx$ (1b)

A is the transversal surface area, Δx is the distance of the body of matter through which the heat is passing, *k* is a conductivity constant (W/(K.m)) and dependent on the nature of the material and its temperature, and Δ*T* is the temperature difference through which the heat is being transferred. See for a simple example of a calculation **More Info**. This law forms the basis for the derivation of the heat equation. The R-value is the unit for heat resistance, the reciprocal of the conductance. Ohm's law (voltage = current·resistance) is the electrical analogue of Fourier's law.
Conductance of the object per unit area U (=1/R) is:

$U = k/\Delta x$ (2)

and so Fourier's law can also be stated as:

$Q = U \cdot A \cdot \Delta T$.

Heat resistances, like electrical resistance, are additive when several conducting layers lie between the hot and cool regions, because heat flow *Q* is the same for all layers, supposing that A also remains the same. Consequently, in a multilayer partition, the total resistance is:
$1/U = 1/U_1 + 1/U_2 + 1/U_2 + \ldots..(3)$
So, when dealing with a multilayer partition, the following formula is usually used:

$$Q = \frac{A \cdot \Delta T}{\Delta_1 x / K_1 + \Delta_2 x / K_2 + \Delta_3 x / K_3 + \ldots} \quad (4)$$

where *K* should be read as *k* and $\Delta_1 + \Delta_2 + \ldots = 1$.
When heat is being conducted from one fluid to another through a barrier, it is sometimes important to consider the conductance of the thin film of fluid, which remains stationary next to the barrier. This thin film of fluid is difficult to quantify, its characteristics depending upon complex conditions of turbulence and viscosity, but when dealing with thin high-conductance barriers it can sometimes be quite significant.

Application

In problems of heat conduction, often with heat generated by electromagnetic radiation and Ultrasound. Generation by radiation can be performed for instance by lasers (e.g. dermatological and eye chirurgy), by microwave radiation, IR and UV (the latter two also cosmetic). Specific applications are thermographic imaging (see Thermography), low level laser therapy (LLLT, Photobiomodulation),

thermo radiotherapy and thermo chemotherapy of cancers. Than, laws of radiation also play a role (see Wien's law. Other fields of application are space and environmental medicine (insulating clothing), cardiochirurgy with a heart-lung machine.
Heat transfer calculations for the human body have several components: internal heat transfer by conduction and by perfusion (blood circulation) and external by conduction and convection. The transport by perfusion complicates calculations substantially.

More info

Newton's law of cooling
Newton's law of cooling states that the rate of heat loss of a body is proportional to the difference in temperatures between the body and its surroundings. This form of heat loss principle, however, is not very precise; a more accurate formulation requires an analysis of heat flow based on the heat equation in an inhomogeneous medium. Nevertheless, it is easy to derive from this principle the exponential decay (or increase) of temperature of a body. If *T*(t) is the temperature of the body as a function of time, and T_{env} the temperature of the environment then its derivative is:

$$dT(t)/dt = -\tau\,(T - T_{env}) \qquad (5)$$

where $1/\tau$ is some positive constant. Solving (5) gives:

$$T(t) = T_{env} + (T_{t=0} - T_{env})e^{-t/\tau}. \qquad (6)$$

The constant τ appears to be the time constant and the solution of (6) is the same as (de)charging a condenser in a resistance-capacitance circuit from its initial voltage $V_{t=0}$ to the final obligatory voltage V_{obl}.

In, fluid mechanics (liquids and gases) the Rayleigh number (see Rayleigh, Grashof and Prandtl Number) for a fluid is a dimensionless number associated with the heat transfer within the fluid. When the Rayleigh number is below the critical value for that fluid, heat transfer is primary in the form of conduction; when it exceeds the critical value, heat transfer is primarily in the form of convection (see Body heat dissipation and related water loss).

An example of heat loss when submerged
The problem is: how many energy passes the skin of a human jumping into a swimming pool, supposing that:

- heat transfer in the body is solely determined by conduction;
- the heat gradient in the body is time invariant;
- the water temperature surrounding the body remains at the initial temperature;
- the mean heat gradient of the body is $\Delta T/\Delta x = 1.5\ ^{o}C/cm = 150\ ^{o}C/m$
- body surface 2.0 m^2 (an athletic body of 75 kg and 188 cm)
- $T_{skin} = 25\ ^{o}C$;
- $T_{water} = 25\ ^{o}C$;
- $k_{water} = 0.6$ W/m.s

Since dQ/dt andT/dx are time and distance invariant, (1a) can directly applied:
$Q = -0.6 \cdot 2.0 \cdot 150 = 180$ W.
For a subject of 25 year basal metabolism is 43.7 W/m^2, implying that Q is about twice basal metabolism. Even this very basic approach illustrates the very high energy expenditure of for instance a swimmer, even in warm water. With $T_{water} = 13\ ^{o}C$, the skin and deeper tissues will cool down fast and soon the temperature gradient in the body is doubled, resulting in a 360 W loss. This makes clear that with normal sized people in such cool water, exhaustion, next hypothermia and finally consciousness and drowning is a process of about an hour.
Several physical multi-compartment models have been developed to calculate heat transfer in the human body (see **References**).

References
ASHRAE, Fundamental Handbook, Ch. 8 Physiological Principles, Comfort and health. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, 1989.
http://www.ibpsa.org/%5Cproceedings%5CBS1999%5CBS99_C-11.pdf

Body heat dissipation and related water loss

Principles

The type of energy, e.g. heat, motion and radiation, produced by the body (nearly) all depends on chemical processes. With a constant body mass (no grow or weight loss) and in rest nearly all energy is transformed to heat, also the small amount of mechanical energy of heart and ventilation. All heat is transferred to the environment and there are various components of this heat transfer.
Calculation of the components can be performed by applying (semi-)empirical equations (curve fitting of experimental data) or by applying physical laws, as will be done here.
Table 1 summarises all components of heat release with their numerical values of a normal shaped man in rest and when performing heavy endurance sport (running). Some of the components are bidirectional, in other words heat can also be collected. This holds for radiation (sun bathing) and convection (as extreme example the hair dryer). The five components are dependent on biometric factors and to calculate numerically their contribution, this factors should by defined. This is done in **More Info**.

Table 1 Components of heat release

Heat release of male, 40 year, 75 kg, 175 height, body area 1.9 m^2		
Components of release (v_{air} < 0.15 m/s)	seated in rest	running 15 km/h
C_{res}, expiration of gas	1.2	41.8
E_{res}, evaporated water in expired air	9.3	332
R_{skin}, radiation	32.8	116.3
C_{skin}, convection along the skin	14.6	50.3#
E_{skin}, perspiration and evaporation via skin	21.1*	685**
Total heat release	79	1226

lower limit (see **More Info**), *sweating 16 mL/h, ** sweating 2 L/h.

Two very small components, heat conduction and sound production (mainly heart) and the small heat conduction (via the buttocks or feet) are neglected.
The dissipated heat of the runner is 15 times that of the seated person in rest. The runner has to dissipate this heat otherwise he will suffer from hyperthermia. A rough calculation yields a body temperature increase of 0.22 K/min. In rest, E_{res} and E_{skin} (for wind speed v < 0.15 m/s) together is 23.4 W, about ⅓ of the total, but during running ca. ¾. R_{skin} is 45 W, about ⅔ of the total loss of 79 Watt in rest. During running loss by radiation and convection become less important. Total water losses by evaporation and perspiration is 46 mL/h for the seated person in rest and 1342 mL/h for the runner (see More Info), 29 times more. The calculations show the enormous difference between both conditions.

Total loss in rest should be nearly the same as the basal metabolism or basal energy expenditure (BEE), which is equal to:

male: $BEE = 66.67 + 13.75W + 5H - 6.76A$ (kCal/day) (1a),

female: $BEE = 665.1 + 9.56W + 1.85H - 4.68A$, (1b)

where H is height (cm), W is weight (kg), A is age (year). With $BEE_{Watt} = 0.0484\ BEE_{kCal/h}$, our model has an BEE of 82.4 Watt. The energy expenditure in rest and reclining is about 2% lower, but seated it is some 8% higher. All together, the calculations of heat losses and heat production are well in accordance.

The components can be calculated as follows. **More Info** gives details and the calculations of the values of Table 1.

Heat release by expiration
$C_{res} = m \cdot c_p \cdot \Delta T$, (2)
with m the mass of gas expired per second, ΔT the temperature difference between inspired and expired gas and c_p the specific heat coefficient (the amount of energy to increase the temperature of one unit of mass of gas with 1^{o}C under constant pressure) of the expired gas.

Heat release by expired water vapor

$E_{res} = m_{H2O} \cdot \Delta H_{H2O}$, (3)

with m_{H2O} the mass of water vapor and ΔH_{H2O} the specific evaporation heat of water:

Heat release by radiation
The peak wavelength of the infrared radiated light by the human body is about 10 µm (calculated from the skin temperature with Wien's law). The law of Stefan-Boltzmann says that a black emitting body radiating in an infinite space with 0 K as background temperature emits $\sigma \cdot A_{body} T^4_{body}$ Watt. The constant σ is the constant of Stefan-Boltzmann, being 5.7 x 10^{-8} W/(K^4.m^2). Extending the law for grey bodies (law of Kirchhoff) and a temperature > 0 K the equation becomes:

$$R_{body} = \varepsilon_{body} \cdot \sigma \cdot A_{body} (T^4_{body} - T^4_{background}) \approx 0.5 \cdot \varepsilon_{body} \cdot \sigma \cdot A_{body} \, \Delta T (T + 2T_{background})^3, \quad (4)$$

where $\Delta T = T_{body} - T_{wall}$ and ε_{body} the emittance coefficient ($0.8 < \varepsilon_{body} < 1.0$, and thin-clothed 0.95).

Convection along the skin
Under many conditions, the human body releases heat to the surrounding air. (But the process can reverse, for instance by entering a warm room).The underlying mechanism is that the air particles colliding with the skin (the "wall") obtain a larger momentum (mass times velocity) at the cost of the velocity of the Brownian motion of the skin particles. And so, the air particles increase their velocity, and consequently the boundary layer of air covering the skin obtains a higher temperature. This heated layer has a lower specific density than the cooler air at a larger distance. In rest, this difference causes a laminar ascend of the boundary layer. This is the process of heat release by convection. Heat release by convection is hard to calculate and various approaches can be found in literature (see **Literature**). With laminar convection the problem is easier than with turbulence, although still complicated.
A laminar gas flow has a Rayleigh number (Ra) between 10^4 and 10^8. (see Rayleigh, Grashof and Prandtl Number). Convection currents with a velocity $v_{air} < 0.15$ m/s (generally indoor) appear to be laminar since Ra is about 1.0×10^6. Under the above conditions the refrigeration law of Newton (see Body heat conduction and Newton's Law of cooling) applies:

$$C_{skin} = \alpha \cdot A \cdot \Delta T, \quad (5)$$

where α the heat convection coefficient, A the area of the body and ΔT the difference in temperature between skin and ambient air. The parameter α is $1.35 \cdot (\Delta T/H)^{1/4}$ at 1 bar and 20 °C, with H the effective height of the subject.

Perspiration and evaporation via the skin
Perspiration is the process of water evaporating though the skin driven by the vapor pressure difference in the outer skin and the lower vapor pressure of the surrounding air. Evaporation is the process of sweat evaporation from the skin surface. The release is calculated by:

$$E_{skin} = m \cdot \Delta H_{skin\ water}, \quad (6)$$

where m is the loss of mass of liquid and ΔH the specific evaporation heat. In rest with a low skin temperature there is no sweat production, but with a high skin temperature there is some sweat production.

Application

These are in aerospace, altitude, and sports medicine, in occupational medicine for heavy exertion and extreme environmental conditions. Further air-conditioning (general, hospitals and commercial aviation) and clothing industry.
A typical application Dehydration during long flights is not caused by the low humidity in the cabin. The extra loss of water due to the humidity is only about 350 mL/day (see **More Info**). The actual reasons are a too low liquid intake by food, at all drinking to few, and acute altitude diuresis.

More info

To clarify the equations the components of heat loss will be calculated in examples. First a human heat-model (standard subject) should be specified:
- 40 years, male.
- Body weight W = 75 kg, length H = 175 cm, body area $A = 0.007184 \cdot W^{0.425} \cdot H^{0.725} = 1.90\ m^2$.
- In rest, seated (on a poorly heat-conduction seat, indoor) and running during heavy endurance sport (running, 15 km/hour).

- Sweat production 0.26 in rest and 33 mL/min when running. Thin clothing.
- RMV (minute respiratory volume of inspiration) is 5.6 L/min in rest, in sitting position. During running 200 L/min.
- FI_{N2}/FE_{N2} = 1.06 (N_2 fraction in inspired air/N_2 fraction in expired gas. From this ratio $RMV_{expiration}$ (= $RMV \cdot FI_{N2}/FE_{N2}$) is calculated.
- Temperature of expired air is 310 K, independent of the ambient air temperature (actually there is a small dependency).
- Temperature of the skin T_{skin} is 303 K (30 ^oC) at rest and 310 K when running.
- Φ_m = 88.2 W (seated some 10% more than lying). Φ_m is total metabolic power. Φ_m is age and sex dependent (see above).
- Air velocity: indoor 0.15 m/s: outdoor, produced by the runner, 4.17 m/s.

C_{res}, *expiration of gas*

$$C_{res} = \{RMV_{exp} \cdot \rho_0 \cdot (273/T) \cdot p/60\} \cdot c_{p,air} \cdot \Delta T, \text{ with:} \quad (2a)$$

$RMV_{exp} = (FI_{N2}/FE_{N2})RMV = 1.06 \cdot RVM$ L/min. RVM is 5.6 and 200 L/min;
- ρ_0 = 1.29 kg/m^3, the specific density of air at 273.15 K and 1 bar;
- T = 310 K, the temperature T of the expired gas is;
- p = 1 bar, the ambient pressure;
- 60, the conversion factor from minute to second;
- $c_{p,air}$ = 1.00 $kJ \cdot kg^{-1}K^{-1}$, the specific heat capacity (at 0 ^oC);
- ΔT = 12 K (ambient temperature is 298 K).

After completing all values, 1.17 and 66.2 W is found for *rest* and *running* respectively.

A completely different, empirical approach (ref. 1) is :

$$C_{res} = 0.0014\ \Phi_m\ (307 - T_{ambient}), \quad (2b)$$

where Φ_m the total metabolic power.
In a commercial aircraft, C_{res} is about 0.8^2 smaller since both ρ_0 and $c_{p,air}$ are reduced with 20%.

E_{res}, *evaporated water in expired air*
Starting from $RMV_{exp} = RMV \cdot FI_{N2}/FE_{N2}$, considering the fraction of evaporated water vapor, correcting for temperature, considering seconds, the volume per second (m^3/s) is found. Via the molecular volume of 22.4 m^3 and the molecular weight of water (m_{H2O}), and ΔH_{H2O} (2260 kJ/kg), the evaporation in kg/s is found.

$$E_{res} = \{[RMV \cdot (FI_{N2}/FE_{N2}) \cdot (FE_{H2O} - FI_{H2O}) \cdot (273/310)/60]/22.4\} m_{H2O} \cdot \Delta H_{H2O} \quad (3a)$$

Between brackets the volume of water vapor in m^3/min at 273 K is found (FE_{H2O} = 0.0618 and FI_{H2O} = 0.003, only about 9% humidity, 310 K is body temperature). For the standard subject in rest E_{res} becomes 9.3 W and 332 W for the runner.
The water loss is 14.8 and 529 mL/h respectively.
There is no pressure dependency (mountaineering, diving) since the alveolar p_{H2O} is always 6.3 kPa.
Consequently, in a commercial aircraft, E_{res} is about the same supposing FI_{H2O} = 0.003, which means very dry air. With a normal humidity (60%), water loss is about 30% less. The example shows that water loss by expiration can be neglected in rest.

An empirical approach, modified after ref. 1, is:

$$E_{res} = 0.0173\ \Phi_m (5.87 - p_{H2O,ambient}/1000). \quad (3b)$$

R_{body}, *Radiation*
Subject in rest within a room Often, the body is surrounded at a finite distance by a wall. Whereas the body radiates in all directions to the walls, each small part of wall basically radiates to the rest of the wall and to the subject. A new parameter C is introduced to control these effects. It comprises the surface of the body and the wall, and the emittance factor of the body and the wall. Using the approximation of (3) the result is:

$$R_{body} \approx C \cdot \sigma \cdot A_{body} \, T(\,) T + 2T_{background})^3/2, \quad (4a)$$

The parameter C is defined as $1/C = 1/\varepsilon_{body} + (A_{body}/A_{wall})(1/T_{wall} - 1)$. The effective area of the sitting body $A_{s\text{-}skin}$ is 1.33 m^2, (70% of A_{skin},) and , $_{wall}$ = 0.9. With a room of 70 m^3, C appears to be 5.4 x 10^{-8} $Wm^{-2}K^{-1}$. With T_{skin} = 303 K and T_{wall} = 298 K the radiant power is 29.4 W/ m^2. In a similar way the radiant power of the chamber wall in the direction of the subject can be calculated. It amounts to 10.1 W/m^2. Since the absorption coefficient of the subject is about 0.65, the net dissipation is 29.4 – 4.8 =

24.6 W/m^2. For the sitting standard subject this finally yields R_{skin} = 32.8 W. When A_{wall}>>A_{skin}, in the definition of C, the second term at the right can be ignored.
Subject running outdoor Equation (4) with T_{skin} = 310 K yields 116.3 W.

A simple approach for most typical indoor conditions (ref. 1) is:

$$R'_{skin} = 4.7 \cdot A \cdot \Delta T \text{ (W)} \qquad \text{(4b).}$$

This results in 31.3 W (sitting subject). For a large ΔT with a high body temperature this linear approximation is less accurate.

C_{skin} Convection along the skin
Subject in rest within a room The dependency of the heat convection coefficient α on the effective height H_e implies a dependency on body posture. For lying, sitting and standing the effective surface is some 65%, 70% ad 73% of total body area respectively, and H_e is some 17, 80 and 91% of actual height. This yields values for α of 2.7, 2.2 and 1.8. With $\Delta T = T_{skin} - T_{air} = 303 - 298 = 5$ K and completing the equation $C_{skin} = \alpha \cdot A \cdot \Delta T$ for reclining, seated and standing posture C_{skin} is 16.8, 14.6 and 12.5 W respectively, values closely together.
Subject running outdoor The runner has a ΔT of 12 K. His effective height is supposed to be 85% of the actual height and his effective surface 100%. This three changed values give an increase of a factor of 4.02 compared to standing in rest. Supposing that the factor α still applies for an air velocity of 15 km/h, than the convection loss of the runner is 50.3 W.
Air velocities v > 1 m/s give a substantial increase in C_{skin}: the factor of proportionality is $(v \cdot p)^{0.6}$, where p is the ambient pressure. From this factor the wind-chill temperature factor can be calculated. The ratio of the air velocities yields a chill ratio of 7.35. This yields 92 W. The original value of 50.3 W is too low and the latter an upper limit. The factor α needs correction, since the thermal diffusivity (see Rayleigh, Grashof and Prandtl Number) is much higher in turbulent convection (the runner) than with laminar convection. The deviation from the upper limit depends on the clothing of the runner (extend of turbulence).
Foggy air augments C_{skin}.
Heat release by laminar convection is pressure dependent: $C_{p\ bar} = p^{¼} C_{1\ bar}$. Consequently in aircrafts and at altitude it is less and in hyperbaric chambers (for hyperbaric oxygen treatment) it increases. The dependency on pressure can be clarified conceptually and only qualitatively as follows. With a higher pressure, there is a higher density and so more collisions with the wall. This effect augments the heat release. But a higher density means that in the gas there are also more collisions, reducing the "diffusion" of the heat. Also the flow of the convection behaves different. This all results in the exponent of ¼.

E_{skin} *Perspiration and evaporation via the skin*

$$E_{skin} = m \cdot \Delta H_{skin\ water.}$$

The mass (m) of sweat and perspirated skin water is 8.7 x 10^{-6} kg/s (0.75 kg/day, supposing both contributions are the same at a skin temperature of 30 °C), and $\Delta H_{skin\ water}$ is 2428 kJ/kg (slightly higher than for pure water due to the dissolved minerals). The calculation yields E_{skin} = 21.1 W.
Supposing that 50% of the sweat production of the runner evaporates, perspiration and evaporation is 282 10^{-6} kg/s ml is consequently E_{skin} = 685 W. With a large sweat production the calculation is actually more complicated since part of the sweat evaporates and the remaining part is cooled and drips off.
E_{skin} reduces with pressure (altitude) since the water particles make less collisions with the air particles which hampers their diffusion in the surrounding air. In an aircraft cabin, the dry air increases the water loss due to perspirated by some 30%, as holds for the water loss of E_{res}. Evaporation also increases some 30%, due to a higher convection.

Literature

1. ASHRAE, Fundamental Handbook, Ch. 8 Physiological Principles, Comfort and health. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, 1989.
2. Bernards J.A. and Bouwman L.N. Fysiologie van de mens. 5ft edition. Bohn, Scheltema & Holkema, 1988.
3. Polytechnisch Zakboekje, 1993, PBNA, Arnhem

Cauterization

Principle

Cauterization by heat
This is burning of tissue with a hot cauter for removing or burning arteries to stop them from bleeding. Cautery can also mean the branding of e.g. a livestock. Formaly, cauterization was used to stop heavy bleeding (amputations) by a piece of heated metal placed onto the wound. Later special medical instruments called cauters were used to cauterize e.g. arteries. The present form is electrocautery. The basics of electrocautory equipment is similar as that of Electrosurgery.
Cauterization by refrigeration
Removal of tissue (warts) and stopping of bleedings can also be performed by a cold probe, cooled by e.g. liquid nitrogen. This technique has been evolved to cryosurgery.
Chemical cautery
Many chemical reactions can destroy tissue and some are used routinely in medicine, most commonly for the removal of small skin lesions (i.e. warts or necrotized tissue) or hemostasis.

Application

Electrocauterization is the process of destroying tissue with electricity and is a widely used technique in modern surgery, as for instance cutting through soft tissue i.e. abdominal fat in a laparotomy or breast tissue in a mastectomy, and especially small vessels bleedings (larger vessels being ligated). Applications of *cryosurgery* respect general surgery, gynecology, otorhinolaryngology and skin oncology.
Chemical cauterizing is performed by e.g. silver nitrate and cantharidin. The former is bounded in a small stick that is dipped into water and pressed onto the lesion to be cauterized for a few moments. Cantharidin, an extract of the blister beetle, causes epidermal necrosis and blistering (warts).

More Info

Nasal Cauterization
Recurrent nose bleeds are most likely caused by an exposed blood vessel. In a bleeding-free period, it can be cauterized. The different methods of cauterization to stop the nose bleeding include burning the affected area with acid, hot metal, Lasers, or silver nitrate. Sometimes liquid nitrogen is used as a less painful alternative, though less effective. Topically applied cocaine make this procedure less uncomfortable and cocaine is the only local anesthetic which also produces vasoconstriction.

Chromatography

Principle

Chromatography is a family of chemical techniques, analytically and preparatively, for the separation of mixtures. It involves passing the mixture containing the analyte, in the "mobile phase", often in a stream of solvent, through the "stationary phase." The stationary phase retards the passage of the sample components. When components pass through the system at different rates they become more and more separated in time. Each component has a characteristic time of passage through the system, called the "retention time." Chromatographic separation is achieved when the retention time of the analyte differs from that of other components in the sample.
The mixture is carried by liquid or gas and is separated into its component parts as a result of differential distributions of the solutes as they flow over a stationary liquid or solid phase. Various techniques rely on the differential affinities of substances for a gas or liquid mobile medium and for a stationary absorbing medium through which they pass, such as paper, gelatin, alumina or silica.
The (chemical) physics underlying all types of chromatography concern various cohesive and adhesive forces (see Cohesion, Adhesion, Capillary forces and Surface tension) and diffusion (see Diffusion: general) with the analyte, mobile and stationary phase playing the key roles.

Application

In many clinical disciplines but especially in internal medicine. Also in many biomedical, biochemical and chemical research and industrial.

More Info

Retention

The retention is a measure of the speed at which a substance moves in a chromatographic system. In continuous development systems where the compounds are eluted with the eluent, the retention is usually measured as the *retention time* R_t or t_R, the time between injection and detection. In interrupted development systems like thin layer chromatography, the retention is measured as the *retention factor* R_f, defined by:

R_f = distance moved by compound / distance moved by eluent.

Since it is hard to standardize retention, a comparison is made with a standard compounds under absolutely identical conditions.
A chromatographic system can be described as the mobile and stationary phases being in equilibrium. The partition coefficient K is based on this equilibrium, defined as K = [solute in stationary phase] / [solute in mobile phase]. K is assumed to be independent of the concentration of the analyte, and can change if experimental conditions are changed, for example temperature. As K increases, it takes longer for solutes to separate. For a column of fixed length and flow, the retention time (t_R) and retention volume (V_r) can be measured and used to calculate K.

Chromatographic techniques

Paper chromatography

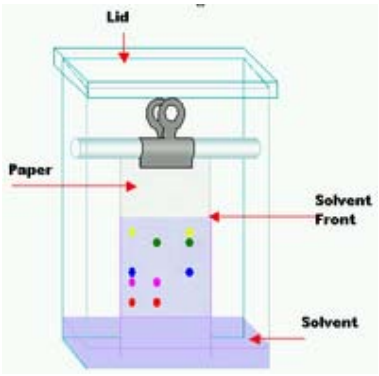


Fig. 1 Principle of a paper chromatograph

In paper chromatography a small spot of solution containing the sample is applied to a strip of chromatography paper (Fig. 1). This sample is adsorbed onto the paper. The paper is then dipped into a

suitable solvent (such as ethanol or water) and placed in a sealed container. As the solvent rises through the paper by capillary forces it meets the sample mixture. This starts to travel up the paper with the solvent. Cohesive and adhesive interactions with the paper make different compounds travel at different rates. The process takes some hours. The final chromatogram can be compared with other known mixture chromatograms for identification. This technique demonstrates very well the principle, but at the moment it has only educational relevance.

Thin layer chromatography (TLC)

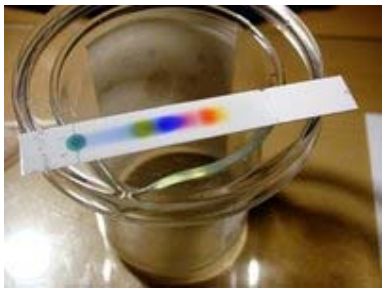


Fig. 2 Separation of black ink on a TLC plate

In TLC the stationary phase is a thin adsorbent layer like silica gel, alumina, etc. on a flat carrier like a glass plate, a thick aluminum foil, etc. (Fig. 2). The process is similar to paper chromatography but runs faster and separate better. It is used for monitoring chemical reactions and analysis of reaction products. Colorless spots of the compounds are made visible by a fluorescent dye (see Fluorescence) in the adsorbent under UV light. R_f values should be the same regardless of the extent of travel of the solvent, and in theory they are independent of a single experimental run. They do depend on the solvent used, and the type of TLC plate. Nowadays relevance is mainly educational.

Column chromatography

Column chromatography utilizes a vertical glass column filled with some form of solid support with the sample to be separated placed on top of this support. The rest of the column is filled with a solvent which, under the influence of gravity, moves the sample through the column. Similarly to other forms of chromatography, differences in rates of movement through the solid medium are translated to different exit times from the bottom of the column for the various elements of the original sample.

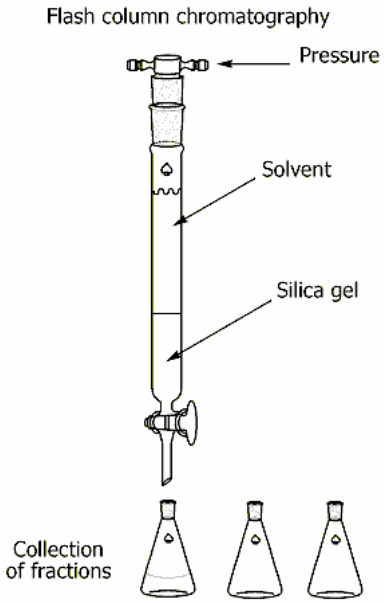


Fig. 3 Principle of a flash column chromatography setup

The flash column chromatography (Fig. 3) is very similar to the traditional column chromatography, except for that the solvent is driven through the column by applying positive pressure. It is faster and gives better separation. Miniaturized (disposable columns), this technique is widely applied.

Gas(-liquid) chromatography G(L)C
Gas-liquid chromatography is based on a partition equilibrium of analyte between a liquid stationary phase and a mobile gas. The mobile phase is a carrier gas, usually an inert gas such as He or N_2, and the stationary phase is a microscopic layer of liquid on an inert solid support inside of a very long and very thin column. It is useful for a wide range of non-polar analytes, but poor for thermally labile molecules. It is often combined with Mass spectrography.

Ion exchange chromatography
Ion exchange chromatography is a column chromatography that uses a charged stationary phase. It is used to separate charged compounds including amino acids, peptides and proteins. The stationary phase is usually an ion exchange resin that carries charged functional groups which interact with oppositely charged groups of the compound to be retained. Bound compounds can be eluted from the column by gradient elution (solvent composition with a time-gradient, e.g. in salt concentration or pH) or isocratic elution (solvent composition constant). Ion exchange chromatography is commonly used to purify proteins using Fast Protein Liquid Chromatography (FPLC).

Immobilized metal ion affinity chromatography, IMAC
IMAC is a popular and powerful way to purify proteins. It is based on the specific covalent binding between histidine or other unique amino acids (either naturally or grafted with recombinant DNA techniques) and various immobilized metal ions, such as copper, nickel, zinc, or iron.

High performance liquid chromatography, HPLC
HPLC is a form of column chromatography used frequently in biochemistry and analytical chemistry. The analyte is forced through a column (stationary phase) by a liquid (mobile phase) at high pressure, which decreases the time the analytes have to diffuse within the column. Diffusion within the column leads to broad peaks and loss of resolution. Less time on the column then translates to narrower peaks in the resulting chromatogram and thence to better resolution and sensitivity (discrimination from ‘’background noise’). Another way to decrease time the analyte stays on the column is to change the composition of the mobile phase over a period of time (a solvent time-gradient). HPLC is often combined within one apparatus with a mass spectrograph or gas(-liquid) chromatograph.

Reversed phase (RP) liquid chromatography
RP-HPLC was developed for large polar biomolecules. Like the name implies the nature of the stationary phase is reversed. RP-HPLC consists of a nonpolar stationary phase and a polar mobile phase. One common stationary phase is special treated silica. The retention time is longer when the mobile phase is more polar. This is the reverse of the situation which exists when normal silica is used as the stationary phase.

Gel permeation chromatography GPC
GPC also known as size exclusion chromatography or Sephadex gel chromatography, separates molecules on basis of size. Smaller molecules enter a porous media and take longer to exit the column, hence larger particles leave the column first. GPC is good for determining polymer molecular weight distribution, but has a low resolution.

Affinity chromatography
Affinity chromatography is based on selective non-covalent interaction between an analyte and specific molecules. It is often used in the purification of proteins (or better protein constructs).

There are many other versions of chromatography, see Wikipedia or textbooks on analytic chemistry.

Diffusion: Fick's laws

Principle

Fick's first law
Fick's first law is used in steady-state diffusion (see Diffusion: general), i.e., when the concentration within the diffusion volume does not change with respect to time ($J_{in} = J_{out}$).

$$J = -D\frac{\partial \phi}{\partial x}, \qquad (1)$$

where
J is the diffusion flux in dimensions of (mol m^{-2} s^{-1});
D is the diffusion coefficient or diffusivity, (m^2 s^{-1});
φ is the concentration (mol m^{-3});
x is the position (m).
D is proportional to the velocity of the diffusing particles, which depends on the temperature, viscosity of the fluid and the size of the particles according to the Stokes-Einstein relation. For biological molecules the diffusion coefficient normally ranges from 10^{-11} to 10^{-10} m^2 s^{-1}.

Fick's second law
Fick's second law is used in non-steady or continually changing state diffusion, i.e., when the concentration within the diffusion volume changes with respect to time.

$$\frac{\partial \phi}{\partial t} = -D\frac{\partial^2 \phi}{\partial x^2}, \qquad (2)$$

Where:
ϕ is the concentration (mol m^{-3});
t is time (s);
D is the constant diffusion coefficient (m^2 s^{-1});
x is the position (m).
With some calculation, it can be derived from the First Fick's law and the mass balance.

Application

Equations based on Fick's law have been commonly used to model transport processes in foods, porous soils, semiconductor doping process, etc.. In biomedicine, transport of biopolymers, pharmaceuticals, in neurons, etc. are modeled with the Fick equations.
Typically, a compound's D is ~10,000x greater in air than in water. CO_2 in air has a diffusion coefficient of 16 mm²/s, and in water, its coefficient is 0.0016 mm²/s.

Physiological examples
For example for the steady states, when concentration does not change by time, the left part of the above equation will be zero and therefore in one dimension and when D is constant, the equation (2) becomes $0 = D\, d\varphi/dx^2$. The solution for the concentration φ will be the linear change of concentrations along x. This is what by approximation happens in the alveolar-capillary membrane and with diffusion of particles from the liquid to liquid phase in the wall of capillaries in the tissues.

Biological perspective
The first law gives rise to the following formula:

$$\text{Rate of diffusion} = KA(P_2 - P_1) \qquad (3)$$

It states that the rate of diffusion of a gas across a membrane is:
- Constant for a given gas at a given temperature by an experimentally determined factor, K
- Proportional to the surface area over which diffusion is taking place, A
- Proportional to the difference in partial pressures of the gas across the membrane, $P_2 - P_1$

It is also inversely proportional to the distance over which diffusion must take place, or in other words the thickness of the membrane. The factor K comprises this thickness.
The exchange rate of a gas across a fluid membrane can be determined by using this law together with Graham's law (see Diffusion: Graham's law.

More Info

Temperature dependence of the diffusion coefficient
The diffusion coefficient at different temperatures is often found to be well predicted by

$$D = D_0 e^{\frac{-E_A}{RT}}, \quad (4)$$

where:
D_0 is the maximum diffusion coefficient at infinite temperature,
E_A is the activation energy for diffusion (J/mol),
T is absolute temperature (K),
R is the gas constant in dimensions of (J/(K·mol)).

3D diffusion
For the case of 3-dimensional diffusion the Second Fick's Law looks like:

$$\frac{\partial \phi}{\partial t} = D\nabla^2 \phi \quad (5)$$

where ∇ is the del operator. In a 3D system with perpendicular coordinates (*x*, *y*, *z*), this is a Cartesian coordinate system $\mathbf{R}^3$, del is defined as:

$$\nabla = i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}, \quad (6)$$

where i, j and k are the unit-vectors in the direction of the respective coordinate (the standard basis in $\mathbf{R}^3$).

Finally if the diffusion coefficient is not a constant, but depends upon the coordinate and/or concentration, the Second Fick's Law looks like:

$$\frac{\partial \phi}{\partial t} = \nabla \cdot (D\nabla \phi). \quad (7)$$

Diffusion: general

Principle

Diffusion is the net action of particles (molecules, atoms, electrons, etc.), heat, momentum, or light whose aim is to minimize a concentration gradient. A concentration gradient is the difference between the high concentration and the low concentration. It also includes the speed of the process. Diffusion can be quantified by measuring the concentrations gradient.
More formally, diffusion is defined as the process through which speed a thermodynamic system at local thermodynamic equilibrium returns to global thermodynamic equilibriums, through the homogenization of the values of its intensive parameters (or bulk parameters, such as viscosity, density, melting point etc.).

In all cases of diffusion, the net flux of the transported quantity (atoms, energy, or electrons) is equal to a physical property (diffusivity, thermal conductivity, electrical conductivity) multiplied by a gradient (a concentration, thermal, electric field gradient). So:
Diffusion ≡ some conductivity times a gradient.
Noticeable transport occurs only if there is a gradient. For example, if the temperature is constant, heat will move as quickly in one direction as in the other, producing no net heat transport or change in temperature.

The process of diffusion minimizes free energy and is thus a *spontaneous* process. An example of diffusion is the swelling of pasta, where water diffuses into the sponge-like structure of the dry and stiff pasta.
The different forms of diffusion can be modeled quantitatively using the diffusion equation, (see **More Info**) which goes by different names depending on the physical situation. For instance - steady-state bi-molecular diffusion is governed by Fick's laws (see Diffusion: Fick's law, steady-state thermal diffusion is governed by Fourier's law. The generic diffusion equation is time dependent, and as such applies to non-steady-state situations as well.

The second law of thermodynamics states that in a spontaneous process, the entropy of the universe increases. Entropy is a measure of how far a spontaneous physical process of smoothing-out differences has progressed, for instance a difference in temperature or in concentration (see further Entropy). Change in entropy of the universe is equal to the sum of the change in entropy of a system and the change in entropy of the surroundings. A system refers to the part of the universe being studied; the surrounding is everything else in the universe. Spontaneous change results in dispersal of energy. Spontaneous processes are not reversible and only occur in one direction. No work is required for diffusion in a closed system. Reversibility is associated with equilibrium. Work can be done on the system to change equilibrium. Energy from the surroundings decrease by the amount of work expended from surroundings. Ultimately, there will be a greater increase in entropy in the surroundings than the decrease of entropy in the system working accordingly with the second law of thermodynamics.

Types of diffusion
Diffusion includes all transport phenomena occurring within thermodynamic systems under the influence of thermal fluctuations (i.e. under the influence of disorder; this excludes transport through a hydrodynamic flow, which is a macroscopic, ordered phenomenon). Well known types are the diffusion of atoms and molecules, of electrons, (resulting in electric current), Brownian motion (e.g. of a single particle in a solvent), collective diffusion (the diffusion of a large number of (possibly interacting) particles), effusion of a gas through small holes, heat flow (thermal diffusion), osmosis, isotope separation with gaseous diffusion.

Application

Application, i.e. occurrence, in nature, dead and alive, can be found nearly everywhere. In living bodies it is found interstitionally as well as in living cells. Many man made apparatus, also for medical purposes, make use of the principle of diffusion. Some very obvious examples of applications are in the techniques of Chromatography, Electrophoresis, Oxygen analysis, Spectroscopy, Thermography, the calculation of body heat conduction and dissipation (see Body heat conduction and Newton's Law of cooling) etc.

Diffusion in biological systems
Specific examples in biological systems are diffusion across biological membranes, of ion through ion channels, in the alveoli of mammalian lungs across the alveolar-capillary membrane. In the latter process the diffusion is Knudson-like diffusion, in other words it is also dependent on space restrictions

(the small size of the providing (alveoli) and recipient (capillaries) volumes. Another type is facilitated diffusion (passive transport across a membrane, with the assistance of transport proteins)

Numeric example

By knowing the diffusion coefficient of oxygen in watery tissue, the distance of diffusion and the diffusion gradient, then the time required for the diffusion of oxygen from the alveoli to the alveolar capillaries can be calculated. It appears to be some 0.6 ms (see **More Info**).

More Info

Diffusion equation

The diffusion equation is a partial differential equation, which describes the density fluctuations in a material undergoing diffusion. It is also used in population genetics to describe the 'diffusion' of alleles in a population.

The equation is usually written as:

$$\frac{\partial \phi}{\partial t} = \nabla \cdot D(\phi) \nabla \phi(\vec{r}, t) \qquad (1)$$

where φ is the density of the diffusing material, *t* is time, D is the collective diffusion coefficient, $\vec{r}$ is the spatial coordinate and the nabla symbol ϕ represents the vector differential operator del (see also Fick's Laws). If the diffusion coefficient depends on the density then the equation is nonlinear. If *D* is a constant, however, then the equation reduces to the following linear equation:

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi(\vec{r}, t)\,, \qquad (2)$$

also called the heat equation.

Diffusion displacement

The diffusion displacement can be described by the following formula:

$$\langle r_k^2 \rangle = 2kD't, \qquad (3)$$

where *k* is the dimensions of the system and can be one, two or three. D' is the diffusion coefficient, but now also per unit of concentration difference of the particles and *t* is time. For the three-dimensional systems the above equation will be:

$$\langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle = \langle r_3^2 \rangle = 6D't\,, \qquad (4)$$

where < > indicates the mean of the squared displacement of the individual particles.

Physiological numeric example

Calculate the time required for O_2 diffusion across the alveolar-capillar membrane. By knowing D' of oxygen in watery tissue (suppose 15×10^{-10} $m^2 s^{-1} bar^{-1}$), the 1D-distance of diffusion x (suppose the distance between alveolar and capillary volume is 0.4 μm) and the O_2 diffusion gradient Δp is 0.09 bar (suppose partial pressures of 0.14 bar in alveoli and 0.05 bar in A. pulmonalis) and supposing that the system is ideal (half infinite at both sides and a constant gradient), then the time required for the diffusion of O_2 from the alveoli to the alveolar capillaries can be calculated from (3) but now defined for a gas:

$$t = x^2/(2D\Delta p) = 0.16 \times 10^{-12}/(2 \times 15 \times 10^{-10} \times 0.09) = 0.6 \times 10^{-3}\ \text{s}. \qquad (5)$$

See further Diffusion: Fick's laws, for the diffusion equations and Diffusion: Grahams Law for particle velocity of diffusing particles.

Diffusion: Graham's law

Principle

Graham's law is also known as Graham's law of effusion. Effusion is the process where individual molecules of a gas flow through a hole without collisions. This will occur if the diameter of the hole is considerably smaller than the mean (collision-)free path of the molecules. The free path is at 1 bar of the oder of 70 nm. The law states that the rate of effusion of a gas is inversely proportional to the square root of the molecular mass of its particles. This formula can be written as:

$$v_1/v_2 = (m_1/m_2)^{0.5}, \quad (1)$$

where:
v_1 is the rate of effusion of the first gas;
v_2 is the rate of effusion for the second gas;
m_1 is the molar mass of gas 1;
m_2 is the molar mass of gas 2.

Graham's law is most accurate for molecular effusion which involves the movement of one gas at a time through a hole. It is only approximate for diffusion of one gas in another or in air, as these processes involve the movement of more than one gas.
According to the kinetic theory of gases, the absolute temperature T (Kelvin) is directly proportional to the average kinetic energy of the gas molecules ($0.5mv^2$). This can be derived from:
$\langle v^2 \rangle = 3RT/(N_A m)$,
where $\langle v^2 \rangle$ the mean of the squared velocity of the particles, R the molar gas constant (= 8315 J/kmol·K), $N_A m$ the gas mass with N_A the Avogadro's number (see Gas laws).
Thus, to have equal kinetic energies and so temperature, the velocities of two different molecules would have to be in inverse proportion to the square roots of their masses. Since the rate of diffusion is determined by the average molecular velocity, Graham's law for diffusion could be understood in terms of the molecular kinetic energies being equal at the same temperature.

Application

The law is of important for processes of diffusion of a gas into another gas or gas mixture (see Diffusion: general). For diffusion across a liquid membrane, gas concentration should be low and the membrane thin.

Graham's Law can also be used to find the approximate molecular weight of a gas if the rates are measured and the molecular weight of one of the gases is a known.

Graham's law was the basis for separating $^{235}UF_6$ from $^{238}UF_6$. Both isotopes of uranium, as element, are found in natural uranium ore, but the 235-isotope about 100 times less. By repeated diffusion through porous barriers the slightly lighter ^{235}U isotope is enriched.

Electrophoresis

Principle

Electrophoresis is the movement of an electrically charged substance under the influence of an electric field. This movement is due to the Lorentz force, which acts on the charge of the particle under study and which is dependent on the ambient electrical conditions. This force is given by:

$$F = qE \quad (1)$$

F (a vector) is the Lorentz force, q is the charge (a scalar) of the particle, *E* is the electric field, a vector. The resulting electrophoretic migration is countered by forces of friction, such that the rate of migration is constant in a constant and homogeneous electric field:

$$F_f = vf\,, \quad (2)$$

where v is the velocity and *f* is the frictional coefficient. Since in the stationary condition forces of friction and Lorentz force become the same, it holds that:

$$qE = vf \quad (3)$$

The electrophoretic mobility μ is defined as:

$$\mu = v/E = q/f\,. \quad (4)$$

The expression (4) above applied only to charged molecules (ions) at a low concentration and in a non-conductive solvent. Poly-ionic molecules are surrounded by a cloud of counter-ions which alter the effective electric field applied on the molecule. This render the previous expression a poor approximation of what really happens in an electrophoretic apparatus.

Application

Electrophoresis is used as an preparative and analytical tool in molecular biology.
Gel-electrophoresis is an application of electrophoresis in molecular biology, especially in DNA techniques. The gel-electrophoresis apparatus uses a positive and a negative charged pole. The (macro)molecule, e.g. DNA is loaded on the negatively charged pole and pulled through the gel toward the positive pole.
The charged of a DNA molecule is provided by negative phosphate groups. The content of the buffers (solutions) and gels used to enhance viscosity greatly affects the mobility of macromolecules. The gel used in the procedure is typically an agarose or a polyacrylamide gel, depending on the type of molecule studied. The thickness of the gel is typically ca. 8 mm. Within the gel is a tightly woven matrix that the molecules must pass through as they are moving from one pole to the other. The smaller molecules can weave in and out of the matrix of the gel with more ease, compared with larger molecules. Wells, or rectangular openings, are formed along one edge of the gel. These wells mark the different lanes in which a sample may be loaded. Agarose is applied to separate large DNA fragments (100-50,000 base pairs) with modest resolution (some 50 base pairs). Polyacrylamide is useful for small fragments (10-1000 pairs) with single base-pair resolution.
Modifications are e.g. gradient (detergent) gel-electrophoresis and (water cooled) rapid agarose gel electrophoresis.

More Info

The mobility depends on both the particle properties (e.g., surface charge density and size) and solution properties, e.g., ionic strength, electric permittivity, and pH. (Permittivity describes how an electric field affects and is affected by a medium, e.g. air or water). For high ionic strengths, an approximate expression for the electrophoretic mobility μ_e is given by the equation:

$$\mu_e = \varepsilon \cdot \varepsilon_0 \cdot \zeta / \eta, \quad (5)$$

where ε is the dielectric constant (relative permittivity) of the liquid, ε_0 is the permittivity of vacuum, η is the viscosity of the liquid, and ζ is the zeta potential (i.e. surface potential) of the particle.

Electrosurgery

Principle

Electrosurgery is the application of a high-frequency electric current to tissue as a means to remove lesions, staunch bleeding, or cut tissue. It is based on the generation of local heat dissipated by a piece of tissue when electric current is flowing through it. The tissue can be considered as an electric resistor. To perform electrosurgery, a voltage source is applied to the fine surgical electrode or probe (electric knife) and a 2nd electrode with the tissue in between. The current can be adjusted by changing the voltage of the source. The dissipated power is $P = V^2/R$ (in Watts) which directly follows from Ohms law ($V = iR$) and the relation $P = iV$).
The electrode does not heat up since the resistance of the metal electrode and metal wire is so much smaller than that of the tissue that very little power is expended inside the metal conductors. Electrosurgery is performed using a device called an Electrosurgical Generator, sometimes referred to as an RF Knife.

The change in temperature of an object is inversely proportional to its (specific) heat capacity (in $J \cdot kg^{-1} \cdot K^{-1}$). For water (near 20 °C) this is 4184 $J \cdot kg^{-1} \cdot K^{-1}$. This value can also be used for watery tissues. Since the heat needed is proportional to the mass of the object, the heated mass is limited by using small probes. By applying a high current density (current/area), which is achieved by a small electrode tip, the resistance of the small volume of tissue adjacent to the tip is subjected to a large current density. The generated heat will now easily burn the tissue at the electrode.

The human nervous system is very sensitive to low-frequency (0 Hz to ca. 1000 Hz) electricity, which stimulates the nervous system. At even low currents low-frequency electricity causes electric shock, which may involve acute pain, muscle spasms, and/or cardiac arrest. The sensitivity decreases with increasing frequency and at frequencies > 100 kHz, electricity does not stimulate the nervous system. To avoid electric shock, electrosurgical equipment operates in the frequency range of 200 kHz to 5 MHz.

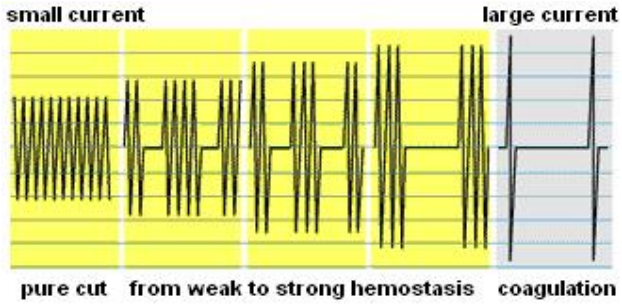


Fig. 1 Modes of operation.

Fig. 1 illustrates the various operation modes for the various aims.

Application

Electrosurgery can be used to cut, coagulate, desiccate, or fulgurate tissue. Its benefits include the ability to make precise cuts with limited blood loss.
Although electrosurgical devices may be used for the cauterization of tissue in some applications (e.g. hemorrhoid surgery), electrosurgery is usually used to refer to a quite different method than that used by many dedicated electrocautery devices. The latter uses heat conduction (see Body heat conduction and Newton's Law of cooling) from a hot probe heated by a direct current (much in the manner of a soldering iron), whereas electrosurgery uses alternating current to directly heat the tissue itself (electric diathermy), while the probe tip remains relatively cool.
Different waveforms of the electric current can be used for different electrosurgical procedures. For cutting, a continuous single frequency sine wave is generated. This produces rapid heating. At the cellular level, rapid heating causes tissue cells to boil and burst. At a larger scale, the ruptured cells create a fine tear in the tissue, creating a clean incision.
For coagulation, the sine wave is modulated by turned on and off in rapid succession. The overall effect is a slower heating process, which causes cells to coagulate. The proportion of on time to on+off time, the duty cycle, can be varied to allow control of the heating rate.

Dermatological applications are removal of skin tags, removal/destruction of benign skin tumors and warts. For several aims it is now often preferred by dermatologists over laser surgery and cryosurgery.

Savety
High power *monopolar* surgery requires a good electrical contact between a large area of the body and the return electrode to prevent severe burns (3rd degree) in unintended areas on the skin and beneath the skin of (anesthetized) patients. To prevent unintended burns, the skin should be clean and dry and a conductive jelly should be used. Proper electrical grounding practices must be followed in the electrical wiring of the building. It is also recommended to use a newer electrosurgical unit that includes alarms for ground circuit interruption.
Safety is essentially improved when the electric circuit is inductively uncoupled. This is performed by a primary coil in the generator and a secondary coil in the circuit of the probe.

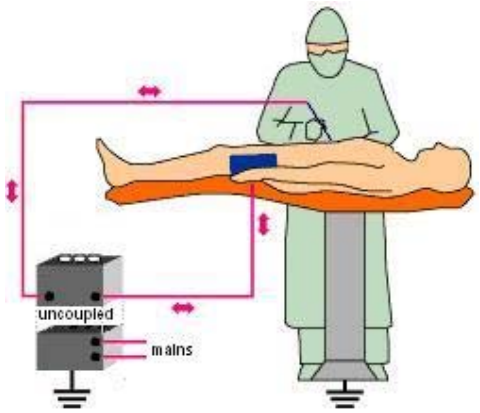


Fig. 2 Bipolar setup with the circuit uncoupled from the mains by coils. The double headed arrow indicates the currents direction which alternates every cycle of the current (A.C. or alternating current).

Medical numerical example
A strongly simplified calculation yields the order of magnitude of the strength of currents and voltages applied. Suppose that a small blood vessel (say 2 mm in diameter) should be coagulated. This is performed by applying a current, which coagulates at 237 ^{o}C a sphere with a diameter of 2.8 mm (volume 0.23 ml). This requires 0.23 g x 4.2 J/(g x K) x 200 K = 10 J (tissue density is 1 g/ml). This is applied in 5 s, so 2 J/s, is 2 Watt. Supposing an electrode resistance of 10 KOhm (bipolar set up), then 140 V (= $P^{0.5}V^{0.5}$ = $(10000 \times 2)^{0.5}$) is needed and the current is 140 V/10 KOhm = 14 mA. This holds for continues current (DC or direct current), whereas the currents is applied intermittent in high frequency bursts. Supposing a duty cycle of 10% and a sinusoidal current, then the amplitude of the voltage is $2^{0.5}$ X 10 X 140/2 is about 1000 V. The 1000 V (generating 140 mA effectively) again underlines the absolute necessity of save grounding and shows that uncoupling is the ultimate save approach.

More Info

Electrosurgical modalities
Monopolar and Bipolar There are two circuit topologies: *monopolar* and *bipolar*. The *bipolar modality* is used less often. Voltage is applied to the patient using a special forceps, with one tine connected to one pole of the A.C. (alternating current) voltage source and the other tine connected to the other pole of the voltage source. When a piece of tissue is held by the forceps, a high frequency electrical current flows from one to the other forceps tine, through the intervening tissue. In the *monopolar modality* the patient lies on top of the *return electrode*, a relatively large metal plate or a relatively large flexible metalized plastic pad (somewhere attached to the body), which is connected to the other wire of the current source. The surgeon uses a single, pointed, probe to make contact with the tissue. The electrical current flows from the probe tip through the body and then to the return electrode. In the monopolar modality the heating is also very precisely confined to the tissue that is near the probe tip since the current rapidly spreads out laterally in the body, causing a quadratic decrease in the current density with the distance.
Spark gap or *fulguration modality* It is a low-powered monopolar electrosurgery performed on conscious outpatients (dermatology). At low power, this technique requires no return electrode or patient-contact-plate since at the very high frequencies and low currents, the parasitic capacitance between the patient's body and the machine's ground potential is large enough to close the electric circuit. If such a spark is small, it can cause relatively minor shocks or burns, but these are not a problem in a low-powered setting with conscious patients because they are immediately noticed. For high-power or surgical anesthesia settings, however, a ground pad is always necessary to insure that all

such stray ground currents enter the machine safely through a large-skin-surface contact, and dedicated wire.

Entrance effect and entrance length

Principle

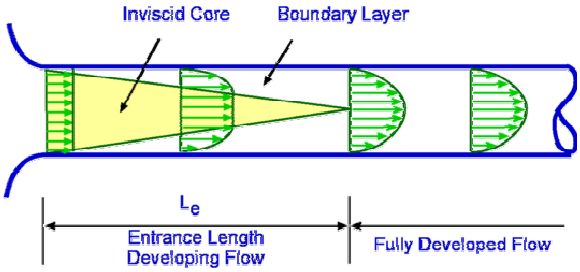


Fig. 1 Flow at the entrance to a tube

Consider a flow entering a tube and suppose that the entering flow is uniform, so inviscid. As soon as the flow 'hits' the tube many changes take place. The most important of these is that viscosity imposes itself on the flow and the friction at the wall of the tube comes into effect. Consequently the velocity along the wall becomes zero (also in tangential direction). The flow in the layer adjacent to the wall decelerates continuously. This layer is what is called the boundary Layer. Viscous effects are dominant within the boundary layer. Outside of this layer is the inviscid core with uniform flow where viscous effects are negligible or absent.
The boundary layer is not a static phenomenon; it is dynamic. It grows meaning that its thickness increases as we move downstream. From Fig. 1 it is seen that the boundary layer from the walls grows to such an extent that they all merge on the centre line of the tube. Once this takes place, inviscid core terminates and the flow is all viscous. The flow is now called a *fully developed flow*; the velocity profile is parabolic or Poiseuille flow (see Poiseuille's Law). Once the flow is fully developed the velocity profile does not vary in the flow direction. In fact in this region the pressure gradient and the shear stress in the flow are in balance. The length of the tube between the start and the point where the fully developed flow begins is called the Entrance Length, denoted by L_e. The entrance length is a function of the Reynolds number Re of the flow.
The distance needed to restore a laminar flow to a parabolic flow is called the entrance length, being:

$$L_{e,laminar} = C \cdot D \cdot Re, \qquad (1)$$

where D is the tube diameter and C a constant, dependent on the diameter change and abruptness of the geometrical change of the entrance. It ranges from about 0.06 (fast conical change) to 0.03 (smooth, slender conical change). The latter value holds rather well for most anatomical geometries, except for the heart-vessel and mouth-trachea transitions.

Application

Hemodynamics and flow in the airways system (see Flow through a stenosis, Flow in curvatures and Flow in bifurcations).
Numerical example Supposing that the aorta diameter is 3 cm and Re is 1400, which holds in rest, then $L_{e,turbulent}$ is about 250 cm. This is much more than the distance of the first large bifurcations after the aortic arch. Also when we take (conceptually) into account the aorta bend and the bifurcations in the aorta bend (which will disturb the process of restoration), the pulsatile character of the aorta flow and the compliance of the wall (which support restoration) Poiseuille flow is not restored.

More info

For turbulent flow holds:

$$L_{e,turbulent} = 4.4D \cdot Re^{1/6}. \qquad (2)$$

At the critical condition, i.e., Re_d =2300, the L_e/D ratio, the entrance length number, for a laminar flow is 138. From this condition, with transitional flow to full turbulent flow at Re = 10000 it *diminishes* to 18.6.

Reference

http://www.aeromech.usyd.edu.au/aero/ pipeflow/node7.html#node9

Flow in a bended tube

Principle

When steady laminar fluid in a long straight tube enters a bend, every fluid element of it must change its direction of motion. It must acquire a component of acceleration at right angles to its original direction, and must therefore experience a force in that direction. This most fast moving elements, those at the axis, will change their original path less rapidly than slower moving ones, on account of their greater inertia ($\frac{1}{2}\rho v^2$). Their original velocity vector obtains an axial (the axis of the bend) component and a component perpendicular to the axis. At the axis within the plane of the bend, this force, perpendicular to the axial force, is directed to the outside wall of the bend. The same holds for neighboring positions, close the central plane of the bend. This results in an outward rather uniform flow. In consequence, the fluid originally at the outside of the bend has a tangential flow component along the wall in the direction of the inside of the bend. And so, a transverse circulation or secondary motion is set up, as shown in Fig. 1a. The further moved into the bend, the original parabolic velocity profile at the entrance becomes more and more skewed in the plane of the bend (Fig. 1b).
Faster moving fluid near the outside of the bend means that the axial velocity profile in the plane of the bend is distorted from its symmetric shape to the M-shape shown in Fig. 1c. The reason is that the fluid being swept back around the side walls is still traveling faster than that pulled towards the center of the tube from the inside wall.
A similar pattern of flow is to be expected far from the entrance in a long, continuously curved tube, like the aorta bend, whatever the initial velocity profile.

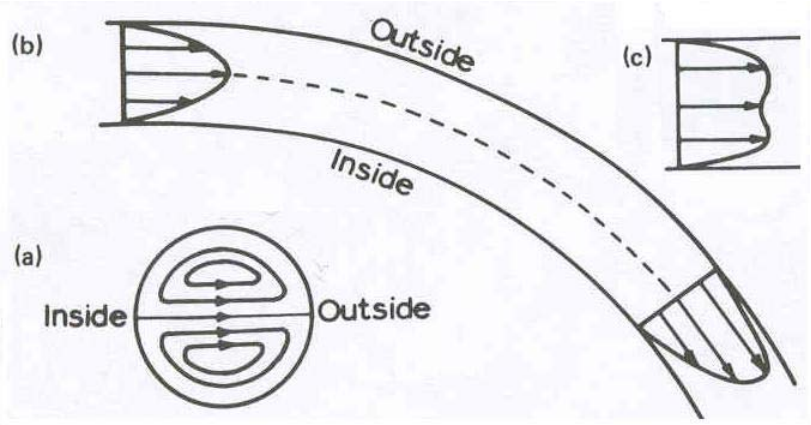


Fig. 1 Laminar flow profile in a curvature. c is perpendicular to b.

Application

In fluid dynamics of blood flow through vessels and flow in the airways. Curvatures in the airways system are generally combined with bifurcation (see Flow in bifurcations). They may increase resistance and pressure drop compared to a bifurcation with straight tubes. In the vascular system bends are very common, for instance in the brain, and can increase resistance considerably. Vessels in joints can be bend easily 90^0 and even more than 135^0. In the bend, the vessel, in particular veins, will be flattened, another effect of increasing resistance pressure drop. Supposing a vessel is compressed to half its diameter, whereas the perimeter remains the same and the new shape is elliptical, than the area is decreased by 16%. When the volume flow remains the same, then the Reynolds number increases by 16%.

With a sharp bend over 90^0 with a curvature ratio of 1.5, as occurs with the *aorta bend*, a curvature coefficient (see **More Info**) ζ_c of 0.17 is reached. In physical rest, with a cardiac output of 6 L/min and v is 0.15 m/s, Reynolds number Re of the aorta is about 1350, supposed the aorta is straight, there is no entrance effect (see Entrance effect and entrance length) of the aortic valve and the flow is steady. Taking the pulsatile blood flow into account the maximal velocity is about 0.6 m/s. Taking the bending and the entrance effect (entrance coefficient ζ_e at least 0.2) also into account, Re will reach a temporarily maximum of Re much higher than 1350. Although with pulsatile flow the critical Re is higher than 2100, the flow in the aorta is supposed to have transitional flow, also at rest. With very high cardiac outputs (30 L/min), as occur during heavy endurance sport, the aortic flow is turbulent and this will proceed in the largest arteries.

In an 180^0 strong bend, as in the aorta, and high flows, which occur during heavy exercise, there arise curling, spiralising, asymmetrically distributed vortices (Dean vortices).

More info

As long as the ratio of the curvature radius (r_c) and tube diameter (D), being r_c/D, is sufficiently large, the flow depends on the value of the parameter De (the Dean number):

$$De = Re\ (D/r_c)^{1/2}.$$

Dean numbers up to 150 hold for the above described flow behavior.
If the Reynolds number is increased sufficiently, the flow will become turbulent. However, the critical Reynolds number for transition is much greater in a curved tube than in a straight one. The transition occurs at critical De numbers, which are r_c/D dependent. Once the flow is turbulent, the dissipation, and hence pressure drop, rises dramatically. However, the fully developed turbulent mean velocity profile is similar to that in laminar flow.

There is no good theory available for large Dean numbers with turbulent flow. With such flow, the velocity profile is blunt and the flow inviscid. At the entrance of the bend, the flow profile, in contrast to that of Poiseuille flow (see Poiseuille's Law), is skewed to the inside. However, the maximum value of the shear rate at the wall, initially on the inside wall, switches over to the outside wall at a distance of only about one diameter from the entrance of the bend. Consequently, the skewed profile flips over to the outside.
At the outside wall of the bend, the rate of energy dissipation by viscosity is high, due to the high velocity gradient. Therefore the pressure gradient required to maintain a given flow rate is also higher than in Poiseuille flow. Since the distortion of the velocity profile increases for increasing Dean number (i.e. for increasing Reynolds number in a given bend), the ratio of actual pressure drop to Poiseuille pressure drop in a given length of tube will also increase.

Analytical calculation of the flow in a bend is very complicated. Here follows a practical technical approach.
The resistance increase due to the curvature can be expressed as the curvature resistance coefficient or loss-factor ζ_c. It is dependent on curvature angle β and the ratio of curvature radius (r_c) and tube diameter (D) r_c/D. Up to β is about 22.5^o there is no r_c/D dependency. At 45^o, normal values in vascular anatomy, a ratio of 10 yields a factor 0.07. In technical handbooks for fluid transport, ζ_c is tabulated for β and curvature ratio.

Literature

Pedley TJ et. al, Gas flow and mixing in the airways, In: West, J.B. (ed.) *Bioengineering Aspects of the Lung*. New York: Marcel Dekker, 1977.

Flow in bifurcations

Principle

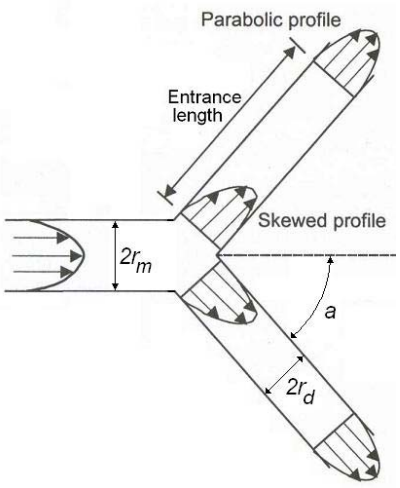


Fig. 1 Bifurcation

As Fig. 1 shows, a bifurcation changes the circular symmetric parabolic fluid flow (of a liquid or gas) in a circular mother tube to a skewed profile in the daughter branches. The radius ratio of mother and daughter branches, r_d/r_m, is energetically optimal for a ratio of $2^{-1/3} = 0.79$. This gives a daughters/mother area ratio of 1.25. For a symmetric bifurcation a semi-bifurcation angle *a* of 37.5^0 (=arctan($2^{-1/3}$)) is optimal. For this angle the resistance increase is only some 10%.

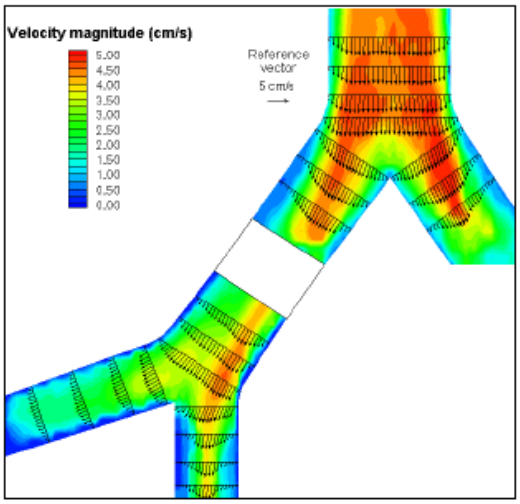


Fig. 2 Model of the flow profile of air in the trachea bifurcation

Application

Vascular and airways flow, and flow in the kidney. They play an important role in the study of the origination and grow of arteriosclerosis and aneurisms.

In the airways tree of the lungs some 26 generations (orders) of bifurcations, from trachea to the most remote alveoli can be distinguished (see Lung gas transport 1: basic principles). In mammals, the bifurcation ratio r_d/r_m, is on average 0.78 and the bifurcation angle 67^o, 8^o less than the optimum. Human midrange generations show higher angles: 79^o. The two semi-bifurcation angles are more similar for higher generations. From the 2^{nd} to the 10^{th} generation the human bifurcation ratio r_d/r_m is

very close to 0.79, but for higher generations the ratio slowly increases and the highest generations hardly show a diameter decrease from mother to daughter. The first generation bifurcation shows daughter diameters which are much smaller than the optimum. However, above optimizations holds for Poisseuille flow (see Poisseuille's Law), and this does not hold for the human trachea.
In the vascular system of mammals an averaged value of 0.68 is found for r_d/r_m. In the myocardium, r_d/r_m goes from 0.79 (the optimum) for the capillaries to about 0.72 for arteries of some mm diameter. Nevertheless, the optimization is poor since the variation is very large with many much too small or large daughters. Also the bifurcations are rather asymmetric, going from 0.8 (= $r_{small\ daughter}/r_{large\ daughter}$) of capillaries to about 0.2 of mm-sized arteries. For energetic reasons, asymmetry in diameter is accompanied by asymmetry in the semi-bifurcation angle. The human brain the mean bifurcation angle is 74^o with tens of degrees of variation. Branching angles in muscles can deviate substantial from their fluid dynamic minimum cost optimum. Classically, minimum cost according to Murray's Law is obtained when:

$$(r_{d1}/r_m)^3 + (r_{d2}/r_m)^3 = 1.$$

With $r_{d1} = r_{d1}$ the smallest minimum (with the ratio 0.79) is obtained. This solution is about the same as that on the basis of calculations with the Womersley number for laminar flow with minimal local wave reflection as criterion. For a given radius ratio, the optimal angles can be calculated.
The experimentally obtained exponents of Murray's law range from 2.5 to 2.8.

More info

Fig. 3 illustrates in more detail the flow patters in a bifurcation.

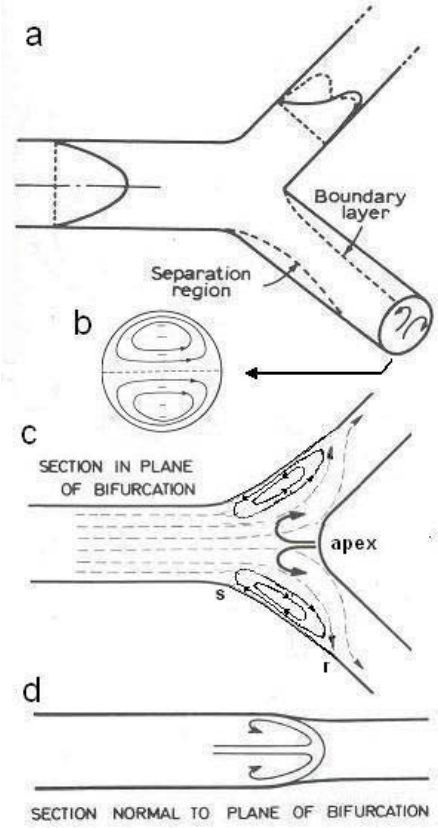


Fig. 3 a. The dashed profile in the upper daughter branch is the velocity profile perpendicular on the plane of the bifurcation. b. the dashed line is in the plane of the bifurcation (the symmetry plane). c. Stream lines with points of separation and reattachment indicated. d Streamline of flow impinging onto the apex.

When the direction of flow is inverted, in the mother tube four secondary loops can arise (Fig. 4). Secondary flows are characterized by a swirling, helical component superimposed on the main streamwise velocity along the tube axis, see also Flow in a bended tube. A complicating factor is wall

compliance, especially of importance with pulsatile flow, another factor which makes the pattern more complicated.

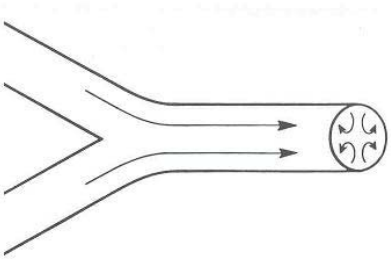


Fig. 4 Inversed flow in a bifurcation.

Literature

Canals M et. al. A simple geometrical pattern for the branching distribution of the bronchial tree, useful to estimate optimality departures. Acta Biotheor. 2004;52:1-16.
Frame MD, Sarelius IH. Energy optimization and bifurcation angles in the microcirculation. Microvasc Res. 1995 Nov;50(3):301-10.
Murray CD. The physiological principle of minimum work. I. The vascular system and the cost of blood volume. Proc Nat Acad Sci 12: 207-214, 1926.
Pedley TJ et. al, Gas flow and mixing in the airways, In: West, J.B. (ed.) *Bioengineering Aspects of the Lung*. New York: Marcel Dekker, 1977.
VanBavel E, Spaan JA. Branching patterns in the porcine coronary arterial tree. Estimation of flow heterogeneity. Circ Res. 1992;71:1200-12.
http://www.vki.ac.be/research/themes/annualsurvey/2002/biological_fluid_ea1603v1.pdf

Flow through a stenosis

Principle

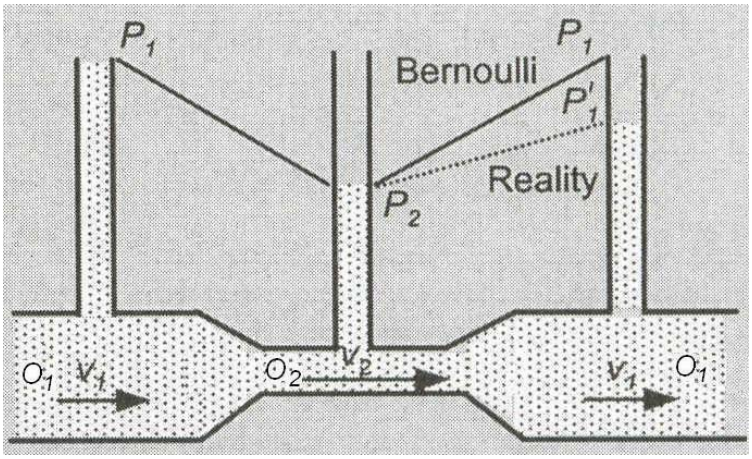


Fig. 1 The pressure after the stenosis P'_1 is smaller than P_1. due to losses at the entrance and exit of the stenosis (see **More Info**). P is pressure, O area and v mean flow velocity.

With incompressible flow of liquid through a stenosis (Fig. 1) in a horizontal tube, after the constriction, according to Bernoulli's law (see Bernoulli's and Pascal's Law) the pressure drops. Since h is constant (horizontal), the gravity term can be neglected and it holds that $\frac{1}{2}\rho v^2 + P = \text{constant}$. Since in the tube the volume flow $\dot{V} = O_1 v_1 = O_2 v_2 = \text{constant}$ and with $O_2 = \beta O_1$, it holds that $v_2 = \beta^{-1} v_1$, and $P_2 = \beta^{-2} P_1$. The constriction can result in a change of the type of flow. For instance, before the constriction it can be laminar, for circular diameters in accordance with Poiseuille's law, and after the constriction slightly or actually turbulent, depending on the Reynolds number. However, since $\dot{V}$ is constant and v applies to the mean flow velocity in the tube, pressure changes are not dependent on the type of flow, in accordance with Bernoulli's law.

Application

In hemodynamics. To maintain the same $\dot{V}$, a stenosis demands a higher driving force. Many stenosis increases substantially energy expenditure, be it cordially or pulmonary and will finally give rise to disorders. The severity of a stenosis is expressed as % area occlusion: $100(1-\beta)\%$.

More info

A constriction increases the tube resistance due to the decreased diameter over the length of the constriction, due to the (sudden) narrowing. For laminar flow this increase is given by Poiseuille's law. To calculate the pressure at the exit of the constriction Bernoulli's law defined as $\frac{1}{2}\rho v^2 + \rho g h + P = \text{constant}$ (ρ is specific density, g gravity constant), is inadequate since it does not account for the entrance effect, and the widening of the constriction. The equation is extended to:

$$\frac{1}{2}\rho v^2 + \rho g h + \Sigma w_i + P = \text{constant}, \qquad (1)$$

where Σw_i is the summed losses in the irregularities in the tube, such as sudden changes in diameter and curvatures. With this extension and elaborating Σw_i it is more common to define Bernouilli's law in terms of equivalent height:

$$\frac{1}{2}v^2/g + h + \Sigma 8\pi l_i v_i \eta / g\rho O_i + \Sigma \frac{1}{2}\zeta_i v_i^2/g + P/\rho g = \text{constant}, \qquad (2)$$

where η is dynamic viscosity. The effects are dependent on the type of flow. The third term accounts for the constriction itself and the tube-approximation of the smooth irregularities of entrance and exit, all with Poiseuille-like flow. l_i is the length of the irregularity and v_i the mean flow velocity. The factor ζ in the fourth term is the resistance coefficient of an irregularity. The effects of narrowing and widening are very strongly dependent on the smoothness of the transition. The range of change of the narrowing is a factor of 10, with a rounded entrance (say angle of narrowing up to 20^o, $\zeta=0.05$ and abrupt, $\zeta=0.45(1-\beta)$, as extremes. Here, β is the contraction coefficient: β = (smaller cross section area)/(larger cross section

area). For a widening $\zeta=(1/\beta - 1)^2$. Then, for a surface ratio of 2 ζ is 1, yielding a pressure drop $\Delta P = -½v^2/g$. This shows that exit effects can be very large. Moreover, it works over a long distance before parabolic laminar behavior is restored (see Entrance effect and entrance length).
A practical equation to calculate the pressure drop over a vascular stenosis with turbulent flow is the approximation:

$$\Delta P = 8\pi l v \eta / O_1 + \zeta v^2 \rho (O_1/O_2 - 1)^2, \qquad (3)$$

where v is the velocity before the stenosis, l the total stenosis length and ζ is about 1.5.

HR_{max}

Principle

HR_{max} is the maximum heart rate (HR) that a person should achieve during maximal physical exertion. Research indicates that it is most closely linked to a person's age; a person's HR_{max} will decline during live. The speed at which it decline over time is related to fitness: the more fit a person is, the more slowly it declines.
HR_{max} is measured during periodically increase of the intensity of exercise (for instance, when a treadmill is being used through increase in speed or slope of the treadmill, or by interval training) until the subject can no longer continue, or until certain changes in heart function are detected in the ECG (at which point the subject is directed to stop). Typical durations of such a test range from 10 to 20 minutes. There are many prediction equations for HR_{max}. One of the most precise prediction equations, published in a meta-analysis (ref. 1), is:

$$HR_{max} = 208 - 0.7 \times age \text{ (beats/min)} \quad (1),$$

where age is in years.
It is independent of sex.
It is slightly too high for people performing frequently endurance sport (a decrease of ¾ point per hour endurance sport/week with a maximum of 10 points). The equation yields probably too high values for people older than 65 years. (A reduction of 1 point for every year above 65 seems likely). People who have participated in sports and athletic activities in early years will have a higher HR_{max} than those less active as children.

Application

HR_{max} is utilized frequently in the fitness industry, specifically during the calculation of target heart rate (ref. 2) when prescribing a fitness regimen.

References

1 Tanaka H, Monahan Kevin D and. Seals DR Age-predicted maximal heart rate revisited2001J Am Coll Cardiol 37, 153-156.
2 Wikipedia

Navier-Stokes equations

Principle

The Navier-Stokes equations are a set of equations that describe the motion of fluids (liquids and gases, and even solids of geological sizes and time-scales). These equations establish that changes in momentum (mass x speed) of the particles of a fluid are simply the product of changes in pressure and dissipative viscous forces (friction) acting *inside* the fluid. These viscous forces originate in molecular interactions and dictate how *sticky* (viscous) a fluid is. Thus, the Navier-Stokes equations are a dynamical statement of the balance of forces acting at any given region of the fluid.
They are one of the most useful sets of equations because they describe the physics of a large number of phenomena of academic and economic interest. They are useful to model weather, ocean currents (climate), water flow in a pipe, motion of stars inside a galaxy, flow around a wing of an aircraft. They are also used in the design of aircraft and cars, the ***study of blood flow***, the design of power stations, the analysis of the effects of pollution, etc.
The Navier-Stokes equations are partial differential equations which describe the motion of a fluid, so they focus on the *rates of change* or *fluxes* of these quantities. In mathematical terms these rates correspond to their derivatives. Thus, the Navier-Stokes for the simplest case of an ideal fluid (i.e. incompressible) with zero viscosity states that acceleration (the rate of change of velocity) is proportional to the derivative of internal pressure. Poiseuille's Law and Bernoulli's equation are special cases of 1D Navier-Stokes.
The fluid motion is described in 3D space, and densities and viscosities may be different for the 3 dimensions, may vary in space and time. Since the flow can be laminar as well as turbulent, the mathematics to describe the system is highly complex. In practice only the simplest cases can be solved and their exact solution is known. These cases often involve non turbulent flow in steady state (flow does not change with time) in which the viscosity of the fluid is large or its velocity is small (small Reynolds number).
For more complex situations, solution of the Navier-Stokes equations must be found with the help of numeric computer models, here called computational fluid dynamics.
Even though turbulence is an everyday experience it is extremely hard to find solutions for this class of problems. Often analytic solutions cannot be found.
Reducing the models to 1D, as is often done in fluid dynamics of blood vessels, makes the problem handsome, see e.g. http://www.math.ist.utl.pt/~mpf2005/abstracts/contributed_moura_vergara.pdf and ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-5052.pdf

Application

Theoretical medical applications in the vascular and airways system, the latter with compressible fluid dynamics.

More Info

Those who like to avoid e.g. the mathematical comprehensive Wikipedia paper on the Navier-Stokes equations can read the more friendly paper on ''Incompressible Navier-Stokes equations reduce to Bernoulli's Law'', http://home.usit.net/~cmdaven/navier.htm.
In this paper the incompressible Navier-Stokes vector-form equation, a nonlinear partial differential equation of second order (in dimensionless variables):

$$v_t + (v \cdot L)v = -Lp + \mu L^2 v + \rho g. \qquad (1)$$

(where v is a vector representing the velocity of an infinitesimal element of mass at a point in 3-D space, p is the scalar pressure at the same point, ρ is the mass density at the point and is assumed constant throughout the medium, μ is the viscosity of the medium, and g is a constant vector acceleration due to some constant external force on the infinitesimal element, usually taken to be gravity) can finally be reduced to Bernoulli's Law in a 4D (3D and time) vector form:

$$\tfrac{1}{2}\rho V^2 = P + \rho g Z + \text{Const} \qquad (2)$$

(where V is the analytic 4-D velocity, P is the 4D analytic vector pressure field, g is a constant acceleration which can be imposed in an arbitrary direction, and Z is a vector representing arbitrary displacement in 4-D space). By reducing this equation to one spatial dimension and taking time-invariant flow the traditional scalar Bernoulli's equation is obtained.

Modified after Wikipedia and http://home.usit.net/~cmdaven/navier.htm.

Pitot tube

Basic principle

A Pitot tube is an instrument to measure flow.

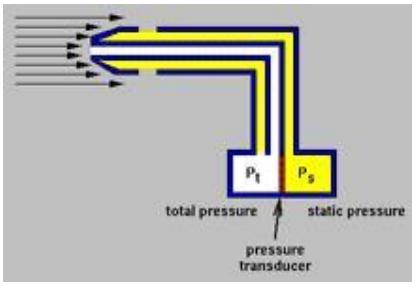


Fig. 1 Pitot tube

The basic instrument, see Fig. 1, consists of two coaxial tubes: the interior tube is open to the flow (i.e. opening perpendicular), while the exterior tube is open at ninety degrees to the flow (i.e. opening parallel). Manometers are used to measure the difference between these two pressures and using Bernoulli's equation the flow rate of the fluid or gas can be calculated.

More Info

The flow field outside the Pitot tube is everywhere homogeneous except around the entrance of the interior tube. In accordance with the law of Bernoulli (see Bernoulli's and Pascal's Law), in the flow field, so at al all stream lines at all positions, it holds that $\frac{1}{2}\rho v^2 + P = \text{constant}$. The term ρgh is omitted since a difference in height can be neglected. At the entrance the law also holds but there is no flow within the internal tube. The velocity pressure or propelling pressure (caused by the force of the fluid which tries to flow into the closed tube interior), being $\frac{1}{2}\rho v^2$, is added to P_{static}, and so:

$$P_{entrance} = \frac{1}{2}\rho v^2 + P_{static}.$$

$P_{entrance}$, the total pressure is also called the stagnation pressure. In the exterior tube, with an opening parallel to the flow, there is also no flow, so no velocity pressure. Here, only the P_{static} will be registered. $P_{entrance} - P_{static}$ is measured by the pressure transducer. When ρ is known v can be calculated.

Application

Pitot tubes are applied in the pulmonary pneumotachograph. Also it is applied in ultrasound flow measurement and fluid flow. Further it is widely applied in technology, like measuring relative velocity (not ground velocity) of a plane (ρ can be calculated by measuring P_{static} and outside temperature).

Poiseuille's Law

Basic Principle

A laminar flow within a pipe or tube, in medicine for instance a blood vessel or a tube of the airways system, has a velocity which increases with the distance from the wall of the tube, provided that the viscosity is everywhere the same. This increase is parabolic with the maximum of the parabola in the centre of the tube and consequently the velocity is there maximal (Fig. 1).

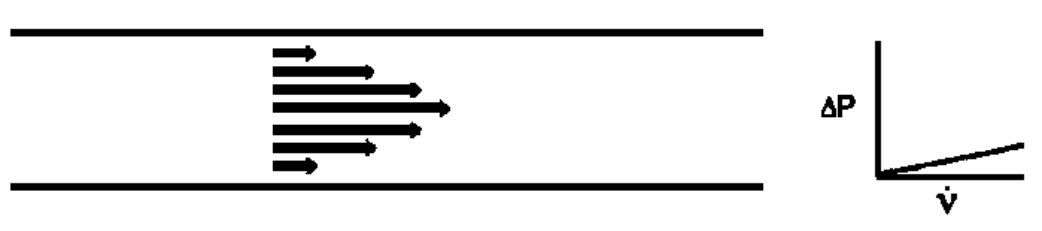


Fig. 1 Laminar flow in accordance with Poiseuille's Law.

Laminar flow is dependent on the viscosity of the gas, but not on its density (ρ). The quantity relating the pressure drop Δp over a certain length of the tube and the stationary volume flow $\dot{V}$ (in m^3/s) is the resistance R of the tube, analogue to the electric law of Ohm (V= i·R). It is given by:

$$\Delta p = \dot{V} \cdot R_{tube}, \qquad (1)$$

with Δp (Pa) the pressure difference over the length (m) of the tube. Δp is the driving 'force' to create flow. R_{tube} of an ideal tube with radius r (m) and length L is:

$$R_{tube} = 8 \cdot \eta \cdot L/(\pi \cdot r^4) \qquad (\text{in Pa/ } m^3). \qquad (2)$$

where η the dynamic viscosity (for air 17.1 and 19.0 x 10^{-6} Pa·s at 0 and 37 ^{o}C respectively, blood ca. 1 x 10^{-3} Pa·s). For a formal derivation of the law see the textbooks of physics.

Substituting 2) in 1) gives the law of Poiseuille:

$$\dot{V} = \pi \cdot \Delta p_{tube} \cdot r^4/(8 \cdot \eta \cdot L), \qquad (3)$$

The flow, as a function of the distance from the centre, the axis of the tube, is parabolic as illustrated in Fig. 1.

Application

Applied in the calculation of flow of blood through the vessels or heart (fluid mechanics of cardiovascular system) and the flow of air and expiratory gas through the airways (see Lung gas transport 1-3).

Some numeric examples
With extreme exercise (heart rate 200 beats/min, stroke volume 100 mL) the mean velocity in the aorta (diameter 3 cm), supposing laminar flow would be 0.47 m/s (or 1.7 km/h). At the axis, the velocity with laminar parabolic flow is exactly twice as much. With 0.47 m/s the current is in between laminar and full grown turbulent since the Reynolds number is about 4200. In the main arterial branches (0.1 cm diameter, with in total 3000 within the human body), Reynolds number is about 5, and consequently the flow is laminar.
A similar calculation holds for the trachea. With extreme exercise, the inspiratory volume of a top-endurance athlete is 200 L/min. With a diameter of the trachea of 2 cm, the peak maximal velocity at the axis supposing laminar flow would be 42.4 m/s (supposing that the duration of inspiration and expiration are equal with a square wave profile in time). However, with a mean velocity of 21.2 m/s this means that the Reynolds number is about 16500, and so turbulent and 8 times larger than laminar flows allows.

More Info

Bifurcations increase the resistance (see Flow in bifurcations). With an optimal angle and laminar flow, this is about 10%. It is caused by the change from an axial symmetric parabolic current profile to an

asymmetric laminar profile. The resistance increase can change the character of the flow: from laminar before the bifurcation to transitional after the bifurcation, and from transitional to turbulent. These so called entrance effects fade away after some distance (see Entrance effect and entrance length and also Lung gas transport 1, basic principles). So, with laminar, symmetric flow before the bifurcation, the profile is again axis symmetric after a distance of some five tube diameters
After bifurcations of arteries or arterioles, the susceptibility for plaque formation is substantially increased when the flow is not anymore laminar.
Constrictions, like that of the glottis or a stenosis (airways and blood vessels) also can change the character of the flow (see Flow through a stenosis).

Literature

Van Oosterom, A and Oostendorp, T.F. Medische Fysica, 2nd edition, Elsevier gezondheidszorg, Maarssen, 2001.

Rayleigh, Grashof and Prandtl Numbers

Principle

These three numbers have to do with heat transport by natural convection, laminar or turbulent, and not with heat conduction. They do not comprise current velocity. For forced convection (e.g. in case of wind), laminar or turbulent, these numbers are irrelevant. Then, Reynolds number and the Nusselt number are applied. For this, the reader is referred to technical textbooks.

Rayleigh number

In fluid mechanics (gases and liquids), the Rayleigh number is a dimensionless number associated with the heat transfer within the fluid. When the Rayleigh number is below the critical value for that fluid, heat transfer is primary in the form of conduction when it exceeds the critical value; heat transfer is primarily in the form of convection. The latter can be laminar or turbulent.

Grashof number

In natural convection the Grashof number plays the same role that is played by the Reynolds number in forced convection. The buoyant forces are fighting with viscous forces and at some point they overcome the viscous forces and the flow is no longer nice and laminar.

Prandtl Number

The Prandtl number is from a conceptual point of view the ratio of the thickness of the velocity boundary layer to thermal boundary layer. The former is the layer where the (macroscopic) particle velocity in the moving medium (e.g. air) varies from zero (at the surface of the e.g. the body) to that of the moving medium (the speed of air surrounding the body). The thermal boundary layer is the layer which comprises the full difference in temperature. When Pr=1, the boundary layers coincide.
Typical values of the Prandtl number are:

Material	Pr
Gases	0.7-1.0
Air 20 ^{o}C	0.71
Water	1.7-13.7
Oils	50-100,000

When Pr is small, it means that heat diffuses very quickly compared to the velocity (momentum). This means the thickness of the thermal boundary layer is much bigger than the velocity boundary layer.

Application

Medical applications are rare since heat transport by convection is seldom of interest. It can play a role in aerospace, environmental and sports medicine.

More info

The *Rayleigh number* is the product of Grashof which describes the relationship between buoyancy and viscosity within a fluid and Prandtl number, which describes the relationship between momentum diffusivity and thermal diffusivity.

For free convection near a vertical smooth surface, the Rayleigh number is:

$$Ra = Gr \cdot \mathrm{Pr} = \frac{g \cdot \beta}{\nu \cdot \alpha} \Delta T \cdot L^3$$

where
Ra = Rayleigh number

Gr = Grashof number
Pr = Prandtl number
g = gravity constant (N/kg)
β = thermal expansion coefficient (1/K)
ΔT = temperature difference between surface and quiescent temperature (K)
L = characteristic length (mostly effective height, for a plate it is the actual height) (m).
v = kinematic viscosity = η/ρ (= dynamic viscosity/density) (m^2/s)
α = thermal diffusivity = $\lambda/(\rho \cdot c_p)$ where λ is the heat conduction coefficient (W/(m·K), ρ the specific density (kg/m^3) and c_p the specific heat coefficient at constant pressure (J/(kg·K).

Free convection over a horizontal surface (plate) also the Nusselt number is relevant (the reader is referred to technical textbooks).

The *Grashof number* is defined as:

$$Gr = \frac{BuoyancyForces}{ViscousForces} = \frac{g \cdot \beta \cdot \Delta T \cdot L^3}{\nu^2}$$

For air of 20 °C, $Gr = 78 \cdot 10^9 \cdot \Delta T \cdot L^3$. For a vertical plate, the flow transitions to turbulent around a Grashof number of 10^9.

The *Prandtl number*, which is specially applied for heat transfer and which comprises some fluid properties, is defined as:

$$\mathrm{Pr} = \frac{\nu}{\alpha} = \frac{\eta \cdot c_p}{\lambda}$$

The Prandtl number is the ratio of momentum diffusivity (kinematic viscosity) to thermal diffusivity.

Literature
http://www.coolingzone.com/Content/Library/Tutorials/Tutorial%201/DNHT.html

Reynolds Number

Principle

The Reynolds number indicates whether the mass transport trough a tube (blood vessel, airways of ventilatory system, etc.) is laminar, for instance parabolic according to Poiseuille's law (Fig. 1a; see *Poiseuille's Law*) or turbulent (Fig. 1b).

The Reynolds number is defined as:

$$Re = 2v \cdot r \cdot \rho/\eta = 2\dot{V} \cdot \rho/(\pi \cdot r \cdot \eta) \qquad (1)$$

with v the mean velocity (m/s) in the tube, r the characteristic (see **More Info**, for circular diameters the radius) tube radius (m), ρ density (kg/m^3), η dynamic viscosity (Pa·s) and $\dot{V}$ the volume flow (m^3/s).

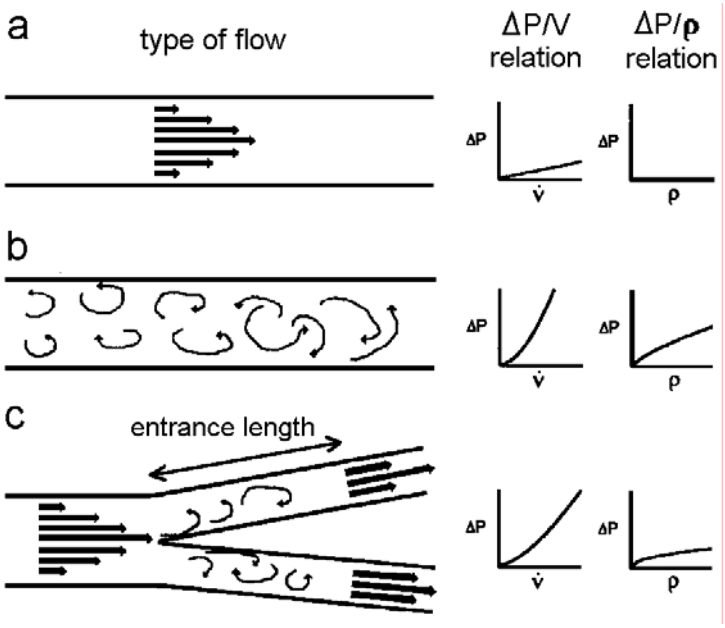


Fig. 1 Types of flow. See **Principle** and **More Info**.

Depending on the Reynolds number, three states of flow can be distinguished. When Re<2000, than the flow is laminar (Fig. 1a) and when Re>10000 the flow is turbulent (Fig. 1b). In between, there is a transitional state. There is a 4th state that occurs at the orifice of a tube in some cavity or after a bifurcation (Fig. 1c).
Fig. 1 visualizes the various types of flow. a) represents laminar flow, here parabolic according to Poiseuille's law. The $\Delta p/\dot{V}$ relation is a strait line with the slope equal to the resistance of the tube, R_{tube} (see Poiseuille's Law). b) presents turbulent flow. The $\Delta p/\dot{V}$ relation is a curved line and the resistance of the tube increases (nearly) quadratic with the flow. An increase in the specific density ρ of the medium (as can hold for an inspiring gas) asks for a higher Δp, the driving force of the flow, to obtain the same flow as with a lower density. This can easily be seen from the $\Delta p/\dot{V}$ relation and Δp/ρ relation. The velocity in the tube is nearly independent of the distance to the wall, so by approximation rectangular (at the wall itself, the velocity is zero). With transitional flow, the velocity is in between a parabola and a rectangle. c) represents the flow in case of a bifurcation. The $\Delta p/\dot{V}$ and Δp/ρ relation is in between those of laminar and turbulent flow. When the turbulent flow after the bifurcation changes to a laminar flow, this laminar flow starts with an axis-asymmetry before becoming axis-symmetric. This asymmetry increases the resistance only some 10%. When the flow is very low, the flow remains laminar along the total entrance length.
Many papers etc. address and model mathematically the flow in the vascular and airways system, also for bifurcations.

Application

Reynolds numbers are used to characterize the flow of blood through the vessels (fluid mechanics of cardiovascular system) and the flow of air and expiratory gas through the airways (see the various chapters about Lung Mechanics). Laminar flow can become turbulent after a stenosis. This substantially increases the resistance, for instance of the vascular system and so increases the load of the heart muscle.
When the diameter of a vessel with Poiseuille flow increases 19% with the same driving pressure, flow vresistance halves ($v \sim r^{-4}$; see Poiseuille's Law), $\dot{V}$ doubles and so does the Reynolds number. With this increase in diameter, the transition from laminar to turbulent occurs at twice the original volume flow. Such a diameter increase can easily be established in arterioles. This point is further addressed in Blood flow .

More info

Reynolds number may be interpreted as the ratio of two forces that influence the behavior of fluid flow in the boundary layer. For simplicity first we will restrict to a plate with characteristic length L. The two forces are the inertia forces and viscous forces:

(2)

$$\mathrm{Re} = \frac{InertiaForces}{ViscousForces} \approx \frac{\rho \cdot v^2 / L}{\eta \cdot v / L^2} = \frac{\rho \cdot v \cdot L}{\eta}$$

When the Reynolds number is large, the inertia forces are in command. Viscous forces dominate the boundary layer when the Reynolds number is small. When the number is low enough the flow over the whole plate is laminar, even when there are tiny local disturbances. As the inertia forces get bigger, the viscosity can no longer maintain order and these tiny disturbances grow into trouble makers and there is a transition to turbulent flow.
Another important quantity of the boundary layer that is influenced by the Reynolds number is its thickness. As the Reynolds number increases, the viscous layer gets squeezed into a smaller distance from the surface.
The value of Reynolds number beyond which the flow is no longer considered laminar is called the *critical* Reynolds number.
Calculation of the Reynolds number is easy as long as you identify the characteristic length and use the right velocity and a consistent set of units. For flow over a flat plate, the characteristic length is the length of the plate and the characteristic velocity is the free stream velocity. For *tubes* the *characteristic length is the tube diameter* and the *characteristic velocity is the average velocity* through the tube obtained by dividing the volumetric flow rate by the cross-sectional area (we are assuming that the tube is full, of course). For tubes with a non-circular cross-section, the characteristic length is the Hydraulic Diameter defined as:

$$L=4A/C, \quad (3)$$

where A is the cross-sectional area of the tube and *C* is the wetted perimeter.
For an ellipse with a and b the half long and short axes a good approximation (that of Ramanujan) of the circumference or perimeter *C* is:

$$C = \pi[3(a+b) - ((3a+b)(a+3b))^{0.5}] \quad (4)$$

It can easily be verified that for a circular tube the hydraulic diameter equals the tube diameter. For non-circular tubes the average velocity is used as the characteristic velocity. Suppose that a vessel or airway tube is compressed to half its diameter, whereas the perimeter remains the same and the new shape is elliptical, than the area is decreased by 16%. When the volume flow remains the same, then according to (1) the Reynolds number increases by a 16%.
With turbulent (gas)flow, tube resistance $R_{tube,turbulent}$ is proportional with:

$$R_{tube,turbulent} \sim L^{-4.25} \quad (5)$$

Consequently, $R_{tube,turbulent}$ is increased by a factor 2.086.

The situation gets complicated when a tube system has many velocity and length scales such as the flow inside the vascular system or the airways system (see Lung gas transport, airways resistance and compliance).

Here are some numerical values of fluids and air:

η_{air} $17.1 \cdot 10^{-6}$ at 0^{o} C (Pa·s)
η_{water} $1002 \cdot 10^{-6}$ ·at 20^{o} C (Pa·s)
η_{blood} $2700 \cdot 10^{-6}$·at 37^{o} C, hematocrit ca. 40% (Pa·s), strongly dependent on hematocrit (with pathological values such as 60-70% ca. 10 mPa·s)
$\eta_{blood\ plasma}$ ca. $1500 \cdot 10^{-6}$·at 37^{o} C
ρ_{air} 1.29 kg/m^3 times 273/T (according to the Gas laws) with the temperature in Kelvin
ρ_{water} 998 kg/m^3 at 20^{o} C (for physiological temperatures independent of the absolute temperature T)

Stokes' law and hematocrit

Principle

Stokes' law describes the motion of a small spherical object in a viscous fluid. The particle is subjected to the downward directed force F_g of gravity and the buoyant force F_A, also known as the force exerted by Archimedes' law. The both forces are:

$$F_g = (4/3)r^3\rho_p g \quad (1a)$$
$$F_A = (4/3)r^3\rho_f g, \quad (1b)$$

where r is the radius (the Stokes radius, see **More Info**) and *g* is the acceleration due to gravity (9.8 m/s^2). The resulting force is:

$$F_r = F_g - F_A = (4/3)r^3 g(\rho_p - \rho_f) \quad (2)$$

The resulting force is directed downward when the specific density of the particle ρ_p is higher than that of the fluid ρ_f (e.g. a glass bead) and upward when it is smaller (a polystyrene sphere). Here it is supposed that the former case holds, and so the sphere will move downward.

As soon as the sphere starts moving there is a third force, the fric
tional force F_f of the fluid. Its direction is opposite to the direction of motion. The total resulting force is:

$$F_{tot} = F_r - F_f. \quad (3)$$

As long as F_{tot} is positive, the velocity increases. However, F_f is dependent on the velocity. Over a large range of velocities the frictional force is proportional to the velocity (v):

$$F_f = 6\pi r\eta v \quad (4).$$

where η is the dynamic fluid viscosity. Expression (4) is Stokes' law.
After some time the velocity does not increase anymore but becomes constant. Then equilibrium is reached. In other words, F_r is canceled by F_f and so $F_{tot} = 0$. From now on the particle has a constant velocity. The equilibrium or setting velocity v_s can be calculated from (2), (3) and (4) with $F_{tot} = F_r - F_f = 0$. The result is:

$$v_s = (2/9)r^2 g(\rho_p - \rho_f)/\eta, \quad (5).$$

The proportionality of v_s with r^2 means that doubling the radius gives a reduction of a factor of four for the setting time.
Equation (5) only holds under ideal conditions, such as a very large fluid medium, a very smooth surface of the sphere and a small radius.

Application

The law has many applications in science, e.g. in earth science where measurement of the setting time gives the radius of soil particles.

Blood cells
In medicine, a well known application is the precipitation of blood cells. After setting, on the bottom of a hematocrit tube are the red cells, the erythrocytes, since they are large and have the highest density. In the middle are the white cells, the leucocytes despite their often larger volume. However, they are less dense and especially less smooth, which slows their speed of setting. On top, and hardly visible, is a thin band of the much smaller platelets, the thrombocytes. The relative height of the band (cylinder) with the red cells is the hematocrit. Although the red cells are not spherical and the medium is not large at a all (a narrow hematocrit microtube) the process still behaves rather well according to the law of Stokes. Red cells can clotter to money rolls, which sett faster (of importance for the hematologist).
The equivalent radius of a red blood cell can be calculated from the various parameters: heamatocrit (assume 0.45 L/L), setting velocity (0.003 m/hour = $0.83\cdot10^{-6}$ m/s), density of red cells (1120 kg/m^3) and plasma (1000 (kg/m^3). Everything filled out in eq. (5) yields r = 3.5 μm. Actually the red cell is disk-shaped with a radius of about 3.75 μm and a thickness of 2 μm. When the volume is by approximated by a cylinder (the inner thickness is less, but the edge is rounded), the equivalent sphere diameter is

2.76 µm. The too large estimate from Stoke's law is due to the strongly deviating shape of the red cell from the sphere, causing a lower setting velocity.

Centrifugation
An important application is the process of centrifugation of a biochemical sample. The centrifuge is used to shorten substantially the setting time. In this way proteins and even smaller particles can be harvested, such as radio nucleotides (enrichment of certain isotopes of uranium in an ultracentrifuge). With centrifugation, the same equations hold, but the force of gravity g should be replaced by the centrifugal acceleration *a*: $a = 4\pi^2 f^2 R$, where f is the number of rotations/s and R the radius of the centrifuge (the distance of the bottom of the tube to the center). In biochemistry, *a* can easily reach 10^4 g and in physics even 10^6 g.

More Info

More formally, Stokes' law is an expression for the frictional force exerted on spherical objects with very small Reynolds numbers (e.g., very small particles) in a continuous fluid with some viscous by solving the small fluid-mass limit of the generally unsolvable Navier-Stokes equations.
The Reynolds number holds for a liquid (liquid or gas) flowing in a pipe, channel etc., but also for an object flowing (blood cell, air craft) in a fluid. This directly follows from the definition of the Reynolds number:

$$\text{Re} = \text{inertial force/frictional force} = \rho v_s L/\eta = v_s L/\nu, \qquad (6)$$

where v_s is velocity, L the characteristic length and ν the kinematical fluid viscosity ($\nu = \mu/\rho$).
With volume flow of a fluid, v_s is the mean flow velocity and for motion of an object in a fluid it is the velocity with respect to the fluid.
For flow in a circular tube, L is the diameter, and for a spherical object it is $2r_h$. The radius r_h is the Stokes radius or hydrodynamic radius. When a hard object is spherical it is (practically) its actual radius. More precisely, the hydrodynamic radius comprises solvent (hydro) and shape (dynamic) effects. It is in between half the largest diameter (the effective or rotational radius) and the equivalent radius, the radius of a sphere with the same volume and weight. Knowing r_h, the diffusion coefficient (D) in the fluid can be calculated. However, calculating r_h itself is not possible.
Going back to the Reynolds number, it appears that Stokes' law holds rather well when Re of the object (particle) is less than or of order 1.

Some typical values of Reynolds number:

Spermatozoa ~ 1×10^{-2}
Blood flow in brain ~ 1×10^{2}
Blood flow in aorta ~ 1×10^{3}
Onset of turbulent flow ~ 2.3×10^{3}(pipe flows) - 10^{6} (boundary layers)
Person swimming ~ 4×10^{6}
Aircraft ~ 1×10^{7}
Blue Whale ~ 3×10^{8}

When an object falls from rest, its velocity v(t) is:

$$v(t) = \frac{mg}{b}(1 - e^{-bt/m}) \qquad (7)$$

where *b*, which is the drag coefficient. v(t) asymptotically approaches the terminal velocity $v_t = mg/b$.
For a certain *b*, heavier objects fall faster.
Stokes law assumes a low velocity of a small falling particle. Stokes drag has a coefficient b equal to $b = 6\pi r\eta$. This is the coefficient used in equation (4). The derivation of b is easy for the parameters r and η, but the problem is the factor 6π. A factor 2 π is caused by a pressure effect and a factor 4π by friction, (For a derivation see ref. 1 or 2).
For example, consider a small sphere with radius *r* = 1 µm moving through water at a velocity v of 10 µm/s. Using 10^{-3} as the dynamic viscosity of water in SI units, we find a drag force of 0.2 pN. This is about the drag force that a bacterium experiences as it swims through water.

Reference
G. K. Batchelor, An Introduction to Fluid Dynamics, Cambridge, 1967, 2000.
www.aemes.mae.ufl.edu/~uhk/STOKESDRAG.pdf

Womersley number

Principle

The largely inviscid and heavy pulsatile character of the flow in large blood vessels and airways is expressed in the Womersley number α, defined as:
$\alpha = r^2 \cdot \omega \cdot \rho / \eta$
where r^2·is the tube radius, ω·the frequency if the oscillations of the flow, ρ the density and η the dynamic viscosity. The physics of pulsatile flow can be derived from the Navier-Stokes equations.
The Womersley number can be considered as the pulsatile version of the Reynolds number. With high numbers (> 7) inertia dominates, yielding a rather well blunted or flat flow front. With low numbers viscosity dominates yielding parabolic-like flows, however skewed towards the outer wall. With α = 2 the flow profile is practically parabolic-like. Fig. 1 illustrates the various types.

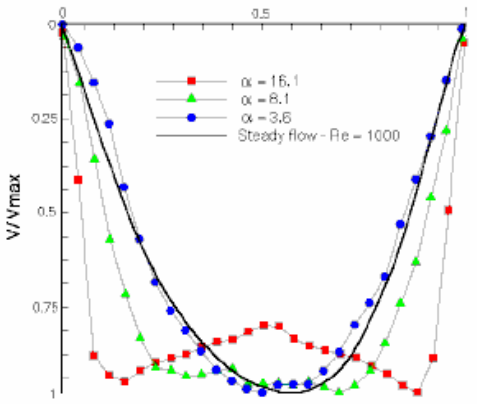


Fig. 1 Data of a model of the pulsatile flow profile of air in the trachea with various Womersley numbers compared to steady Poiseuille flow with Reynolds number Re = 1000. Relative scales. From http://www.vki.ac.be/research/themes/annualsurvey/2002/biological_fluid_ea1603v1.pdf

Application

Vascular and airways flow, described by the oscillatory flow theory. This theory is only relevant in the conduit tubes with there large diameters and turbulent flow. For the small tubes laminar flow holds.

More info

Since the heart does not generate a single sine wave but a series of harmonics (see Fourier analysis) the flow profile is found by adding the various harmonics. The relation between pressure drop and flow as given above is the so-called longitudinal impedance of a vessel segment.
In the periphery with small blood vessels or airways (small r) and little oscillation (α ca. 1), there is no need for the oscillatory flow theory and we can describe the pressure-flow relation with Poiseuille's law.
For the very large conduit arteries and airways, where α > 10, friction does not play a significant role and the pressure-flow relation can be described with inertance alone. For α values in between, the combination of the resistance plus the inductance approximates the oscillatory pressure-flow. Models of the entire arterial and airways system have indicated that, even in intermediate size arteries, the oscillatory flow effects on velocity profiles are mediocre. The main contributions of the tree to pressure and flow wave shapes are due to branching, non-uniformity and bending of the tubes. Thus for global transport phenomena, a segment of a tube can be described, in a sufficiently accurate way, by an inertance in conduit tubes, and a resistance in peripheral arteries and bronchioles and alveolar tubes. The oscillatory flow theory is, however, of importance when local phenomena are studied. For instance, detailed flow profiles and calculation of shear stress at the tube wall require the use of the oscillatory flow theory.

Gases and the Lungs

Adiabatic compression and expansion

Principle

To understand the concept of adiabatic phenomena, first Boyle's law is applied to the simple set-up of Fig. 1. It is assumed that both compartments have constant volume and temperature.

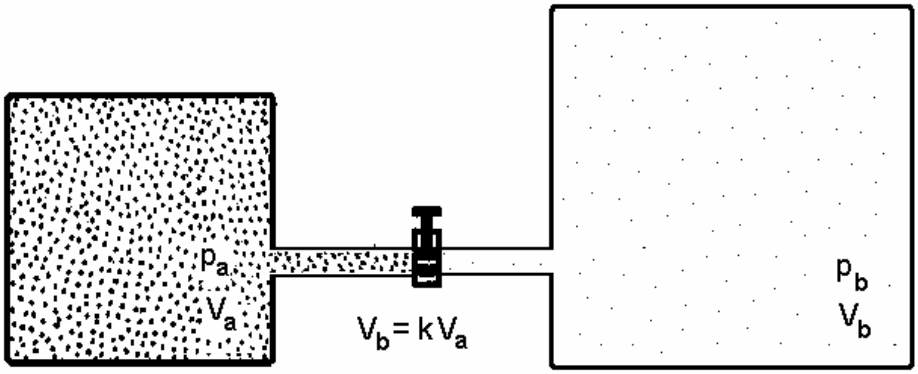


Fig. 1
The amount of gas in both compartments together is proportional $p_a \cdot V_a + p_b \cdot V_b$ which holds with the closed valve (at time zero) as well as after equilibration of the pressures (at time t) when the valve has been opened. So, it holds that :

$$p_{a,0} \cdot V_a + p_{b,0} \cdot V_b = p_t \cdot (V_a + V_b). \quad (0)$$

With given $p_{a,0}$, $p_{b,0}$, V_a and $V_{b=}$, p_t can be calculated. However, this is only true when T in the whole system remains constant, i.e. isothermic compression and expansion holds. In daily life, a bicycle pump becomes warm when a tire is inflated. This is due to heat transfer from the heated, compressed air in the pump. When the heat produced by compression is not given off to the environment but "absorbed" by the gas it self, or when the energy needed for expansion is not provided by the environment but provided by the expanding gas itself, than Boyle's law does not hold. So, a compressing gas heats and an expanding gas cools down. These processes are called adiabatic compression and expansion. The deviation from Boyle's law can be very substantial (Fig. 2), as will be explained conceptually with the example of Fig. 1.

When the valve has a very narrow opening, the process of equilibration takes so much time (tens of minutes) that the whole process is isothermic. There is enough time for heat release and uptake between the system and the environment. When the equilibration is very fast (about a second; pipe and valve diameter large) the process is practically adiabatic and the changes of temperature can rather well be calculated.

For ideal gases under pure adiabatic conditions the p-V relation is:
$p \cdot V^{\gamma}$ = constant, (1)
with γ a numerical value greater than 1 and depending on the type of gas. γ is the so-called c_p/c_V ratio, the ratio of heat capacity of the gas (c_p) with constant p and the specific heat capacity of the gas (c_V) with constant V. It depends on the number of atoms in the gas molecule. For mono-atomic gases, e.g. the noble gases, γ is 5/3. For diatomic gases, for instance air, γ is 7/5 (= 1.4). More precisely, c_p/c_V =is :

$$c_p/c_V = c_p/(c_p - R), \quad (2)$$

where R the universal gas constant (see Gas laws).

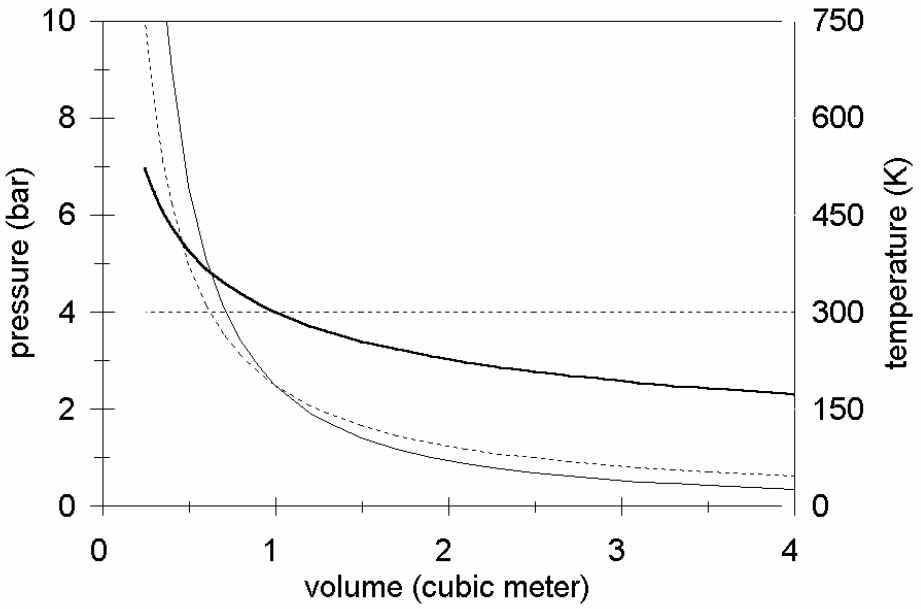


Fig.2 Isothermic and adiabatic p-V relation. The dashed curve gives the p-V relation according to Boyle's law and the dashed straight horizontal line the temperature belonging to it. The solid curve represents the adiabatic p-V relation and the thick solid curve gives the adiabatic temperature. n = 0.1 kmol.

As follows from Eq (1), reducing a volume by a factor k^1 gives a pressure increase much higher than k (being $k^{1.4}$). Conceptually, one can say that a factor k is due to the volume decrease according to Boyle and that the remaining factor, $k^{0.4}$, is caused by the temperature increase.
An isothermal p-V curve according to Boyle and an adiabatic p-V curve are depicted in Fig. 2. Numerically, the resulting temperature effect can be very substantial. A further description of adiabatic processes can be found in More info. Since textbooks seldom give examples how to calculate the adiabatic effect, an example is given in More Info.

Application

Adiabatic effects are seldom applied and in contrary, are generally unwanted. So, prevention is required. (Semi)-artificial breathing apparatus (home care pulmonary diseases, pulmonology, IC, surgery theatre) should have valves, tubes etc. designed in such a way that valves are not blocked by freezing and the breathing gas is not cooled down by adiabatic effects. Adiabatic effects also play a role in the technology of (clinical) hyperbaric chambers, high pressure gas tanks and diving regulators.
In the human (and mammalian) body), compression and expansion of gas filed cavities, adiabatic effects play now role. Even when one dives in the water from some tens of meters altitude, the temperature rise in the cavities is physiologically irrelevant since the process is sufficiently isotherm.
A medical imaging method, Optoacoustic Imaging is based on adiabatic processes.

More Info

The exponent γ is the c_p/c_V ratio, the ratio of heat capacity of the gas (c_p) with constant p and the specific heat capacity of the gas (c_V) with constant V. The complicated derivation can be found in physical textbooks. Here, a conceptual explanation is given.
The process of compression always has the same energetic cost, so independent of the velocity of the compression. This external energy is added to the intrinsic energy of the gas, which is actually the kinetic energy of the particles. The kinetic energy of the particles is proportional with the temperature. When the external energy is supplied instantaneously, it is completely added to the kinetic energy of the particles. When the compression is slow, then during the time of compression, the particles have the time to transfer a part of their kinetic energy to the vessel wall. The result is that the temperature of the gas raises less and the temperature of the wall increases. When the compression is extremely slow, the gas remains its temperature since all external energy is (indirectly) transferred to the environment. Now, the compression is isothermic and Boyle's law holds.
For instantaneous compression, the increase can be calculated as follows. Suppose that volume V with pressure p is compressed to V' yielding a new pressure p'. Applying Eq. (1) means that:

$$p \cdot V^{\gamma} = p'(V')^{\gamma} . \qquad (3).$$

Before compression and after compression the universal gas law (see Gas laws) holds, so that pV = nRT and p'V' = nRT' (T' is the new temperature), and some more calculation yields:
$\Delta T = T' - T = T((V/V')^{(\gamma -1)} - 1)$ (4)

This relation also holds for volume expansion.
In a similar way, for a given change of pressure it can be derived that:

$\Delta T = T((p/p')^{(\gamma-1)/\gamma} - 1)$. (5)

Example of calculation
Suppose, everything ideal, that a gas volume at 300 K (27 ^{o}C) is instantaneously reduced in volume by a factor 2.5 and that there is no heat transport. According to Eq. (4) this yields $\Delta T = 300 \times (2.5^{0.4} - 1) =$ 133 K. However, due to the temperature increase, a much higher factor of pressure increase is obtained (being $2.5^{1.4} = 3.6$) rather than the factor of 2.5 according to Boyle's law. Actually, a compression-factor of 1.92 ($= 2.5^{1/1.4}$) is required to obtain p = 2.5 bar. Then, the temperature increase becomes $300 \times (1.92^{0.4} - 1) = 90$ K.
The equation $p \cdot V^{\gamma} =$ constant works also the other way around. Suppose that the pressure p is increased instantaneously to $k \cdot p$ by a factor of k. Then, according to Boyle, the volume V reduces with a factor k, but now, the reduction is only a factor $k^{1/1.4}$. Again, in the pressure increase, a "Boyle factor" $k^{1/1.4}$ and a "temperature factor" $k^{0.4/1.4}$ are comprised. Of course, these two factors together have to yield k ($= k^{1/1.4} \cdot k^{0.4/1.4}$). Now, with k = 2.5 and T = 300 K the temperature increase is again 90 K ($2.5^{0.4/1.4} \times 300$ K).

In practice, it appears that the processes of compression and expansion are seldom purely adiabatic or purely isothermic. When the process is in between the calculation is complicated and, moreover in practice valves may block by freezing, for instance the valve of Fig.1. This may result in oscillatory behavior between blocking and unblocking by freezing and unfreezing.

Capnography

Principle

A capnograph is an instrument used to measure the CO_2 concentration of the expired air. Generally it is based on infrared (IR) analysis with a single beam emitter and non-dispersive IR absorption measurement with a solid state mainstream sensor and ratiometric measurement of red/IR absorption, similar as with oximetry (see Pulse oximetry).

Molecular Correlation Spectrography (MCS with IR)
Laser-based technology, i.e., MCS, is used to generate an IR emission that precisely matches the absorption spectrum of CO_2 molecules. The high emission efficiency and extreme CO_2 specificity and sensitivity of the emitter-detector combination allows for an extremely short light path which allows the use of very small sample cell (15 µL). This in turn permits the use of a very low flow rate (50 mL/min) without compromising accuracy or responsive time. This is in contrast to conventional CO_2 IR method, where the sampling flow rate is 150 mL/min.

Raman Spectrography
Raman Spectrography uses the principle of Raman scattering for $C0_2$ measurement. The gas sample is aspirated into an analyzing chamber, where the sample is illuminated by a high intensity monochromatic argon laser beam. The light is absorbed by molecules which are then excited to unstable vibrational or rotational energy states (Raman scattering). The Raman scattering signals (Raman light) are of low intensity and are measured at right angles to the laser beam. The spectrum of Raman scattering lines can be used to identify and quantify all types of molecules in the gas phase.

Mass Spectrography
The mass spectrograph separates molecules on the basis of mass to charge ratios (see Mass spectrography).

Applications

Capnography is widely used in clinical practice, such as pulmonology, anesthesiology and in the IC-unit. Mass spectrometers are quite expensive and bulky to use at the bedside. They are either "stand alone," to monitor a single patient continuously, or "shared," to monitor gas samples sequentially from several patients in different locations (multiplexed). Tens of patients may be connected to a multiplexed (with a rotary valve) and monitored every 2 or 3 min.

More Info

In addition to this ‘’classical’’ methods new techniques are photo-acoustic and magneto-acoustic technique for CO_2 monitoring in the open unintubated airway.

Acoustic impedance measurement (PAS) PAS is a still experimental approach (see ref. 1), based on an acoustic impedance measurement of the air with an electro-acoustic sensor coupled to an acoustic resonator. When IR energy is applied to a gas, the gas will try to expand which leads to an increase in pressure. The applied energy is delivered in pulses causing pulsation of the pressure. With a pulsation frequency lying within the audible range, an acoustic signal is produced. This is detected by a microphone. The impedance characteristic is depending on the sound velocity within the gas mixture contained in the acoustic resonator. The relation between the acoustic impedance and the CO_2 concentration is approximately linear. Since the sound velocity (and so the acoustic impedance) is also is dependent on temperature and humidity, the outcome should be corrected for these parameters. Potential advantages of PAS over IR spectrometry are:

- higher accuracy;
- better reliability;
- less need of preventive maintenance;
- less frequent need for calibration.

Further, as PAS directly measures the amount of IR light absorbed, no reference cell is needed and zero drift is nonexistent in PAS. The zero is reached when there is no gas present in the chamber. Despite being a superior method of measurement of CO_2, photoacoustic spectrography did not gain as much popularity as IR spectrography.

Literature

1. http://www.mrtc.mdh.se/publications/1006.pdf
2. http://www.capnography.com/Physics/Physicsphysical.htm

Compression and expansion

Principle

To understand the concept of diabatic phenomena, first Boyle's law is applied to the simple set-up of Fig. 1.

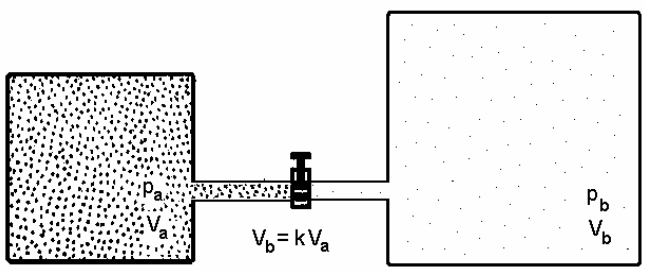


Fig. 1

The amount of gas in both compartments together is $p \cdot V_a + p \cdot V_b$ which holds with the closed valve (at time zero, t_0) as well as after equilibration of the pressures (at time t) when the valve has been opened. So, it holds that:

$$p_{a,0} \cdot V_a + p_{b,0} \cdot V_b = p_t \cdot (V_a + V_b). \quad (1)$$

With given $p_{a,0}$, $p_{b,0}$, V_a and V_b at t_0, p_t can be calculated. When $V_b = jV_a$ and $p_{b,0} = kp_{a,0}$ then p_t becomes:

$$p_t = p_a(1+jk)/(1+j). \quad (2)$$

However, this is only true when the temperature T in the whole system remains constant during the whole process of pressure equilibration, i.e. the process is isotherm or diabatic. When this is not the case, then the left compartment cools down and the right one heats up. For more info about this adiabatic process, see Adiabatic Compression and expansion.
Another condition is that the gas is ideal, i.e. the particles do not interact and they have no size. When pressures (so densities) are low, both conditions apply well and (1), which is actually based on Boyle's law (see Gas Laws), can be used. When they do not hold, the Van der Waals corrections are necessary.

Application

Innumerous in science, technology and so indirectly in medicine, e.g. in breathing apparatus and especially in pulmonology.
The Van der Waals corrections are used for mass calculations of commercials gases in high-pressure tanks e.g. applied in medicine, especially with the expensive helium.

More Info

The Van der Waals correction has two constants *a* and *b* (Table 1). The correction factor "*a*" is needed for the interaction between the gas particles (attraction coefficient), and a correction factor "*b*" for the volume occupied by the gas particles.

Table 1.

Molecule, atom or mixture	Van der Waals constants *a* $10^3 J.m^3/kmol^2$	*b* $10^{-3} m^3/kmol$
He	3,5	22
H_2	25	26
O_2	140	31
N_2	140	39
CO_2	360	44
H_2O	550	30.5
air	140	37.4

b of air interpolated from weighed *b*'s of O_2 and N_2

According to Van der Waals:

$$(p + an^2/V_m^2)\,(V_m - nb) = nRT, \qquad (3a)$$

where V_m the total volume and n the number of kmoles in the volume V. They increase and reduce pressure, respectively. Table 1 gives for some gases the numerical values of the two constants. When a and b are zero, then the van der Waals equation degenerates to Boyle's law. To calculate p, (3) can be rewritten:

$$p = nRT/(V_m - nb) - an^2/V_m^2, \qquad (3b)$$

With normal temperatures and pressures of about 100 bar or more, the Van der Waals correction makes sense. For instance, a tank filled with air 300 bar and 290 K comprises 7.8% less. (assumed that $a_{air} = 37.4 \times 10^{-3}$). At 200 bar the interaction effect dominates the particle-volume-effect, but at 300 bar the situation is reversed. The p-V diagram of Fig. 2 also illustrates the rather complex behavior of the correction. Since Boyle's law is independent of the type of particles, the straight line of Boyle (log-log diagram) holds for both air and helium.

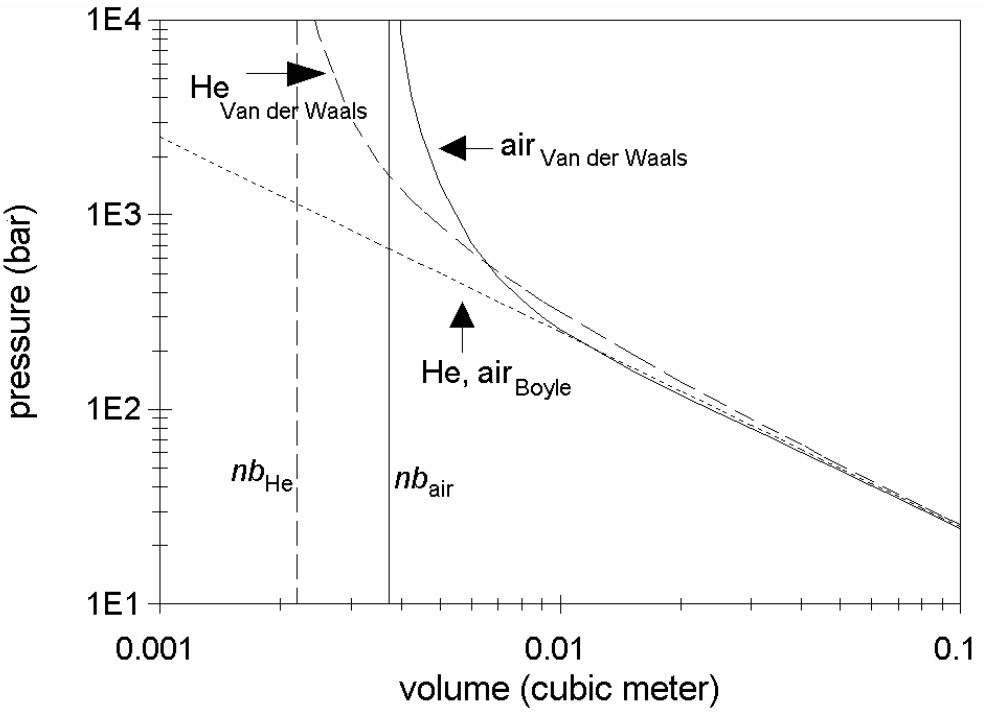


Fig. 2 Comparison of p/V relation according to the Law of Boyle and according to the Van der Waals equation for air and He. The curves are calculated for n = 0.1 kmol and T = 300 K (1E4 = 10^{14}).

The Van der Waals curve of He shows already strong deviations (about 10%) from Boyle's law at 100 bar, since the interaction effect is weak compared to the particle-volume-effect. The p-V curve approaches the straight Boyle line from above, but the air-curve first crosses the Boyle line and then approaches the line from below. For low pressures the Boyle line is the asymptote for the Van der Waals curves. The rather surprising behavior of the Van der Waals equation is due to the fact that it is a cubic equation in n and in V.
For very precise calculations in gas mixtures at pressures beyond 50 bar, this correction is even not precise enough. Then, the theory of virial coefficients, taking also into account the (first and higher order) interactions between the types of particles in a mixture (), is applied (see textbooks of physics).

Gas laws

Laws of Boyle, Gay-Lussac, Avogadro and Dalton, and the universal gas law

Basic principles

A gas is one of the three aggregation states of a substance, solid, fluid and gaseous. A gas is compressible and often a mixture. It occupies all available space uniformly and completely, has a small specific mass, diffuses and mixes rapidly, and is mono- (He), di- (N_2, O_2, CO), tri- (CO_2, O_3), or poly-atomic (NH_3, methane etc). In a so-called ideal gas, the particles (being molecules or atoms) have no size and do not influence each other since they show no mutual attraction (no cohesion).
Gas particles move at random with velocities of many hundreds m/s (Table 1). The mean particle-particle distance is in the nm-range. Gas volume is empty for ca. 99.9% and therefore the gas particles compared to liquids infrequently collide with each other. At 300 K and 1 bar the mean free path of H_2 is ca. 66 nm, more than 10 times than in the liquid phase. Collisions with constant temperature (isotherm) are pure elastic, also with a wall at the same temperature.
The gaslaws are those of *Boyle* (and Mariotte), *Gay-Lussac* (and that of Charles), *Avogadro* and *Dalton*, and the *Universal gas law*.

Table 1

Particle	diameter of particle nm	velocity v (273 K) m/s	km/h	molecular mass m	c_p/c_V ratio or γ *
He	0.027	1304	4690	4.003	1.66
H_2		1838	6620	2.016	1.41
O_2	0.034	461	1740	32.00	1.4
N_2	0.037	493	1860	44.011	1.29
H_2O(vapor)	0.027			18.016	1.33

* $c_p/c_V = c_p/(c_p\text{-}R)$ (see Boyle's law, The universal gas law and Compression and expansion).

Medical applications

These laws are fundamental for anesthesiology, the (patho)physiology of pulmonology, the medicine of diving, especially with respect of the occurrence of decompression sickness, hyperbaric medicine (HBO, the application of pure O_2 as breathing gas under *h*yper*b*aric conditions), aviation (also recreational types) and aerospace medicine, mountaineering. The law of Dalton is basic for the physiology of respiration.

More info

These laws can only be applied when there is heat exchange with the environment. Then the process is called isothermic or diabatic. In practice this seldom holds. For instance when a high-pressure helium tank is filled in the factory it raises in temperature (adiabatic compression). This compression is so fast that the produced heat has not the time to be transferred to the environment. For adiabatic (non-isothermic compressions or expansions see Adiabatic compression and expansion). The ideal gas laws assume an isothermic process and, self-evident with an invariable amount of gas mass.

The law of Boyle (and Mariotte)

In the derivation of Boyle's law it is assumed that the gas particles make elastic collisions with the wall surrounding the gas. The collisions exert a force on the wall, which is proportional to the number of particles per unit of volume, their velocity and their mass. Pressure (p) is defined as force (F) per area (A). By doubling the density of the gas (this is doubling the number of particles in a given volume), the number of collisions doubles, and hence the exerted force and so the pressure. Hence, the equation:

$$p_1 \cdot V_1 = p_2 \cdot V_2 = \text{constant} \quad \text{or}$$
$$p_1/p_2 = V_2/V_1, \text{ or } p = \text{constant} \cdot V^{-1}, \qquad (1)$$

is obtained. This law holds well for moderate gas densities (pressures < 300 bar, with regular temperatures) and under isotherm conditions, i.e. the process is diabatic. For higher pressures the condition that the total volume of the particles can be neglected and that the particles do not influence

each other does not hold any longer. The Law of Boyle is refined with the van der Waals corrections, see Compression and expansion.

The law of Gay-Lussac

This law states that the ratio of pressure and absolute temperature is constant provided that volume of the gas is constant:

$$p_1/ p_2 = T_1/T_2 = \text{constant}. \quad (2a)$$

It has been proved that the squared velocity $\langle v^2 \rangle$ is proportional with T and reciprocal to the gas mass:

$$\langle v^2 \rangle = 3RT/(N_A m), \quad (2b)$$

where $\langle v^2 \rangle$ the mean of the squared velocity of the particles, R the molar gas constant (= 8315 J/kmol·K), $N_A m$ the gas mass with N_A the Avogadro's number and m the particle mass.
Conceptually, the correctness of the law can be understood by realizing that $\frac{1}{2} \cdot m \cdot v^2$ is kinetic energy of a particle. So, for a certain type of particle an increase in T gives an increase of v^2 and consequently of p. When p and *n* are constant it holds that:

$V_1/T_1 = V_2/T_2 = \text{constant}$, *the law of Charles*. (2c)

The universal gaslaw (of Boyle and Gay-Lussac)

This is a combination of the law of Boyle and the law of Gay-Lussac:

$$pV = n \cdot R \cdot T, \quad (3)$$

with *n* the number of kmoles of the gas and with R the molar gas constant (= 8315 J/kmol·K).
It holds that $p \cdot V = \frac{1}{3} \cdot N_A \cdot m \cdot \langle v^2 \rangle = \text{constant}$ and that $\langle v^2 \rangle = 3RT/(N_A m)$. After substitution of $\langle v^2 \rangle$ and applying this for *n* moles the law follows.

The laws of Avogadro

$$V_1/n_1 = V_2/n_2 = \text{constant} \quad (4)$$

Since equal volumes of ideal gasses at equal pressure and equal temperature comprise an equal number of particles, the law follows directly.

The law of Dalton

The pressure of a mixtures of gasses is the sum of the pressure of the individual gasses (defined as the partial pressures) since the kinetic energy ($\frac{1}{2} \cdot m \cdot \langle v^2 \rangle$) of all types of particles, irrespective their type, is the same: $m\langle v^2 \rangle = 3RT/N_A = \text{constant}$ (see law of Gay-Lussac). So:

$$p_{total} = p_1 + p_2 + p_3 \ldots. \quad (5)$$

Gas volume units, STPD, BTPS and ATPS

Since the mass of gas in a unit volume is dependent on pressure and temperature (see Gas laws) they have to be specified with their pressure and temperature.
Three quantities are used.

- STPD: Standard Temperature and Pressure Dry, so at 1 atm (760 mm Hg), 0^o C and $p_{HO2} = 0$;
- BTPS: Body Temperature Pressure Saturated, defined at 37^o C, ambient pressure and saturated with water vapor ($p_{H2O} = 6.3$ kPa or 47 mm Hg);
- ATPS: Ambient Temperature Pressure Saturation, so at ambient temperature and pressure, with p_{HO2} saturated.

Respiratory volumes are usually reported at BTPS. On the other hand, moles of gas (i.e. O_2, CO_2 production) are commonly reported at STPD.
Sometimes volume measurements are performed neither at STPD nor at BTPS, but at ATPS.
Conversion among these conditions can be easily performed by applying the Gas laws for ideal gases, with volume proportional to absolute temperature T, and inversely proportional to pressure P.
As an example, if a spirometer is used to measure the tidal volume (V_{tidel}) of a subject in a mountain region where barometric pressure P=450 mm Hg, and the spirometer is at 23^o C (knowing that the pressure of water vapor at saturation at 23^o C is 21 mm Hg, and that at a body temperature of 37^o C is 47 mm Hg) ATPS can be converted to BTPS. For convenience pressures are all in mm Hg, so recalculation in bar or Pascal is not required.

$$V_{BTPS}/V_{ATPS} = ((T_{BTPS} / T_{ATPS})) \cdot ((P_{ATPS}) / P_{BTPS}))$$
$$= ((273+37)/ (273+23)) \cdot ((450-21)/(450-47))$$

hence,

$$V_{tidel,BTPS} = 1.1149\ V_{tidel,ATPS}\ .$$

Similarly, a BTPS volume can be converted to a STPD volume:

$$V_{STPD} = ((P_{BTPS} -47)/760) \cdot (273/310) \cdot V_{BTPS} = 0.00116 \cdot V_{BTPS}$$

When $P_{BTPS} = 760$ mm Hg, then:

$$V_{STPD} = ((760-47)/760) \cdot (273/310) \cdot V_{BTPS} = 0.826\ V_{BTPS}$$

Hot wire anemometry

Principle

In the hot wire anemometry an electrically heated wire (or screen or film) is placed in the gas pathway, which is cooled by the gas flow (Fig. 1). The degree of cooling depends on the gas flow rate, which can thus be calculated.

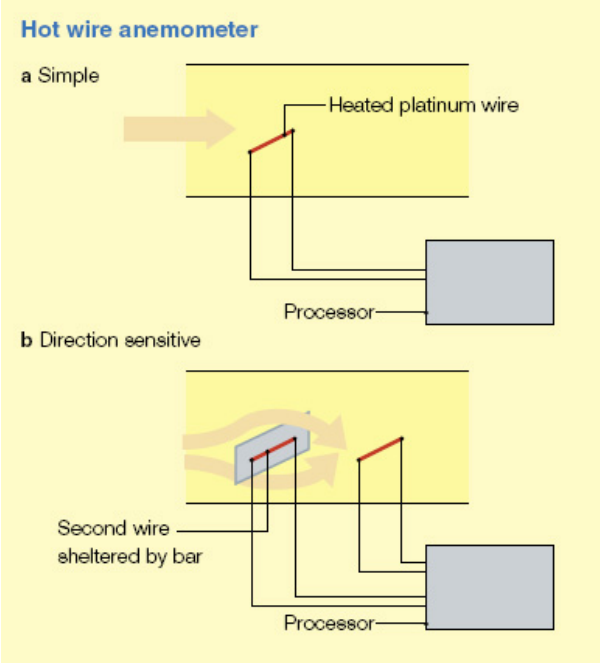


Fig. 1 (from http://www.frca.co.uk/article.aspx?articleid=100390)

Application

Especially in pulmonology, also neonatal, and IC unit.

More Info

The hot wire (usually Pt) is incorporated into a balanced Wheatstone bridge circuit. Cooling the wire changes its resistance and unbalances the bridge. Most designs work on the constant temperature principle, whereby a correcting current is applied through the hot wire to compensate for the cooling effect of the gas, maintaining a constant wire temperature and thus restoring the balance in the Wheatstone bridge. This current is measured and from it the gas flow rate is determined. To compensate for changes in the gas temperature, a second wire is usually incorporated, which is maintained at ambient temperature. Minor corrections are also made according to the gas composition to accommodate the variation in specific heat capacity. Hot wire anemometry is generally extremely accurate.

The cooling effect occurs with flow in either direction, and so to measure exhaled tidal volume the hot wire anemometer is placed in the expiratory branch of the circuit. It can be modified to provide information about the direction of flow by using an additional heated wire placed just downstream from a small bar, as shown in Fig. 1b. This bar shelters the wire from the full cooling effects of flow in one direction but not the other, and thus inspiratory and expiratory flows can be calculated separately. For this purpose the sensor must be placed in the Y-mouth-piece before the bifurcation. This technique is particularly useful to control neonatal ventilation.

Lung gas transport 1, basic principles

Theorie of gas flows in the airways system
Pulmonary gas transport is afflicted with a huge number of physiological and clinical variables. Possibly this is a reason for its step motherly position in the medical curriculum.
Streams of gas and liquid in pipes behave on a gliding scale from orderly, i.e. laminar, to chaotic, i.e. turbulent, with a transitional stage in between; see Fig. 1 of Poiseuille's Law. Types of flow are classified according to Reynolds number, Re. For Re < 2200 the flow is laminar and for Re > 10000 the flow is turbulent. For laminar flow in circular tubes Poiseuille's Law holds (parabolic velocity profile).
The airways can be described as a system of many-fold bifurcating tubes as illustrated in Fig. 1.

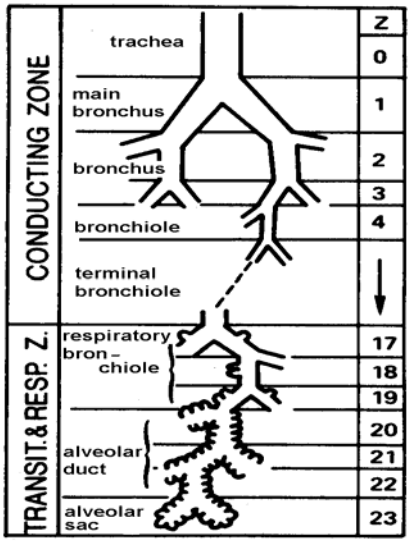


Fig. 1 Weibel model of airways with its generation numbers indicated (see text). Conduction zone: transport; transition zone: transport and gas exchange; the respiratory zone (alveoli) gas exchange. (Modified from Weibel and Gil, 1977).

The trachea is called the 0^{th} generation, the bifurcation of the trachea is of the 1^{st} order and the two main bronchi are the 1^{st} generation. The trachea always has the highest Re and the glottis still more. Even for highest inspiratory volumes (RMV, Respiratory Minute Volume, > 120 L/min), the high generations have laminar flow.

Basic lung gas transport
The driving force of volume flow is the transpulmonal pressure difference Δp_{tp}, between the pressure in the tissue directly surrounding the lung and the pressure in the mouth cavity. For simplicity flow is considered as stationary and not pulsatile (by approximation sinusoidally) as holds for normal breathing. Generally it is measured in the esophagus proximal of an intra-esophageal inflated balloon. When the pressure needed to inflate the alveoli is subtracted, the resulting difference Δp varies from 100 Pa (= 1 cm H_2O; RMV about 6 L/min) to 800 Pa (high flows) and under special conditions (e.g. airways disorders) it can go up to about 3000 Pa. Δp is distributed over the whole bifurcating network from mouth up to the last generation.

Application

The Weibel model is widely applied for flow calculations. It is useful to obtain insight in flow behaviour with disorders in the upper and lower airways of patients of all ages.

More Info

The driving force of volume flow (i.e. volume/time) $\dot{V}$ in a tube of length L is the pressure difference Δp_{tube} over L. ($\dot{V}$ = 2·RMV/60 in L/s, the factor 2 is often, but not always, already implied in the equations.) When Poiseuille's law holds, then the flow-velocity v increases linear with Δp_{tube}. Laminar flow decreases linear with the gas viscosity (but is independent of its density ρ). Therefore, the

resistance of laminar flow is called the viscous resistance. Note that laminar flow is not always parabolic. Poiseuille's Law only holds for parabolic flow.

With $\Delta p_{trachea}$ = 1 Pa, Poiseuille's law gives a $\dot{V}_{trachea}$ of about 70 L/min. Is this flow still laminar? $Re_{trachea}$ appears to be 5560 (at 35 ^{o}C), so the flow is in between laminar and turbulent. Therefore the actual $\dot{V}_{trachea}$ is considerably lower. Imperfections (non-smooth wall, slightly varying diameter, curvature etc.) further diminish $\dot{V}_{trachea}$. However, most effective is the constriction of the glottis causing substantial pressure loss and non-laminar behavior due to the constriction effect (see Flow through a stenosis) and the entrance effect (see Entrance effect and entrance Length). In conclusion, even for low flows ($\dot{V}$ = 0.3 L/s; nearly rest) Poiseuille's law is not valid in the trachea. A higher Δp than given by Poiseuille's law is needed to obtain this flow.

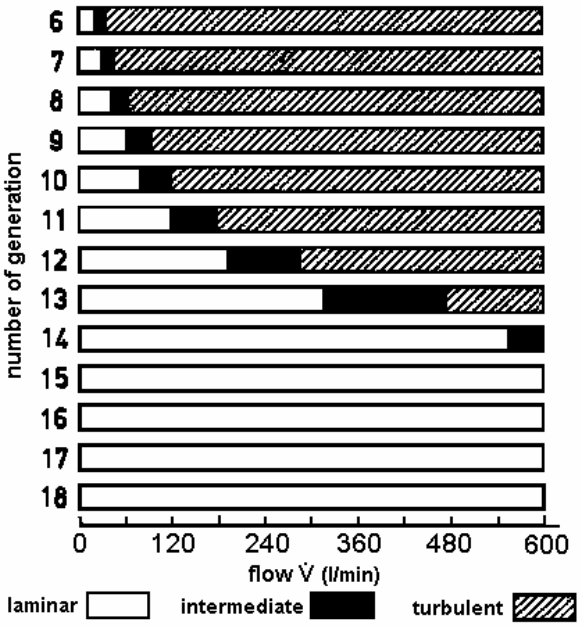


Fig. 2 Types of flow in the airway generations (after Hrnčíř, 1996, with increased generation numbers by 5-9 depending on the flow, to account for the entrance effects). A flow at the maximum of ventilation of 600 L/min can be reached by an endurance top-athlete during short lasting extreme exertion.

Also bifurcations disturb Poiseuille flow (see Flow in bifurcations). Heavy disturbance can give intermediate or turbulent flow.

After the entrance length, dependent on the diameter of the daughter branch and its Re, in the daughter branch the flow regains a laminar profile. Subsequent orders of bifurcations are so close together that for the upper generations the current fails to change to laminar before the next bifurcation is reached. However, from a certain critical generation laminar flow is always realized.

For ½ L/s (as with walking) this is at about generation 10. The two effects, the entrance phenomenon and the flow-profile dependency on Re, results in a critical generation with a lower generation number the higher the flow. For the lowest flow (0.2L/s) this is about generation 5 and for the highest flow (10 L/s) generation 15.

All together, the P-V relation of the airways system is non-linear. The non-laminar behavior implies that Δp is proportional to V^e with 1<e<7/4. This is discussed in Lung gas transport 2, pressure-flow relation.

References

Hrnčíř, E. Flow resistance of airways under hyperbaric conditions. Physiol. Res., 45, 153-158, 1996. 1962.

Weibel E.R. and Gil J. Structure-function relationships at the alveolar level. In: Bio-engineering aspects of the lung, West, J.B. (ed), Lung Biology in health and Disease, Vol. 3, Marcel Dekker Inc., New York and Basel, 1977

Lung gas transport 2, pressure, volume and flow

Principle

The volume-pressure relationships

$\dot{V}$ (L/s) is the time derivative of volume, V. V and $\dot{V}$ are measured with a spirometer, generally as function of time (see Spirometry). Pressures are measured with a barogram, as function of time (Fig. 1). The peak-peak value of the barogram and spirogram, dynamically obtained (quite breathing), are Δp_{tp} and V_{tidal} (=0.5 $\dot{V}$ /respiratory frequency) respectively. Combining both, the V-P relation is obtained (Fig. 1, right), that is not linear but elliptic caused by non-equal airways resistances of in- and expiration. The expiratory airflow is limited by dynamic airway compression, the downstream distally narrowing of the airways up to a *choke point* or equal pressure point. But the main cause of the hysteresis is pneumatic behaviour of the alveoli. During one breathing cycle the elliptic V-P loop is completed. With a high breathing frequency the ellipse becomes rounder and with a high V_{tidal} larger.
The V-P relation can also be statically measured with forced ventilation, generally with anaesthetised subjects (Fig. 2). With a high V_{tidal} the resulting 'loop' becomes flatter and strongly sigmoid when V_{tidal} = TLC-RV (total lung capacity minus residual volume).
Another diagram obtained with spirometry is the V- $\dot{V}$ diagram obtained during maximal forced expiration. Its maximum provides insight in the mechanical properties of the airways and its shape about flow limitations.
In the P-V diagram, the dashed axis of the ellipse gives the dynamic lung compliance ($C_{L,dyn}$) and the surface of the ellipse the work done to overcome the viscous airways resistance R_{aw}.

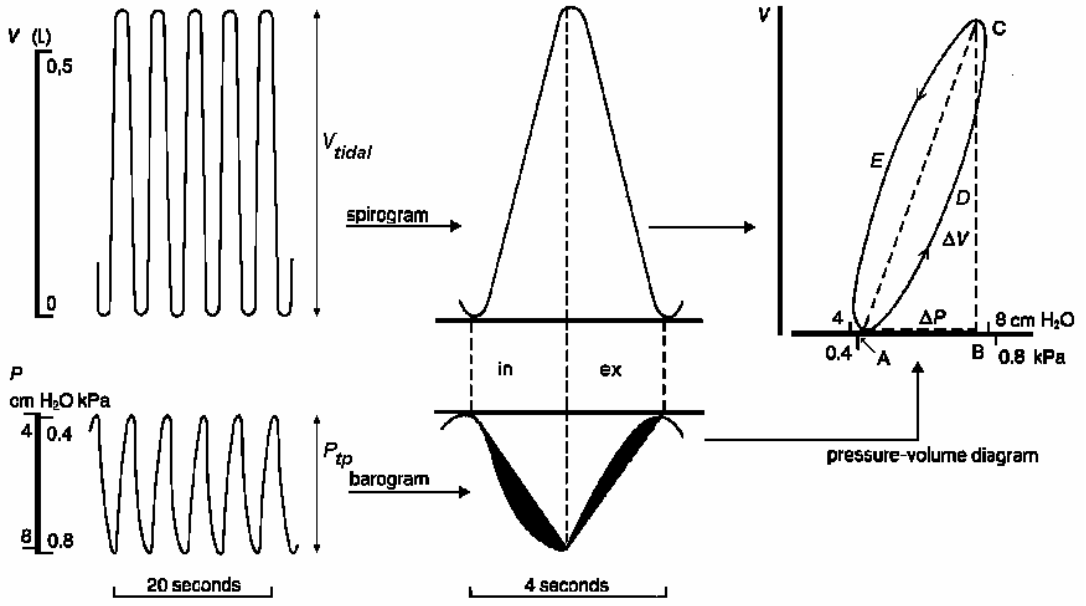


Fig. 1. The black area in the middle lower diagram accounts for the elliptic shape of the P-V loop (right diagram). It gives the dynamic lung compliance $C_{L,dyn} = \Delta V/p\Delta$. ($P_{tp}$ is Δp_{tp}).

Application

Performing heavy work requires an RMV of about 37.5 L/min. This is nearly 7 times the value in rest (RMV = 5.6 L/min, or $\dot{V}_{O2}$ = 0.32 O_2/min). The supposed Δp is 600 Pa, and consequently Δp_{tp} is about 1000 Pa when the effect of the compliance (mainly alveoli) is added. This value can be maintained for a long time.
What does this means for (mild) allergic asthma? Suppose, that (as a model) due to (very) cold inspired air there is an acute constriction of 15% of the bronchioles and smaller airways. Then **R_{aw}** (the airways resistance) increases by a factor of about 1.6 ($\approx 0.5 + 0.5x0.85^{-19/4}$, see Fig.1 and More Info of Lung gas transport 3, resistance and compliance), yielding a Δp of 1060 Pa. Adding the requested pressure to overcome the elastance **E** (the resistance of the alveoli to be inflated), Δp_{tp} becomes about 1500 Pa. This is much too fatiguing to maintain e.g. running and it has a high risk of escalation (more constriction and hypercapnia).

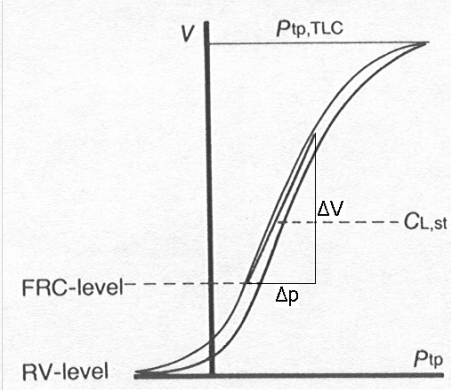


Fig. 2 P_{tp}-V diagram recorded under static conditions (passive ventilation). $\Delta p_{tp} = p_{mouth}-p_{oesophagus}$. $C_{L,stat}$ (static lung compliance) = $\Delta V/\Delta p$. FRC functional residual capacity, TLC total lung capacity, RV the residual volume, all in L.

Chronic smokers (>10 cigarettes/day, > 15 years) have poor gas transport since their $\mathbf{R}_{aw}$ is at least 25% higher (resulting in $\Delta p \approx 1100$ Pa). They have many bullae, which, according to the changed surface tension, decreases the elastance **E**. However, it is mostly found that **E** is increased since the alveolar tissue stiffness is substantially increased. They have a reduced alveolar volume. Conclusion: chronic smoking and endurance sport are pulmonary incompatible.

The shape of the hysteresis loop of the static and dynamic p-V diagram is diagnostically of great importance. In healthy lungs, $C_{L,dyn}$ halfs when the breathing frequency (f) changes from 10/min to maximal (>40/min). The static narrow sigmoid P-V diagram changes to pear-shaped with f-30/min and V_{tidal} maximal.
With disorders the shape of the loop is more complicated, especially with bronchoconstriction and dilatation induced by drugs.
With emphysema and normal breathing, the static Δp-V loop is shifted to left and for maximal ventilation the loop is wider. TLC and static C_L are increased and Δp_{tp} decreased. With bronchodilatation all these effects are reversed. Pulmonary fibrosis causes a wider and more horizontal loop and so an increased p_{tp} and decreased TLC and $C_{L,stat}$.

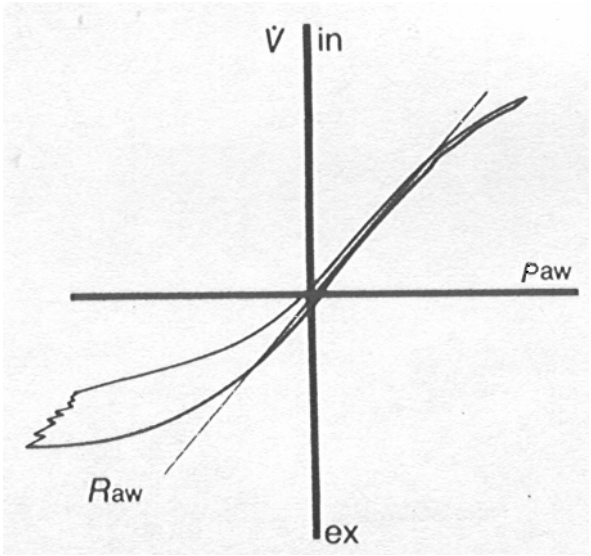


Fig. 3 p_{aw}-$\dot{V}$ diagram with $p_{aw} = p_{mouth}$—$p_{alveolar}$--. The latter is measured with body Plethysmography.

Another diagnostically important diagram obtained with Spirometry is p_{aw}-$\dot{V}$ diagram (Fig. 3), measured during quiet breathing. It is useful to find abnormal changes of R_{aw} (= $d(\dot{V})/d(p_{aw})$).

In general, obstructive and restrictive disorders have different outcomes for the various volume measures, as Table 1 shows.

Table 1

	Obstructive	restrictive
FEV1/FVC ratio	↓	↑
FEV1	↓	↓
FVC	↓	↓
TLC	↑	↓
RV	↑	-
FRC	-	↓

More Info

The airways have a mixed laminar and turbulent behaviour, yielding the classical equation:

$$\Delta p = \text{resistance times current} = (K_1 + K_2 \cdot \dot{V}) \dot{V} = K_1 \cdot \dot{V} + K_2 \cdot \dot{V}^2, \quad (1)$$

with the laminar term linear and turbulent term quadratic with $\dot{V}$. K_1 represents the laminar flow-resistance. Simplified, it comprises the viscosity of the gas, and the dimensions of the airways. K_2 accounts for the extra resistance due to irregularities of the walls of bronchi etc., to constrictions and bifurcations, resulting together in turbulent flow. The behaviour of the alveoli with their surface tension and the counteracting effect of the alveolar surfactant (see Surface Tension) are not yet incorporated. In addition to this, there is a small force caused by the elasticity (the stiffness to prevent strain, for hollow organs the elastance **E** (see Compliance (hollow organs)) of the tissue of alveoli and bronchioli. Surface tension and elasticity together constitute the static elasticity **E** of the system. For a further treatment of **E** see Lung gas transport 3, airways resistance and compliance. The semi-theoretical approximation with **E** implied is given by the equation:

$$\Delta p_{tp} = \mathbf{E} \cdot V_{tidal} + K_1 \cdot \dot{V} + K_2 \cdot \dot{V}^2. \quad (2)$$

(The product $\mathbf{E} \cdot V_{tidal}$ has the well-known electric analogue $Q/C = V$, with Q the charge of a capacitor and analogous to V, *C* the capacitance (compliance) of a capacity and analogous to 1/**E**, and V the electric potential, analogous to pressure.) Since Eq. (2) is very sensitive for the value of **E** it is inappropriate to calculate the constants or Δp_{tp}. Ignoring **E** of airways and alveoli, and with low, constant flows (or slowly changing), such that the effects of inertia can also be ignored, analogue to the electric law of Ohm ($V = i \cdot \mathbf{R}$), it holds that:

$$\Delta p = \mathbf{R}_{aw} \cdot \dot{V}, \quad (3a)$$

where $\mathbf{R}_{aw}$·the airways resistance. With turbulent flow, tube resistance is proportional to $\dot{V}^{3/4}$ (see Lung gas transport 3, airways resistance and compliance). This reduces (3a) to:

$$\Delta p = k \cdot \dot{V}^{7/4}. \quad (3b)$$

Including mouth and glottis, the constant k is nearly 120, and replacing by RMV, the equation becomes:

$$\Delta p' = 0.3 \cdot RMV^{7/4}. \quad (3c)$$

For intermediate flow the exponent is in between 1 and 7/4. Eq. (3b) is theoretically a better alternative for (1).

Literature

Nunn, J.F. Applied Respiratory Physiology. 2nd edition. Butterworths, London-Boston, 1977.
Tammeling, G.J. and Quanjer, Ph.H. Contours of Breathing, Boehringer Ingelheim BV, 1978, Haarlem.

Lung gas transport 3, resistance and compliance

Principle

The ventilatory system is characterized by a resistance, an inert (mass) component and a compliance. The latter is 1/stiffness = 1/elastance=1/**E**. **E** is measured in kPa/L. A high elastance counteracts inflation of the lungs. The airways, the alveoli, the lung tissue and chest wall all contribute. The inert airways and alveolar component is always neglected. The same holds for the lung tissue and wall inertia up to moderate ventilation.
Resistances and compliances can be measured and modelled, but both approaches are not trivial.

Resistance of the airways tree
Empirically, $\mathbf{R}_{aw}$ is related to the total lung capacity TLC (in litre) and age (year) according to Tammeling and Quanjer (1978):

$$\mathbf{R}_{aw} = 410((0.0021A+0.4)TLC)^{-1}, \qquad (Pa{\cdot}L^{-1}{\cdot}s) \tag{1}$$

where TLC is 10.67L – 12.16 (male) with L the length (m). $\mathbf{R}_{aw}$ can also be measured from the dynamically recorded P_{aw} - $\dot{V}$ curve (see Lung gas transport 2, pressure, volume and flow). For a subject of 40 years and L=1.75 m this yields a $\mathbf{R}_{aw}$ of 130 $Pa{\cdot}l^{-1}{\cdot}s$.
Finally, $\mathbf{R}_{aw}$ can be estimated from the Weibel model of the airways system (see Fig. 1 of Lung gas transport 1, basic principles) by adding $\mathbf{R}_{tube}$ of all tubes together (see **More Info**). The airways tree appears to have a resistance **R** of 29 $Pa{\cdot}L^{-1}{\cdot}s$ Since $\mathbf{R}_{aw}$ is proportional to $\dot{V}^{3/4}$, $\mathbf{R}_{aw}$ becomes:

$$\mathbf{R}_{aw} = k{\cdot}\mathbf{R}{\cdot}\dot{V}^{3/4} = k{\cdot}29\,\dot{V}^{3/4}. \tag{2a}$$

With k is 4 (see More Info) and rounded $\mathbf{R}_{aw}$ is:

$$\mathbf{R}_{aw} = 120\,\dot{V}^{3/4} \tag{2b}$$

For $\dot{V}$=1 L, (2b) is close to that found with the empirically equation (1). Chest wall resistance is about twice $\mathbf{R}_{aw}$ (lung tissue resistance can be neglected), which brings total pulmonary resistance at about 400 $Pa{\cdot}L^{-1}{\cdot}s$, about the mean of literature data (Nunn 1977).

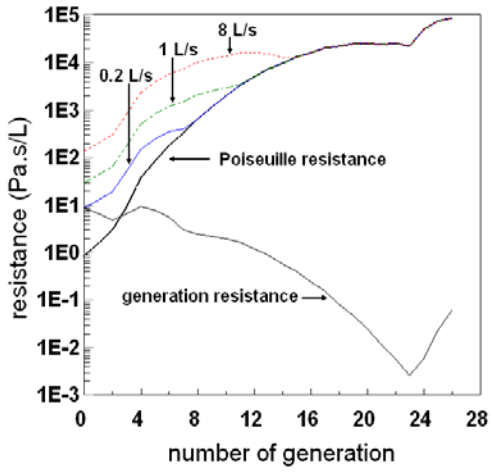


Fig. 1 The curves indicated by flows are the resistances of individual branches according to (3a) and the Poisieulle resistance according to (3b). The generation resistance, obtained by dividing the tube resistance by the number of tubes in a single generation, holds for $\dot{V}$=0.2 L/s (awake, in rest). $\dot{V}$=1 L/s and $\dot{V}$=8 L/s are equivalent to normal and extreme endurance sport respectively. Generation 0 is the trachea.

Compliance of the airways system

The determining factors for the alveolar compliance are the surface tension of the alveoli (and alveolar ducts) and the counteracting effect of the alveolar surfactant (see SurfaceTension).
The healthy $C_{L,dyn}$ is about 1.8 and 1.0 L/kPa of lung and chest wall respectively. $C_{L,dyn}$ can be measured from the dynamic p_{tp}-V diagram. These values are about 20% larger than the static compliances.
The measurement of $C_{L,stat}$ is done passively (anaesthetized subjects or cadavers) with aid of the static P_{tp}-V diagram (see Lung gas transport 2, pressure, volume and flow). At an age of 25 year, $C_{,stat}$ (lung and chest wall together) is 2.7 L/kPa. It decreases logarithmically to 1.8 L/kPa at 75 year (*Respiration* 1989;55:176-80). This is in line with the age dependent increase of the bulk modulus K (see Elasticity 3: compressibility and bulk modulus) and shear modulus G (Elasticity 2: shear modulus) (*J Appl Physiol* 89: 163–168, 2000). The age effect increases with p_{tp}. The change of K and G is mainly an alveolar wall property. Mostly it is found that smoking decreases $C_{L,stat}$.

Application

$\mathbf{R}_{aw}$ can be measured with several methods, such as body Plethysmography. Constrictions in the alveolar ducts (emphysema) considerably increase $\mathbf{R}_{aw}$. Many other disorders affect the lower airways, as chronic bronchitis, bronchiectasis, cystic fybrosis, asthma and bronchiolitis. Generally they produce such severe constrictions that the $\mathbf{R}_{aw}$ increases substantially. $\mathbf{R}_{aw}$ of smokers is about 20% enlarged. The age effect of K and especially G considerably attributes to the increase of stiffness of lung parenchyma. Lungs with chronic obstructive pulmonary disease and lung emphysema with a1-antitrypsin deficiency have an increased K. The effect is consistent with the behavior found in other organs, such as systemic arteries.

More Info

Modelling $\mathbf{R}_{aw}$
With turbulent flow, i.e. a high $\dot{V}$, the tube resistance $\mathbf{R}_{tube}$ is (modified after Clarke and Flook, 1999, for 1 bar air pressure):

$$\mathbf{R}_{tube} = 3.2 \cdot 9.9 \cdot L \cdot D^{-19/4} \cdot \dot{V}^{3/4} \qquad (\text{Pa} \cdot \text{L}^{-1} \cdot \text{s}) \qquad (3a)$$

where L and D (diameter) in cm and $\dot{V}$ the stationary flow in L/s. The constant 9.9 comprises the dynamic viscosity η (18.8 µPa.s at 310 K) and ρ (1.14 g/L). The constant 3.2 comprises the effect of entrance length (=0.06Re·D) and tube imperfections (not perfectly cylindrical, round and smooth, and with curvatures). For Poiseuille flow, i.e. a low $\dot{V}$, $\mathbf{R}_{tube}$ is:

$$\mathbf{R}_{tube} = 1.2 \cdot 0.76 \cdot L \cdot D^{-4} \qquad (3b)$$

The transition from weak turbulent to laminar flow occurs at the generation number where the coloured curves of Fig. 1 merge the Poiseuille resistance curve. The lower curve indicates that only the first, say 10, generations (the upper airways) contribute substantially to $\mathbf{R}_{aw}$.

Due to the complicated geometry of the system, numeric results of modelling of $\mathbf{R}_{aw}$ only give rough approximations. $\mathbf{R}_{aw}$ can be found by summation of all tube resistances $\mathbf{R}_{tube}$ over all generations for a given $\dot{V}$. In all generations there is pressure drop. The resistance of the airways tree is:

$$\mathrm{R} = \mathrm{k} \cdot 2^{5/2} \cdot \pi^{-7/4} \cdot \eta^{1/4} \cdot \rho^{3/4} \cdot \sum_{i=0}^{i=26} (L_i \cdot D_i^{-9/4} \cdot (v/2^i)^{3/4}) / 2^i , \quad (4)$$

where k a constant, η the dynamic viscosity, ρ the gas density and i the generation number. $v/2^i$ (= $\dot{V}/2^i$) is the flow per tube and 2^i the number of tubes in the i-th generation. The nominator 2^i is implied since in the i-th generation 2^i tubes are parallel. L_i and D_i can be obtained from the Hansen Ampaya non-scaled model with downscaled tube numbers for generation 24-26 (since paths did already end in sacs). The first five generations contribute most to the resistance. For $\dot{V}$ = 1 L/s, and all constants, lengths and diameters completed, **R** is about 29 $\text{Pa} \cdot \text{l}^{-1} \cdot \text{s}$ with k=1.
Under the assumption that some bifurcations are missing and due to tube imperfections **R** is larger. Also mouth and glottis should be included. Especially constriction and bifurcations (entrance effects) substantially increase **R**. Finally the flow is not stationary but dynamically (sinusoidal breathing). All together, this result in a correction factor k estimated at 4.

Literature

Clarke JR and Flook V. Respiratory function at depth, in: The Lung at Depth, Lundgren C.E.G. and Miller J.N. (Ed), Marcel Dekker Inc., New York, Basel, pp 1-71,1999.
Nunn JF, Applied Respiratory Physiology, 2nd ed. Butterworths, London-Boston, 1977.
Tammeling, GJ and Quanjer, PhH. Contours of Breahting, Boehringer Ingelheim BV, 1978, Haarlem.

Lung gas transport 4, cost of breathing

Principle

Theoretically a well as experimentally, the cost of respiration P_{res} is hard to estimate due to the non-linear behaviour of the resistance of the airways-alveolar-long tissue-chest wall system. Empirically it can be performed indirectly from O_2 consumption and from the area of the hysteresis loop of the V-p_{tp} diagram (see Fig. 1, 2 of Lung gas transport 2, pressure, volume and flow). As a rule of thumb it holds that $P_{res,rest}$ =0.1 Watt and at MVV, the Maximal voluntary ventilation (maintained for ca. 15 s) 0.5-1.0 Watt/breath. With a maximal breathing frequency (ca. 60/min), the maximal P_{res} is 30-60 Watt (e.g. Hesser et al. 1981). These estimates hold for adult, healthy subject up to an age of about 60 year.
The direct application of the general law 'power ≡ resistance times current^2 ' is the basis of two models. There are various models to estimate the resistance. Model [1], the classical one, divides the airways resistance $\mathbf{R_{aw}}$ in a constant and a term linear dependent on $\dot{V}$: $\mathbf{R_{aw}} = K_1 + K_2 \cdot \dot{V}$ (see Lung gas transport 2, pressure, volume and flow). In model [2] airways resistance is $\mathbf{R_{aw}} = k \cdot \dot{V}^{3/4}$. Model [3] utilizes the before mentioned histeresis loop of the V-p_{tp} diagram

When the visco-elastic and plasto-elastic losses in the lung tissue and chest wall tissue are taken into account, although hard to estimate, the total cost P_{res} is assumed to be a factor k' higher (twice for low flow to about triple for high flow) than $P_{vis,aw}$. The three models yield the following.

Model [1]

$$P_{res} = 0.06 + 0.0011 \cdot RMV^2 + 8.8 \cdot 0^{-6}\, RMV^3 \qquad (k'=3) \qquad (1a)$$

This equation is recalculated from empirical measurements (Milic-Emili and D'Angelo, 1997) and RMV calculated from $\dot{V}$. The constant 0.06 was added to obtain 0.1 W at rest.

Model [2]

$$P_{res} = 0.1 + 31 \cdot 10^{-6} \cdot RMV^{11/4} \qquad (k'=3) \qquad (1b)$$

The 0.1 was added to obtain 0.1 W at rest. To obtain this equation the viscous cost of the airways was multiplied by 3 to account for other resistive losses.

Model [3]:

This model is imprecise since the shape of the loop changes considerably with V_{tidal} and frequency. Actually, viscous wall cost should be implemented too. Total cost is the area ADCEA of the above mentioned Fig. 1, multiplied by ca. 3.0 to account for viscous wall losses. Approximations systematically yield too low values (about factor of 5). Experimentally, the reported values are mostly determined by the integral of $\Delta p \cdot \dot{V}$, which can give useful results.
Fig. 1 gives the results of models [1] and [2].

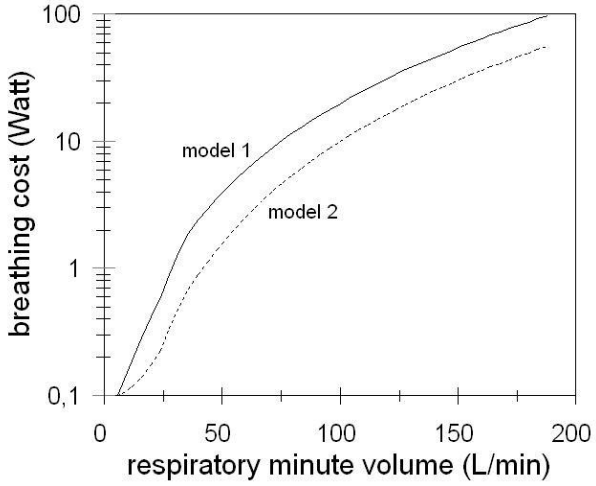


Fig. 1 Breathing cost of subject with normal height (175 cm) and weight (75 kg).

Application

Fig. 1 shows that for healthy lungs the cost is irrelevant except for the artificial condition of maximal breathing without other exertion and for the highest RMVs (nearly 10% of total energy expenditure). For most patients in (nearly) rest, the cost of breathing is such low that it can nearly always be afforded. However, moderate and high RMV can be impossible.
A low temperature of the air results in a higher airway resistance (especially for asthma patients). The reduced MVV reduces the maximal exertion. Only in extreme cases (e.g. nearly blocking of the airways) it can become a limiting factor. The same holds for smokers, but here the main cause is their reduced diffusing capacity of the lung epithelium. Acute respiratory distress syndrome (ARDS; inflammation of the of the lung parenchyma) of adults and Infant respiratory distress syndrome (IRDS; lack of surfactant) gives rise to a severe increase of breathing cost. These patients may die due to exhaustion.

More Info

Model [1] The power of breathing with strict laminar conditions is $P_{res} = \Delta p \cdot \dot{V}$. For the airways system with its non-laminar flow, the classical approach is to substitute Δp by $\mathbf{E} \cdot V_{tidal} + K_1 \cdot \dot{V}_{in} + K_2 \cdot \dot{V}_{in}^2$ (eq. (2) in Lung gas transport 2, pressure, volume and flow), and doubling the cost due to the expiration. With a low V_{tidal}, the "elastic energy" stored during inspiration in the alveoli is used during the expiration (elastic recoil), although there is always some loss. This can be minimized by chosen an optimal breathing frequency and V_{tidal} for a given RMV. With RMV is 14 L/min, the optimal V_{tidal} is close to 1.2 litre. Completing the constants, P_{res} becomes (see Milic-Emili and D'Angelo, 1997):

$$P_{aw} = 3.95 \cdot \dot{V}_{in}^2 + 1.95 \cdot \dot{V}_{in}^3 \text{ (W)} \qquad (2a)$$

Replacing $\dot{V}_{in}$ (in L/s) by RMV and accounting for the other resistances, (2a) becomes (1a).
Model [2] An approach via the general law 'power ≡ pressure times current' is

$$P_{aw} = 120 \cdot 1000^{1.75} \cdot \dot{V}^{1.75}, \qquad (3a)$$

obtained by substitution of Δp by $\Delta p = 120\, \dot{V}^{3/4}$ (Δp in Pa s/L and $\dot{V}$ in L/s) and converting L in m^3. Accounting for the other losses (k'=3), adding the constant 0.1 W and expressing $\dot{V}$ in RMV (L/min) P_{res} becomes (1b).
Model [3] Breathing cost can also be calculated with a more theoretically based model that also implies the tissues surrounding the lung (Tammeling and Quanjer, 1978). When a volume, e.g. a spirometer or a plastic back is filled with air, the work E is proportional with p·V (Joule). E is comprised of two parts, a viscosity (or resistance) part E_{vis}, given by the viscosity of the gas, and an elastic part (reciprocal to the compliance *C*) E_{el}. Both comprise a component of the thoracal wall E_W and of the airways system, E_L. Total work E for one inspiration then becomes:

$$E_{ins} = E_{W,vis} + E_{W,el} + E_{L,vis} + E_{L,el} \qquad (4a)$$

For $E_{L,el}$ it holds that:

$$E_{L,el} = \tfrac{1}{2} V_{tidal} \cdot \Delta p_{L,el}\,, \qquad (4b)$$

with $\Delta p_{L,el}$ the pressure to overcome the elastance of the alveoli and lung tissue during quit breathing. Eq. (4b) is actually the area of the triangle ABCA of Fig. 1 of Lung gas transport 2, pressure, volume and flow. However, $E_{L,el}$ is regained when the expiration takes place. The same recoil holds for $E_{W,el}$, which mainly comprises the compliance of the chest wall. At maximal airflow or during the early part of the expiratory maneuver, the main driving pressure is the intrinsic elastic recoil of the lungs. What remains for the cost are both viscous parts. $E_{L,vis}$ is the area of the hysteresis loop of Fig. 1. For very large flows the cost of the $P_{L,el}$ of inspiration and that of expiration do not longer cancel; there is a net loss which increases progressively with V and the same holds for $P_{W,el}$.

References

Hesser CM, Linnarsson D, Fagraeus L. Pulmonary mechanisms and work of breathing at maximal ventilation and raised air pressure. J Appl Physiol. 1981, 50:747-53.
Milic-Emili j and D'Angelo E, Work of breathing. In: The Lung, Crystal RG, West JB, Barnes PJ and Weibel ER, Lipencott-Raven, Philadelphia, New York, 1437-1446, 1997.
Tammeling, GJ and Quanjer, PhH. Contours of Breathing, Boehringer Ingelheim BV, 1978, Haarlem.

Oxygen analysis

Principle

Oxygen is the most vital gas for almost all biological species and a key-element in many chemical reactions. There are many physical and chemical principles to measure O_2 in a gas mixture and dissolved in fluid.
In gas mixtures the following techniques are most common:

- Paramagnetic principle The analyzer measures the paramagnetic susceptibility of O_2.
- High temperature zirconia sensor A type of electrochemical analyzer.
- O_2 analysis using electrochemical cells The electrochemical O_2 sensor operates more or less like a battery.
- Coulometric Based on the same principle as conventional electrochemical analyzers.

In liquids the following techniques are most applied:

- Hersch cell It is based on electrolytic chemical reactions.
- Polarographic A galvanic process with O_2 permeating through a membrane.
- Optical fluorescence In this fluorescence application the presence of O_2 slows or reduces the amount of fluorescence.
- Absorption spectroscopy The sensor uses laser diode absorption spectroscopy in the visible spectrum.

See **More info** for a detailed description.

Application

Numerous in experimental medicine, pulmonology, clinical physiology, cardiology etc., space, environmental and sport medicine and in biochemistry, industrial microbiology etc.

More Info

In gas mixtures

Paramagnetic principle
The analyzer measures the paramagnetic susceptibility of O_2 (see Paramagnetism, diamagnetism and magnetophoresis) in the sample gas mixture by means of a magneto-dynamic measuring cell. The physical property which distinguishes O_2 from other gases is its much higher paramagnetism compared to other common gases. Advantages are the fast response time, high flow rates (up to 1 L/min), small sample volumes (2 mL), the extremely low drift, the absolute linearity and the negligible cross sensitivity against other sample gas components.
The measuring cell with a small mirror at its center is mounted in a strong inhomogeneous magnetic field. The paramagnetic O_2 strengthens the forces on the diamagnetic measuring cell and causes a shifting which is detected by a system consisting of light beam, mirror and a photo cell.
A compensation current is induced via the feedback coil on the measuring cell that leads to a zero voltage. The required current is linearly proportional to the O_2 concentration.

High temperature zirconia sensor
A zirconia O_2 sensor uses a stabilized zirconia ceramic. Platinum electrodes are painted on the outside and inside of the O_2 sensor. The sensor is heated above 600^{o} C. At this temperature, the crystal lattice structure expands allowing O_2 ions to migrate through the O_2 sensor. Oxygen breaks down into O_2-ions at the electrodes of the sensor and travels through the sensor between the electrodes. The flow of O_2-ions is either outward or inward of the O_2 sensor depending on the O_2 concentration in the sample gas compared to the O_2 concentration in a reference gas. Advantages of this type of O_2 sensor are a wide measuring range and fast response. The sensor gives mV output, which is converted to $\%O_2$.

O_2 analysis using electrochemical cells
The electrochemical O_2 sensor operates much like a battery. Oxygen gas flows by an electrode and becomes a negatively charged hydroxyl OH-ion. This ion moves through an electrolyte in the O_2 sensor to a positively charged electrode typically made of lead. The OH-ion reacts with Pb and releases electrons. The electron flow is measured and be converted to an O_2 concentration. Advantages of this type of O_2 sensor include ability to measure O_2 in hydrocarbon or solvent streams, accurate, and inexpensive. Low O_2 measurement down to 0.01ppm is possible.

Coulometric The non-depleting coulometric cell (referring to Coulomb) operates on essentially the same principle as conventional electrochemical analyzers. However, the non-depleting electrodes provide the capability for continuous O_2 measurement in the sample gas with no degradation of the electrodes (no consumable lead anode), and, thus, false low O_2 readings due to electrode degradation have been eliminated. Sample gas diffuses through a simple gas diffusion barrier to the cathode in an electrolyte solution. Oxygen is reduced at this electrode to hydroxyl ions. Assisted by a KOH electrolyte, the ions migrate to the anode, where they are oxidized back to O_2. Unlike replaceable galvanic electrodes used as the driving mechanism for the reaction, an external electromagnetic force of 1.3 V DC drives the reaction. The resulting cell current is directly proportional to the O_2 concentration in the gas stream. Because of the inherent stability of the electrodes, sensitivity levels to less than 5ppb (parts per billion) of O_2 can be achieved.

Liquid dissolved O_2

Hersch cell It is based on electrolytic chemical reactions and detects O_2 at ppb level. The galvanic cell (or Hersch cell) is composed of essentially two electrodes immersed in an electrolytic fluid, usually concentrated KOH. No external potential is applied.
In the membrane version, O_2 diffuses into the sensor, usually through a Teflon membrane, where it is reduced electrochemically to hydroxide at the silver cathode. The HO-ions migrate to the Pb anode, where Pb is reacted to form lead oxide. The current generated from the reduction/oxidation reaction is proportional to the O_2 concentration of the sample gas. Therefore, the sensor has an absolute zero.
In the version without a membrane the electrodes are exposed to the wastewater, which is used as the electrolyte. The sensor is calibrated in O_2 saturated water. As water has different pH and conductivity values to that of wastewater, there can be no certainty that the measurement, especially below 2 ppm (parts per million) is accurate.

Polarographic (membrane) (See Polarography) This is sometimes called amperometric. The anode and cathode are immersed in an electrolyte, into which O_2 permeates through a membrane. It differs from a galvanic sensor in that the anode has to be polarized at the specific galvanic voltage (of the electro potential series) of O_2, being –650 mV. Without O_2 in the solution, also a small current is flowing. The polarographic sensor needs calibration with liquids with known O_2 concentrations. The voltage-concentration relation is linear. As the sensor ages, its zero offset changes.

Optical fluorescence This is sometimes also called luminescent. In this Fluorescence application the presence of O_2 slows or reduces the amount of fluorescence. As light is constantly produced in a bandwidth rather than an absolute wavelength, there is no absolute zero. However, as the measurement is frequency (rate of change) based, there is no drift as long as the signal strength is reasonable. Measurement in absolute dark (contamination, sensor damage) is a prerequisite. A major problem is in that the fluorophore is soluble in water. So the material needs to be bonded to another material, e.g. a ruthenium (fluorophore)/silicon (bond) matrix, which impedes or stops the degradation of the fluorescing material.

Absorption spectroscopy The sensor uses Laser diode absorption Spectroscopy in the visible spectrum, similar to the absorption method used to measure CO_2, N_2O, and anesthetic agents in the IR (infra red) spectrum.
However, O_2 absorption is just visible (760 nm) without interference or absorption by other ventilation or anesthetic gases. The emission and the absorption line width of O_2 are very narrow, < 0.01 nm, compared to some 100 nm for the CO_2 absorption band at atmospheric pressure.
As the O_2 concentration increases, the light intensity is attenuated, and the photodetector response (thermally adjusted emitted frequency) varies linearly with the O_2 concentration.

Almost all dissolved O_2 analyzers are calibrated using air as the reference. At 25°C, the saturation value of O_2 in water is 8.4 ppm.

In addition to this ''classical'' methods new techniques are for instance photo-acoustic and magneto-acoustic techniques for O_2 monitoring in the open un-intubated airway.

Plethysmography

Principle

Plethysmograpy provides the most accurate measure of volumes, such as that of the whole body (submerged in a water filled chamber), the lungs and extremities.

Use for lungs

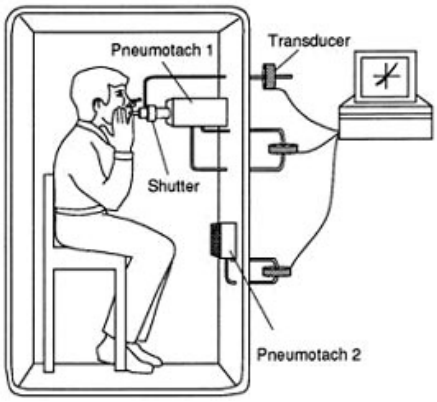


Fig 1. The lung-plethysmograph and the principle of measurement

It measures the functional residual capacity (FRC) as well as the total lung capacity (TLC).

Use for limbs

Some plethysmograph devices are attached to arms, legs or other extremities and used to determine circulatory capacity.

Application

Plethysmography is the most accurate measure of lung volumes. The difference in full versus empty lungs can be used to assess diseases.
Since with a plethysmograph also pressures and flows can be measured, it can be used for precise measurement of resistance and compliances. Therefore it is ideal for assessing airway passage restrictions and airways obstruction reversibility. An obstructive disease will show increased FRC because some airways do not empty normally, while a restrictive disease will show decreased FRC. Body plethysmography is particularly appropriate for patients who have air spaces which do not communicate with the bronchial tree; in such patients the gas dilution method (see Pulmonary tests and instruments) would give an incorrectly low reading.

More info

At the end of normal expiration (FRC remains) and, during a second measurement, at the end of maximal inspiration (lung contents is V_{lung} or TLC), the mouthpiece is closed. The patient is then asked to make an inspiratory effort. As the patient tries to inhale the glottis is closed and the lungs expand due to decreasing pressure. This expands the lung volume (Fig. 2). This, in turn, increases the pressure with ΔP_{box} within the box since it is a closed system and the volume of the body compartment has increased by $\Delta V_{box} = -\Delta V_{lung}$ (provided that temperatures are constant). This is measured with pneumotach2 (see Pulmonary tests and equipment) after opening it. Since the initial pressure P_{box} and the ΔP_{box} (manometer measurements) are also known ΔV_{lung} can be found with Boyle's Law:

$$\Delta V_{lung} = -\Delta V_{box}\Delta P_{box}/(\Delta P_{box}+P_{box}),$$

where V_{box} the empty box volume minus the patient volume (estimated from weight and body fat measurement).

By applying the law again, now for the lung, FRC and TLC is found:

$$V_{lung} = \Delta V_{lung}(\Delta P_{lung}+P_{lung})/\Delta P_{lung},$$

with both pressures known from the manometer in the mouthpiece.

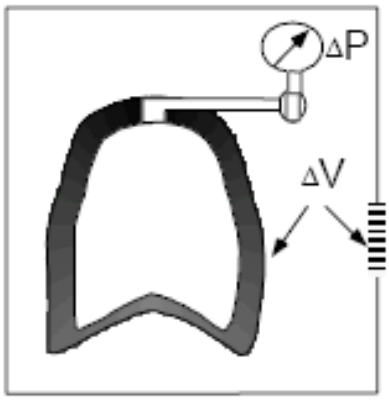


Fig. 2 Principle of determination

Pneumotachography

The pneumotachograph or the pneumotach provides accurate flows over the entire physiological range in the most demanding applications and exercise, pulmonary function testing, and nutritional assessment. The measurement is often based on a bi-directional differential pressure Pitot tube. The output is unaffected by condensation, turbulence, gas viscosity, temperature or position.

Another approach is to measure a pressure drop across a resistance in the gas pathway. The measurement is rapidly and accurately using a differential pressure transducer, from which flow rate and volume are calculated. The resistance is usually designed to produce laminar flow, so that the flow rate is directly proportional to the measured pressure drop. This is achieved using a series of small-bore tubes arranged in parallel (Fig. 1), through which the gas flow must pass. A heating element is sometimes incorporated to prevent the build-up of condensation that could compromise accuracy. The total resistance added by the pneumotachograph should be small so that it can be used in a spontaneously breathing patient.

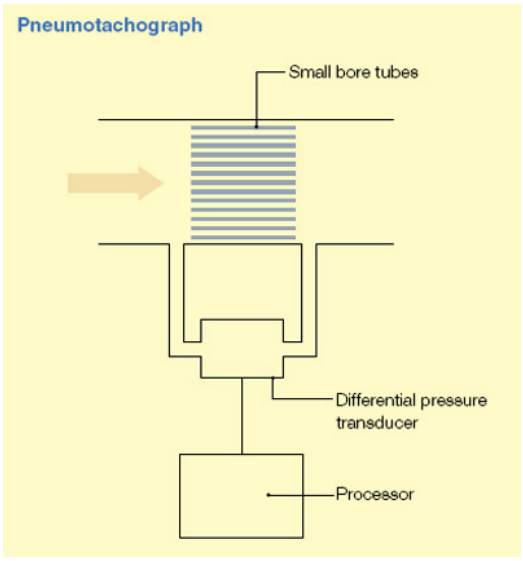


Fig. 1 (from http://www.frca.co.uk/article.aspx?articleid=100390)

Measurement can be made at various points in the breathing system (or ventilator), and a pair of sensors is often used so that inspired and expired tidal volumes can be measured independently. In addition to the differential pressure across the chamber, the absolute pressure in the airway can be easily measured. When linked to the recorded tidal volume, compliance can be calculated (see Lung gas transport 1, basic principles) and displayed in real time.

Pulmonary tests and instruments

Working principles of volumetric and flow measurements

Direct measurement
Gas volumes (and associated flows) can be measured directly using bulk filling of an enclosed space of known volume. Instruments using direct measurement include the industrial gas meter, vitalograph and water-displacement spirometer. Because of their intractability, their use in clinical practice is limited.

Indirect measurement
For clinical use, measurement is usually made indirectly, using a property of the gas that changes in parallel to flow or volume and which can be more easily determined.
Pressure drop across a resistance
– as flow occurs through a resistance, a pressure drop occurs. This effect can be used to calculate flow by keeping the resistance constant and measuring the pressure change as the flow varies, as in a pneumotachograph.
Mechanical movement
– flowing gas has kinetic energy related to its velocity, which can be converted into a measurable value by rotation of a vane (e.g. a spirometer) or bending a flexible obstruction, transducing this to produce an electrical signal.
Heat transfer
– gas flowing past a heated element acts to cool it, as in a hot-wire anemometer.
Ultrasound interference
– the velocity of an ultrasound signal is increased by a gas flowing alongside it in the same direction, and decreased if the gas is flowing against it.

Volume Units
Since the mass of gas in a unit volume is dependent on pressure and temperature (see Gas laws) they have to be specified with their pressure and temperature. Common are STPD, BTPS and ATPS. See for their definitions Gas volume units, STPD, BTPS and ATPS.

Volumetric and flow measurements

Vitalograph
The vitalograph is used specifically to record a single vital capacity breath. Its design uses an expanding bellows. Volume is displayed on the vertical axis and time on the horizontal axis, so that the pattern of expiration is shown as well as the volume.

***F**orced **E**xpiratory **V**olume in 1 second*
The FEV_1 is the most widely used parameter to measure the mechanical properties of the lungs and is measured with e.g. a spirometer (see Spirometry). FEV1accounts for the greatest part of the exhaled volume from a spirometric maneuver and reflects mechanical properties of both the large airways and medium-sized airways. In a normal flow-volume loop, the FEV1occurs at about 75% of the forced vital capacity (FVC). This parameter is reduced in both obstructive and restrictive disorders. In obstructive diseases, FEV1 is reduced disproportionately to the FVC and is an indicator of flow limitation. In restrictive disorders FEV1, FVC and total lung volume are all reduced, and in this setting FEV1 is a measure of volume rather than flow.

Gas dilution method
The method measures lung volumes and is based on:
– starting at the functional reserve capacity (FRC),
– known volume of tracer gas at known concentration,
– measurement of final concentration after rebreathing.
It is easily performed, but less accurate when an obstruction is present.

Hot wire anemometry see Hot wire anemometry

Mechanical flow transducer
A device using mechanical movement is the flow transducer, also used in IC-ventilators. The gas flow is split so that measurement is made in a small side channel. This comprises a thin metal disc supported on a flexible pin, which is mounted perpendicular on the flow direction (Fig. 1). This results in it being bent backwards by the flow. A strain gauge (comprising piezo crystals, see Piezoelectricity) situated immediately behind the pin is compressed as it is bent, with a force dependent on the flow. The resulting electrical signal is processed to calculate the flow rate with a high degree of accuracy.

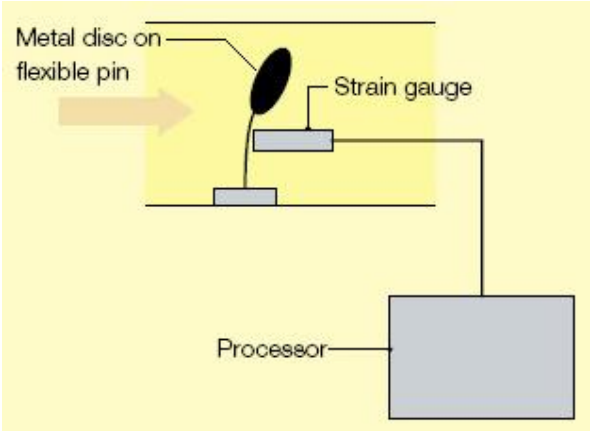


Fig. 1 Mechanical flow transducer

Peak Flow Meter
A peak flow meter is a small, hand-held device used to measure the degree of restriction in the airways. It measures the Peak Expiratory Flow, PEF.
The meter is a specialized vane meter that measures the maximum flow rate only, without calculation of the associated volume. It is modified so that movement of the vane by expiratory gas results in a steadily enlarging pathway for gas escape. The final position to which the bar has moved corresponds to the peak expiratory flow rate.
Its use (in the clinic and at home) is to assess conditions such as asthma, where the problem is largely confined to airway resistance, which limits the expiratory flow rate. Although its use is limited in this respect, it is a very simple and reliable bedside test.

Plethysmography, see Plethysmography

Pneumotachography, see Pneumotachography

Spirometer see Spirometry

Vane meter
The most common vane meter is the Wright's spirometer, in which the gas flow is directed tangentially to strike a rotating vane in the gas pathway (Fig. 2). Originally the design was strictly mechanical, but modern versions use a light source and photodetector positioned across the vane to count its rotation. It is lower at low flows (because of friction) and higher at high flows than a linear relationship would predict because of the momentum (mass times velocity).

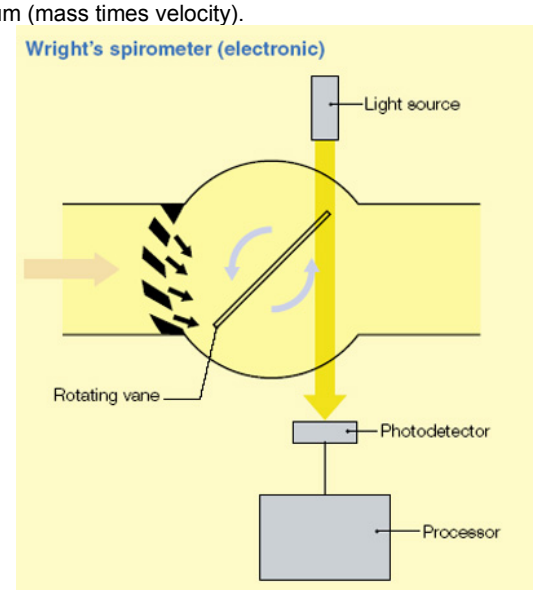


Fig. 2

Ultrasonic flow meter

Ultrasonic flow meters work on the principle that when an ultrasound signal is being transmitted within a flowing gas, the medium in which the sound wave propagates, its velocity changes in proportion to that of the gas flow. When the gas flow and ultrasound signal are in the same direction, an increase in signal velocity occurs. Conversely, when the signal is against the direction of gas flow, its velocity decreases.

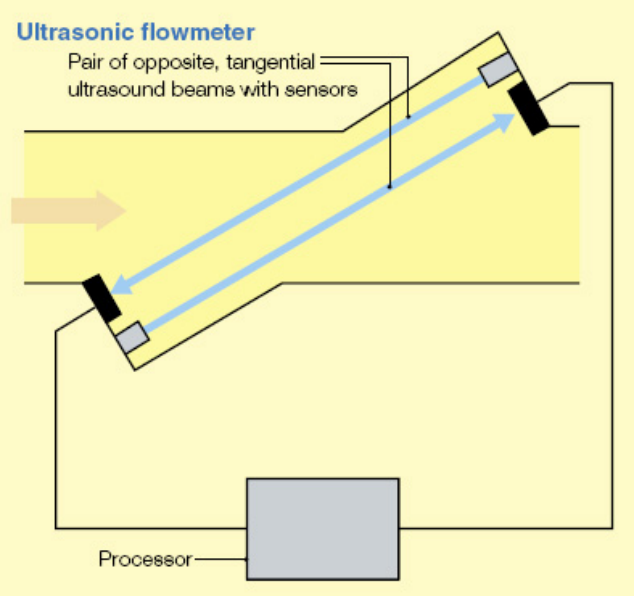


Fig. 3

The usual design is to incorporate a pair of ultrasound beams aimed in opposite directions, each with a sensor. The beams can be situated either directly along the line of flow (within the lumen of the tubing) or tangentially across it (Fig. 3). When no flow is present the velocity of the two beams is equal, and pulses of ultrasound arrive at the sensors simultaneously. When flow occurs there is a time difference between signal detection at the sensors, from which the gas velocity and flow rate can be calculated.

Some other instruments

Capnography See Capnography

Diffusing Capacity for CO, DLCO
DLCO measures transfer of a soluble gas (CO) from the airspace to the blood. It is presented as mL $CO \cdot min^{-1} \cdot mm\ Hg^{-1}$ (STPD). The alveolar volume (VA) at which the DLCO is measured is also commonly reported (BTPS, Gas volume units, STPD, BTPS and ATPS). The ratio of DLCO to VA is also commonly reported as the DL/VA or simply D/VA.
The method is based on a single deep breath with a known concentration of CO for a defined amount of time and measurement of exhaled CO. The CO taken up is calculation. The uptake is proportional to surface area for gas exchange. 'Confounders' are Hb available to bind CO, CO present in blood and thickness of alveolar membrane. It is possible to correct for them. The outcome is evaluated in relation to the measured VC.
The outcome is reduced in case of destruction of alveoli, emphysema, pulmonary fibrosis and lung infiltration, loss of pulmonary vascular bed, pulmonary hypertension, and pulmonary embolism.
It is normal with asthma and elevated in increased pulmonary blood flow or volume (exercise, mild congestive heart failure). See for applications, (contra)indications etc.
http://www.rcjournal.com/online_resources/cpgs/sbcmdc99cpg.html

Measurement of resistance and compliance with the interrupter technique
Measurements of airway *r*esistance using the *int*errupter technique (Rint) is useful for assessing lung function, especially in children of preschool age since cooperation of the subject can be minimal. The principle is that, following an instantaneous interruption of airflow at the airway opening (by closing a valve), there is an instantaneous equilibration of pressure between the alveoli and the airway opening, the mouth (behind the valve). The technique further assumes that there is a single value of alveolar pressure. Following the occlusion, a rapid change in pressure is seen which is equal to the airways resistive fall in pressure between the alveoli and the airway opening. Dividing this pressure change by the flow occurring immediately before the occlusion gives Rint. The initial rapid pressure change is followed by a second slower change in pressure which is related to the visco-elastic properties of the

respiratory tissues, together with any gas redistribution following the occlusion. This effects are dependent on the timing of the occlusion in the phase of respiration.
The equipment includes a rapidly closing valve that occludes the airway for 100 ms before allowing normal respiration to resume. In practice, Rint is rather constant during the tidal cycle. Generally, Rint is measuring during the expiratory phase of respiration at peak tidal flow.
The high frequency (100-2000 Hz) version of this technique is suitable to estimate airway wall resistance and compliance, especially in infants with wheezing. The impedance data (comprising resistance and compliance, see Impedance), spectrally analysis yield also information about airway geometry (diameters), e.g. to quantify bronchoconstriction, or bronchodilation. The measure is only slightly influenced by airway by lung and chest wall tissues.
The forced oscillation technique (FOT) is a method to assess resistances and compliances. A typical application is assessment of bronchial hyperresponsiveness. FOT employs small-amplitude pressure oscillations superimposed on normal breathing.
See for more info http://thorax.bmjjournals.com/cgi/content/full/58/9/742 and http://erj.ersjournals.com/cgi/content/full/22/6/1026

Pulse Oximetry see Pulse oximetry

Oxygen analyzers see Oxygen analysis

Literature

http://www.frca.co.uk/article.aspx?articleid=100390 (Most figures are adopted from this source)
http://thorax.bmjjournals.com/cgi/content/full/58/9/742.)
http://erj.ersjournals.com/cgi/content/full/22/6/1026

Pulse oximetry

Principle

Pulse oximetry is a non-invasive method to monitor the blood oxygenation in the capillary bed, expressed as arterial oxy-Hb/total Hb in % and called SaO_2. (SpO_2, mostly used instead of SaO_2, is an undefined, sloppy acronym). Modern meters also measure pulse rate and pulse strength.

The principle of pulse oximetry is based on the red and infrared light absorption characteristics of oxygenated and deoxygenated Hb. Oxy-Hb absorbs more infrared light and allows more red light to pass through (Fig. 1). Deoxy-Hb absorbs more red light and allows more infrared light to pass through.

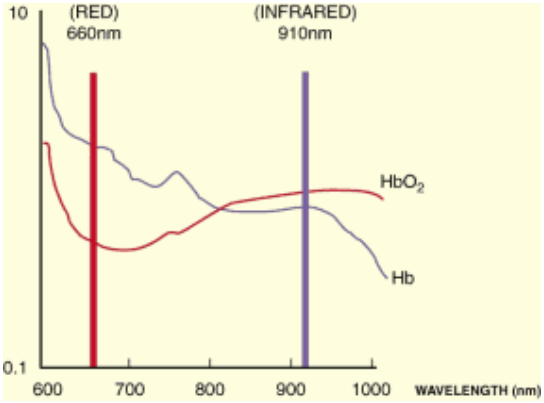


Fig. 1 (from http://www.frca.co.uk/article.aspx?articleid=100390)

Pulse oximetry uses a light emitter with red and infrared light (IR) emitting diodes (LEDs, at 660 nm and 910-940 nm, respectively) that shines through a reasonably translucent site with good blood flow. A red and IR light emitter is (usually) placed on top of a fingertip, toe, pinna or lobe of the ear and a photodetector (red and IR) just opposite at the other side. The diodes flash at a rate of approximately 30/s. The diodes are switched on in sequence, with a pause with both diodes off. This allows compensation for ambient light. The microprocessor analyses the changes in light absorption during the arterial pulsatile flow and ignores the non-pulsatile component of the signal (which results from the tissues and venous blood).

Application

It is especially useful in an intensive care setting, for assessment of emergency patients, determining the effectiveness of or need for supplemental O_2, and monitoring during anesthesia.
Still, conventional pulse oximetry accuracy suffered greatly during motion and low perfusion and made it difficult to depend on when making medical decisions. Arterial blood gas tests have been and continue to be commonly used to supplement or validate pulse oximeter readings. Recent pulse oximetry technology shows significant improvement in the ability to read through motion and low perfusion; thus making pulse oximetry more useful.
Falsely low readings may be caused by hypoperfusion (cold or vasoconstriction, e.g. due to vasopressor agents); incorrect sensor application; and movement (such as shivering), especially during hypoperfusion. Falsely high or falsely low readings will occur when Hb is bound to CO or cyanide.

It should be noted that this is a measure solely of oxygenation and is not a substitute for blood gas analyses since it gives no indication of CO_2 levels, blood pH, or $NaHCO_2$.

More Info

The oxygen saturation is estimated by measuring the transmission of light through the pulsatile tissue bed. This is based on the Beer-Lambert law: the intensity of transmitted light through a material (here the tissue) decreases exponentially as a function of the concentration of the absorbing substance *and* decreases exponentially as a function of the path length through the material. The law is applied for both types of molecules *and* for both wavelengths, yielding two equations with two unknown concentrations and the unknown path length. Since not concentrations themselves are relevant but the concentration ratio, this ratio can basically be solved. However, there are some complications.

The major problem is to differentiate between the arterial pulsation, the light absorption and the absorption in skin, tissue and venous blood. However, with each heart beat there is a surge of arterial blood, which momentarily increases arterial blood volume across the measuring site. This results in more light absorption during the surge.
The light absorbed by non-pulsatile tissues is constant (DC). The alternating absorption component (AC) is the result of pulsatile blood pulsations. The photodetector generates a voltage proportional to the transmitted light. The AC component of the wave accounts for only 1-5% of the total signal. The high frequency of the diodes allows the absorption to be calculated many times per second. This reduces movement effects on the signal.
Since peaks occur with each heartbeat or pulse, the term “pulse oximetry” was coined.
The microprocessor analyses both the DC and AC components at 660 nm and 940 nm. Modern pulse oximeters may use more than two wavelengths.

There are two methods of sending light through the measuring site: transmission and reflectance. In the transmission method, the emitter and photodetector are opposite of each other with the measuring site in-between. The light can then pass through the site. In the reflectance method, the emitter and photodetector are next to each other on top of the measuring site. The light bounces from the emitter to the detector across the site. The transmission method is the most common type used and for the discussion below the transmission method will be implied.
After the transmitted red (R) and infrared (IR) signals pass through the measuring site and are received at the photodetector, the R/IR ratio is calculated. The R/IR is compared to a “look-up” table (made up of empirical formulas) that convert the ratio to a SaO_2 (SpO_2) value. Typically an R/IR ratio of 0.5 equates to approximately 100% SaO_2, a ratio of 1.0 to approximately 82% SaO_2, while a ratio of 2.0 equates to 0% SaO_2.

Spirometry

Basic Principles

Spirometry, the most common of the Pulmonary Function Tests (PFTs), is the measurement of lung function, specifically the measurement of the amount (volume) and speed (flow) of air that can be inhaled and exhaled. Spirometry is an important tool used for assessing lung diseases such as asthma, cystic fibrosis, and COPD. Results are usually given in both raw data (e.g. L/s) and percent predicted - the test result as a percent of the "predicted values" for the patients of similar characteristics (height, weight, age, sex, and sometimes race).
Spirometer comes in many different varieties. Many produce a Flow-Volume diagram, see Fig. 1. Most spirometers also display a Volume-Time curve, with volume (L) along the vertical axis.

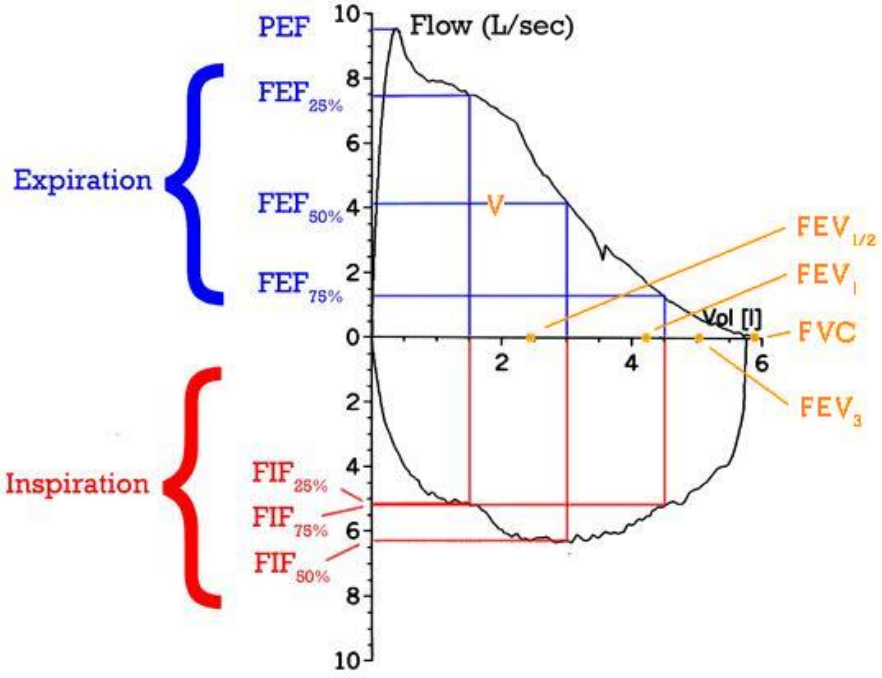


Fig. 1 Flow-Volume diagram. Positive values represent expiration, negative values represent inspiration. The trace moves clockwise for expiration followed by inspiration. (Note the FEV_1, $FEVA_{1/2}$ and FEV_3 values are arbitrary in this graph and just shown for illustrative purposes, they must be recorded as part of the experiment). For the definitions of the various parameters see the list below.

The basic spirometry test itself is simple, and varies slightly depending on the equipment used. Generally, the patient is asked to take the deepest breath they can, and then exhale into a machine as hard as possible, for as long as possible, followed by a rapid deep breath in. Sometimes, the test will be preceded by a period of quiet breathing in and out from the machine, or the rapid breath in will come before the hard exhalation. During the test, soft clips are used to prevent air escaping through the nose. Sometimes, filter mouthpieces are used to prevent the spread of germs, or in other cases a length of tubing is placed between the patient and the machine. The test is repeated, usually at least three times and often up to as many as eight times, to insure that the results are accurate and repeatable. With provocation tests some drug is administered to study a lung disorder or to study hypersensitivity to some allergic agent.
Due to the patient cooperation required, spirometry can only be used on children old enough to comprehend and follow the instructions given (typically about 4-5 years old), and only on patients that are able to understand and follow instructions - thus, this test is not suitable for patients that are unconscious, heavily sedated, or have limitations that would interfere with vigorous respiratory efforts. Other types of PFTs are available for infants and unconscious persons.

Explanation of Common Test Values
FVC: *Forced Vital Capacity* - This is the total amount of air that you can forcibly blow out after full inspiration, measured in liters.

FEV 1: *Forced Expiratory Volume in 1 second* - This is the amount of air that you can forcibly blow out in one second, measured in liters. Along with FVC it is considered one of the primary indicators of lung function.
FER: This is the ratio of FEV1/FVC, which showing the amount of the FVC that can be expelled in one second. In healthy adults this should be approximately 0.8.
PEF: *Peak Expiratory Flow* - This is the speed of the air moving out of your lungs at the beginning of the expiration, measured inL/s.
FEF 25-75% or 25-50%: *Forced Expiratory Flow 25-75% or 25-50%* - This is the average flow (or speed) of air coming out of the lung during the middle portion of the expiration (also sometimes referred to as the MMEF, for maximal mid-expiratory flow).
FIF 25-75% or 25-50%: *Forced Inspirtory Flow 25%-75% or 25%-50%* - This is similar to FEF 25%-75% or 25%-50% except the measurement is taken during inspiration.

Application

F-V diagrams are of great help to diagnose low or high located constrictions in the airways.

More Info

Over the years volumetric spirometers with a water bell were the most common ones. For flow-spirometry for instance the Fleisch-pneumotach (differential pressure measurement (on the basis of Bernoulli's principle). The turbine (rotation velocity of a hand-held small turbine screw), the Pitot tube (principle of the water jet steam) or the hot-wire anemometer can also be the instrumental basis of a flow-spirometer.

Thermodynamic equations for an ideal gas

$$PV^n = Constant$$

Process	Isobaric	Isochoric	Isothermal	Adiabatic
Variable =>	Pressure	Volume	Tempemature.	No Heat Flow
Quantity Constant =>	$\Delta P = 0$	$\Delta V = 0$	$\Delta T = 0$	$Q = 0$
n	0	∞	1	$\gamma = C_p/C_v$
First Law	$\Delta U = Q - W$	$\Delta U = Q$ $W = 0$	$\Delta U = 0$ $Q = W$	$\Delta U = -W$ $Q = 0$
Work $W = \int P dV$	$P(V_2 - V_1)$	0	$P_1 V_1 \ln\left(\frac{V_2}{V_1}\right)$	$\frac{P_1V_1 - P_2V_2}{\gamma - 1}$
Heat Flow Q	$m C_p (T_2 - T_1)$	$m C_v (T_2 - T_1)$	$P_1 V_1 \ln\left(\frac{V_2}{V_1}\right)$	0
Heat Capacity	C_p	C_v	∞	0
Internal Energy $\Delta U = U_2 - U_1$	$m C_v (T_2 - T_1)$	$m C_v (T_2 - T_1)$	0	$m C_v (T_2 - T_1)$
Enthalpy $\Delta H = H_2 - H_1$ $H = U + PV$	$m C_p (T_2 - T_1)$	$m C_p (T_2 - T_1)$	0	$m C_p (T_2 - T_1)$
Entropy $\Delta S = S_2 - S_1$ $= \int dQ/T$	$mC_p \ln \frac{T_2}{T_1}$	$mC_v \ln \frac{T_2}{T_1}$	$nR \ln \frac{V_2}{V_1}$	0*
Ideal Gas Relations $\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$ $PV = NkT$	$P_1 = P_2$ $\frac{V_1}{T_1} = \frac{V_2}{T_2}$ $\frac{T_1}{T_2} = \frac{V_1}{V_2}$	$V_1 = V_2$ $\frac{P_1}{T_1} = \frac{P_2}{T_2}$ $\frac{T_1}{T_2} = \frac{P_1}{P_2}$	$T_1 = T_2$ $P_1V_1 = P_2V_2$ $\frac{P_1}{P_2} = \frac{V_2}{V_1}$	$Q = 0$ $(S_1 = S_2)$ * $P_1V_1^\gamma = P_2V_2^\gamma$ $\frac{T_1}{T_2} = \left(\frac{V_2}{V_1}\right)^{\gamma-1}$

* For Reversible Processes

$n c_v = m C_v$ **$c_p - c_v = R$** **$n R = N k$** **$n c_p = m C_p$**
$\gamma = C_p/C_v = c_p/c_v$ = Ratio of Specific Heats
C_p = Constant Pressure Specific Heat Capacity (J/kg/ ^oC)
C_v = Constant Volume Specific Heat Capacity (J/kg/ ^oC)
c_p = Molar Constant Pressure Heat Capacity (J/mole/oC)
c_v = Molar Constant Volume Heat Capacity (J/mole/oC)

VO_{2max}

Principle

VO_{2max} is the maximum amount of oxygen, in mL, one can use in 1 min per kg of bodyweight, so in mL/kg·min. It is also called the (mass) specific VO_{2max}. VO_{2max} is also known as maximal O_2 consumption or maximal O_2 uptake of the whole body, the absolute VO_{2max}. Then, it is mostly expressed in L/min for the whole body. For experimental purposes and for comparing the aerobic performance of for instance endurance sports athletes, the former expression is highly preferred.

Measuring VO_{2max}
Accurately measuring VO_2 max involves a ramp test (treadmill or cyclo-ergometer) in which exercise intensity is progressively increased until exhaustion while measuring the rate and O_2 concentration of the inhaled and exhaled air.
A simple but adequate test for a rough estimate is the well-known Cooper test in which the distance (in km) covered by running or swimming 12 minutes is measured. For running the estimate of VO_{2max} is:

$$VO_{2max} = (\text{distance} - 0.505)45 \text{ (mL/min·kg)}. \quad (1)$$

There exist many other more or less reliable tests and VO_{2max} models (calculators) to estimate VO_{2max} (see ref. 1, 2 and 3). The parameters of the calculators are always age and BMI (body mass index, mass/height^2 in kg/m^2), and often also sex and a measure of endurance sport activity, generally hours/week. Sometimes HR_{max} is used (ref. 3) by applying the Fick principle (see **More Info**).
VO_{2max} declines with age, as illustrated in Table 1, which presents norm values as function of age (calculated from ref. 1).

Table 1 VO_{2max} (mL/min·kg)

Age (year)	<30	30-40	40-50	50-60	60-70
men	39	36½	33	31½	29½
women	35	32½	30	28	27

or in formula:

$$VO_{2max,man} = 48.4 - 0.43A + 0.0020A^2$$
$$VO_{2max,vrouw} = 44.1 - 0.42A + 0.0024A^2,$$

where A is age (year, > 20).

Application

VO_{2max} is utilized in sports medicine, and in daily life in the fitness industry and sports, especially with top athletes.

More info

.
Fick Equation
VO_2 (L/min) is properly determined by the Fick Equation (see Diffusion: Fick's laws):

$$VO_2 = Q(C_aO_2 - C_vO_2),$$

where Q is the cardiac output (beats/time x stroke volume, i.e. HR·SV) of the heart, C_aO_2 is the arterial oxygen content, and $C_vO_2$is the venous oxygen content. By applying the principle two times, for HR_{max} and HR_{rest}, finally the following expression for VO_{2max} is obtained:

$$VO_{2max} = (HR_{max}/HR_{rest})\cdot(SV_{max}/SV_{rest})\cdot((C_aO_2 - C_vO_2)_{max}/(C_aO_2 - C_vO_2)_{rest})\cdot VO_{2rest} \text{ (L/min)}.$$

Since the ratio's (SV_{max}/SV_{rest}) and $(C_aO_2 - C_vO_2)_{max}/(C_aO_2 - C_vO_2)_{rest}$, and the specific VO_{2rest} are rather constant over subjects, they are about 3.4, 1.3 and 3.4 (mL/kg/min) respectively, one obtaines:

$$VO_{2max} = (HR_{max}/HR_{rest})\cdot 3.4\cdot 1.3\cdot 3.4\cdot W = 15\cdot W(HR_{max}/HR_{rest}), \quad \text{(mL/min)}$$

where W is body weight (kg).
For the specific VO_{2max} one obtaines:

$VO_{2max} = 15(HR_{max}/HR_{rest})$, (mL/min·kg)

The equations hold in between 20 and 50 years and are sex invariant.

VO_{2max} levels
VO_{2max} varies considerably in the population. The scores improve with training. In endurance sports, such as cycling, rowing, cross-country skiing and running, VO_{2max} values above 75 mL/kg/min are rather common. World class cyclists and cross-country skiers typically exceed 80 mL/kg/min and a rare few may exceed 90 for men. Women top athletes generally exceed 70 mL/kg/min. A competitive club athlete might achieve a VO_{2max} of ca. 70 mL/kg/min.

References

MArdle, WD., Katch FI, Katch VL, , Exercise Physiology: Energy, Nutrition & Human Performance, Lippincott Williams and Wilkins, Section 3, Chap 7, 11 and 17, 2001.

Uth N,Sorensen H, Overgaard K, Pedersen PK.. Estimation of VO2max from the ratio between HRmax and HRrest--the Heart Rate Ratio Method. 2004 Eur J Appl Physiol 91 111-115.

Light and Optics

CCD camera

Principle

A charge-coupled device (CCD) is an image sensor, consisting of an integrated circuit containing an array of linked, or coupled, light-sensitive capacitors. This device is also known as a Color-Capture Device.
The capacitors are the classical components of the CCD camera, but more and more photodiodes are the fundamental collecting units of the CCD.
Physics of operation An image is projected by a lens on the capacitor array (actually a 2-d array or matrix), causing each capacitor to accumulate an electric charge proportional to the light intensity at that location. A one-dimensional array, used in line-scan cameras, captures a single slice of the image, while a two-dimensional array, used in video and still cameras, captures the whole image or a rectangular portion of it. Once the array has been exposed to the image, a control circuit causes each capacitor to transfer its contents to its neighbor. The last capacitor in the array dumps its charge into an amplifier that converts the charge into a voltage. By repeating this process, the control circuit converts the entire contents of the array to a varying voltage, which it samples, digitizes and stores in memory.

Application

CCDs are used in medical fluoroscopy, optical and UV spectroscopy and in all kind of basic cellular research. Frequent science applications are in astrophysics. Daily life applications are digital photography and "1D" CCD grids are applied in fax machines.
CCDs containing grids of pixels are used in digital cameras, optical scanners and video cameras as light-sensing devices. They commonly respond to 70% of the incident light (meaning a quantum efficiency of about 70%) making them more efficient than photographic film, which captures only about 2% of the incident light.

More info

CCDs are typically sensitive to infrared light, which allows infrared photography, night-vision devices, and zero lux (or near zero lux) video-recording/photography. Because of their sensitivity to infrared, CCDs used in astronomy are usually cooled to liquid nitrogen temperatures, because infrared black body radiation (see Body heat dissipation and related water loss and Wien's displacement law) is emitted from room-temperature sources. One other consequence of their sensitivity to IR is that infrared from remote controls will often appear on CCD-based digital cameras or camcorders if they don't have infrared blockers. Cooling also reduces the array's dark current, improving the sensitivity of the CCD to low light intensities, even for ultraviolet and visible wavelengths.
Thermal noise, dark current, and cosmic rays may alter the pixels in the CCD array. To counter such effects, an average of several exposures is made. The average of images taken with the shutter closed is necessary to lower the random noise. Once developed, the "dark frame" average image is then subtracted from the open-shutter image to remove the dark current and other systematic defects in the CCD (dead pixels, hot pixels, etc.).

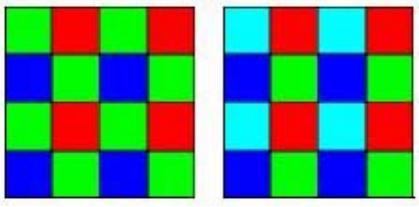


Fig. 1 Left the Bayer filter with twice the green cells and right the RGBE filter with as fourth color cyan. All four colors have the same frequency of occurrence.

Digital color cameras generally use a Bayer or RGBE filter s over the CCD. These filters are color filters in a matrix arrangement. After passing the filer matrix, the light is detected by a matrix of photosensors. Better color separation can be reached by three-CCD devices (3CCD) and a dichroic beam splitter prism, (see Dichroism) and Light: beam splitter) that splits the image into red, green and blue components. Each of the three CCDs is arranged to respond to a particular color. Some semi-professional digital video camcorders (and all professionals) use this technique.

Chemoluminescence and Bioluminescence

Principle

Chemoluminescence

Chemoluminescence is the emission of light as the result of a chemical reaction. The reaction may occur in the liquid phase or in the gas phase. Most simply, given reactants **A** and **B**, with an excited intermediate ◊, the reaction is:

[**A**] + [**B**] → [◊] → [*Products*] + light.

The decay of the excited state [◊] to a lower energy level is responsible for the emission of light. In theory, one photon of light should be given off for each molecule of reactant, so Avogadro's number of photons per mole. In actual practice, non-enzymatic reactions seldom exceed 1% quantum efficiency. For example, in the liquid phase, if [A] is luminol and [B] is hydrogen peroxide in the presence of a suitable catalyst the reaction is:

luminol + H_2O_2 → 3-aminophthalate[◊] → 3- aminophthalate + light.

A standard example of chemoluminescence in the laboratory setting is found in the luminol test, where evidence of blood is taken when the sample glows upon contact with iron. A daily live example is a lightstick.

Fig. 1 Lightsticks

Enzymatic chemoluminescence (ECL) is a common technique for a variety of detection assays in biology. An horseradish peroxidase molecule (HRP) is tethered to the molecule of interest (usually by immunoglobulin staining). This then locally catalyzes the conversion of the ECL reagent into a sensitized reagent, which on further oxidation by hydrogen peroxide, produces an excited triplet (a set of three quantum states of a system, each with total spin *S* = 1), e.g. of carbonyl which emits light when it decays to the singlet (S = 0). The result is amplification of antibody detectability.
When chemoluminescence takes place in living organisms, the phenomenon is called *bioluminescence*.

Applications

Analysis of organic species: useful with enzymes, where the substrate isn't directly involved in chemoluminescence reaction, but the product is a reactant of the chemoluminescence reaction. Further environmental gas and liquid analysis for determining small amounts of impurities or poisons in air. Typical example is NO determination with detection limits down to 1 ppb.

Bioluminescence

Bioluminescence is the production and emission of light by a living organism as the result of a chemoluminescence reaction during which chemical energy is converted to light energy. Bioluminescence is really a form of "cold light" emission; less than 20% of the light is generated by thermal radiation. It should not be confused with fluorescence, phosphorescence or refraction of light.

The most striking example is bioluminescence by dinoflagellates at the surface of seawater when the surface is agitation, e.g. by a swimmer or a copepode. The λ_{max} is at ca. 472 nm and the emittance has the remarkable efficiency of more than 50%,

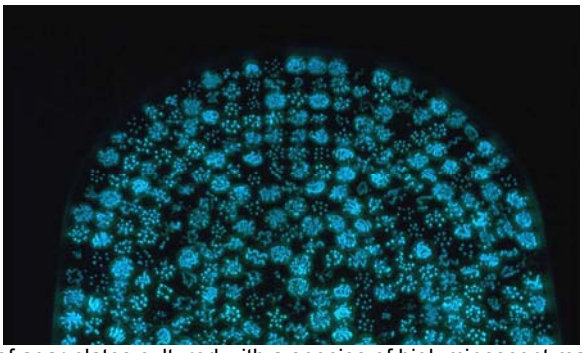


Fig. 2 Image of hundreds of agar plates cultured with a species of bioluminescent marine bacteria.

Bioluminescence may be generated by symbiosis of organisms carried within a larger organism. It is generated by an enzyme-catalyzed chemoluminescence reaction, wherein a luceferin (a kind of pigment) is oxidized by a luceferase (a kind of enzyme). ATP is involved in most instances. The chemical reaction can be either external to cells, or an intracellular process. The expression of genes related to bioluminescence in bacteria is controlled by the lux operon.

Application

Luciferase systems are widely used in the field of genetic engineering as reporter genes (green fluorescent protein, see Fluorescence).
The structure of photophores, the light producing organs in bioluminescent organisms, are being investigated by industry (glowing trees, organisms needing watering), food quality control, detecting bacterial species and studies into potential applications as for tagging domestic animals.

More Info

All cells produce some form of bioluminescence within the electromagnetic spectrum, but most is neither visible nor noticeable to the naked eye. Every organism's bioluminescence is unique in wavelength, duration, timing and regularity of flashes.
90% Of deep-sea marine life is estimated to produce bioluminescence in one form or another. Many marine invertebrates have bioluminescence, like planktons, microbes, corals, clams, jelly fish, nudibranchs, crustaceans (lobsters, squids etc.), echinoderms (sea stars, sea urchins etc.). Most marine light-emission belongs in the blue and green light spectrum, the wavelengths that have the most powerful penetrating power in water. However, certain jawless fish emit red and IR light.
Non-marine bioluminescence is less widely distributed, but with more color variety. Well-known forms of land-bioluminescence are fireflies and New Zealand glow worms. Other insects (and larvae), worms (segmented), arachnoids, fish and even species of fungi have bioluminescent abilities.
Most forms are brighter (or only exist) at night, following a circadian rhythm.
It is thought to play a direct role in camouflage, attraction, repulsion and communication. It promotes the symbiotic induction of bacteria into host species, and may play a role in colony aggregation.

Dichroism

Principle

Dichroism has two related but distinct meanings in optics. With the first one, a dichroic material causes visible light to be split up into distinct beams of different wavelengths (colors), not to be confused with dispersion as happens in a prism (see Light and Light: refraction). With the 2nd one, light rays having different polarizations (see Light: polarization), are absorbed by different amounts.
Which meaning of *dichroic* is intended can usually be inferred from the context. A mirror, a filter or beam splitter (see Light: beam splitter) is referred to as *dichroic* in the color-separating first sense; a dichroic crystal or material refers to the polarization-absorbing second sense.

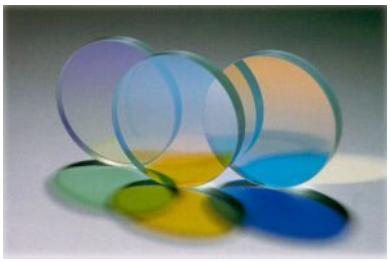


Fig 1 Dichroic filters

Application

General
The most common example is the dichroic filter. Dichroic filters operate using the principle of interference (see Light: diffraction and Huygens' principle). Alternating layers of an optical coating are built up upon a glass substrate, selectively reinforcing certain wavelengths of light and interfering with other wavelengths. By controlling the thickness and number of the layers, the wavelength of the bandpass can be tuned and made as wide or narrow as desired. Because unwanted wavelengths are reflected rather than absorbed, dichroic filters do not absorb much energy during operation and so become much less warm as absorbance filters (see for absorbance Lambert-Beer law).
Other examples of the wavelength type of dichroism are the dichroic mirror and the dichroic prism. The latter is used in some camcorders (miniaturized video cameras), which uses several coatings to split light into red, green and blue components. This is also applied in the CCD camera.
Medicine and food industry
The dichroism of optically active molecules (see **More Info**) is used in the food industry to measure syrup concentration, and in medicine as a method to measure blood sugar (glucose) in diabetic people.

More Info

The original meaning of *dichroic* refers to any optical device, which can split a beam of light into two beams with differing wavelengths.
Basically, a dichroic filter has more than one transmission peak, with transmission frequencies harmonically related. However, in practice one needs nearly always a filter with one transmission peak. These are the filters, which are called interference filters. The higher harmonics, with have much lower transmission, are mostly attenuated by an absorbance filter. Side bands (similar of those of Fig. 1 in Light: diffraction) have very small transmission. They can strongly be diminished by stacking two identical filters, but on the cost of transmission of the principal peak.

The second meaning of dichroic refers to a material in which light in different polarization states, traveling through it, experience a varying absorption. The term comes from observations of the effect in crystals such as tourmaline. In these crystals, the strength of the dichroic effect varies strongly with the wavelength of the light, making them appear to have different colors when viewed with light having differing polarizations. This is more generally referred to as pleochroism, and the technique can be used to identify minerals.
Dichroism also occurs in optically active molecules, which rotate linearly polarized light (see Light: polarization). Depending on the 3-D molecular structure the rotation is left (levorotatory) or right (dextrorotatory). This is known as circular dichroism. Glucose is dextrorotatory and fructose strongly levorotatory. However, basically an optic active substance has a dextrorotatory and levorotatory version.
Dichroism occurs in liquid crystals (substances with properties between those of a conventional liquid, and those of a solid crystal) due to either the optical anisotropy of the molecular structure (resulting in more than one refractive index, see Light: Snell's law) or the presence of impurities or the presence of dichroic dyes.

Endoscopy

Principle

Endoscopy refers to looking inside the human body using an *endoscope*. Endoscopy is a minimally invasive diagnostic technique used to assess the interior surfaces of an organ by inserting a tube into the body. The instrument may have a rigid (borescope) or flexible tube (fiberscope) and not only provide an image for visual inspection, photography and video-scopy, but also enable taking biopsies and retrieval of foreign objects. Endoscopy is the vehicle for minimally invasive surgery.

An endoscope comprises an eyepiece, the ocular, producing a parallel exit bundle or virtual image and light source to illuminate the object on one end. At the other end is an objective lens producing a real image (see Light: the ideal lens). Both are linked by a tube mounting an optical fiber system (see Fiber optics). So, the fundamental principle of operation is transmitting optical information through a bundle of optical fibers such that an image can be observed. However, a classical *boroscope* may comprise instead of the fiber system a whole series of lenses as transmission system. Basically, an endoscope is a kind of microscope. The light source may provide wide band visible light, whether or not spectrally scanned, and for specific applications narrow band, (near) IR light. For Fluorescence (natural or artificial), e.g. applied in examination the esophagus, UV light can be used. Often, the light is provided by a Laser. An additional channel allows entry of air, fluid, as well as remote control of medical instruments such as biopsy forceps or cytology brushes.

Fig. 1 A flexible endoscope.

Application

Endoscopy is applied in nearly every medical discipline, also for outdoor patients. Endoscopes can be divided into two distinct categories according to their medical application. These are the regular, macroscopic endoscopes such as the gastroscope, colonscope, and bronchoscope to inspect epithelial surfaces. The second category comprises the miniaturized types. They includes ultrathin endoscopes for use as ophthalmic endoscopes, angioscopes, robotic surgery and needlescopes. The latter, with a diameter less than one mm, have been developed to examine very small parts of internal organs. The images of *ultrathin needlescopes* contain 2,000 to 6,000 pixels with a smallest resolution of about 0.2mm. They can be inserted into for instance mammary glands to detect breast cancer at early stages. A borescope is used in arthroscopy (and also in engineering).

Non-medical uses are in architectural design (pre-visualization of scale models) and internal inspection of complex technical systems and examination of improvised explosive devices by bomb robots.

More Info

The type of fibers used is dependent on the type of the illuminating light and the image specifications. Often, the fibers to deliver the light, the light guide, (mostly with coherent light, so light with the same frequency and intensity) and those to transmit the image information, the image guide, are of different types of fibers (see Fiber optics).
Recent developments are fiber-optic fluorescence imaging systems. Until recently, fiber-based fluorescence imaging was mainly limited to epifluorescence and scanning confocal modalities (confocal micro-endoscopy) (see Light microscopy: confocal). New classes of photonic crystal fibers (see Fiber optics) facilitate ultra-short pulse delivery for fiber-optic two-photon fluorescence imaging. This can be combined with two-photon fluorescence and second harmonic generation microscopy, miniaturized in a nonlinear optical endoscope based on a double-clad photonic crystal fiber to improve the detection efficiency and a MEMS (MicroElectroMechanical System) mirror to steer the light at the fiber tip (see Fiber optics, Light microscopy: two-photon fluorescence, and the chapters about light microscopy).

Another new application is combining laser holographic interferometry with an endoscope (see Holography and Huygens'principle). Another combination is laser Doppler imaging (see Doppler principle) of blood flow with endoscopy.
With the application of robotic systems, telesurgery was introduced as the surgeon could operate from a site physically removed from the patient.
Wireless capsule endoscopy is another emerging technology. This technique uses an ingestible capsule comprising a miniature camera with a MEMS mirror for scanning and a transmitter. In this way some part of the gastrointestinal tract can be visualized. Nowadays, this application in the esophagus is more or less standard, but other parts of the tract are still experimentally due to peristaltic movements. MEMS technology may provide a solution for this.

Fiber optics

Principle

An optical fiber is a cylindrical light-isolated waveguide that transmits light along its axis by the process of total internal reflection (see Light: Fresnel equations). The fiber consists of a *core* surrounded by a *cladding* layer. To confine the optical signal in the core, the refractive index of the core must be greater than that of the cladding.
Optical fibers may be connected to each other by connectors or by *splicing* that is joining two fibers together to form a continuous optical waveguide. The complexity of this process is more difficult than splicing copper wire.

Multimode fiber (MMF) and multi-mode fiber

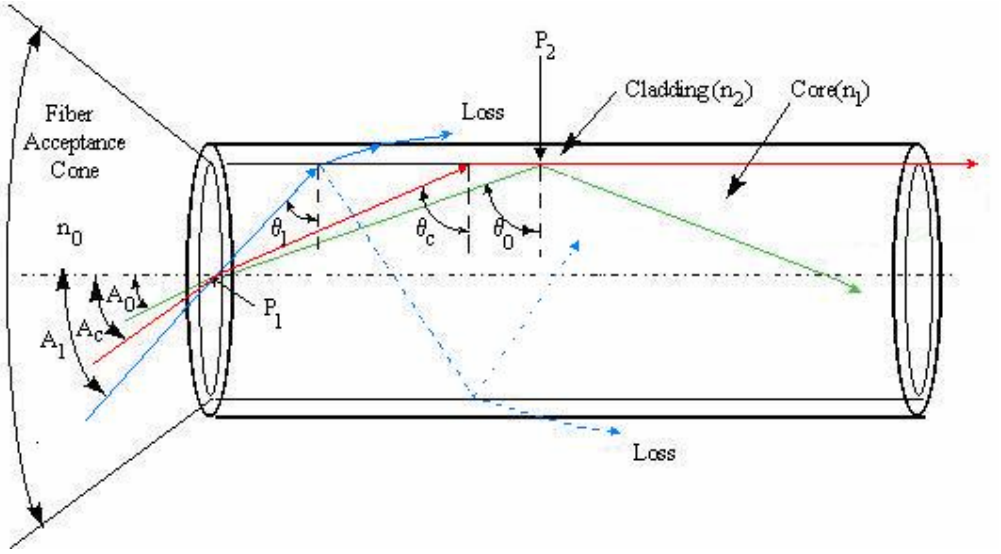


Fig. 1 The propagation of light through a SI-MMF.

A fiber with large core diameter (some tens of μm up to hundreds of μm) behaves in accordance with geometric optics. It is called a *multimode fiber* (MMF) since there are various modes of vibration given by the wave equations (see textbooks of physics).
When the boundary between the core and cladding is abrupt, the MMF is called a *step-index* (SI) fiber. When it is gradual it is a *graded-index* (GRIN) MMF.
In a SI MMF, rays of light are guided along the fiber core by total internal reflection. Rays (for instance the green one in Fig. 1) are completely reflected when they meet the core-cladding boundary at a higher angle (measured relative to a line normal to the boundary) than the critical angle Θ_c, the minimum angle for total internal reflection. The red ray in Fig. 1 impinges with the angle Θ_c. Θ_c is $n_{gladding}/n_{core}$.
Rays that meet the boundary at a lower angle (the blue one in Fig. 2) are lost after many repeated reflections/refractions from the core into the cladding, and so do not convey light and hence information along the fiber.
Θ_c determines the acceptance angle of the fiber, also expressed in the numerical aperture NA ($\equiv n_0 \cdot \Theta_c$). A high NA allows light to propagate down the fiber in rays both close to the axis and at various angles, allowing efficient sending of light into the fiber. However, this high NA increases the amount of dispersion as rays at different angles have different path lengths and therefore take different times to traverse the fiber. This argues for a low NA.

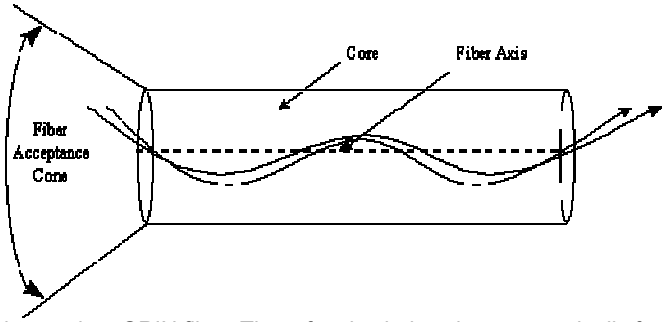


Fig. 2. Paths of light rays in a GRIN fiber. The refractive index changes gradually from the center to the outer edge of the core.

In a GRIN fiber, n_{core} decreases continuously from the axis to the cladding. This causes light rays to bend smoothly as they approach the cladding, rather than reflecting abruptly from the core-cladding boundary. The resulting curved paths reduce multi-path dispersion because the undulations (see Fig.2) diminish the differences in path lengths. The difference in axial propagation speeds are minimized with an index profile which is very close to a parabolic relationship between the index and the distance from the axis.
FI and GRIN fibers suffer from Rayleigh scattering (see Light: scattering), which means that only wavelengths between 650 and 750 nm can be carried over significant distances.

Singlemode fiber (*SMF*)
This single glass fiber (core diameter generally 8 - 10 µm) has the axial pathway as solely mode of transmission, typically at near IR (1300 or 1550 nm). It carries higher bandwidth than multimode fiber, but requires a light source with a narrow spectral width. Single-mode fiber have a higher transmission pulse rate and cover up to 50 times more distance than multimode, but it also costs more. The small core and single light-wave virtually eliminate any distortion that could result from overlapping light pulses (little pulse dispersion), providing the least signal attenuation and the highest transmission speeds of any fiber cable type.

Image transmission
It is impossible for a single fiber to transmit an image. An individual fiber can transmit only a spot of a certain color and intensity. To transmit an image, a large number of single fibers must be aligned and fused together. This means assembly of optical fibers in which the fibers are ordered in exactly the same way at both ends of the bundle to create an image. This type of fiber bundle is called a coherent bundle or image guide bundle. On the other hand, the assembly of optical fibers that are bundled but not ordered is called an incoherent bundle. An optical fiber which is incapable of producing an image is used in medical endoscopes, boroscopes, and fiberscopes as a light guide. The light guide, as well as the image guide, is essential to construct an image in any optical instrument. Light guides are much less expensive and easy to produce compared to image guides and are designed to maximize light carrying ability. In an image guide, the amount of image detail (resolving power) depends on the diameter of each fiber core. Generally, the individual fibers of a light guide are much thicker than fibers (MMFs) in image guides because resolution is not a factor.

Application

Medical
Optical fibers are used in transducers and bio-sensors for the measurement and monitoring of for instance body temperature, blood pressure, blood flow and oxygen saturation levels. In medical applications, the fiber length is so short (less than a few meters) that light loss and fiber dispersion are not of concern. Glass optical fibers are used in most endoscopes and are made with SI fibers.
Optical fibers are also used as transmission lines in equipment that is very sensitive to disturbance by electric fields, such as EEG amplifiers. At the other hand, they are applied to prevent the generation of a magnetic field due to current flowing in an electric cable. Even very small current produce magnetic fields strong enough to disturb MEG recordings (see Magnetoencephalography (MEG)). All these applications are based on SI MMFs.
The principle of fiber optics is also found in nature. Slender rods, as found in for instance frogs, and the visual sensory cells of arthropods, the ommatidia, act as wave guides.
General
Optical fiber cables are frequently used in ICT applications (such as cable television and all kind of data transport). For far distance transmission SMFs are used. MMFs can only be used for relative short distances, e.g. for ICT applications in a building.

More info

Consider Fig. 2 again. The ray incident on the face of the SI fiber at angle A_0 will be refracted inside the core and refracted into the cladding. At angle A_1 a ray will be refracted along the boundary of the core and the outside medium. The angle A_c is referred to as the maximum acceptance angle and θ_c is the critical angle for internal reflection. The angles A_c and θ_c are determined by the refractive indices of core and cladding. Therefore, a ray incident on the core-cladding boundary at an angle less than θ_c will not undergo total internal reflection and finally will be lost. However at an angle greater than θ_c, a ray will propagate inside the core by a series of internal reflections.
In Fig. 2 at the point P_1 it holds that:

$$n_o \sin A_c = n_1 \sin (90-\theta_c) \quad (1)$$

Also at the point P_2:

$$n_1 \sin \theta_c = n_2 \sin (90) = n_2 \text{ or } \Theta_c = \arcsin(n_2/n_1) \quad (2)$$

Together they give:

$$n_0 \sin A_c = n_1 \cos \theta_c = (n^2_1 - n^2_2)^{1/2} = NA, \text{ or}$$

$$A_c = \arcsin(1-\cos^2\Theta_c)^{0.5} = \arcsin(n_1 \sin\Theta_c). \quad (3)$$

NA is the numerical aperture of the SI fiber and is defined as the light-gathering power of an optical fiber. When the face of the fiber is in contact with air ($n_0 = 1$ for air), $NA = \sin\theta_c$. When $n_2/n_1 = 0.99$, then θ_c is 8.1^o and A_c is 12.2^o.

When total internal reflection occurs, there is also light transmission in the gladding, the evanescent wave (a very nearby standing wave). This can cause light leakage between two adjacent fibers even when the diameter of a fiber is many times greater than the wavelength. In SMFs, the energy transmitted via the evanescent wave is a significant fraction.
In SI fibers, the light rays zigzag in straight lines between the core/cladding on each side of the fiber axis. In GRIN fibers, the light travels in a curved trajectory, always being refracted back to the axis of the fiber. At angles > Θ_c, light never reaches the outer edge of the fiber. At angles< Θ_c, the light enters the adjacent fiber, traverses the guide and is absorbed on the periphery of the guide as in the case of the SI guide.

Glass optical fibers are mostly made from silica (SiO_2) with a refractive index of about 1.5. Typically the difference between core and cladding is less than one percent. For medical applications, due to the required properties (optical quality, mechanical strength, and flexibility) also plastic optical fibers are used. Plastic fibers have the advantages of much simpler and less demanding after-processing and plastic fibers are lighter and of lower cost than glass fibers.
Plastic is common in step-index multimode fiber with a core diameter of 1 mm and have more propagation losses than glass fiber (1 dB/m or higher).

A fiber with a core diameter < $10 \cdot \lambda$ cannot be modeled using geometric optics, but must be analyzed as an electromagnetic structure, by solution of the electromagnetic wave equation, which describes the propagation of electromagnetic waves (see textbooks on physics).
The number of vibration modes in a SI MMF can be found from the V number:

$$V = (2\pi r/\lambda) (n^2_1 - n^2_2)^{1/2}, \quad (4)$$

where r is the core radius and λ wavelength. When $n_0 = 1$, than V becomes $(2\pi r/\lambda)NA$. When $V < 2.405$ only the fundamental mode remains and so the fiber behaves as a SMF.
The electromagnetic analysis may also be required to understand behaviors such as speckles that occur when coherent light (same frequency and intensity) propagates in a MMF. (A speckle pattern is a random intensity pattern produced by the mutual interference of coherent wave fronts that are subject to phase differences and/or intensity fluctuations. See also Huygen's Principle.) Speckles occur in optical coherence tomography and laser Doppler imaging.

A new type of crystals, photonic crystals led to the development of photonic crystal fiber (PCF). (Photonic crystals are periodic optical (nano)structures that are designed to affect the motion of photons in a similar way that periodicity of a semiconductor crystal affects the motion of electrons.) These fibers consist of a hexagonal bundle of hollow microtubes embedded in silica with in the center the fiber of photonic crystal. A PCF guides light by means of diffraction from a periodic structure, rather than total internal reflection. They can carry higher power than conventional fibers.

Fluorescence

Principle

Fluorescence, as other types of luminescence, is mostly found as an optical phenomenon in cold bodies (in contrast to incandescence, a process with a flame), in which a molecule absorbs a high-energy photon, and re-emits it as a lower-energy photon with a longer wavelength (λ). The energy difference between the absorbed and emitted photons ends up as molecular vibrations, finally in the form of heat. Usually the absorbed photon is in the UV, and the emitted light is in the visible range, but this depends on the absorbance curve and the shift to higher emitted λ (Stokes shift: Δλ) of the particular fluorophore (the molecule with the fluorescent structure).
The process can be described by:

$$S_1 \rightarrow \ S_2 + h/\lambda_{emitted}.$$

The system starts in state S_1, and after the fluorescent emission of a photon with energy hv, it is in state S_2 where h is Planck's quantum mechanical constant, being $6.626 \cdot 10^{-34}$ Js.

Applications

There are many natural and synthetic compounds that exhibit fluorescence, and they have a number of medical, biochemical and industrial applications (fluorescent lighting tubes).
The fluorophore attached by a chemical reaction to bio-molecules enables very sensitive detection of the molecule.

Examples are:

Automated sequencing of DNA by the chain termination method; each of four different chain terminating bases has its own specific fluorescent tag. As the labeled DNA molecules are separated, the fluorescent label is excited by a UV source, and the identity of the base terminating the molecule is given by the wavelength of the emitted light.
DNA detection The compound ethidium bromide, when free to change its conformation in solution, has very little fluorescence. Ethidium bromide's fluorescence is greatly enhanced when it binds to DNA, so this compound is very useful in visualizing the location of DNA fragments in agarose gel electrophoresis (see Electrophoresis).

The DNA microarray.

Immunology and immonohistochemistry An antibody has a fluorescent chemical group attached, and the sites (e.g., on a microscopic specimen) where the antibody has bound can be seen, and even quantified, by fluorescence.

FACS, fluorescent-activated cell sorting.

Fluorescence resonance energy transfer and similar techniques has been used to study the structure and conformations of DNA and proteins. This is especially important in complexes of multiple biomolecules.

Calcium imaging Aequorin, from the jellyfish *Aequorea victoria*, produces a blue glow in the presence of Ca^{2+} ions (by a chemical reaction). Other fluorescent dyes are calcium orange and the intracellular indicator Indo-1. It has been used to image calcium in cells in real time, especially in neurobiological applications. This technique has a long history in research of hippocampus slices. Imaging at the light microscopic and confocal level (see Confocal microscopy) is also used to explore the contribution of inward calcium currents and calcium release in relation to synaptic transmission in neurons. Specific applications are analyses of neuronal networks and synaptic plasticity, often studied with the patch-clamp technique and voltage clamp technique. This techniques may use the voltage sensitive Ca^{2+} dyes Fluo, Ca-green en Fura.

More Info

The success with aequorin has led to the discovery of Green Fluorescent Protein (GFP), an important research tool. GFP and related proteins are used as reporters for any number of biological events including sub-cellular localization. Levels of gene expression are sometimes measured by linking a gene for GFP production to another gene. Fluorescent calcium indicator proteins [FCIPs]) are Ca^{2+}-sensitive GFP varians. Fig. 1 shows an example of light-evoked Ca^{2+} responses in retinal ganglion cells.

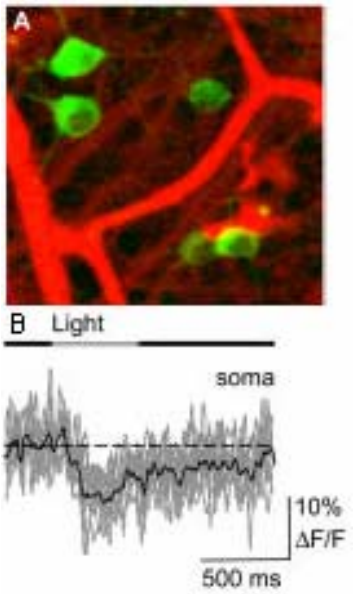


Fig. 1. Intact, light-sensitive retinal whole mount. *A* Blood vessels are red and active retinal ganglion cells are green (FCIP-positive). *B* Light-stimulus-evoked Ca^{2+} response (black trace; gray traces are single trials) measured in the soma with ΔF/F the relative fluorescence changes. After *PLoS Biol.* 2004; 2(6): e163.

Also, many biological molecules have an intrinsic fluorescence that can sometimes be used without the need to attach a chemical tag. Sometimes this intrinsic fluorescence changes when the molecule is in a specific environment, so the distribution or binding of the molecule can be measured. Biliburin, for instance, is highly fluorescent when bound to a specific site on serum albumin. Zinc protoporphyrin, formed in developing red blood cells instead of hemoglobin when iron is unavailable or lead is present, has a bright fluorescence and can be used to detect these abnormality.

Fluoroscopy

Principle

Fluoroscopy is an imaging technique commonly used to obtain real-time images of the internal structures. In its simplest form, a fluoroscope consists of an X-ray source and fluorescent screen between which the object, the patient, is placed. Modern fluoroscopes couple the screen to a CCD video camera allowing still images or images to be played on a monitor or an X-ray image intensifier for digital imaging. (A CCD camera is a charge-coupled device is an image sensor, consisting of an integrated circuit containing a matrix of coupled, light-sensitive capacitors).

Fig. 1 A modern fluoroscope

The x rays are attenuated in dependence on the type of structure of the body. They cast a shadow of the structures on the fluorescent screen. Images on the screen are produced as the unattenuated x rays are absorbed by atoms, which process gives rise to the emission of free electrons with a high kinetic energy (the photoelectric effect). While much of the energy given to the electrons is dissipated as heat, a fraction of it is given off as visible light by exiting atoms in the fluorescent molecules. Then, by ''de-excitation'' light is emitted (see Fluorescence), the fluorescent process and this produce the image.

Application

Common fields of application are the gastrointestinal tract (including administration of barium, and enteroclysis), orthopedic surgery (operation guidance), angiography of the leg, heart and cerebral vessels, urological surgery (e.g. retrograde pyelography), implantation of cardiac rhythm devices (pacemakers, implantable cardioverter defibrillators and cardiac resynchronization devices).

Risks
The risk on radiation damage by ionizing should be balanced with the benefits of the procedure to the patient. Although the length of a typical procedure often results in a relatively high absorbed dose, digitization of the images captured and flat-panel detector systems has reduced the radiation dose. Radiation doses to the patient depends especially on length of the procedure, with typical skin dose rates quoted as 20-50 mGy/min (Gy is Gray, the applied dose. 1 Gy is 1 J/kg tissue). Because of the long length of some procedures, in addition to standard cancer-inducing stochastic radiation effects, deterministic radiation effects have also been observed ranging from mild erythema, equivalent of a sun burn, to more serious burns. While deterministic radiation effects are a possibility, they are not typical of standard fluoroscopic procedures. Most procedures sufficiently long in length to produce radiation burns are part of necessary life-saving operations.

More Info

X-ray Image Intensifiers
At present, the original X-ray image intensifiers are replaced by CCD cameras or modern image intensifiers, which no longer use a separate fluorescent screen. Instead, a cesium iodide phosphor is

deposited directly on the photocathode of the intensifier tube. The output image is approximately 10^5 times brighter than the input image. This *brightness gain* is comprised of a *flux gain* (amplification of photon number) and *minification gain* (concentration of photons from a large input screen onto a small output screen). Each of them approximates a gain of a factor of 100. This gain is such that quantum noise, due to the limited number of X-ray photons, is now a significant factor limiting image quality.

Flat-panel detectors
Also flat-panel detectors replace the image intensifier in fluoroscope design. They have increased sensitivity to X-rays, and therefore reduce patient radiation dose. They have also a better temporal resolution, reducing motion blurring. Contrast ratio is also improved: flat-panel detectors are linear over a very wide latitude, whereas image intensifiers have a maximum contrast ratio of about 35:1. Spatial resolution is approximately equal.
Since flat panel detectors are considerably more expensive they are mainly used in specialties that require high-speed imaging, e.g., vascular imaging and cardiac catheterization.

Imaging concerns
In addition to spatial blurring factors, caused by such things as the Lubberts effect, (non-uniform response of an imaging system at different depths), K-fluorescence reabsorption (reabsorption in the K-orbit of the atom) and electron range, fluoroscopic systems also experience temporal blurring due to system lag. This temporal blurring has the effect of averaging frames together. While this helps reduce noise in images with stationary objects, it creates motion blurring for moving objects. Temporal blurring also complicates measurements of system performance for fluoroscopic systems.

Holography and mass image storage

Principle

Holography is an advanced form of photography that allows an image to be recorded in 3-D. This technique can also be used to optically store, retrieve, and process information.

Fig. 1 Identigram as a security element in a German Identity card (Personalausweis)

Several types of holograms can be made. The first holograms were "transmission holograms", which were viewed by shining laser light through them and looking at the reconstructed image of the other side. A later refinement, the "rainbow transmission" hologram allowed viewing by white light and is commonly seen today on *credit cards* as a security feature and on product packaging. These versions of the rainbow transmission holograms are now commonly formed as surface relief patterns in a plastic film, and they incorporate a reflective aluminum coating which provides the light from "behind" to reconstruct their imagery. Another kind of common hologram is the true "white-light reflection hologram" which is made in such a way that the image is reconstructed naturally using light on the same side of the hologram as the viewer.

Technical description
The difference between holography and photography is best understood by considering what a black and white photograph actually is: it is a point-to-point recording of the intensity of light rays that make up an image. Each point on the photograph records just one thing, the intensity (i.e. the square of the amplitude of the electric field) of the light wave that illuminates that particular point. In the case of a color photograph, slightly more information is recorded (in effect the image is recorded three times viewed through three different color filters), which allows a limited reconstruction of the wavelength of the light, and thus its color. Recent low-cost solid-state lasers are performed to make holograms

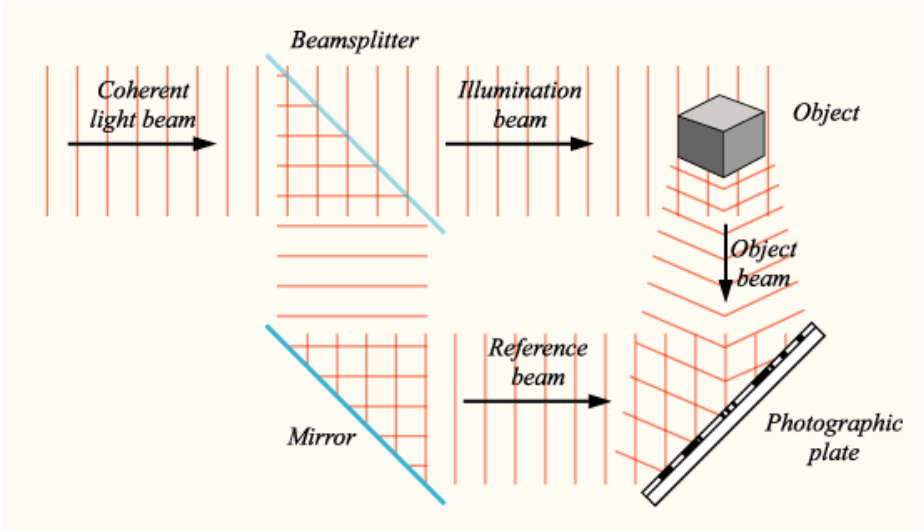


Fig. 1 Principle of making a hologram. (See Light: beam splitter for its working principle.)

Light, being a wave phenomenon, is characterized also by its phase. In a photograph, the phase of the light from the original scene is lost, and with it the three-dimensional effect. In a hologram, information from both the intensity and the phase is recorded. When illuminating the hologram with the appropriate light, it diffracts part of it into exactly the same wave (up to a constant phase shift invisible to our eyes) which emanated from the original scene, thus retaining the three-dimensional appearance. Also color holograms are possible.
Holographic recording process

To produce a recording of the phase of the light wave at each point in an image, holography uses a *reference beam* (see Fig. 1) which is combined with the light from the object (the *object beam*). If these two beams are coherent, optical interference (see Huygens'principle and Light: diffraction) between the reference beam and the object beam, due to the superposition of the light waves, produces a series of intensity fringes that can be recorded on standard photographic film. These fringes form a type of diffraction grating on the film, which is called the hologram. The central miracle of holography is that when the recorded grating is later illuminated by a substitute reference beam, the original object beam is reconstructed, producing a 3D image.

These recorded fringes do not directly represent their respective corresponding points in the space of a scene (the way each point on a photograph represents a single point in the scene being photographed). Rather, a small portion of a hologram's surface contains enough information to reconstruct the entire original scene, but only what can be seen from that small portion as viewed from that point's perspective. This is possible because during holographic recording, each point on the hologram's surface is affected by light waves reflected from all points in the scene, rather than from just one point. It is as if, during recording, each point on the hologram's surface were an eye that could record everything it sees in any direction. After the hologram has been recorded, looking at a point in that hologram is like looking "through" one of those eyes.
To demonstrate this concept, you could cut out and look at a small section of a recorded hologram; from the same distance you see less than before, but you can still see the entire scene by shifting your viewpoint laterally or by going very near to the hologram, the same way you could look outside in any direction from a small window. What you lose is the ability to see the objects from many directions, as you are forced to stay behind the small window.

Holographic reconstruction process
When the processed holographic film is illuminated once again with the reference beam, diffraction from the fringe pattern on the film reconstructs the original object beam in both intensity and phase (except for rainbow holograms). Because both the phase and intensity are reproduced, the image appears three-dimensional; the viewer can move his or her viewpoint and see the image rotate exactly as the original object would.

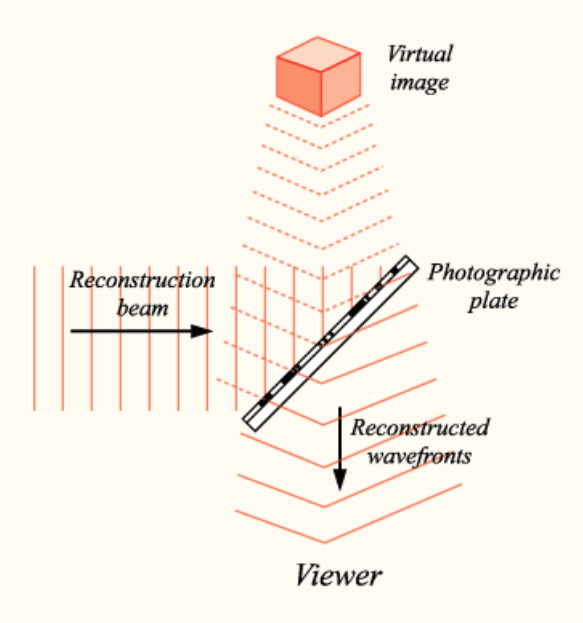


Fig. 2 Principle of reconstruction of the image.

It is possible to store the diffraction gratings that make up a hologram as phase gratings or amplitude gratings of various specific materials.

Application

High tech applications are numerous in (astro)physics, but large scale application in mass storing medical images etc. is coming soon. On bank notes, credit cards etc. it are normal safely features. Because of the need for coherent interference between the reference and object beams, Laser light is used to record holograms. But formerly other coherent light sources such as Hg-arc lamps, see Light: sources). In simple holograms, the coherence length of the beam determines the maximum depth the image can have. The coherence length L is:

$$L = \lambda^2/(n\,\Delta\lambda).$$

where λ is the central wavelength of the source, *n* is the refractive index of the medium, and Δλ is the spectral width of the source. In sunlight and incandescent holograms on credit cards have depths of a few mm. A good holography laser will typically have a coherence length of several meters.

Holography can be applied to a variety of uses other than recording images.
Holographic data storage stores information at high density inside crystals or photopolymers and has the potential to become the next generation of popular storage media with possibly 1 gigabit/s writing speed and 1 terabit/s readout speed. Four terabyte disks are nearly commercially available.
An alternate method to record holograms is to use a digital device like a CCD camera instead of a conventional photographic film. This approach is often called *digital* holography.

More Info

Dynamic holography
The discussion above describes static holography, with sequentially recording, developing and reconstructing. A permanent hologram is produced.
There exist also holographic materials which don't need the developing process and can record a hologram in a very short time (optical parallel processing of the whole image). Examples of applications of such real-time holograms include phase-conjugate mirrors ("time-reversal" of light), optical cache memories, image processing (pattern recognition of time-varying images) and optical computing.
The fast processing compensates the fact that the recording time. The optical processing performed by a dynamic hologram is much less flexible than electronic processing. On one side one has to perform the operation always on the whole image, and on the other side the operation a hologram can perform is basically either a multiplication or a phase conjugation. But remember that in optics, addition and Fourier transform (see Fourier analysis) are already easily performed in linear materials, the second simply by a lens. This enables some applications like a device that compares images in an optical way.

Holonomic brain theory
The fact that information about an image point is distributed throughout the hologram, such that each piece of the hologram contains some information about the entire image, seemed suggestive about how the brain could encode memories. The fact that spatial frequency encoding displayed by cells of the visual cortex was best described as a Fourier transform of the input pattern. Also the cochlea makes a Fourier transform. This holographic idea leaded to the term "holonomic".

Huygens' principle

Principle

The Huygens principle is a method of analysis applied to problems of wave propagation. This holds for macroscopic phenomena (optical devices such as lenses, prisms etc. very much larger than the wavelength). It recognizes that each point of an advancing wave front is in fact the center of a source of a new train of waves and that the advancing wave as a whole may be regarded as the sum of all the secondary waves arising from points in the medium already traversed.

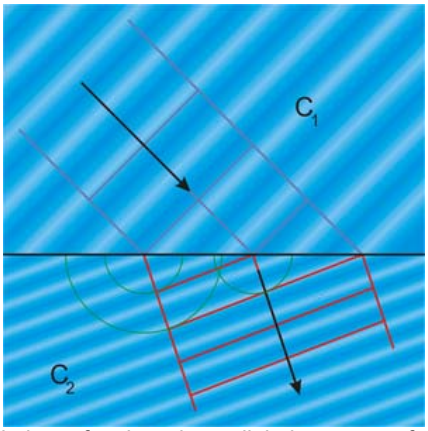


Fig. 1 Huygens' principle applied to refraction when a light beam goes from medium with a high speed of propagation C1 and consequently low refractive index to medium C2 with a low speed and high refractive index.

The Huygens' principle also holds for (near-)microscopic phenomena (optical devices such as apertures and slits of the order of a wavelength). It simply states that a large hole can be approximated by a collection of many small holes so each is practically a point source (whose contribution is easy to calculate). A point source generates waves that travel spherically in all directions). Similarly a relatively wide slit is composed of many narrow ones (subslits), and adding the waves produced by each produces the diffraction pattern (see Light: diffraction). For example, if two rooms are connected by an open doorway and a sound is produced in a remote corner of one of them, a person in the other room will hear the sound as if it originated at the doorway. As far as the second room is concerned, the vibrating air in the doorway is the source of the sound. The same is true of light passing the edge of an obstacle, but this is not as easily observed because of the short wavelength of visible light.

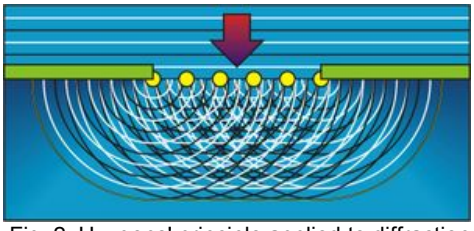


Fig. 2 Huygens' principle applied to diffraction.

The above views of wave propagation helps better understand a variety of wave phenomena, such as refraction and diffraction. The former is visualized in Fig. 1 and the latter in Fig. 2.

Lambert-Beer law

Principle

The Lambert-Beer law, also known as Beer's law or the Beer-Lambert-Bouguer law relates the absorption of light to the properties of the material through which the light is traveling.

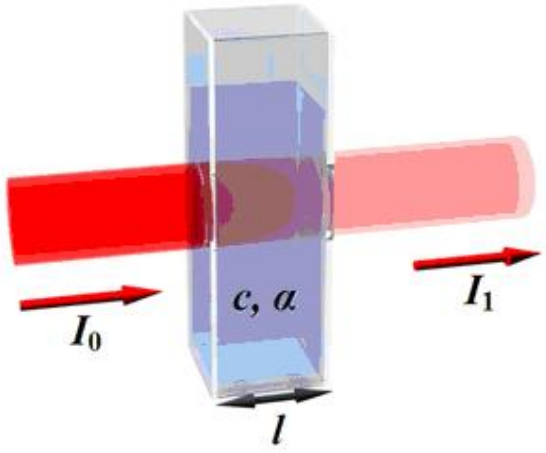


Fig. 1 Beer-Lambert absorption of a beam of light as it travels through a cuvette of size *l*.

There are several ways in which the law can be expressed:

$A = \alpha l c$,
$I_1/I_0 = e^{-\alpha/c}$, and
$A = -\log (I_1/I_0)$

with $\alpha = 4\pi k/\lambda$

where:
A is absorbance or extinction
I_0 is the intensity of the incident light
I_1 is the intensity after passing through the material
l is the distance that the light travels through the material (the path length)
c is the concentration of absorbing species in the material
α is the absorption coefficient or the molar absorptivity of the absorber, the logarithmic decrement per unit path length per mol
λ is the wavelength of the light
k is the extinction coefficient.

In essence, the law states that there is an exponential dependence between the transmission of light through a substance and the concentration of the substance, and also between the transmission and the length of material that the light travels through. Thus if l and α are known, the concentration of a substance can be deduced from the amount of light transmitted by it.
The units of *c* and α depend on the way that the concentration of the absorber is being expressed (see More info).

Application

The law's link between concentration and light absorption is the basis behind the use of Spectroscopy to identify substances.
Further the law is applied in many medical, scientific and technical instruments.

More info

If the material is a liquid, it is usual to express the absorber concentration *c* as a mole fraction (, the number of moles divided by the total number of moles in the liquid, so the mole fraction is

dimensionless. The units of α are thus reciprocal length (e.g. cm^{-1}). In the case of a gas, *c* may be expressed as a density (units of reciprocal length cubed, e.g. cm^{-3}), in which case α is an *absorption cross-section* and has units of length squared (e.g. cm^2). If concentration *c* is expressed in moles per unit volume, α is a molar absorptivity usually given in units of mol cm^{-2}. In spectroscopy, often extinction (E = c/ k) is used.
The value of the absorption coefficient α varies between different absorbing materials and also with wavelength for a particular material. It is usually determined by experiment.
The law tends to break down at very high concentrations, especially if the material is highly scattering (see Light scattering). If the light is especially intense, nonlinear optical processes can also cause variances.

Laser

Principle

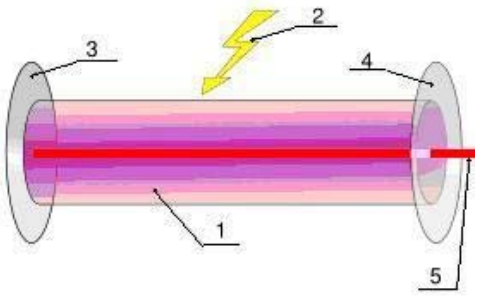
A laser (**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation) is an optical source that emits photons in a coherent beam (i.e. when the beam is split in many tiny waves, actually consisting of a single photo, they are all in phase, have the same polarization and form together a single well-defined wave front, see Waves). Laser light is typically near-monochromatic, i.e. consisting of a single wavelength or color, and emitted in a narrow beam. This is in contrast to common light sources, such as the incandescent light bulb, which emit incoherent photons in almost all directions, usually over a wide spectrum of wavelengths. Laser action is based on quantum mechanics and thermodynamics theory.

Application

In industry, technology, science and consumer electronics etc. the number of applications is enormous. In medicine, the laser scalpel is used for surgical techniques. Lasers are also used for dermatological procedures including removal of birthmarks, tattoos etc. Laser types used in dermatology include ruby (694 nm), alexandrite (755 nm), pulsed diode array (810 nm), and the various YAG laser. The YAG lasers emit infrared light. The YAG lasers are also applied in ophthalmology.
LASIK surgery is **La**ser-assisted in **si**tu **k**eratomileusis. First a low-power laser creates a topographic map of the cornea. Then an excimer laser (193 nm) is used to remodel the corneal stroma. The laser vaporizes tissue without causing damage.
Lasers are also used as the light delivery system in a fibro scope (see Fiber optics) for endoscopy.
Safety The coherence and low divergence of laser light means that it can be focused by the eye into an extremely small spot on the retina, resulting in localized burning and permanent damage in seconds or even faster. Lasers are classified into safety classes numbered I, inherently safe, to IV. Even scattered light can cause eye and/or skin damage. Laser products available for consumers, such as CD players and laser pointers are usually in class I, II, or III.

More info

A laser is composed of an active laser medium and a resonant optical cavity. Fig. 1 gives the principle of producing a laser beam.



1. Active laser medium, 2. Laser pumping energy, 3. Mirror, 4. Partial mirror, 5. Laser beam.

The gain medium is a material of controlled purity, size, and shape. The material itself is for instance a helium-neon gas or rubidium gas. The gain medium uses a quantum mechanical effect called stimulated emission to amplify the beam. This is the process by which, when perturbed by a photon, matter may lose energy resulting in the creation of another photon. The perturbing photon is *not* destroyed in the process (as with absorption), and the second photon is created with the same phase and frequency (or wavelength) as the original. These are the photons of the laser beam.
For a laser to operate, the gain medium must be "pumped" by an external energy source, such as electricity or light (from a classical source such as a flash lamp, or another laser). The pump energy is absorbed by the laser medium to produce exited states. (An exited state exists of an electron in a higher orbit than the lowest possible one, the ground state, so with higher energy than the ground state that is more energy than the absolute minimum). The mirrors enable multiple reflection so that the photons remain for a longer time in medium. In this way more easily excited particles are created and the number of particles in one excited state considerably exceeds the number of particles in some lower state. In this condition, an energy providing optical beam passing through the medium produces more stimulated emission than stimulated absorption. The process can be thought of as *optical amplification*, and it forms the basis of the laser (or for radio waves the maser).

Light

Principle

Light is the part to the electromagnetic spectrum that is visible to the animal eye. The study of light and the interaction of light and matter is termed optics.
The elementary particle that defines light is the photon. The three basic dimensions of light (or better all electromagnetic radiation) are:

- intensity (or amplitude), which is related to the human perception of brightness of the light;
- frequency (or wavelength), perceived by humans as the color of the light, and
- polarization (or angle of vibration), which is only weakly perceptible by humans under ordinary circumstances.

Due to the wave-particle duality of matter light simultaneously exhibits properties of both waves and particles.
Here follows a description of the most important features of light.

Speed of light The speed of light in a vacuum is exactly 299,792,458 m/s (fixed by definition).
Refraction When light goes from the one to another medium, it is refracted (see Light: refraction) and reflected (see Fresnel equations).
Dispersion Since refraction is frequency dependent, the refracted beam is decomposed in its various frequencies (or wavelengths) which all have their own angle of refraction. The classical way to achieve this is with a prism, see Fig. 1.

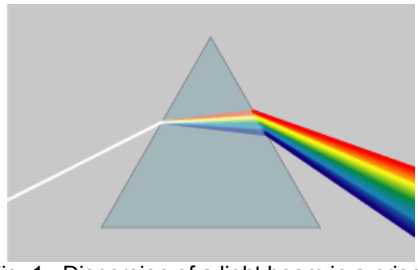


Fig. 1 Dispersion of a light beam in a prism.

The visible spectrum (see Fig. 1).
Electromagnetic radiation from 400 to 700 nm is called visible light or simply light. However, some people may be able to perceive wavelengths from 380 to 780 nm. A light-adapted eye typically has its maximum sensitivity at around 555 nm, which is in the green region (see: Luminosity function). In day light, so with photopic vision, the different wavelengths are detected by the human eye and then interpreted by the brain as colors. The spectrum does not, however, contain all the colors that the human eyes and brain can distinguish. For instance, brown and pink are absent. See Color vision to understand why.
The *optical spectrum* includes not only visible light, but also ultraviolet (UV) at the short wavelength (high frequency) end and infrared (IR) at the long wavelength end. Some animals, such as bees, can see UV radiation while others, such as pit viper snakes, can see IR light.

Fig. 1 The part of the optical spectrum visible to the human eye.

Polarization With reflection and refraction light is also polarized to some extent (see Light: polarization). Polarization describes the direction of the electric oscillation in the plane perpendicular to the direction of propagation.
Diffraction This refers to phenomena associated with wave propagation, such as the bending, spreading and interference of waves emerging from an aperture on the order of the wavelength (pinhole and narrow-split experiments). For more explanation see Light: diffraction.
Absorption When light propagates trough a medium some of its energy is absorbed by the medium (see Lambert-Beer law). In general, all or most of the absorbed energy is transformed to heat. The part that is not transformed to heat can be emitted as radiation (see Chemoluminescence and

Bioluminescence, Fluorescence, Phosphorescence) or transformed to electric current (the photoelectric effect, see **More Info**).
Scattering Scattering of is a process whereby light (and sound or moving particles), are forced to deviate from a straight trajectory by one or more localized non-uniformities in the medium through which it passes. This also includes deviation of reflected radiation from the angle predicted by the law of reflection (called *diffuse* reflections). An example is scattering of light in the eye lens and intraretinal scatter. See further Light: Scattering.

Theories about light

Classical particle theory (Newton)
Light was assumed to be composed of corpuscles (particles of matter) which were emitted in all directions from a source. This theory cannot explain many of the properties of light. It wrongly assumed a higher speed in a denser medium. The classical particle theory was finally abandoned around 1850.

Classical wave theory (Huygens)
Light was (and is) assumed to be emitted in all directions as a series of waves in a medium (Fig. 3). As waves are not affected by gravity, it was assumed that they slowed down upon entering a denser medium. It can explain phenomena such as refraction, polarization, dispersion and diffraction. It was wrongly assumed that light waves would need a medium for transmission (like sound waves indeed need).

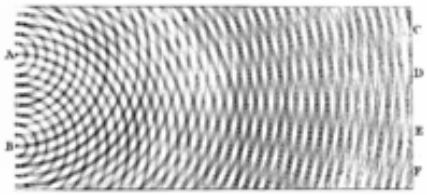


Fig. 3 Interference of the waves emitted by two sources.

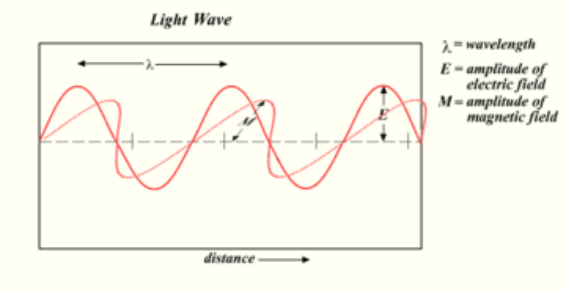


Fig. 4 A linearly-polarized light wave frozen in time and showing the two oscillating components of light; an electric field and a magnetic field perpendicular to each other and to the direction of motion.

Electromagnetic theory
The angle of polarization of a beam of light as it passed through a polarizing material could be altered by a magnetic field, an effect now known as Faraday rotation. It is one of the arguments that light is a high-frequency electromagnetic vibration, which could propagate even in the absence of a medium such as the “ether”. The frequency range of light is only a very small part of the whole electromagnetic range. Other parts of the electromagnetic spectrum are applied in e.g. radio, radar, television, electromagnetic imaging (see Spectroscopy).

Application

Medical Many medical instruments and apparatus are based on light for imaging, as do also prostheses like spectacles etc. All this things will not be discussed. Here some applications based on UV and IR light are mentioned. .
UV radiation is not normally directly perceived by humans except in a much delayed fashion, as overexposure of the skin. UV light can cause sunburn, or skin cancer. Underexposure can cause vitamin D deficiency. However, because UV is a higher frequency radiation than visible light, it very easily can cause materials to fluorescence visible light.
Thermography is performed with a camera using IR light. In general heating of the skin or the whole body by radiation is caused by IR light. However, any intense radiation can have the same effect. Other examples are UV and IR spectroscopy (see Spectroscopy).

Technical IR cameras convert IR light to visible light. Depending on their application we distinguish night-vision binoculars, cameras. These are different from image intensifier cameras, which only amplify available visible light.

More info

The Special Theory of Relativity
The wave theory explains nearly all optical and electromagnetic phenomena, but some anomalic phenomena remained that could not be explained:
- the constant speed of light,
- the photoelectric effect,
- black body radiation.

The constant speed of light contradicted the mechanical laws of motion, which stated that all speeds were relative to the speed of the observer. This paradox was resolved by revising Newton's laws of motion into Einstein's special theory of relativity.
The photoelectric effect, being the ejection of electrons when light strikes a metal surface, causing an electric current to flow out. The explanation is given by the wave-particle duality and quantum mechanics.
A third anomaly involved measurements of the electromagnetic spectrum emitted by thermal radiators, or so-called black bodies (see Wien's displacement law and Body heat dissipation and related water loss). The explanation is given by the *Quantum theory*. The theory of black body radiation says that the emitted light (and other electromagnetic radiation) is in the form of discrete bundles or packets of energy. These packets were called quanta, and the particle of light was given the name photon, just as other particles, such as an electron and proton. A photon has an energy, E, proportional to its frequency, f:

$$E = hf = hc/\lambda,$$

where h is Planck's constant (= $6{,}623 \cdot 10^{-34}$ Js), λ is the wavelength and *c* is the speed of light. Likewise, the momentum (mass times speed) p of a photon is also proportional to its frequency and inversely proportional to its wavelength:

$$p = E/c = hf/c = h/\lambda.$$

Wave-particle duality and of *quantum electrodynamics*
The modern theory that explains the nature of light is the wave-particle duality, founded by quantum theory. More generally, the theory states that everything has both a particle nature and a wave nature, and various experiments can be done to bring out one or the other. The particle nature is more easily discerned if an object has a large mass, but also particles, such as electrons and protons exhibited wave-particle duality. The quantum mechanical theory of light and electromagnetic radiation culminated with the theory of quantum electrodynamics, or QED.

Light: beam splitter

Principle

A beam splitter is an optical device that splits a beam of light in two. In its most common form, it is a cube, made from two triangular glass prisms, which are glued together by resin (Fig. 1). The thickness of the resin layer is adjusted such that (for a certain wavelength) half of the light incident through one "port" (i.e. face of the cube) is reflected and the other half is transmitted. Polarizing (see Light: polarization) beam splitters, use birefringent materials (two instead of one refractive index), splitting light into beams of differing polarization.

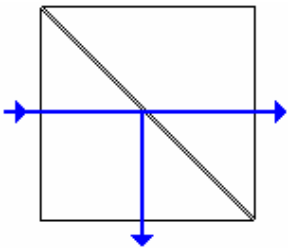


Fig. 1 Schematic representation of a beam splitter cube

Another design is the use of a *half-silvered mirror*. This is a plate of glass with a thin coating of aluminum. The coating is such thick that light incident at 45^{o} is half transmitted and half reflected.

Application

Beam splitters are found in all kind of (medical) equipment, especially in microscopes, spectroscopes, instruments with laser and ophthalmic instruments.

More Info

A third version of the beam splitter is a trichroic mirrored prism assembly which uses trichroic optical coatings to split the incoming light into three beams, one each of red, green, and blue (see Dichroism). Such a device was used in multi-tube color television cameras and in 3-color film movie cameras. Nowadays it is often applied in CCD cameras. Other applications are LCD screens and LCD projectors.

Light: diffraction

Principle

Diffraction refers to various phenomena associated with wave propagation, such as the bending, spreading and interference of waves emerging from an aperture. It occurs with any type of wave, including sound waves, water waves, electromagnetic waves such as light and radio waves, and matter displaying wave-like properties according to the wave–particle duality (see Light). While diffraction always occurs, its effects are generally only noticeable for waves where the wavelength is on the order of the size of the diffracting object (e.g. slit) or aperture.
Much theory has been developed to describe diffraction patterns of a pinhole and of one, tow and more slits. They all rely on the Huygens' principle. Fig. 1 gives the pattern of a single slit.
In slit experiments, narrow enough slits can be analyzed as simple wave sources. A slit reduces a wave 3-D problem into a simpler 2-D problem.

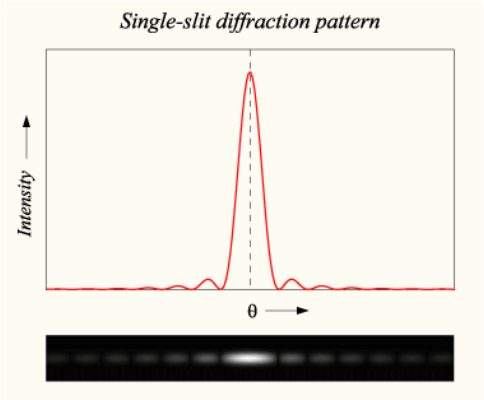


Fig. 1 Graph and image of single-slit diffraction

It is possible to calculate exactly the intensity of the diffraction pattern as a function of angle in the case of single-slit diffraction. Since the angles of the maxima and minima in the pattern are dependent on the wavelength λ, diffraction gratings impart angular dispersion (decomposed in wavelengths or “colors”, see Light) on a beam of light.

Application

There are many applications in specific fields of physics, and technology, and so indirectly in medicine. Diffraction plays a major role in light- and electron microscopy.
Diffraction is used in X-ray crystallography, called *Bragg diffraction*, (see **More Info**) to deduce the structure of a crystal from the angles at which X-rays are diffracted from it. The most common demonstration of Bragg diffraction is the spectrum of colors seen reflected from a compact disc: the closely-spaced tracks on the surface of the disc form a diffraction grating, and the individual wavelengths of white light are diffracted at different angles from it.

More info

The angels of the maxima and minima are inversely proportional to the slit or aperture diameter. So, the smaller the diffracting object the 'wider' the resulting diffraction pattern. When the diffracting object is repeated, for example in a diffraction grating the effect is to create narrower maximum on the interference fringes.
It is mathematically easier to consider the case of so called far-field or *Fraunhofer diffraction*, where the slits or pinholes are far from the point where the pattern is observed, so in the far-field, i.e. distance >> λ. The more general case is known as near-field or Fresnel diffraction, and involves more complex mathematics. Here, far-field diffraction is considered, which is commonly observed in nature.
Diffraction through a circular aperture
For diffraction through a circular aperture, there is a series of concentric rings surrounding a central disc (together called the Airy pattern illustrated in Fig. 2)).
A wave does not have to pass through an aperture to diffract. Any beam of light undergoes diffraction and spreads in diameter. This effect limits the minimum size d of a spot of light formed at the focus of a lens, known as the diffraction limit:

$$d = 1.22\,\lambda f/a$$

where λ is the wavelength of the light, f is the focal length of the lens, and a is the diameter of the beam of light, or (if the beam is filling the lens) the diameter of the lens. The spot contains about 70% of the light energy.
By use of Huygens' principle, it is possible to compute the diffraction pattern of a wave from any shaped aperture. If the pattern is observed at a sufficient distance from the aperture, in the far-field, it will appear as the two-dimensional Fourier transform of the function representing the aperture.

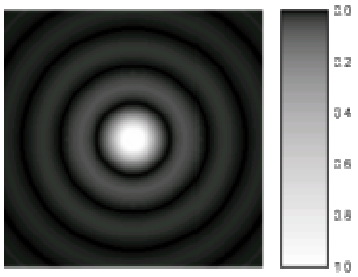


Fig. 2 The image of an Airy pattern. (The grey scale intensities have been adjusted to enhance the brightness of the outer rings of the pattern.)

Diffraction through slits
Fig. 3 illustrates the principle of interference with two slits In the far-field description. The two “rays” passing the two slits of Fig. 3 are supposed to be parallel (the prerequisite of the Fraunhofer diffraction). In this case they have an angle such that there is a path length difference of 0.5 λ. This creates the first minimum in the interference pattern, since the both waves extinct each other (destructive interference). Fig. 4 gives all maxima (with crossings of both spherical wave fronts) and minima (black crossings). The angular positions of the multiple-slit minima correspond to path length differences of an odd number of half wavelengths:

$$a\sin\theta = \tfrac{1}{2}\lambda(2m+1),$$

where m is an integer that labels the order of each minimum, a is the distance between the slits and θ is the angle for destructive interference. The maxima are at path differences of an integer number of wavelengths:

$$a\sin\theta = m\lambda.$$

We can calculate that for three waves from three slits to cancel each other the phases of slits must differ 120^0, thus path difference from the screen point to slits must be λ/3, and so forth. In the limit of approximating the single wide slit with an infinite number of subslits the path length difference between edges of slit must be exactly λ to get complete destructive interference (and so a dark stripe on the screen).

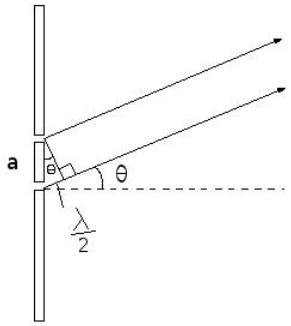


Fig. 3 Principle of interference. The slit are perpendicular on the plane of the screen.

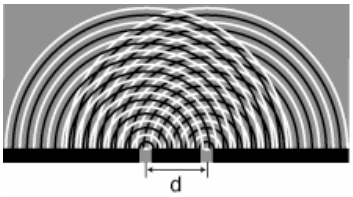


Fig. 4 Double-slit diffraction and interference pattern. The slits are perpendicular on the plane of the paper.

Bragg diffraction
Diffraction from multiple slits, as described above, is similar to what occurs when waves are scattered from a periodic structure, such as atoms in a crystal or rulings on a diffraction grating. Each scattering center (e.g., each atom) acts as a point source of spherical wave fronts; these wave fronts undergo constructive interference to form a number of diffracted beams. The direction of these beams is described by *Bragg's law*:

$$m\lambda = 2L\sin\theta$$

where L is the distance between scattering centers and m is an integer known as the *order* of the diffracted beam.

Diffraction of particles
The experimental prove of the diffraction of "particles," such as electrons, was one of the powerful arguments in favor of quantum mechanics and the existence of the wave-particle duality. The wavelengths of these particle-waves are small enough that they are used as probes of the atomic structure of crystals.

Light: Fresnel equations

Principle

The Fresnel equations, describe the behavior of light when moving between media of differing refractive indices (see Light: refraction). The ***reflection*** of light that the equations predict is known as Fresnel reflection and the refraction is described by Snell's law.
When light moves from a medium of a given refractive index n_1 into a second medium with refractive index n_2, both reflection and refraction of the light may occur.

The fraction of the intensity of incident light that is reflected from the interface is given by the *reflection coefficient R*, and the fraction refracted by the *transmission coefficient T*. The Fresnel equations assume that the two media are both *non-magnetic*.

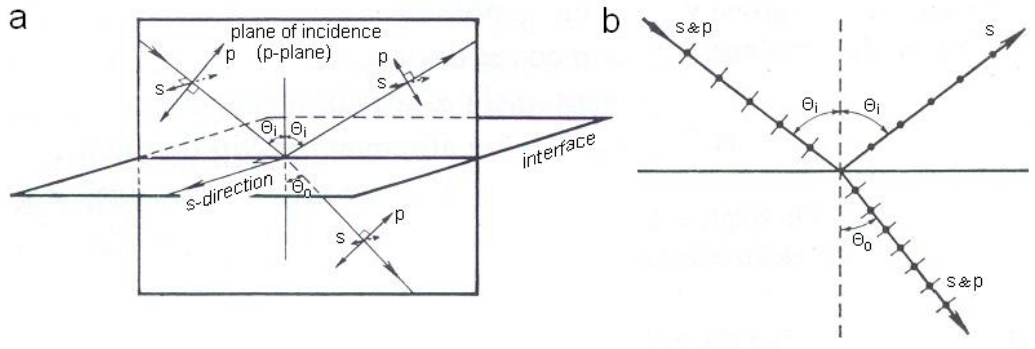


Fig. 1. a. Visualization of the decomposition of the unpolarized beam is the p and s-polarized parts. b. The angle of incidence is Brewster's angle. The reflected beam is purely s-polarized.

The calculations of R and T depend on the polarization of the incident ray. If the light is polarized with the electric field of the light perpendicular to the plane of Fig. 1a (s-polarized), the reflection coefficient is given by R_s. If the incident light is polarized in the plane of the diagram (p-polarized), the R is given by R_p.
A graphical example of the calculation of R_s and R_p is given in Fig. 2.

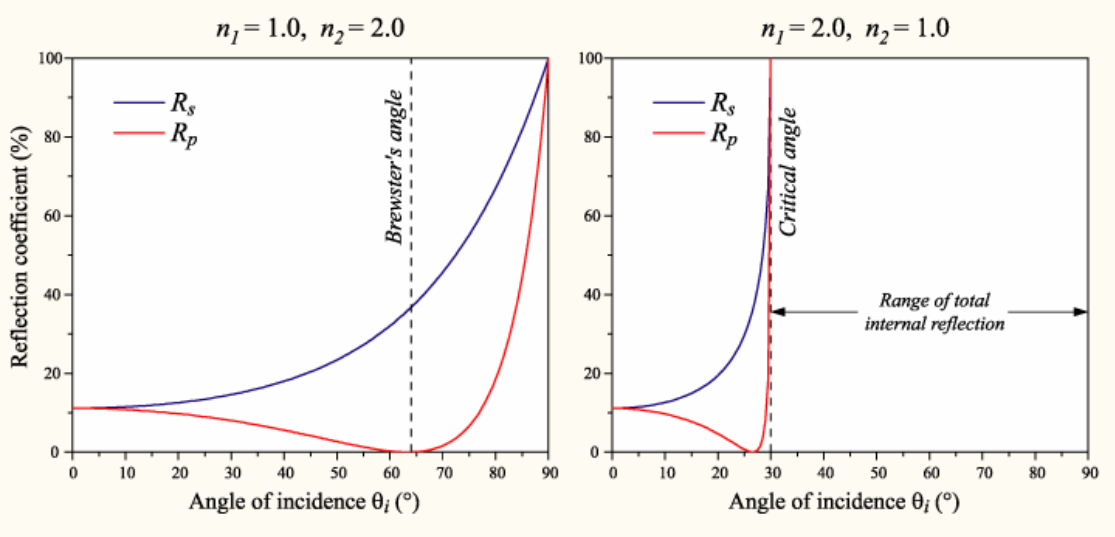


Fig. 2 Left, R_s and R_p for a light beam going from a medium with a low refractive index to one with a high index. Right, R_s and R_p for a light beam going from high to low refractive index.

At one particular angle for a given n_1 and n_2, the value of R_p goes to zero and a s-polarized incident ray is purely refracted. This angle is known as Brewster's angle, and is around 56° for a glass medium in air or vacuum.
When moving from a more dense medium into a less dense one (i.e. $n_1 > n_2$), exceeding an incidence angle known as the critical angle, all light is reflected and $R_s = R_p = 1$. This phenomenon is known as total internal reflection (see Snell's law). The critical angle is approximately 41° for glass in air.
If the incident light is unpolarized (containing an equal mix of s- and p-polarizations), the reflection coefficient is $R = (R_s + R_p)/2$.

When the light is about perpendicular to the interface ($\theta_i \approx \theta_t \approx 0$), R and T are given by:

$R = \{(n_1 - n_2)/(n_1 + n_2)\}^2$, and
$T = 1 - R$.

For common (clean!) glass, R is about 4%. Note that reflection by a window is from the front side as well as the back side, and that some of the light bounces back and forth a number of times between the two sides. The combined reflection coefficient for this case is 2R/(1 + R).

Application

Applying Brewster's angle is a way to produce pure polarized light. Repeated reflection and refraction on thin, parallel layers is responsible for the colors seen in oil films on water, used in optics to make reflection free lenses and perfect mirrors, etc.

More info
The general equations for R_s and R_p are:

$$R_s = \left(\frac{n_1 \cos(\Theta_i) - n_2 \cos(\Theta_t)}{n_1 \cos(\Theta_i) + n_2 \cos(\Theta_t)} \right)^2$$

and:

$$R_t = \left(\frac{n_1 \cos(\Theta_t) - n_2 \cos(\Theta_i)}{n_1 \cos(\Theta_t) + n_2 \cos(\Theta_i)} \right)^2$$

where θ_t can be derived from θ_i by Snell's law.

In each case T is given by $T_s = 1 - R_s$ and $T_p = 1 - R_p$.
R and S correspond to the ratio of the *intensity* of the incident ray to that of the reflected and transmitted rays. Equations for coefficients corresponding to ratios of the electric field amplitudes of the waves can also be derived, and these are also called "Fresnel equations". Here, we use squared amplitude (just as in the above equations n^2). Therefore, the intensity is expressed in Watts, more precisely in W/steradian (see Light: photometric and radiometric units of measure). The above equations are approximations. The completely general Fresnel equations are more complex (see e.g. .Wikipedia).

Light: the ideal lens

Principle

The ideal lens has the following properties:
- All light rays emanating from a point of light, an object point, at the object distance from the lens, are refracted by the lens through one point, the image point.
- Rays hitting the lens parallel to its optic axis (like the red one below) are refracted through its focus.
- Rays hitting the lens at its centre (the blue one) are not refracted.

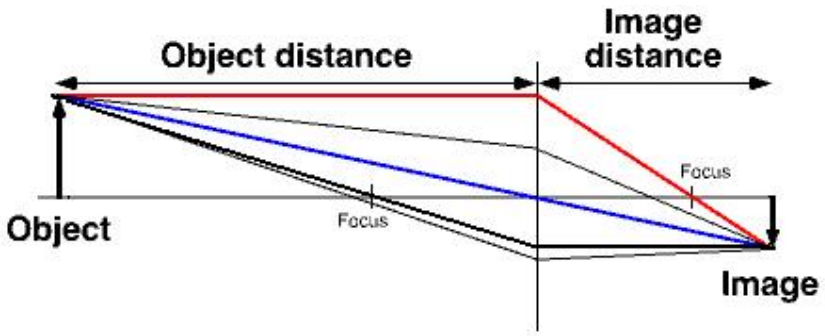


Fig. 1 Simple formal description of a positive lens, presented by a single principal plane (the vertical line). The image is real, i.e. can be made visible at a piece of paper.

The lens equation relates object distance, image distance and lens power:

$$1/d_{object} + 1/d_{image} = 1/f = \varphi, \quad (1)$$

where d is distance and φ lens power. It holds for *thin* lenses.

Note that if $d_{object} < f$, d_{image} becomes negative, the image is positioned on the same side of the lens as the object. This imaginary image cannot be projected on a screen, but it can be observed when looking from the right through the lens. A magnifying-glass creates this kind of image. The magnification is given by:

$$M = - d_{image}/d_{object} = f/(f - fd_{object}), \quad (2)$$

where M is the magnification factor. If |M|>1, the image is larger than the object. Notice the sign convention here shows that, if M is negative, as it is for real images, the image is upside-down with respect to the object (Fig. 1) and smaller. For virtual images, M is positive, the image is upright and larger than the object.
In(1), negative lenses 1/f obtains a minus sign. A negative lens cannot make an image at all. It is always practiced together with a positive lens, generally to compensate for lens errors (see **More Info**).
When two ideal lenses are put next to each other, their powers add. In formula: $\varphi = \varphi_1 + \varphi_2$. This does not apply if the lenses are not close together. Then, $\varphi = \varphi_1 + \varphi_2 - d_{12}\varphi_1\varphi_2$ where d_{12} is the distance between both lenses.
Real lenses do not behave exactly like ideal lenses but thin lenses are fairly ideal. Therefore, the term thin lens is often used as a synonym for ideal lens. However, the most ideal lenses available in practice (such as camera and microscope objectives) are often very thick, being composed of several lenses with different refractive indices.

Application

A lens shows its most ideal behavior when only its central part is used. An example is the photopic adapted eye with its very large φ but small pupil, since good imaging is only needed at the fovea centralis. Therefore, ideal lens optics can be useful in elementary eye optics (see Optics of the eye).

Lenses are applied in innumerous medical apparatus and research equipment.
Lenses and the theory of lenses play a fundamental role in all versions of light microscopy and in ophthalmology.

The human cornea-lens system has comparatively small lens errors (see below), seen its enormous refractive power (about 58 dptr). Only astigmatism can be corrected well.

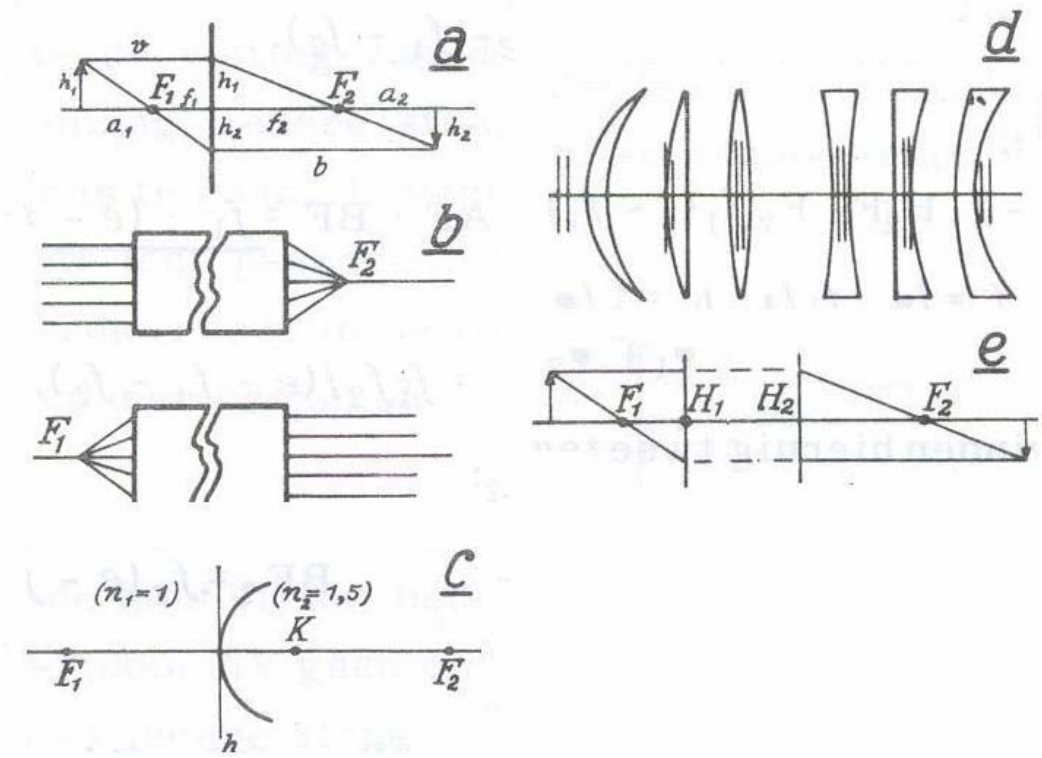


Fig. 2 Principle planes a the lens concept with one main plane, b and e with two main planes, c two media separated by a spherical interface, for instance the cornea (with K the center of the sphere. K is also called the principal point). An object in the medium with n=1.5 produces a diverging beam at the left. This is related to the fact that $F_2 > F_1$. d position of the two main planes for various types of lenses.

More info

Strenght
If thickness of the lens, *d* is small compared to the both radii of curvature, r_1 and r_2, then the thin lens approximation can be made, and *f* is then estimated as:

$$1/f = (n_{lens}/n_{medium} - 1)(1/r_1 - 1/r_2)$$

Principal planes
A more precise description is applying two principal planes. Fig. 2 illustrates how to handle this concept. Spectacle lenses are generally convex-concave (positive, left one Fig. 2e) or convex-convex (negative, right one Fig. 2e). As Fig. 2d indicates, their principal planes lay outside the lens itself. Therefore, for accurate calculations, the single-principal-plane concept is generally not adequate. The theory of the two principal planes, illustrated in Fig. 2e, holds for any system of lenses.

Lens errors
Even a thin lens does not produce an image *point* as suggested by Fig. 1, but a small blurred spot, which intensity falls off with eccentricity. This holds also for monochromatic light and the phenomenon is called spherical aberration, one of the monochromatic aberrations. The diameter of the circle increases with d^3.

Other monochromatic aberrations are:

- Coma This is an imperfection, which remains when spherical aberration is solved in a lens system. It is the comet-like blurring spot (with the tail pointing outwards) of the image of a parallel beam which hits the lens not-parallel to its optical axis.

- Astigmatism This happens when for two (perpendicular) planes through the optical axis the power φ is different. It produces an ellipse with blurred and sharp sectors when a circle is imaged.

The monochromatic aberrations are 3rd order errors, since their size is dependent on the summed powers of lens diameter d and object size h (e.g. d^2h), yielding power of 3.
Another error, occurring with large objects and large image angle is the curvature of the plane if the image. Depending on the system, the image is cushion shaped or ton shaped when displayed on a straight plane. The error is proportional with h^4.
A non-monochromatic error is chromatic aberration. This is caused by the wavelength dependency of the refractive index, called dispersion (see Light: refraction), resulting in a φ dependent on wavelength. Finally this results in an image for each wavelength with a shorter d_{image} the shorter the wavelength. It can be minimized in an achromatic doublet lens, consisting of two lenses of different types of glass.

Since various errors occur simultaneously the image of a single point produces a complicated figure as image.
When a small aperture is hold in front of a positive lens, then a parallel axial beam also produces a circular spot. This is due to the wave character of light. The diameter of the spot is distance between the two first order minima of the single-slit diffraction pattern:

$$d = 2.44(\lambda/d_{lens})(f/n_0),$$

where λ is the wavelength and n_0 the refractive index of the medium between lens and spot.

References
For more info see student textbooks of physics, Wikipedia chapters and books of optics.

Light: polarization

Principle

In electrodynamics, polarization is the property of electromagnetic waves, such as light, that describes the direction of their transverse electric field. More generally, the polarization of a transverse wave describes the direction of oscillation in the plane perpendicular to the direction of travel. Longitudinal waves, such as sound waves, (see Sound and Acoustics) do not exhibit polarization, because for these waves the direction of oscillation is along the direction of propagation.

The simplest manifestation of polarization to visualize is that of a plane wave, which is a good approximation to most light waves (a plane wave is a wave with infinitely long and wide wave fronts). All electromagnetic waves propagating in free space or in a uniform material of infinite extent have electric and magnetic fields perpendicular to the direction of propagation. Conventionally, when considering polarization, the electric field vector is described and the magnetic field is ignored since it is perpendicular to the electric field and proportional to it. The electric field vector may be arbitrarily divided into two perpendicular components labeled *x* and *y* with z indicating the direction of travel, see Fig. 1. The x vector is at the left (red) and the y-vector at the right (green). For a simple harmonic wave (sinus), the two components have the same frequency, but the amplitude and phase may differ. In the example of Fig. 1, the projection of the movement of the electric vector (blue in Fig. 1) on the horizontal x-y plane creates the purple line (the most simple Lissajous figure).

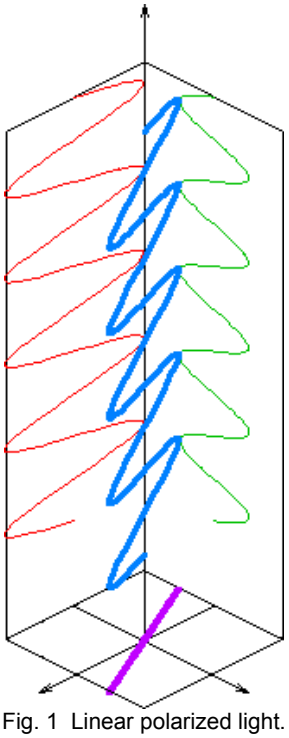


Fig. 1 Linear polarized light.

In Fig. 1, the two orthogonal (perpendicular) components are in phase. Since the ratio of the strengths of the two components is constant, the direction of the electric vector (the vector sum of these two components) is constant. And since the tip of the vector traces out a single line on the plane perpendicular on the direction of propagation, this special case is called *linear polarization*. The direction of this line depends on the relative amplitudes of the two components.

Application

Light reflected by shiny transparent materials is partly or fully polarized, except when the light is normal (perpendicular) to the surface. A polarizing filter, such as a pair of polarizing sunglasses, can be used to observe this by rotating the filter while looking through. At certain angles, the reflected light will be

reduced or eliminated. Polarizing filters remove light polarized at 90° to the filter's polarization axis. If two polarizers are placed atop one another at 90 ° angles to one another, no light passes through. Linear polarization is used in polarization microscopy.
Polarization by scattering is observed as light passes through the atmosphere. The scattered light produces the brightness and blue color in clear skies. This partial polarization of scattered light can be used to darken the sky in photographs, increasing the contrast (Fig. 2). This effect is easiest to observe at sunset, on the horizon at a 90 ° angle from the setting sun.

Fig. 2 Left unpolarized, right partly polarized. The effects of a polarizer on the sky in a color photograph. The right picture has the polarizer, the left does not.

Brightness of images of the sky and clouds reflected from horizontal surfaces are drastic reduced, which is the main reason polarizing filters are often used in sunglasses.
Rainbow-like patterns, visible through polarizing sunglasses, are caused by color-dependent *birefringent* (double refraction, see Snell's Law) effects, for example in toughened glass (e.g. car windows) or items made from transparent plastics. The role played by polarization in the operation of liquid crystal displays (LCDs) is also frequently apparent to the wearer of polarizing sunglasses, which may reduce the contrast or even make the display unreadable.
[
Biology
The naked human eye is weakly sensitive to polarization. Polarized light creates a very faint pattern near the center of the visual field, called Haidinger's brush. With sun light this pattern, a yellow horizontal figure 8 shaped spot with a bleu spot above and below the figure 8, is very difficult to see, but with practice one can learn to detect polarized light with the naked eye.
Many animals are apparently capable of perceiving the polarization of light, which is generally used for navigational purposes, since the linear polarization of sky light is always perpendicular to the direction of the sun. This ability is very common among the insects, including bees, which use this information to orient their communicative dances. Polarization sensitivity has also been observed in species of octopus, squid, cuttlefish, and mantis shrimp. This ability is based on the polarizing effect of the anisotropic microstructure of the photoreceptor cells of these animals.

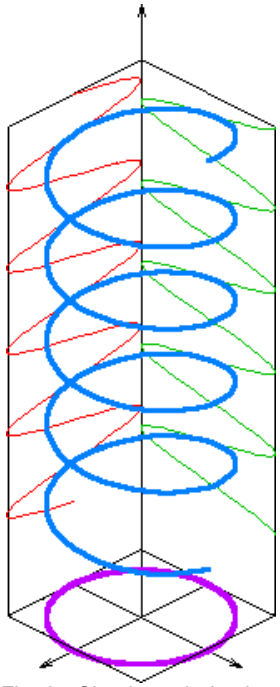


Fig. 3 Circular polarization.

More info

In Fig. 3 the two orthogonal components have exactly the same amplitude and are exactly (plus or minus) 90^o out of phase. In this special case the electric vector traces out a circle in the plane of projection, so this special case is called circular polarization. The direction of rotation depends on the sign of the phase difference.
All other cases, that is where the two components are not in phase and either do not have the same amplitude and/or are not 90^o out of phase are called elliptical polarization because the electric vector traces out an ellipse in the horizontal plane.

In nature, electromagnetic radiation is often produced by a large number of individual radiators, producing waves independently of each other. In general there is no single frequency but rather a spectrum of different frequencies present, and even if filtered to an arbitrarily narrow frequency range, there phases and planes of polarization are different. This type of light is described as *incoherent*. However, this does not mean that polarization is only a feature of coherent radiation. Incoherent radiation may show statistical correlation between the components of the electric field, which can be interpreted as *partial polarization*. In general it is possible to describe a beam of light as the sum of a completely incoherent part (no correlations) and a completely polarized part. One may then describe the light in terms of the *degree of polarization*, and the parameters of the polarization ellipse.

A more complete description, fundamental and mathematically can be found in Wikipedia.

Light: refraction

Principle

When light passes through a transparent substance, such as air, water or glass, its speed is reduced, and it undergoes refraction. The reduction of the speed of light in a denser material can be indicated by the refractive index, n, which is defined as:

$$n = c/v,$$

where c is the speed in vacuum and v the speed in the medium.
Thus, n = 1 in a vacuum and n > 1 in (any) matter.
When a beam of light enters a medium from vacuum or another medium, it keeps the same frequency. So, its "color" remains the same. However, its wavelength changes since speed changes. If the incident beam is not orthogonal to the edge between the media, the direction of the beam will change.
If the refractive indices of two materials are known for a given frequency, then one can compute the angle by which radiation of that frequency will be refracted as it moves from the first into the second material from Snell's law.

Application

The refractive index of a material is the most important property of any optical system that uses refraction. It is used to calculate the focusing power of lenses, and the dispersive power of prisms. Since refractive index is a fundamental physical property of a substance, it is often used to identify a particular substance, confirm its purity, or measure its concentration. Refractive index is used to measure solids (glasses and gemstones), liquids, and gasses. Most commonly it is used to measure the concentration of a solute in an aqueous solution. A refractometer is the instrument used to measure refractive index. For a solution of sugar, the refractive index can be used to determine the sugar content. Refraction of light by lenses is used to focus light in magnifying glasses, spectacles and contact lenses (see The ideal lens) and microscopes (see Light microscopy).

More info

The refractive index n of a material is defined by:

$$n = (\varepsilon_r \mu_r)^{0.5},$$

where ε_r is the medium's dielectric constant (relative permittivity), and μ_r is its relative magnetic permeability. For a non-magnetic material, μ_r is very close to 1, therefore n is approximately $\varepsilon_r^{0.5}$.
Remind that ε_r is strongly frequency dependent. So for water, one should not take its static ε_r, which is 81!
The effect of the frequency dependency of n (except in vacuum, where all frequencies travel at the same speed, c) is known as *dispersion*, and it is what causes a prism to divide white light into its constituent spectral colors. It explains rainbows, and is the cause of chromatic aberration in lenses (see The ideal lens).
Anisotropy
The refractive index of certain media may be different depending on the polarization and direction of propagation of the light through the medium. This is known as birefringence or anisotropy (see Snell's Law).
There are some other phenomena related to refraction like non-linear behavior of n (e.g. Kerr effect), inhomogeneity etc. For this, the reader consults physical textbooks.

Light: scattering

Principle

Scattering is a process whereby some forms of radiation, such as light or moving particles, are forced to deviate from a straight trajectory by non-uniformities in the medium. It occurs also with sound. In conventional use, this also includes deviation of reflected radiation from the angle predicted by the law of reflection (see Light: refraction). Reflections that undergo scattering are often called *diffuse* reflections and unscattered reflections are called *specular* (mirror-like) reflections.
The types of non-uniformities that can cause scattering, the *scatterers* or *scattering centers* are for instance particles, bubbles, droplets, cells in organisms, density fluctuations in fluids and surface roughness.
Scattering can be distinguished between two broad types, *elastic* and *inelastic*. Elastic scattering involves no change of radiation energy, but inelastic scattering does. If the radiation loses a significant proportion of its energy, the process is known as *absorption*. This is governed by the Lambert-Beer law.
Major forms of elastic light scattering are *Rayleigh scattering* and so called *Mie scattering*. Inelastic EM scattering effects include Brillouin scattering (see **More Info**).
With Rayleigh scattering of light, or EM radiation, it holds that particle diameter $d < 0.1\lambda$ (λ is wavelength) of the light. It occurs when light travels in transparent solids and liquids, but especially in gases. Rayleigh scattering is proportional to λ^{-4}.
If $d > \lambda$, light is not separated in all its wavelengths and the scattered light appears white, as do salt and sugar.

Light scattering is one of the two major physical processes that contribute to the visible appearance of most objects. The other is absorption. Surfaces described as *white* owe their appearance almost completely to the scattering of light by the surface. The absence of surface scattering leads to a shiny or glossy appearance. Light scattering can also give color to some objects, usually shades of blue (as with the sky, the human iris, and the feathers of some birds), but resonant light scattering in nanoparticles can produce different highly saturated and vibrant hues.

Application

In medicine
Scattered light is the image forming light in dark field microscopy. Scattered sound plays the same role in Echography. However, in many applications of computer-generated imagery one tries to minimize the disturbance influence of scattering.
Some specific way of scattering is of importance for detecting DNA, proteins and for Fluorescence.
In ophthalmology it is of importance with respect of the quality of the eye media. Scattering in the eye media, especially in the eye lens disturb clear vision, especially under scotopic conditions. In vision research diffusers (see More Info) are often applied.
In daily life
Why is the sky blue? This effect occurs because blue photons hit the air molecules in the earth's atmosphere and are scattered down to the earth's surface. Red photons are not affected by the particles and pass on through the earth's atmosphere. This causes blue light to be scattered down to the earth's surface which makes the sky appear blue.
During sunrise and sunset the Sun's light must pass through a much greater thickness of the atmosphere to reach an observer on the ground. This extra distance causes multiple scatterings of blue light, but relatively little scattering of red light. This is seen as a pronounced red-hued sky in the direction towards the sun: an orange-red sun, which is yellow during daytime.
For the sun high overhead, sunlight goes through a much smaller atmospheric layer, so little scattering takes place. This is why the sky close to the overhead sun in midday appears mostly white, the sun's color.

More info

When radiation is only scattered by one localized scattering center, this is called *single scattering*, but mostly scattering centers are grouped together, and *multiple scattering* occurs. The main difference between the effects of single and multiple scattering is that single scattering can usually be treated as a random phenomenon and multiple scattering is usually more deterministic. Single scattering is often described by probability distributions.
With multiple scattering, the final path of the radiation appears to be a deterministic distribution of intensity as the radiation is spread out. This is exemplified by a light beam passing through thick fog.
Multiple scattering is highly analogous to diffusion and the terms *multiple scattering* and *diffusion* are

interchangeable in many contexts. Optical elements designed to produce multiple scattering are thus known as *diffusers* or Lambertian radiators or reflectors.

Not all single scattering is random. A well-controlled laser beam can be exactly positioned to scatter off a microscopic particle with a deterministic outcome. Such situations are encountered in radar scattering from e.g. a car or aircraft.

Rayleigh scattering
The inherent scattering that radiation undergoes passing through a pure gas or liquid is due to microscopic density fluctuations as the gas molecules move around.
The degree of Rayleigh scattering varies as a function of particle diameter d, λ, angle, polarization (see Light: polarization), and coherence (see Light). The intensity *I* of light scattered by a single small particle from a beam of unpolarized light of intensity I_0 is given by:

$$I = I_0 \frac{(1+\cos^2\theta)}{2l^2}\left(\frac{2\pi^2}{\lambda}\right)^4\left(\frac{n^2-1}{n^2+2}\right)^2\left(\frac{d}{2}\right)^6$$

where ***l*** is the distance to the particle, θ is the scattering angle, n is the refractive index (see Light: refraction) of the particle.
The angular distribution of Rayleigh scattering, governed by the $(1+\cos^2\theta)$ term, is symmetric in the plane normal to the incident direction of the light, and so the forward scatter equals the backwards scatter.

Mie scattering
For larger diameters the shape of the scattering center becomes much more significant and the theory only applies well to spheres, spheroids (2 equal axes) and ellipsoids (3 unequal axes).

Both Mie and Rayleigh scattering of EM radiation can undergo a Doppler shift (see Doppler principle) by moving of scattering centers.
At values $d/\lambda > 10$ the laws of geometric optics are mostly sufficient to describe the interaction of light with the particle, and at this point the interaction is not usually described as scattering.
Another special type of EM scattering is coherent backscattering. A description of this phenomenon is beyond the scope of this compendium.

Tyndall effect
This is the effect of light scattering on particles in colloid systems, such as emulsions (see Colloid and Emulsion). The effect distinguishes between these types of colloids. It is proportional to d^6 and hardly on λ.

Brillouin scattering
This occurs when light in a medium (such as water or a crystal) interacts with density variations and changes its path. When a medium is compressed n changes and the light's path necessarily bends. The density variations may be due to acoustic modes, vibration phenomena in crystals (phonons) or temperature gradients.

Snell's law

Principle

Snell's law calculates the refraction of light when traveling between two media of differing refractive index n.

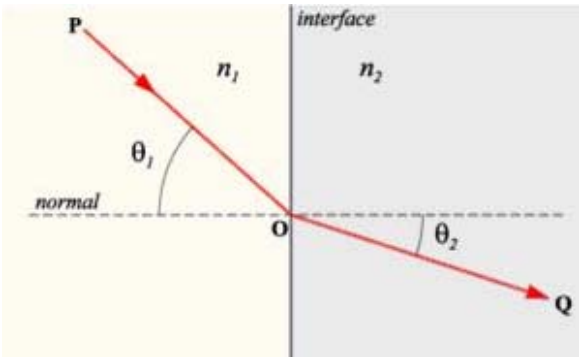


Fig. 1 Refraction of light at the interface between two media of different refractive indices with $n_2 > n_1$.

A ray entering a medium perpendicular to the surface is never bent.
Snell's law gives the relation between the angles θ_1 and θ_2:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \quad (1)$$

A qualitative rule for determining the direction of refraction is that the ray in the denser medium is always closer to the normal. A handy way to remember this is to visualize the ray as a car crossing the boundary between asphalt (the less dense medium) and mud (the denser medium). The car will swerve in the direction of that front wheel that first becomes into the mud.
Snell's law may be derived from *Fermat's principle*, which states that the light travels the path which takes the least time. In a classic analogy, the area of lower refractive index is replaced by a beach, the area of higher refractive index by the sea, and the fastest way for a rescuer on the beach to get to a drowning person in the sea is to run along a path that follows Snell's law.

Snell's law is only generally true for *isotropic* media (see **More Info**) such as glass and water.

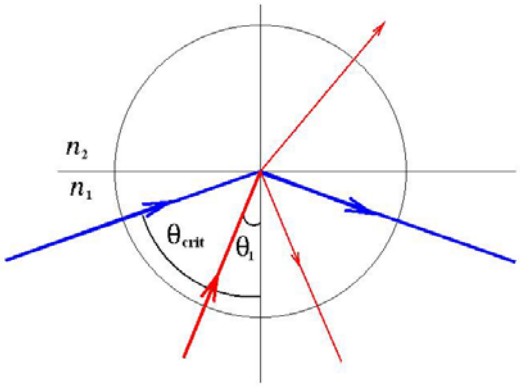


Fig. 2 Total internal reflection

Total internal reflection When moving from a dense to a less dense medium (i.e. $n_1 > n_2$), it is easily verified that the above equation (1) has no solution when θ_1 exceeds a value known as the critical angle:

$$\theta_{crit} = \arcsin(n_2/n_1) \quad (2)$$

When $\theta_1 > \theta_{crit}$, no refracted ray appears, and the incident ray undergoes *total internal reflection* from the interface (Fig. 2).

See for a more detailed description of refraction Light: refraction and for reflection Light: reflection.

Application

The law is applied innumerous in (medical and daily life) optical apparatus, like glasses, binoculars, microscopes etc.
Birefringence (see below) is widely used in optical devices, such as LCDs (liquid crystal displays) in notebooks, light modulators, color filters, wave plates, optical axis gratings, etc. It is also utilized in *medical diagnostics*. Needle biopsies of suspected gouty joints will be negatively birefringent if urate crystals are present.
In situ amyloidal deposits in Alzheimer's disease plaques also show birefringence.

More info

A medium is *anisotropic* when n is directionally dependent, such as occurs in some crystals. This causes birefringence, or double refraction. This is the decomposition of a ray of light into two rays (the *ordinary ray* and the *extraordinary ray*), depending on the polarization of the light (see Light: polarization). The plane of polarization of the ordinary and *extraordinary* or *e*-ray may are generally perpendicular to ach other. With a single axis of anisotropy, (i.e. it is uniaxial) birefringence can be formalized by assigning two different n's to the material for different polarizations. The birefringence magnitude is then defined by:

$$\Delta n = n_e - n_o, \qquad (3)$$

where n_o and n_e are the refractive indices for polarizations perpendicular (ordinary) and parallel (extraordinary) to the axis of anisotropy respectively. Fig. 3 gives an example.

Fig. 3 A calcite crystal laid upon a paper with some letters showing the double refraction.

Biaxial birefringence, also known as trirefringence describes an anisotropic material that has more than one axis of anisotropy.

Light: sources

Principle

The most common processes of emission of light are based on:
- Burning;
- heating (thermal);
- line-emission;
- radio activity.

Heating The most common light sources are thermal: a body at a given temperature emits a characteristic spectrum of black body radiation. Examples include sunlight, glowing solid particles in flames and incandescent light bulbs. These bulbs emit only 10% of their energy as visible light, 20% is lost by heat conduction and convection and the remainder as infrared. The peak of the blackbody spectrum is in the infra red (IR) for relatively cool objects like human beings (see Wiens displacement law, Thermography). As the temperature increases, the peak shifts to shorter wavelengths, producing first a red glow, then a white one (metal heating from "red hot" or "white hot), and finally a blue color as the peak moves out of the visible part of the spectrum and into the ultraviolet ".
Line emission Atoms emit and absorb light at characteristic energies. This produces emission lines (a bright line in a continuous spectrum) in the atomic spectrum. Emission can be spontaneous, as in light-emitting diodes, gas discharge lamps (such as Ne-lamps, Hg-vapor lamps, etc.), and flames (light from the hot gas itself—so, for example, Na in a gas flame emits characteristic yellow light). Emission can also be stimulated, as in a laser or a maser (applied in the kitchen microwave).
Certain substances produce light by fluorescence (fluorescent lights). Some substances emit light slowly after excitation by more energetic radiation, known as phosphorescence. Phosphorescent t materials can also be excited by bombarding them with subatomic particles, a process known as cathodoluminescence (used in cathode ray tube televisions). Certain chemicals produce visible radiation by chemoluminescence and bioluminescence.
Certain other mechanisms can produce light: electroluminescence (a light emitting material in response to an electric current passed through it, or to a strong electric field), sonoluminescence (emission of short bursts of light from imploding bubbles in a liquid when excited by sound, an example is the snapping a specialized shrimp claw), visible cyclotron radiation (acceleration of a free charged particle, such as an electron), scintillation (a flash of light produced in certain materials when they absorb ionizing radiation), radioactive decay, etc.

Table 1 lists overall luminous efficacy and efficiency for various light sources:

Category	Type	Overall luminous efficacy (lm/W)	Overall luminous efficiency
Combustion	candle	0.3	0.04%
Incandescent	40 W tungsten	12.6	1.9%
	quartz halogen	24	3.5%
	high-temperature	35	5.1%
Fluorescent	5-24 W compact	45-60	6.6%-8.8%
	28 W tube (T5)	104	15.2%
Light-emitting diode	white LED	26-70	3.8%-10.2%
Arc lamp	Xe-arc lamp	30-150	4.4%-22%
	Hg-Xe arc lamp	50-55	7.3%-8.0%
Gas discharge	high pressure Na lamp	150	22%
	low pressure Na lamp	183 - 200	27%

Radio active processes Particles produced by radioactive decay, moving through a medium faster than the speed of light in that medium can produce visible *Cherenkov radiation* (characteristic "blue glow" of nuclear reactors). Other examples are radioluminescent paint and self-powered tritium lighting, formerly and nowadays used on watch and clock dials.

Application

Medical applications are multiple in apparatus, instruments, etc. In Thermography the human body can be considered as an IR light source.
I daily live and industry, applications are innumerous.
For two reasons Na-lamps for road lighting are optimal. Their efficacy is high and since they emit monochromatic light, the eye does not suffer from chromatic aberration (see Light: the ideal lens).

More info

See Wikipedia and the internet.

Light: units of measure

Principle

Photometry is measuring the *perceived* brightness to the human eye. It is distinct from radiometry, see below. The human eye is not equally sensitive to all wavelengths. Photometry attempts to account for this by weighting the measured power at each wavelength with a factor that represents how sensitive the eye is at that wavelength. The standardized model of the eye's response to light as a function of wavelength is given by the photopic luminosity function (see Vision of color). Photometry is based on the eye's photopic response, and so photometric measurements will not accurately indicate the perceived brightness of sources in dim lighting conditions.

Many different units are used. The adjective "bright" can refer to a lamp which delivers a high luminous flux (measured in lumens, lm), or to a lamp which concentrates the luminous flux it has into a very narrow beam (in candelas, cd). Because of the ways in which light can propagate through three-dimensional space, spread out, become concentrated, reflects off shiny or matte surfaces, and because light consists of many different wavelengths, the number of fundamentally different kinds of light measurement is large, and so are the numbers of quantities and units that represent them.

Table 1 gives the quantities and units are used to measure the quantity or “brightness" of light.

Table 1 SI photometry units

Quantity	Symbol	Abbr SI unit.	Notes
Luminous energy	Q_v	lm·s	units are sometimes called Talbots
Luminous flux	F	lm	also called *luminous power*
Luminous intensity	I_v	cd	an SI base unit
Luminance	L_v	Cd/m^2	units are sometimes called nits
Illuminance	E_v	lx	Used for light incident on a surface
Luminous emittance	M_v	lx	Used for light emitted from a surface
Luminous efficacy		lm/W	ratio of luminous flux to radiant flux; theoretical maximum is 683

lx is lux. 1 lx = 1 $cd.sr/m^2$. 1 lm = 1 cd.sr (sr is steradian, a solid angle of about 66^o, see Steradian).

Radiometry is measuring the power emitted by a source of electromagnetic radiation (see Light). Radiometric quantities use unweighted *absolute power*, so radiant flux is in watts (versus luminous flux in lumens).

Photometric versus radiometric quantities Every quantity in the photometric system has an analogous quantity in the radiometric system. Some examples of parallel quantities include:

- Luminance (photometric) and radiance (radiometric)
- Luminous flux (photometric) and radiant flux (radiometric)
- Luminous intensity (photometric) and radiant intensity (radiometric)

A comparison of the watt and the lumen illustrates the distinction between radiometric and photometric units.

Light bulbs are distinguished in terms of power (in W) but this power is not a measure of the amount of light output, but the amount of electricity consumed by the bulb. Because incandescent bulbs all have fairly similar characteristics, power is a guide to light output. However, there are many lighting devices, such as fluorescent lamps, LEDs, etc., each with its own efficacy. Therefore, power is also used for the *amount of emitted light*. See for more details Light: sources.

The lumen is the photometric unit of light output and defined as amount of light given into 1 sr (steradian) by a point source of 1 Cd (the candela). The candela, a base SI unit, is defined as the I_v of a monochromatic source with frequency 540 10^{12}Hz, and I = 1/683 W/sr. (540 THz corresponds to about 555 nm, the wavelength to which the human eye is most sensitive. The number 1/683 was chosen to make the candela about equal to the standard candle, the unit which it superseded).

Combining these definitions, we see that 1/683 W of 555 nm green light provides 1 lm. The relation between watts and lumens is not just a simple scaling factor. The definition tells us that 1 W of pure green 555 nm light is "worth" 683 lm. It does not say anything about other wavelengths. Because lumens are photometric units, their relationship to watts depends on the wavelength according to how visible the wavelength is. Infrared (IR) and ultraviolet (UV) radiation, for example, are invisible and do

not count. According to photopic spectral luminous function, 700 nm red light is only about 4% as efficient as 555 nm green light. Thus, 1 W of 700 nm red light is "worth" only 27 lm.
Table 2 gives the quantities and units are used to measure the output of electromagnetic sources.

Table 2 SI radiometry units

Quantity	Symbol	Abbr. I unit	Notes
Radiant energy	Q	J	energy
Radiant flux	Φ	W	radiant energy per unit time, also called *radiant power*
Radiant intensity	I	$W \cdot sr^{-1}$	power per unit solid angle
Radiance	L	$W \cdot sr^{-1} \cdot m^{-2}$	power per unit solid angle per unit *projected* source area. Sometimes confusingly called "intensity".
Irradiance	E	$W \cdot m^{-2}$	power incident on a surface. Sometimes confusingly called "intensity".
Radiant emittance or exitance	M	$W \cdot m^{-2}$	power emitted from a surface. Sometimes confusingly called "intensity".
Spectral radiance	L_λ or L_ν	$W \cdot sr^{-1} \cdot m^{-3}$ *or* $W \cdot sr^{-1} \cdot m^{-2} \cdot Hz^{-1}$	commonly measured in $W \cdot sr^{-1} \cdot m^{-2} \cdot nm^{-1}$
Spectral irradiance	E_λ or E_ν	$W \cdot m^{-3}$ *or* $W \cdot m^{-2} \cdot Hz^{-1}$	commonly measured in $W \cdot m^{-2} \cdot nm^{-1}$

Application

Photometric measurement is based on photodetectors. The lighting industry uses more complex forms of measurement, e.g. obtained with spherical photometers and rotating mirror photometers.

More Info

Non-SI photometry units for luminance are the Footlambert, Millilambert and Stilb and for illuminance the foot-candle and the Phot.

Phosphorescence

Principle

Phosphorescence is a specific type of photoluminescence, related to fluorescence, but distinguished by slower time scales associated with quantum mechanically forbidden energy state transitions of electron orbits.
Phosphorescence is a process in which energy stored in a substance is released very slowly and continuously in the form of glowing light. The reason of the characteristically slow rate of energy release is that on a microscopic level, the probability for the process of emitting light to occur is very low, almost being forbidden by quantum mechanics. A Phosphorescent substance is at the same time also fluorescent as is illustrated in Fig. 1.

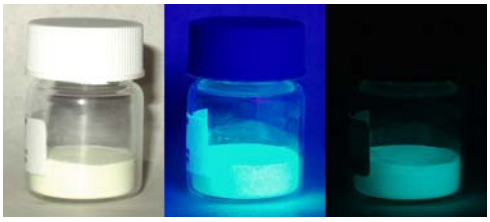


Fig. 1 Phosphorescent powder under visible light (left), ultraviolet light, and total darkness (right).

Applications

Industrial (cathode ray tubes, analog TV sets, watch dials and dials of other consumer equipment), tryptophan phosphorescence spectroscopy and phosphorescence in (pathogenic) bio-aerosols are bio-medical applications. Further labeling of biomolecules by a phosphorophore, a carrier of phosphoresce. The delayed emission and the temperature dependency promises future biomaterial identification.

More Info

Most photoluminescent events, in which a chemical substrate absorbs and then re-emits a photon of light, are fast, on the order of 10 ns! However, for light to be absorbed and emitted at these fast time scales, the energy of the photons involved (i.e. the wavelength of the light) must be carefully tuned according to the rules of quantum mechanics to match the available energy states and allowed transitions of the molecule. In the special case of phosphorescence, the absorbed photon energy undergoes an unusual transition into a higher energy state (usually a triplet state, a set of three quantum states of a system, each with total spin $S = 1$). This state is metastable and transition to the initial state is forbidden. As a result, the energy can become trapped in a state with only quantum mechanically "forbidden" transitions available to return to the lower energy state. Emission occurs when thermal energy raises the electron to a higher state from which it can de-excite. Therefore, phosphorescence is temperature dependent.
Most phosphorescent compounds are still relatively fast emitters, with triplet lifetimes on the order of milliseconds. However, some compounds have triplet lifetimes up to minutes or even hours, allowing these substances to effectively store light energy in the form of very slowly degrading excited electron states. If the phosphorescent quantum yield is high, these substances will release significant amounts of light over long time scales, creating so-called "glow in the dark" materials.

Radiation

Angiography and DSA

Principle

Angiography or arteriography is a medical imaging technique in which an X-ray picture is taken to visualize the inner volume of blood filled structures, including arteries, veins and the heart chambers. The X-ray film or image of the blood vessels is called an angiograph, or more commonly, an angiogram. Angiograms require the insertion of a catheter into a peripheral artery, e.g. the femoral artery.
As blood has the same radiodensity (see CT scan (dual energy)) as the surrounding tissues, a radiocontrast agent (which absorbs X-rays, see Spectroscopy) is added to the blood to make angiographic visualization possible. The image shows shadows of the inside of the vascular structures carrying blood with the radiocontrast agent. The tissue of the vessels or heart chambers themselves remain largely to totally invisible on the X-Ray image.
The images may be taken as either still images, displayed on a fluoroscope (see Fluoroscopy and Fluorescence) or film, useful for mapping an area. Alternatively, they may be motion images, usually taken at 30 frames/s, which also show the speed of blood (i.e. the speed of radiocontrast).

D digital subtraction angiography (DSA) is the procedure to visualize blood vessels with contrast medium in a bony environment by subtracting the pre-contrast image (the mask) from the image with contrast medium (see **More Info**).
Intravenous DSA (IV-DSA) compares an X-ray image before and after radiopaque iodine based dye has been injected intravenously. (Radiopacity is the ability of electromagnetic radiation to pass through a particular material.) Tissues and blood vessels on the first image are digitally subtracted from the second image, leaving a clear picture of the artery which can then be studied independently and in isolation from the rest of the body.

Application

With a catheter (in groin or forearm) the radiocontrast agent is administrated at the desired area. X-ray images of the transient radiocontrast distribution visualize of the inner size of the arteries. Presence or absence of atherosclerosis or atheroma within the walls of the arteries cannot be clearly determined. The most common angiogram is that of the coronary arteries.
Angiography is also commonly performed to identify vessel narrowing in patients with retinal vascular disorders, such as diabetic retinopathy and macular degeneration.
Other common clinical applications consider the cerebrum, extremities, liver, kidneys, lungs and lymph system.

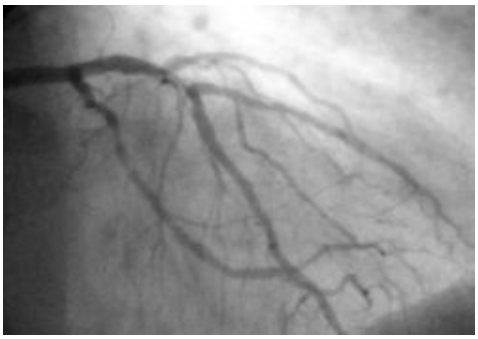


Fig. 1 Coronary angiogram with a stenosis in the left branch.

DSA is useful in diagnosing arterial occlusion, including carotid artery stenosis and pulmonary artery thrombosis, and in detecting renal vascular disease. After contrast material is injected into an artery or vein, fluoroscopic images are produced.

IV-DSA can be used for studying the vessels of the brain and heart, detecting carotid artery obstruction and lesions, and mapping cerebral blood flow. IV-DSA has also been useful in assessing patients prior to surgery and after coronary artery bypass surgery and some transplant operations. However, IV-DSA is unsuitable for patients with diabetes or renal insufficiency (due to the high dose of dye).

More Info

The subtraction technique
In traditional angiography, we acquire images of blood vessels on films by exposing the area of interest with time-controlled X-ray energy while injecting contrast medium into the blood vessels. The images thus obtained would also record other structure besides blood vessels as the X-ray beam passes through the body. In order to remove these distracting structures to see the vessels better, we need to acquire a mask image (image of the same area without contrast administration) for *subtraction*. So, using manual darkroom technique, clear pictures of blood vessels are obtained by taking away the overlying background.
In DSA, the images are acquired in digital format through the computer. Then all images would be recorded into the computer and subtracted automatically. As a result, a near-instantaneous film shows the blood vessels alone.

Diaphanography and optical mammography

Principle

Diaphanography or transillumination is a method to detect breast tumors by compressing the breast between glass plates and using red light to shine through the tissue (see Fig. 3a in **More Info**). Nowadays, diaphanography or better, optical mammography is based on ultra-short (10^{-12} regime) laser pulses, which are used as an alternative to traditional X-ray.

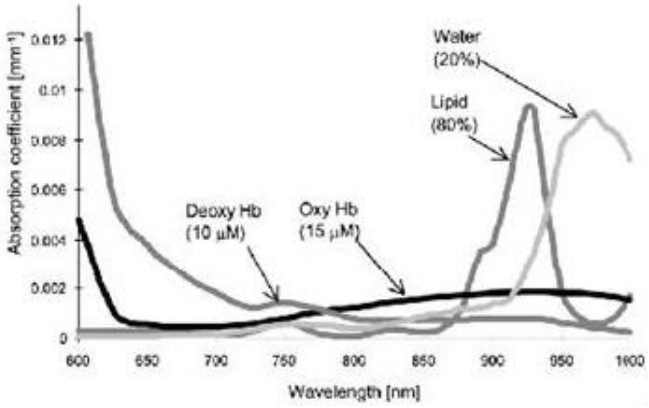


Fig. 1 Absorption coefficients (see Lambert-Beer law) of hemoglobin (Hb) (see also Pulse oximetry), water and fat (lipid) at concentrations typical for female breasts.

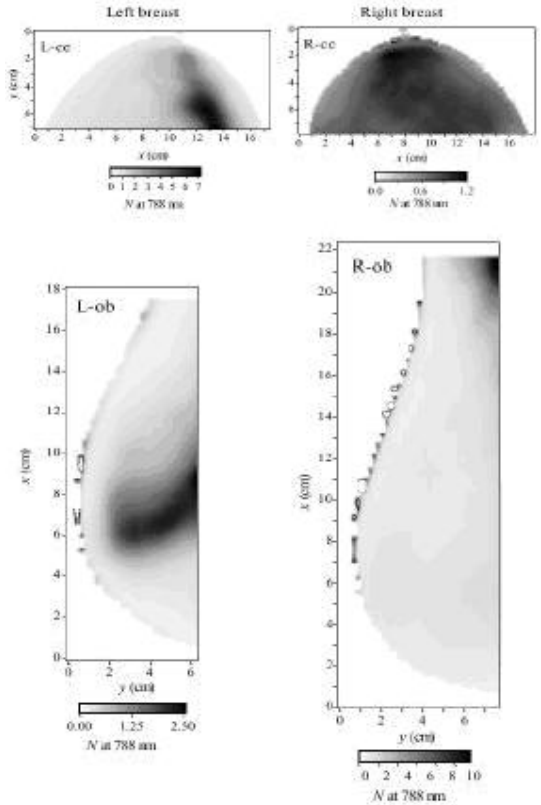


Fig. 2 N-images obtained with 788 nm. These mammograms refer to a 58 year old woman affected by breast cancer (tumor size: 30mm) in the left breast.

Tissue optical properties

The main absorbers in breasts at visible and near-IR light are hemoglobin, water and fat (see for absorption spectra Fig. 1). The 650-950 nm region is most suited for deep-structure detection, and the penetration depth is a few cm. The scattering is caused by small differences in refractive index (see Snell's Law and Light: refraction) at the microscopic level, mainly of fat droplets and structures inside the cell nucleus. The scattering increases for shorter wavelengths (λ), approximately obeying the λ law of Mie scattering theory (see Light: scattering). The scattering, however, is strong even for λ > 1μm.

The main advantage of using short-pulse lasers for deep-structure is that by using time-resolved detection it is possible to select light, which arrives at the detector at different time slots. The light that arrives early has traveled a shorter and straighter path than the late light. The object of the new generation of optical tomography is to reconstruct an image of both the absorption (direct pathway) and the scattering properties inside the tissue, based on measurements at several light-source and detector position on the skin. Actually, the whole shape of the detected pulses can be of use.

Oxygen as a key
On the basis of the 3 different wavelengths the kind of tissues can be detected. A tumor wants to grow and therefore induces growth of a lot of little blood vessels around itself, (particularly an invasive one). But vessels growth is delayed and therefore the tumor gets more oxygen out of the blood than normally. So the volume of blood increase, but the oxygenation level of the blood goes down. By detecting both at the same time one can distinguish between benign and malignant tumors.

Application

Diaphanography has been developed to overcome the draw back of missing tumors in classical screening mammography due to a too low resolution. Fig. 2 shows an example of a scan.
Optical mammography is not yet full-grown commercial and in many western countries X-ray screening is still the practice, the more since today doses are very low and seldom causes the metastasizing of an existing cancer. Despite this, nowadays, retrorespective detection rate is about 90% of histological validated malign tumors.

More Info

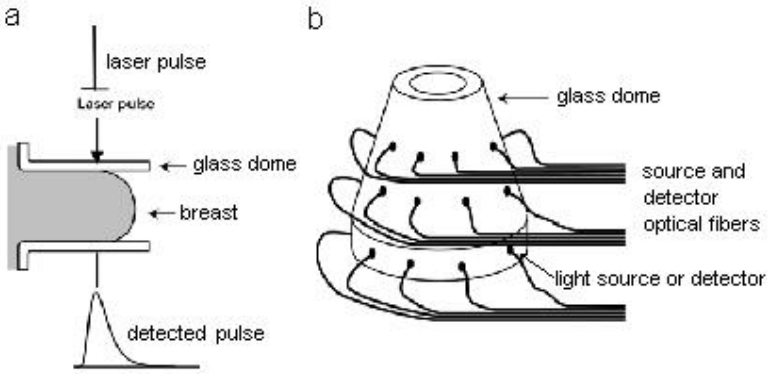


Fig. 3 a Principle of operation. b Principle of the design of the scanner.

In addition to the use of the whole detected light pulse (see Fig. 3a), the lasers may be chosen at λ's that enhance the contrast between the tumor and surrounding tissue, as in classical spectroscopy, where different compositions of the absorbers give rise to different absorption at various wavelengths. The scattering can also vary between tissue types and wavelengths, which can also be used for contrast enhancement. The choice of wavelengths and knowledge for the biological processes that determine the tissue composition is crucial.

Pulsed laser diodes with so called time-correlated single-photon-counting detection are used to get time-resolved measurements with a sub-ns resolution. The small light sources and detector are placed in a conical configuration, more resembling the conical geometry of conventional tomography (Fig. 3b). The breast is compressed, and the source-detector optical fibers are scanned, for a series of point-wise measurements. Acquiring data for a whole breast with about 1000 points takes a few minutes. Because of this relatively long time, the breast is not compressed as hard as for X-ray mammography.

Finding the best geometry is related to the second problem, development of reconstruction algorithms. For example, should the breast be compressed or not, and it is possible to get better results using fluids

to match the optical properties between the skin and the detectors, are some of the questions under investigation.

The principle of the reconstruction algorithm is to divide the problem into two parts, the forward and the inverse problem. The forward problem deals with computation of what the detected signal would be, given that the absorption and scattering are known. The inverse problem is the most likely anatomy which matches best the signals measured with the detectors. Since the problem is highly non-linear, the reconstruction is based on iterations. The solution strongly depends on contrast function that discriminates the tumors. The relative compositions of water, fat and hemoglobin vary not only between tumors and healthy tissue, but also with for example age and hormonal cycles. Furthermore, there is not a single type of breast cancer. Tumors vary a lot in composition and structure. All these parameters have to be understood and quantified.

References

Dehghani H. et al. Multiwavelength Three-Dimensional Near-Infrared Tomography of the Breast: Initial Simulation, Phantom, and Clinical Results, Applied Optics, 2003, 42, 135-145.
Paola Taroni P. et al. Time-resolved optical mammography between 637 and 985 nm: clinical study on the detection and identification of breast lesions Phys. Med. Biol. 2005, 50, 2469-2488.
Rinneberg H. et al. Scanning time-domain optical mammography: detection and characterization of breast tumors in vivo. Technol Cancer Res Treat. 2005, 4, 483-96.

CT scan (dual energy)

Principle

.

Fig. 1 CT apparatus

Computed tomography (CT), originally known as computed axial tomography (CAT or CT scan) and body section röntgenography (see X-ray machine), is a medical imaging method (techniques and processes used to create images of parts of the human body for clinical purposes). It employs tomography (imaging by sections) where digital geometry processing (design of algorithms for 3D-modeling) is used to generate a 3D image of the internals of an object from a large series of 2D X-ray (electromagnetic radiation with 10-0.01 nm wavelengths, see also Spectroscopy) scans. The data of CT scans taken around a single axis of rotation can be manipulated by a process known as windowing (see Imaging: windowing). Standard CT generates images in the axial or transverse plane which can be reformatted in various planes or as volumetric (3D) representations of structures.

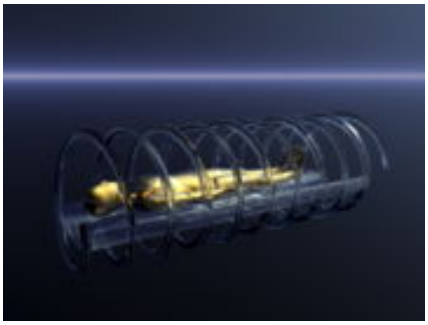


Fig. 2 Principle of spiral CT

The principle of operation is an X-ray source that rotates around the object with X-ray sensors positioned on the opposite side of the circle from the X-ray source. X-ray slice scans are taken one after the other as the object is gradually passed through the gantry. The data of the slices are combined together by the mathematical procedure known as tomographic reconstruction, for instance to generate 3D volumetric information (3D-CT scan), which are viewable from multiple different perspectives on CT workstation monitors. This can also be done with helical or spiral CT machines, which generate high-resolution 3D images from the data of the moving individual slices.

Application

Although most common in medicine, CT is also used in other fields, for example in experimental earth sciences, material sciences and archeology.

CT has become an important tool in medical imaging to supplement X-rays and ultrasonography or echography. Although it is still expensive, it is the gold standard in diagnostic. It is also used in preventive medicine or screening. Some important fields of application are described below.

Neurology CT is frequently used in neurology and related disciplines: cerebrovascular accidents, intracranial hemorrhage (stroke). increased *intracranial pressure*, facial and skull *fractures*, surgical planning for craniofacial and dentofacial deformities, ophthalmology (orbita), trough, nose and ear medicine, brain *tumors* with IV contrast (but less sensitive than MRI).
Pulmonology Non-contrast CT scans are excellent for acute and chronic changes in the lung parenchyma (pneumonia, cancer). Thin section CT scans are used for emphysema, fibrosis, etc., IV-contrast for the mediastinum and hilar regions. CT angiography (CTPA) is applied for pulmonary embolism and aortic dissection (helical scanners).
Cardiology Dual Source CT scanners and high resolution (multi-slice) and high speed (subsecond rotation) CT scanners are used for imaging of the coronary arteries.
Abdominal and pelvic region Cancer, acute abdominal pain, organ injury, disorders resulting in morphological changes of internal organs, often with contrast (barium sulfate for fluoroscopy, (see also Fluorescence), iodinated contrast for pelvic fractures).
Further CT is useful for (complex) fractures in the *extremities*.
CT is inadequate for osteoporosis (radiation doses, costs) compared to DXA scanning for assessing bone mineral density (BMD), which is used to indicate bone strength.

More Info

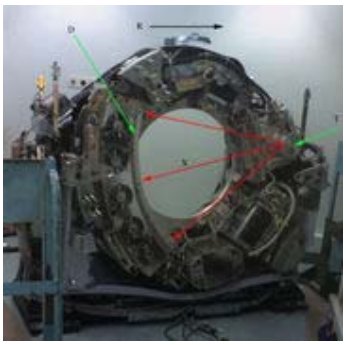


Fig.3 ..CT scanner with cover removed to show the principle of operation

Working principle

In conventional CT machines, an X-ray tube (see X-ray machine)and detector are physically rotated behind a circular shroud (see Fig. 3). In electron beam tomography (EBT) the tube is far larger and higher power to support the high temporal resolution. The electron beam is deflected in a hollow funnel shaped vacuum chamber. X-rays are generated when the beam hits the stationary target. The detector is also stationary. Contrast materials are used to highlight structures such as blood vessels and to obtain functional information.
The data stream, representing the varying radiographic intensity sensed, reaches the detectors on the opposite side of the circle during each sweep. Then, it is computer-processed to calculate cross-sectional estimations of the radiographic density, expressed in Hounsfield units (HU).

Hounsfield unit
This unit is defined as:

$$HU_{material} = 1000(\mu_{material} - \mu_{water})/\mu_{water},$$

where μ is the attenuation coefficient, which is analogue to the parameter A (absorbance or extinction) of the Lambert-Beer law. HU is dependent of the beam intensity of the X-ray source (mostly given in kilo-electron volt, keV).
Since μ_{water} is by definition zero, the *radiodensity* of distilled water at STPD conditions (see Gas volume units, STPD, BTPS and ATPS) is defined as 0 HU. At 80 keV, that of air at STPD is ca.−1000 HU, cancellous bone amounts to 400 HU and cranial bone to 2000 HU. The attenuation of metallic implants (dental, extremities) depends on the element's atomic number. Titanium has ca. 9000 HU (at 80 keV) and iron steel ca. 24500, which can completely extinguish the X-ray and is therefore responsible for well-known line-artifacts in computed tomograms.
Pixels in an image obtained by CT scanning are displayed in terms of HUs (from -1024 to +3071). When the CT slice thickness is also factored in, the volumetric unit is known as a voxel, which is a cubical

pixel. The phenomenon that one part of the detector cannot differ between different tissues is called the Partial Volume Effect. That means that a big amount of cartilage and a thin layer of compact bone can cause the same attenuation in a voxel as hyperdense cartilage alone.
Dual source (or energy) scanning offers the potential of differentiating materials beyond the visualization of morphology – for example, direct subtraction of either vessels or bone during scanning. Dual Source CT scanners, allow higher temporal resolution so reduce motion blurring at high heart rates, and potentially require a shorter breath-hold time. Manufacturers are now actively developing 256-slice MSCT (multi-slice CT), true 'volumetric' scanners, primarily for their improved cardiac scanning performance.

Advantages and hazards

Advantages over Projection Radiography (see Radiography)
- CT completely eliminates the superposition of images of structures outside the area of interest.
- Because of the inherent high-contrast resolution of CT, differences between tissues that differ in physical density by less than 1% can be distinguished.
- Data from a single CT imaging procedure consisting of either multiple contiguous or one helical scan can be viewed as images in the axial, coronal, or sagittal planes. This is referred to as multiplanar reformatted imaging.

Radiation exposure
CT is regarded as a moderate to high radiation diagnostic technique. Unfortunately the newer CT technology requires higher doses for better resolution. For instance, CT angiography avoids the invasive insertion of an arterial catheter and guide wire and CT colonography may be as good as barium enema for detection of tumors in the large intestines, but at the cost of a higher dose.
Cardiac MSCT is equivalent of 500 chest X-rays in terms of radiation. The risk on breast cancer is not well established. The positive predictive value is approximately 82% while the negative prediction is ca. 93%. Sensitivity is ca. 81% and the specificity is about 94%. The real benefit in the test is the high negative predictive value.
The radiation dose for a particular study depends on multiple factors: volume scanned, patient build, number and type of scan sequences, and desired resolution and image quality.

Table 1 Typical scan doses

Examination	**Typical effective dose (mSv*)**
Chest X-ray	0.02
Head CT	1.5
Abdomen	5.3
Chest	5.8
Chest, Abdomen, Pelvis	9.9
Cardiac CT angiogram	6.7-13
CT colongraphy	3.6 - 8.8

*The sievert (Sv) has the same dimensions as the gray (i.e. 1 Sv = 1 J/kg = 1 $m^2 \cdot s^{-2}$), but the former measures the biological effect and the later the radiated dose.

Adverse reactions to contrast agents
Because CT scans often rely on IV-contrast agents, there is a low but non-negligible level of risk associated with the contrast agents themselves, like (life-threatening) allergic reactions to the contrast dye (e.g. causing kidney damage).

Electron Spin Resonance (ESR)

Principle

Electron spin resonance (ESR) or electron paramagnetic esonance (EPR) is a spectroscopic technique (see Spectroscopy) which detects atoms that have unpaired electrons. It means that the molecule in question is a free radical if it is an organic molecule. Because most stable molecules have a closed-shell configuration of electrons without a suitable unpaired spin, the technique is less widely used than nuclear magnetic resonance, NMR. (Spin refers to rotating of an object around some axis trough the object.)
The basic physical concepts of the technique are analogous to those of NMR, but instead of the spins of the atom's nuclei, electron spins are *excited*. Because of the difference in mass between nuclei and electrons, some 10 times weaker magnetic fields and higher frequencies are used, compared to NMR. For electrons in a magnetic flux field (or magnetic induction) of 0.3 T (tesla), spin resonance occurs at around 10 GHz. (1 T = 1 N/(Am) = 10.000 Gauss = 10^9 gammas).

Application

ESR is used in solid-state physics for the identification and quantification of radicals (i.e., molecules with unpaired electrons), in (bio)chemistry to identify reaction pathways. Its use in biology and medicine is more complicated. This group can be used to couple the probe to another molecule, e.g. a bio-molecule).
Since radicals are very reactive, they do not normally occur in high concentrations in biological environments. With the help of specially designed non-reactive, so stable free radical molecules carrying a functional group that attach to specific sites in a biological cell, it is possible to quantify structures comprising these sites with these so-called spin-label or spin-probe molecules.
To detect some subtle details of some systems, high-field-high-frequency electron spin resonance spectroscopy is required. While ESR is affordable for a medium-sized academic laboratory, there are few scientific centers in the world offering high-field-high-frequency electron spin resonance spectroscopy.

More Info

An electron has a magnetic moment (is torque/magnetic field strength, in Am^2) and spin quantum number s = 1/2, with magnetic components $m_s = +1/2$ and $m_s = -1/2$. When placed in an external magnetic field of strength B_0, this magnetic moment can align itself either parallel ($m_s = -1/2$) or antiparallel ($m_s = +1/2$) to the external field. The former is a lower energy state than the latter (this is the Zeeman effect), and the energy separation between the two is:

$$\Delta E = g_e \mu_B B_0, \qquad (1)$$

where g_e is the gyromagnetic ratio or g-factor of the electron, a dimensionless quantity. It is the ratio of its magnetic dipole moment to its angular momentum. The magnetic moment in a magnetic field is a measure of the magnetic flux set up by the rotation (spin and orbital rotation) of an electric charge in a magnetic field. Further, μ_B is the Bohr magneton (a constant of magnetic moment $\mu_B = 9.27 \times 10^{-24}$ Am^2) and B_0 the magnetic field strength (N/A·m).

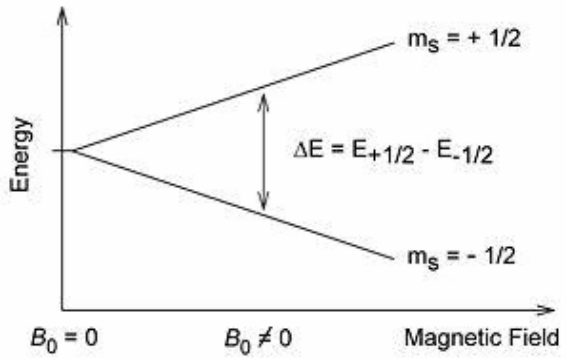


Fig. 1

An unpaired electron can move between the two energy levels by either absorbing or emitting electromagnetic radiation of energy ε = *h*ν (Planck's constant times frequency) such that the resonance condition, ε = ΔE, is obeyed. Substituting in ΔE = *h*ν in (1) gives:

$$h\nu = g_e \mu_B B_0. \qquad (2)$$

This is the fundamental equation of EPR spectroscopy.
This equation implies that the splitting of the energy levels is directly proportional to the magnetic field's strength, as shown in Fig. 1.

A collection of paramagnetic molecules (molecules with the property to align in a magnetic field, see Diamagnetism and paramagnetism), such as free radicals, is exposed to microwaves at a fixed frequency. By increasing an external magnetic field, the gap between the $m_s = +1/2$ and $m_s = -1/2$ energy states is widened until it matches the energy of the microwaves, as represented by the double-arrow in Fig. 1. At this point the unpaired electrons can move between their two spin states. Since there typically are more electrons in the lower state, there is a net absorption of energy, and it is this absorption which is monitored and converted into a spectrum.

A free electron (on its own) has a *g* value of about 2.0023 (which is g_e, *the electronic* g factor). This means that for radiation at the commonly used frequency of 9.75 GHz (known as X-band microwave radiation, and thus giving rise to X-band spectra), resonance occurs at a magnetic field of about 0.34 tesla.

ESR signals can be generated by resonant energy absorption measurements made at different electromagnetic radiation frequencies ν in a constant external magnetic field (*i.e.* scanning with a range of different frequency radiation whilst holding the field constant, like in an NMR experiment). Conversely, measurements can be provided by changing the magnetic field *B* and using a constant frequency radiation; due to technical considerations, this second way is more common. This means that an ESR spectrum is normally plotted with the magnetic field along the horizontal axis, with peaks at the field that cause resonance (whereas an NMR spectrum has peaks at the frequencies that cause resonance).

For more information see e.g. the Wikipedia chapter on ESR.

Image processing: 3D reconstruction

Principle

To apply one of the reconstruction techniques, a preliminary step is *windowing* of the 2D slices. It is the process of using the measured radiographic densities, i.e. Hounsfield units, (HU, see CT scan (dual energy)) to make an image. The various HU amplitudes are mapped to 256 gray-shades. They are distributed over a wide range of HU values when an overview of structures is needed. They can also be distributed over a narrow range (called a *narrow window*) centered over the average HU value of a particular structure in order to discern subtle structural details. This image processing technique is known as *contrast compression*.
There exist various 3D reconstruction techniques. Briefly some are mentioned. Of some of them, details can be found in. Important are

Because contemporary CT scanners offer (nearly) isotropic (the property of being independent of direction) resolution in space, a software program can 'stack' the individual slices one on top of each other to construct a volumetric or 3D image. The image is obtained by taking slices through the volume in a different plane (usually orthogonal). The procedure is called *multiplanar reconstruction* (MPR).

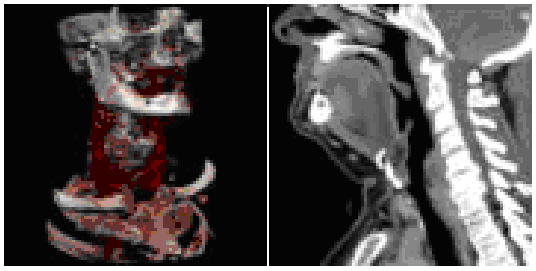


Fig. 1 Typical screen layout for diagnostic software, showing a 3D (left, torso) and a MPR (right, vertebrate column) view.

A special projection method *maximum-intensity projection* (MIP) or minimum-intensity projection (mIP) can be used to build the reconstructed slices. With MIP a transparent 3D view is visualized, which can be rotated.
With *rendering* techniques boundaries or volumes can be highlighted.
Segmentation is the technique to separate kinds of tissues, for instance with head MRI skull, liquor and brain from each other.

Application

3D-reconstruction techniques are basic tools for all imaging techniques in medicine, so it is not limited to CT (see CT scan (dual energy)), but also used for PET, SPECT, MRI, fMRI and optical imaging techniques such as thermography, diaphanography and optical mammography. They all rely on windowing as the preliminary step. To explain how windowing works in practice some examples are given.
Windowing liver and bones
For example, to evaluate the abdomen in order to find subtle masses in the liver, a good liver window is 70 HU as an average with the shades of gray distributed over a narrow window of ± 85 HU. Any HU value below -15 would be pure black, and any HU value above 155 HU would be pure white in this example. Using this same logic, bone windows would use a wide window since bone contains dense cortical bone as well as the low dense fatty marrow. The process will take between some minutes and one hour.
Visualization of spines
Axial images through the spine will only show one vertebral body at a time and cannot reliably show the intervertebral discs. By reformatting the volume with MPR it is possible to visualize the position of one vertebral body in relation to the others.

More Info

A maximum intensity projection (MIP) is a computer visualization method for 3D data that projects in the visualization plane the voxels (a voxel is the 3D analog of the 2D pixel) with maximum intensity that fall

in the way of parallel rays traced from the viewpoint to the plane of projection. This implies that two MIP renderings from opposite viewpoints are symmetrical images.
This technique is computationally fast, but the 2D results do not provide a good sense of depth of the original data. To improve the sense of 3D, animations are usually rendered of several MIP frames in which the viewpoint is slightly changed from one to the other, thus creating the illusion of rotation. This helps the viewer's perception to find the relative 3D positions of the object components. However, since the projection is orthographic the viewer cannot distinguish between left or right, front or back and even if the object is rotating clockwise or anti-clockwise.
Retrieved from "http://en.wikipedia.org/wiki/Maximum_intensity_projection" where also an animation can be found.

Modern software allows reconstruction in non-orthogonal (oblique) planes so that the optimal plane can be chosen to display an anatomical structure (e.g. for bronchi as these do not lie orthogonal to the direction of the scan).
For vascular imaging, *curved-plane reconstruction* can be performed. This allows bends in a vessel to be 'straightened' so that the entire length can be visualized on one image. After 'straightening', measurements of length and diameter can be made, e.g. for planning surgery.
Since MIP reconstructions enhance areas of high radiodensity, they are useful for angiography. mIP reconstructions enhance air spaces, so they are useful for assessing *lung structure*.

3D rendering techniques
Surface rendering A threshold value of radiodensity is chosen by the operator (e.g. for bone). A threshold is set, using edge detection image processing algorithms. In edge detection, small groups of pixels (e.g. forming a line piece) with strong intensity differences from neighboring ones are marked. In this way, finally 3D-surface is extracted and a 3D model with contour is constructed. (Actually, complicated mathematics is used to realize edge detection.) Multiple models can be constructed from various different thresholds, allowing different colors to represent each anatomical component such as bone, muscle, and cartilage. Surface rendering only displays surfaces meeting the threshold, and only displays the surface most close to the imaginary viewer. Therefore, the interior structure of each element is not visible.
Volume rendering In volume rendering, transparency and colors are used to allow a better representation of the various structures within a volume, e.g. the bones could be displayed as semi-transparent, so that even at an oblique angle, one part of the image does not conceal another.

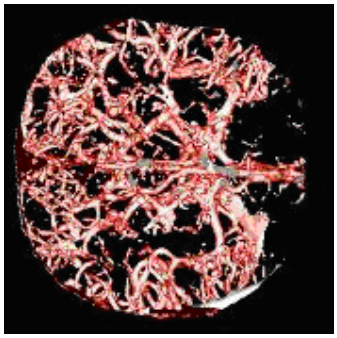


Fig.2 Brain vessels reconstructed in 3D after bone has been removed by segmentation

Image segmentation
Where different structures have similar radiodensity, it is impossible to separate them with volume rendering. The solution is called segmentation, a manual or automatic procedure that can remove the unwanted structures from the image.
After using a segmentation tool to remove the bone of the skull, now the previously concealed vessels can be demonstrated.

Raman spectroscopy

Principle

Raman spectroscopy is a technique used in (mainly) solid state physics and in chemistry to study vibrational, rotational, and other low-frequency modes in a system. Raman techniques are frequently applied to biological material (cells, proteins etc.). It relies on inelastic scattering, or Raman scattering of monochromatic light, usually from a laser in the visible, near-IR, or near-UV range. IR spectroscopy yields similar, but complementary information. Raman scatter is proportional to the 4th power of wavelength.

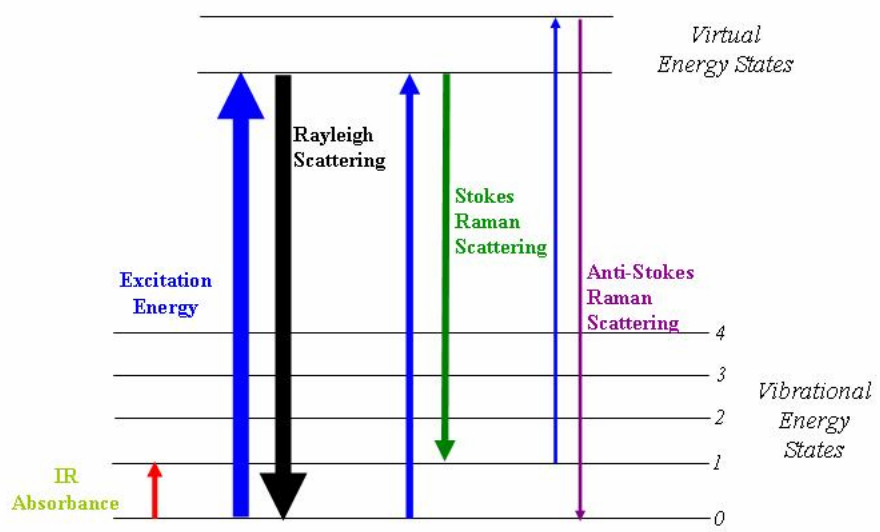


Fig. 1 Rayleigh and Raman scattering visualized with the energy sates of the atom.

When light is scattered from an atom or molecule, most photons are elastically scattered (Rayleigh scattering, see Light: scattering) as illustrated in Fig. 1. The scattered photons have the same energy (frequency) and, therefore, wavelength, as the incident photons. However, a small fraction of scattered light (approximately 1 in 1000 photons) is scattered from excitations with optical frequencies different from, and usually lower (the green arrow in Fig. 1) than the frequency of the incident photons. This type of scattering, with a shift towards higher wavelengths, is Raman scattering. In a gas, Raman scattering can occur with a change, mostly an increase, in vibrational, rotational or electronic energy of a molecule. The following is amore precise description of the process.
The Raman effect occurs when light impinges upon a molecule and interacts with the electron cloud of the bonds of that molecule. A molecular polarizability change, or amount of deformation of the electron cloud, with respect to the vibrational coordinate is required for the molecule to exhibit the Raman effect. The amount of the polarizability change will determine the intensity, whereas the Raman shift, the shift in wavelength, is equal to the vibrational level that is involved. The incident photon (light quantum), excites one of the electrons into a virtual state. For the spontaneous Raman effect, the molecule will be excited from the ground state (0) to a virtual energy state, and relax into a vibrational excited state (e.g. 1), which generates Stokes Raman scattering (Fig. 1). If the molecule was already in an elevated vibrational energy state and relaxes in a lower state, the Raman scattering is then called anti-Stokes Raman scattering (shift to shorter wavelengths).

Application

In science
Raman spectroscopy is commonly used in chemistry, since vibrational information is very specific for the chemical bonds in molecules. It therefore provides a fingerprint by which the molecule can be identified. The fingerprint region of organic molecules is in the range 5-20 μm. Another way that the technique is used is to study changes in chemical bonding, e.g. when a substrate is added to an enzyme.
In solid state physics, spontaneous Raman spectroscopy is used to, among other things, characterize materials and measure temperature. Raman scattering by an anisotropic crystal gives information on the crystal orientation. The polarization of the Raman scattered light can be used to find the orientation of the crystal.

In medicine
Raman gas analyzers are for instance they are used for real-time monitoring of anaesthetic and respiratory gas mixtures during *surgery*.
Spatially Offset Raman Spectroscopy (SORS), which is less sensitive to surface layers than conventional Raman, can be used for non-invasive monitoring of biological tissue.
Raman microspectroscopy
The advantage of Raman spectroscopy, since it is a scattering technique, is that specimens do not need to be fixed or sectioned. Raman spectra can be collected from a very small volume (< 1μm in diameter); these spectra allow the identification of species present in that volume. Water does not interfere very strongly. Therefore it is suitable for the microscopic examination of cells and proteins.
In *direct imaging*, the whole field of view is examined for scattering over a small number of wavelengths (Raman shifts). For instance, a wavelength characteristic for cholesterol to record the distribution of cholesterol within a cell culture.
The other approach is *hyperspectral imaging*, in which thousands of Raman spectra are acquired from all over the field of view. The data can then be used to generate images showing the location and amount of different components. The image could show the distribution of cholesterol, as well as proteins, nucleic acids, and fatty acids. Sophisticated signal- and image-processing techniques can be used to ignore the presence of water, culture media, buffers, and other interferents.
Raman microscopy
Lateral and depth resolutions with confocal Raman microscopy (see Microscopy; confocal) may be as low as 250 nm and 1.7 μm, respectively. Since the objective lenses of microscopes focuses the laser beam to several microns in diameter, the resulting photon flux is much higher than achieved in conventional Raman setups. This has the added benefit of enhanced fluorescence quenching. Since the high photon flux can also cause sample degradation, a thermally conducting substrate for heat absorption is achieved.
In vivo time- and space-resolved, Raman spectra of microscopic regions of samples can be measured. As a result, the fluorescence of water, media, and buffers can be removed. Consequently *in vivo* time- and space-resolved Raman spectroscopy is suitable to measure cells, proteins, organs, and erythrocytes.
Raman microscopy for biological and medical specimens generally uses near-IR lasers. Due to the low efficacy of Raman scatter, the laser needs a high intensity. To prevent tissue damage and improve efficacy, near-IR is used ($E \sim 1/\lambda$ and the intensity of the scattered beam is proportional with λ^4).

More Info

Measurement principle
Typically, a sample is illuminated with a laser beam. Light from the illuminated spot is collected with a lens and sent through a monochromator. Wavelengths close to the laser line (due to elastic Rayleigh scattering) are filtered out and those in a certain spectral window away from the laser line are dispersed onto a detector.
Molecular structure analysis
Phonons or other excitations in the system are absorbed or emitted by the laser light, resulting in the energy of the laser photons being shifted up or down. (A phonon is a quantized mode of vibration occurring in a rigid crystal lattice, such as the atomic lattice of a solid.) The shift in energy gives information about the phonon modes in the system.
Raman microscopy
A Raman microscope begins with a standard optical microscope, and adds an excitation laser, a monochromator, and a sensitive detector (such as a charge-coupled device (CCD) or photomultiplier tube).
Spontaneous Raman scattering is typically very weak, and as a result the main difficulty of Raman spectroscopy is separating the weak inelastically scattered light from the intense Rayleigh scattered laser light. Raman spectrometers typically use holographic diffraction gratings (see Holography) and multiple dispersion stages to achieve a high degree of laser rejection. A photon-counting photomultiplier tube or, more commonly, a CCD camera is used to detect the Raman scattered light. In the past, photomultipliers were the detectors of choice for dispersive Raman setups, which resulted in long acquisition times. However, the recent uses of CCD detectors have made dispersive Raman spectral acquisition much more rapid.
Raman spectroscopy has a stimulated version, analogous to stimulated emission, called stimulated Raman scattering.
In addition to the discussed types of Raman spectroscopy there exist some other seven types (see Wikipedia).

Spectroscopy

Basic principle

Spectroscopy is the study of spectra of electro-magnetic radiation (Fig. 1) of some physical quantity (e.g. emission) performed with a spectrometer.

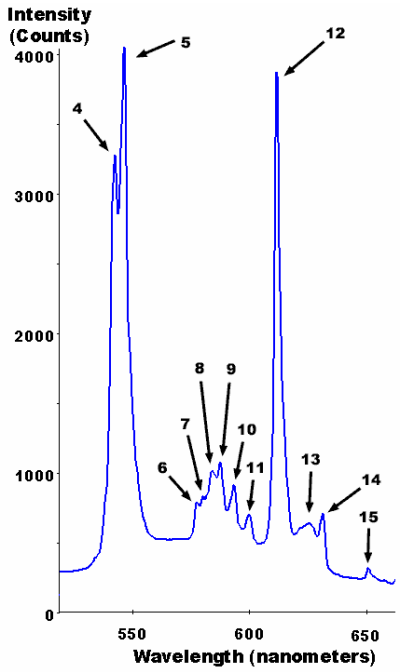


Fig.1 Fluorescence spectrum showing prominent mercury peaks

Spectroscopy can be classified according to the physical quantity which is measured or calculated, or according to the method of measurement. The general principle is that a (biological) sample is bombarded by radiation or by particles and that the output is radiation or particles and that holds for both types of input. The physical process is absorbance and/or emittance. The electro-magnetic spectra are basically line-spectra, but in practice they are often a continuous spectrum with many peaks (Fig. 1). In the context of medicine it is mostly performed with visible light (Vis), infrared (IR) light and UV light. Spectroscopy is often used for the identification of substances by emission or absorption. Fig. 2 shows the basic set up.

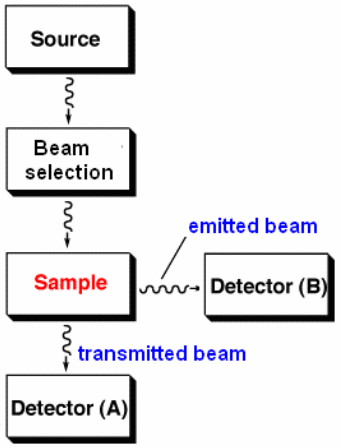


Fig. 2 Note that the Beam selection device becomes unnecessary if a Laser is used as light source (laser light is monochromatic).

Physical quantity measured
The type of spectroscopy depends on the physical quantity measured; normally the amount emitted or absorbed electro-magnetic radiation or the amplitude of small vibrations (acoustic spectroscopy).

Spectroscopically useful radiation covers the whole electromagnetic spectrum from high energy X-rays to low energy radio frequencies (Fig. 3). A vast number of different radiation sources has accordingly been developed, in many cases for the specific purpose of generating radiation for spectroscopic applications.

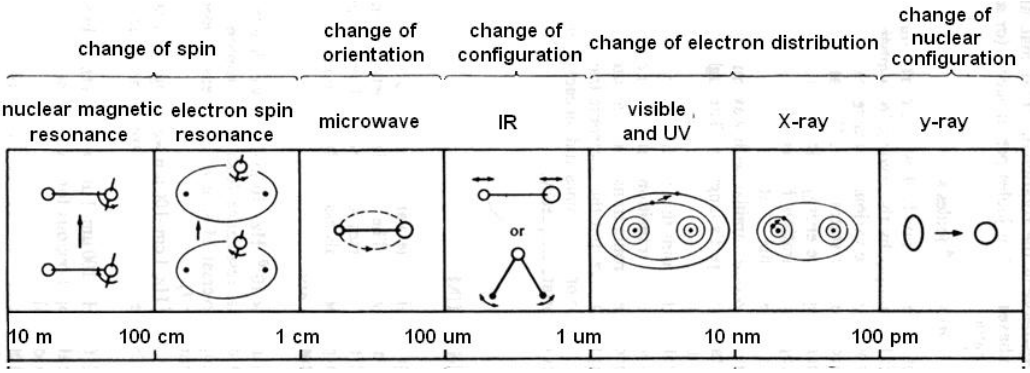


Fig. 3 The spectroscopic methods with their electro-magnetic wavelengths.

With nuclear magnetic and electron spin resonance the signal is the change of the spin. The microwave principle is the interaction between the oscillating electric or magnetic field and molecules that can undergo a change in dipole moment. Vibration or bending of a bond gives rise to vibrational spectra (IR- and Raman spectra). Visible and UV spectroscopy are based on absorption and emission due to a change of the energy states of outer electron orbitals. With X-ray (emission) spectroscopy the energy state of inner electron orbitals provide the signal of interest. Gamma spectroscopy is a radiochemistry measurement method that determines the energy and count rate of gamma rays emitted by radioactive substances.

Application

Applications are multitude in basic medical research and (bio)chemistry.
There are many spectroscopic techniques for element analysis, which are also useful for biomaterials, like atomic absorption, atomic fluorescence, Vis and UV spectroscopy (the latter two often combined), which causes often photo-ionization. IR and NMR spectroscopy are typically used for structure-analysis. However, the most relevant application of NMR is medical imagery (MRI).

More Info

Main types of spectroscopy

Absorption spectroscopy uses the range of electromagnetic spectra in which a substance absorbs. In atomic absorption spectroscopy, the sample is atomized and then light of a particular frequency is passed through the vapor. The amount of absorption can be related to the concentrations of various metal ions through the Lambert-Beer law, such as Na, K and Ca in blood. UV/Vis absorption spectroscopy is performed on liquid samples to detect molecular content and IR spectroscopy is performed on liquid, semi-liquid (paste or grease), dried, or solid samples to determine molecular information, including structural information.

Table 1 Characteristics of the techniques (After I.R. Beattie, *Chem. Soc. Rev.*, 1975, *4*, 107)

Technique	Nature of the effect	Information	Sensitivity	Potential Problems
X-ray diffraction	Scattering, mainly by electrons, followed by interference (λ = 0.01-1 nm)	Electron density map of crystal	crystal ~10^{-3} cm^3	Difficult to locate light atoms in presence of heavy atoms. Difficult to distinguish atoms with similar number of electrons
Neutron diffraction	Scattering, mainly by nuclei, followed by interference (λ = 0.1 nm)	oriented internuclear distances	crystal ~1 cm^3	Best method to locate hydrogen atoms. Allows determination of magnetic ordering. Requires large crystals (1mm and up)
Electron diffraction	Diffraction mainly by nuclei, but also by electrons (λ = 0.01-0.1 nm)	Scalar distances due to random orientation	100 Pa (1 Torr)	Thermal motions cause blurring of distances. Molecular model required. Gas phase and surface only
Microwave Spectroscopy	Absorption of radiation due to dipole change during rotation (λ = 0.1-30 cm; 300-1 GHz in frequency)	Mean value of r^{-2} terms; potential function	10^{-2} Pa (10^{-4} Torr)	Mean value of *r*-2 does not occur at *r*e even for harmonic motion. Dipole moment necessary. Only one component may be detected. Analysis difficult for large molecules of low symmetry
IR Spectroscopy	Absorption of radiation due to dipole change during vibration (λ = 10^{-1}-10^{-4} cm)	Symmetry of molecule Functional groups	100 Pa (1 Torr)	Useful of characterization. Some structural information from number of bands, position and possibly isotope effects. All states of matter
Raman Spectroscopy	Scattering of radiation with changed frequency due to polarizability change during a vibration (λ = visible usually)	Symmetry of mole-cule. Functional groups	10^4 Pa (100 Torr)	Useful for characterization. Some structural information from number of bands, position, depolarization ratios, and possibly isotope effects. All states of matter
Electronic Spectroscopy (UV-Vis)	Absorption of radiation due to dipole change during an electronic transition (λ = 10-10^2 nm)	Qualitative for large molecules	1 Pa (10^{-2} Torr)	All states of matter
Nuclear Magnetic Resonance Spectro-scopy	Interaction of radiation with a nuclear transition in a magnetic field (λ = 10^2-10^7 cm; 3 KHz to 300 MHz	Symmetry of molecule through number of magnetically equivalent nuclei Many many others	10^3 Pa	Applicable to solutions and gases. In conjunction with molecular weight measurements may be possible to choose one from several possible models
Mass Spectroscopy	Detection of fragments by charge/mass	Mass number,frag-mentation patterns	10^{-9} Pa (10^{-11} Torr)	Gas phase only. Fragmentation pattern changes with energy of excitation
Extended X-ray absorption fine structure (EXAFS)	Back scattering of photoelectrons off ligands	Radial distances, number, and types of bonded atoms	Any state	Widely used for metallo enzymes and heterogeneously supported catalysts

Emission spectroscopy uses the range of electromagnetic spectra in which a substance radiates. The substance first absorbs energy and then radiates this energy as light, evoked by for instance chemical reactions (e.g. bioluminescence) or light of a lower wavelength (fluorescence).
Scattering spectroscopy measures certain physical properties by measuring the amount of light that a substance scatters at certain wavelengths, incident angles, and polarization angles. One of the most useful applications of light scattering spectroscopy is Raman spectroscopy.

Other common types of spectroscopy (see also Table 1)

Fluorescence spectroscopy uses higher energy photons to excite a sample, which will then emit lower energy photons. This technique has become popular for its biochemical and medical applications, and can be used for confocal microscopy.

Mass spectrometry, see Mass spectrography.

X-rays spectroscopy and X-ray crystallography are specialized types to unravel structures of e.g. biomolecules.
Flame Spectroscopy Liquid solution samples are aspirated into a burner or nebulizer/burner combination, desolvated, atomized, and sometimes excited to a higher energy electron-state. The use of a flame during analysis requires fuel and oxidant, typically in the form of gases. These methods analyses metallic element in the part per million or billion (ppm or ppb), or even lower concentration ranges. Light detectors are needed to detect light with the analysis information coming from the flame.
Atomic Emission Spectroscopy This method uses flame excitation; atoms are excited from the heat of a very hot flame to emit light. A high resolution polychromator (a device to disperse light into different directions to isolate parts of the spectrum), can be used to produce an emission intensity vs wavelength spectrum over a range of wavelengths showing multiple element excitation lines. So, multiple elements can be detected in one run. Alternatively, a monochromator (device that transmits a selectable narrow band of wavelengths of light chosen from a wider range of wavelengths available at the input) can be set at one wavelength to concentrate on analysis of a single element at a certain emission line. Plasma emission spectroscopy is a more modern version of this method.
Nuclear Magnetic Resonance spectroscopy NMR spectroscopy analyzes certain atomic nuclei to determine different local environments of H, C, O_2 or other atoms in the molecule of an (organic) compound, often for structure-analysis at a microscopic or macroscopic scale (medical NMR).

These are various modifications of the above basic types.

Literature
http://131.104.156.23/Lectures/CHEM_207/structure_and_spectroscopy.htm

Thermography

Principle

Thermography, or digital infra red thermal imaging (DITI), is a type of infra red (IR) imaging. Thermographic cameras detect IR radiation (roughly 0.9-14μm) and produce images of the radiating body. Infrared radiation is emitted by all objects based on their temperature, according to the laws of (black body) electromagnetic radiation (see e.g. Wien's Law). Thermography makes it possible to "see" one's environment with or without visible illumination. The amount of radiation emitted by an object increases with temperature. Therefore thermography allows one to see variations in temperature, hence the name. An IR scanning device is used to convert infrared radiation emitted from the object surface into electrical impulses that are visualized in color on a monitor. This visual image graphically maps the surface temperature and is referred to as a thermogram. With a thermographic camera warm objects stand out well against cooler backgrounds. Humans and other warm-blooded animals become easily visible against the environment day or night, hence historically its extensive use can be ascribed to military and security services.

Medical DITI is a noninvasive diagnostic technique that allows the examiner to visualize and quantify changes in skin surface temperature. Since there is a high degree of thermal symmetry in the normal body, subtle abnormal temperature asymmetry's can be easily identified.
Medical DITI's major clinical value is in its high sensitivity to pathology in the vascular, muscular, neural and skeletal systems and as such can contribute to the pathogenesis and diagnosis made by the clinician. Attractive is its completely non-invasive nature and the use of a body-generated signal.

Application

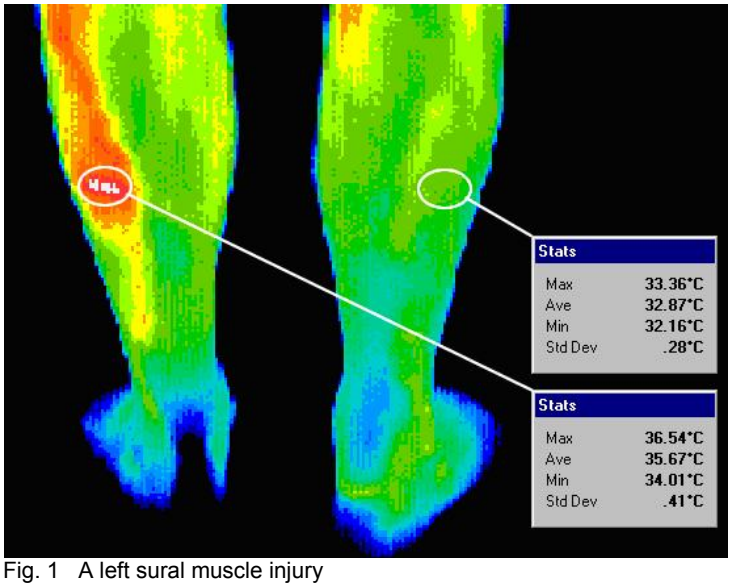


Fig. 1 A left sural muscle injury

Clinical uses for DITI include:
- To define the extent of a lesion of which a diagnosis has previously been made;
- To localize an abnormal area not previously identified, so further diagnostic tests can be performed;
- To detect early lesions before they are clinically evident;
- To monitor the healing process before the patient is returned to work or training.

Skin blood flow is under the control of the sympathetic nervous system. In healthy people there is a symmetrical dermal pattern which is consistent and reproducible for any individual. This is recorded in precise detail with a temperature sensitivity of 0.1°C by DITI. The neuro-thermography application of DITI measures the somatic component of the sympathetic response by assessing dermal blood flow. This response appears on DITI as a localized area of altered temperature, up to even 10° C with specific features for each anatomical lesion.

Regular Imaging applications are e.g. early diagnostics of breast cancer and superficial neuro-musculo-skeletal examinations). Further, applications in rheumatology and dermatology and sports medicine.

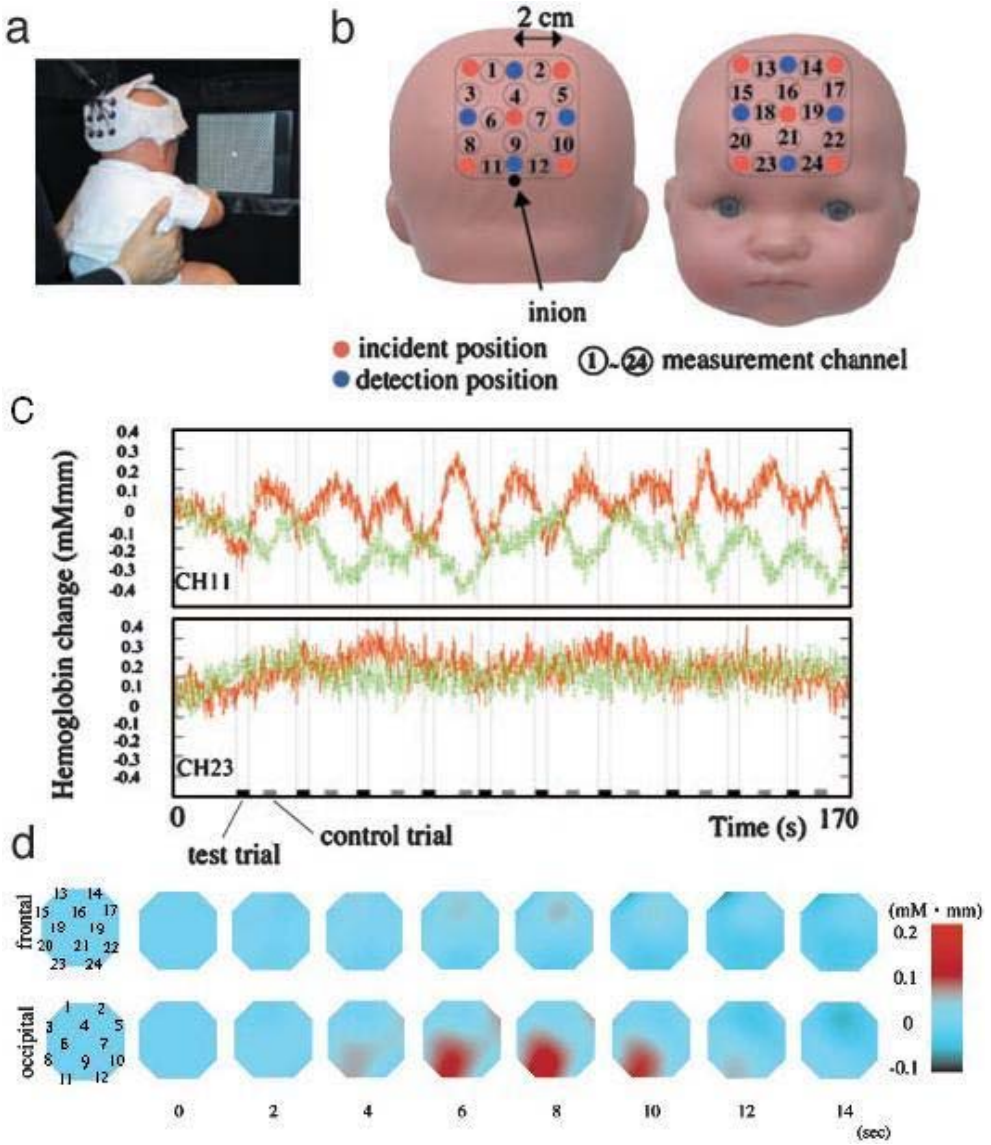


Fig. 2 Functional near-IR optical topography a). Child seated for a PC screen. b) channel configuration with emitting and receiving channels. c) Time series of relative changes in [oxy-Hb] (red line) and [deoxy-Hb] (green line) at channel 11 over the occipital cortex and at channel 23 over the frontal cortex of a 2-month-old infant. Black and gray bars indicate the test and control trials, respectively. d) response to visual stimulation. Epoch-averaged images of [oxy-Hb] over the occipital and frontal cortex of a 4-month-old infant (S6) are illustrated at 2-s intervals. Scale is relative [oxy-Hb] change from an arbitrary zero baseline. From ref. 1.

A limitation of the latter is the restricted depth of imaging. Yet, there are techniques to measure deep venous thrombosis.
Some other physiological activities can also be monitored with thermographic imaging, for instance brain activity, especially in infants as an alternative of the expensive fMRI technique. Actually it is an application in the field of spectroscopy, more precisely a multichannel near-IR optical topography measuring time courses of the levels of oxy-Hb (780 nm) and deoxy-Hb (830 nm) (Fig. 2).

Literature

1) http://www.pnas.org/cgi/reprint/100/19/10722
2) http://www.meditherm.com/therm_page1.htm

Wien's displacement law

Principle

Wien's displacement law is a law that states that there is an inverse relationship between the wavelength of the peak of the emission of a black body and its temperature. A black body is an object that absorbs all electromagnetic radiation that falls onto it. No radiation passes through it and none is reflected. Despite the name, black bodies are not actually black as they radiate energy as well since they have a temperature larger than 0 K. The law states:

$$\lambda_{max} = b/T, \quad (1)$$

where
λ_{max} is the peak wavelength (nm),
T is the temperature of the blackbody (K) and
b is Wien's displacement constant, $2.898 \cdot 10^6$ (nm·K).
Wien's law states that the hotter an object is, the shorter the wavelength at which it will emit most of its radiation. For example, the surface temperature of the sun is on average 5780 K. Using Wien's law, this temperature corresponds to a peak emission at a wavelength of 500 nm. Due to a temperature gradient in the surface boundary layer and local differences the spectrum widens to white visible light. Due to the Rayleigh scattering (see light scattering) of blue light by the atmosphere this white light is separated somewhat, resulting in a blue sky and a yellow sun.

Application

In medicine the application, together with Planck's law of black body radiation (see below) and Stefan-Boltzmann law is calculating heat transfer of body radiation (see Heat dissipation) in e.g. space and environmental medicine, and medical thermographic imaging (see Thermography). Stefan-Boltzmann law calculates the total radiated power per unit surface area of a black body. In the above applications the body generates the radiation.
As human donor of thermal radiation, applications are thermo radiotherapy and thermo chemotherapy of cancers, microwave/thermo therapies of tumors, low level laser therapy (LLLT, Photobiomodulation), and laser surgery (see Laser). Energy transfer and wavelength dependency are the primary interests. Further the technology of IR physical therapy equipment and application is (medical) cosmetic equipment (skin) and costumer's applications (IR and UV apparatus, sunglasses for UV A and B).

More Info

A light bulb has a glowing wire with a somewhat lower temperature, resulting in yellow light, and something that is "red hot" is again a little less hot.

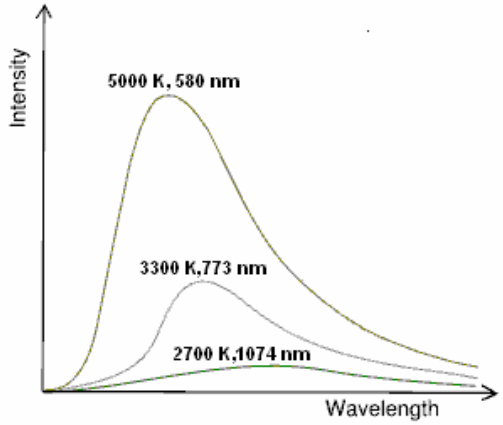


Fig. 1 As the temperature decreases, the peak of the black body radiation curve moves to lower intensities and longer wavelengths.

The shape of the spectrum in λ-notation is given by Planck's law of black body radiation:

$$I(\lambda,T)=\frac{2hc}{\lambda^3}\frac{1}{e^{\frac{hc}{\lambda kT}}-1}, \qquad (2)$$

where
I is spectral radiance ($J\cdot s^{-1}\cdot m^{-2}\cdot sr^{-1}\cdot Hz^{-1}$; sr is steradian, the SI unit of solid angle. and the 3-D equivalent of the 2-D radian. A sphere comprises 4π steradians, since its surface is $4\pi r^2$. A steradian can also be called a squared radian);
λ is wavelength (m);
h is Planck's constant (J/Hz);
c is speed of light (m/s);
k is Boltzmann's constant (J/K).

In terms of frequency *f* (Hz), Wien's displacement law becomes:

$$v_{max} = \alpha\cdot k\cdot T/h \approx 58.79\cdot 10^9\cdot T \text{ (Hz/K)},$$

where v_{max} is the peak frequency (Hz) and $\alpha \approx 2.821$ is a constant resulting from the numerical solution of the maximization equation.
Because the spectrum resulting from Planck's law of black body radiation takes a different shape in the frequency domain from that of the wavelength domain, the frequency location of the peak emission does **not** correspond to the peak wavelength using the simple relationship between frequency, wavelength, and the speed of light.
For more info, for instance equations in the frequency domain and Planck's law of black body radiation see e.g. Wikipedia.

Sound and ultrasound

Acoustic impedance

Principles

Acoustic impedance Z (or sound impedance) is the ratio of sound pressure *p* to particle velocity v. Also it is the product of the density of air ρ (rho) and the speed of sound c. The acoustic impedance Z is expressed in rayl (from Rayleigh, in $N \cdot s \cdot m^{-3} = Pa \cdot s/m$):

$$Z = \frac{p}{v} = \frac{J}{v^2} = \frac{p^2}{J} = \rho \cdot c \qquad (1)$$

with p = sound pressure, in N/m^2 = Pa = Pascal,
v = particle velocity in m/s,
J = sound intensity in W/m^2,
ρ (rho) = density of the medium (air) air in kg/m^3,
c = speed of sound (the acoustic wave velocity) in m/s.

v is the acoustic analogue of electric current, and p the analogue of voltage. Table 3 gives the densities (ρ), sound velocities (c) and acoustic impedance (Z) is of some bio-materials.

Table 1

material	density ρ (kg/m^3)	speed c (m/s)	impedance Z (rayl)
air (20 °C)	1.20	343	412
water (37 °C)	$1 \cdot 10^3$	1525	$1.52 \cdot 10^6$
brain	$1.02 \cdot 10^3$	1530	$1.56 \cdot 10^6$
muscle	$1.04 \cdot 10^3$	1580	$1.64 \cdot 10^6$
fat	$0.92 \cdot 10^3$	1450	$1.33 \cdot 10^6$
trabecular bone	$0.9 \cdot 10^3$	1540	$1.39 \cdot 10^6$
cortical bone	$1.9 \cdot 10^3$	4040	$7.68 \cdot 10^6$

In dry air at 20°C (68°F) the speed of sound is approximately 343 m/s, about 1 m each 3 ms. The velocity in brain, muscle and other bio-materials with a high water content, and trabecular bone is slightly higher. Fat has a slightly lower value of 1450 m/s and cortical bone with its much higher density ($1.9 \cdot 10^3$ kg/m3) 4040 m/s.

Application

Ultrasound Sound speed is a basic parameter in ultrasound applications, especially for moving objects (like the valves of the heart) which movement is determined by the Doppler effect (see Doppler principle).

More info

Hearing For an equal sound pressure in two materials, v is reciprocally with Z. For instance $Z_{water} \approx 4000\ Z_{air}$, and so the particle velocity in water is 4000 smaller than that in air. Therefore, also the particle velocity of a sound impinging from a source in the air onto the head is 4000 times smaller than in the air. The resulting vibration of the head gives rise to bone conduction, but with respect to the sound sensation evoked by the pressure impinging onto the eardrum it is irrelevant.

Speed of sound Under normal conditions air is nearly a perfect gas, and so its speed does hardly depend on air pressure and on humidity.
Sound travels slower with an increased altitude, primarily as a result of temperature and humidity changes. The approximate speed (in m/s) is:

$$c_{air} = (331.5 + 0.6 \cdot t_C) \quad (2)$$

where t_C is the temperature in ^{o} C. A more accurate expression is

$$c = \sqrt{\gamma \cdot R \cdot T} \quad (3)$$

where γ is the adiabatic index or c_p/c_V ratio, the ratio of heat capacity of the gas (c_p) with constant p and the specific heat capacity of the gas (c_V) with constant volume, T the absolute temperature (K), and R (287.05 J/(kg·K) for air) the universal gas constant (see Gas laws) In this case, the gas constant R, which normally has units of J/(mol·K), is divided by the molar mass of air. The derivation can be found in various textbooks. For air γ = 1.402.

See further Sound and acoustics, Ultrasound and the Doppler principle.

Contrast enhanced ultrasound, CEU

Principle

CEU is administrating *gas-filled microbubbles* intravenously to the systemic circulation in echography. Microbubbles in fluid and subjected to an ultrasound field show compressions and rarefactions, so they oscillate strongly and consequently reflect the waves. They are highly echogenetic, due to the large acoustic impedance difference between gas and liquid. There characteristic echo generates the strong and unique echogram of CEU. CEU can be used to image blood perfusion and flow rate in organs.

Bubbles in blood are thought to be covered by a *surfactant* of blood-macromolecules. However, often bubbles with an artificial skin (coating, e.g. phospholipids) are injected. This coating also improves stability and prevents dissolution.

The gas bubbles the most important part of the ultrasound contrast signal and determines the echogenicity. Surfactants lower the threshold for cavitation (the collapse of the bubble during rarefaction). The reflected signal as well as the signal emitted during cavitation can be used.
Common practiced gases are of N_2, or heavy gases like sulfurhexafluoride (F_6S), perfuorocarbon and octafluoropropane, which are all inert. Heavy gases are less water-soluble so they leak less to the medium, guarantying long lasting echogenicity. Regardless of the shell or gas core composition, microbubble size is fairly uniform, ranging *1-4 µm* in diameter. F_6S bubbles have a mean of 2,5 µm and 90%<6 µm. With these sizes, they flow easily through the microcirculation.
Selection of shell material determines how easily the microbubble is taken up and how long the bubbles survive.

Applications

Targeting ligands that bind to receptors characteristic of intravascular diseases can be conjugated to the microbubble skin, enabling the microbubble complex to accumulate selectively in areas of interest. However, the targeted technique has not yet been approved for clinical use; it is currently under preclinical research and development.
Genetic drugs can incorporated into ultrasound contrast agents. Gene-bearing microbubbles can be injected IV and ultrasound energy applied to the target region. As the microbubbles enter the region of insonation, the microbubbles cavitate, locally releasing DNA. Cavitation also likely causes a local shockwave that improves cellular uptake of DNA. By manipulating ultrasound energy, cavitation and so delivery can be visualized in the vessels with bubbles. Imaging can be performed before, just after IV and during cavitation, each with a different energy, to control the process of delivery.

Untargeted CEU is currently applied in echocardiography. Microbubbles can enhance the contrast at the interface between the tissue and blood. When used in conjunction with Doppler (see Doppler principle) Ultrasound, microbubbles can measure myocardial flow rate to diagnose valve problems. The relative intensity of the microbubble echoes can also provide a quantitative estimate on blood volume.
In vascular medicine, bubbles visualize perfusion. Therefore this technique is crucial for tracking down a stenosis.

Targeted CEU is being developed for a variety of medical applications. Microbubbles targeted with ligands are injected systemically in a small bolus. The ligands bind to certain molecular markers that are expressed by the area of imaging interest. Microbubbles theoretically travel through the circulatory system, eventually finding their respective targets and binding specifically. Ultrasound waves can then be directed on the area of interest.
Specific applications are to visualize for instance inflammatory organs (Crohn's disease, arteriosclerosis, heart attacks).
Microbubbles-targeted ligands can bind receptors like VEGF to depress angiogenesis in areas of cancer. Detection of bound targeted microbubbles may show the area of expression. This can be indicative of a certain disease state, or may identify particular cells in the area of interest.
Drugs can be incorporated into the microbubble's lipid shell. The microbubble's large size relative to other drug delivery vehicles like liposomes may allow a greater amount of drug to be delivered per vehicle. By targeted the drug-loaded microbubble with ligands that bind to a specific cell type, microbubble will not only deliver the drug specifically, but can also provide verification that the drug is delivered if the area is imaged using ultrasound. Local drug delivery is used for angiogenesis, vascular remodeling andr tumor destruction.
The force associated with the bursting may temporarily permeabilize surrounding tissues and allow the DNA to more easily enter the cells. This can further be facilitated by coating materials of the shell, e.g.

liposomes, positively charged polymers, and viruses (as they do already for millions of years for delivering genetic materials into living cells).

More info

Gas bubbles in a liquid are characterized by a resonance frequency *f*, which is directly related to their diameter R_0. *f* can by approximated (liquid surface tension not included; S. De, 1987) by:

$$f = 0.5 \cdot \pi^{-1} \cdot R_0^{-1} \left[(3\gamma / \rho_l) P_0 \right]^{0.5} \quad (1)$$

with γ the specific heat ratio of the gas (=1.40 for N_2; see Gas laws), ρ_l the liquid density (=1050 kg/m^3 for blood) and P_0 the ambient pressure (here 114.7 kPa). With R_0=2.5 µm, f is about 1.36 MHz. With R_0>100 µm (1) is accurate but with $R_0 \leq 10$ µm f is about 3.5% too small. Implying the surface tension of blood (=0.058 N/m) would increase f with 7.7%, but bubbles of this size will be surrounded by a surfactant skin, which counteracts the effect of the surface tension (see **More info** of Surface tension). In pure water, bubbles have an extremely sharp resonance peak (quality Q about 70) but in blood and with the skin surfactant, this is much lower due to the skin-shear and skin-stiffness. With the heavy multi-atomic gases *f* is smaller since γ is smaller (for F_6S γ =1.10) and ρ_l considerably larger.

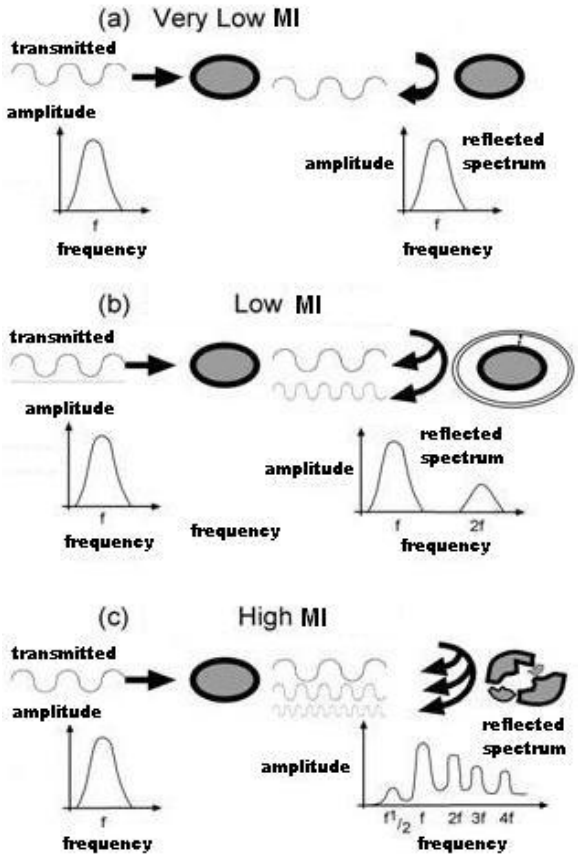


Fig. 1 Different MIs produce different reflected spectra.

Micro-bubbles oscillate (expand and contract) in the ultrasound field. The pattern and nature of their oscillation, and thus the nature of the backscatter signal, differs, depending on the transmitted acoustic power. The power of the ultrasound field is expressed as the mechanical index (MI; see Ultrasound). With *very low* MI (< 0.1), micro bubbles demonstrate linear oscillation (reflected frequency equals impeding frequency). *Low* MIs (0.1 – 0.6), generates nonlinear oscillation of the micro bubble whereby expansion is greater than contraction. In addition to the usual backscatter of the fundamental frequency, the bubbles also produce backscatter of harmonics (see Signal analysis and Fourier). When exposed to *high* MI (> 0.6, i.e. the MI used for standard imaging) the bubbles oscillate wildly and burst. Upon destruction, micro-bubbles produce a brief, high amplitude signal, with good resolution, which is rich in harmonic signals, containing backscatter at the second third and fourth harmonics etc (Fig. 1).

The most important limitation of this technique is motion artifact from tissue, because tissue motion will be expressed like bubble destruction, potentially showing perfusion when none is present (a false negative). If Doppler frequency is increased, pulse separation is reduced, so tissue movement between pulses can be minimized. However, if the pulses are too close, not all the gas within the bubble will have dissipated before arrival of the next pulse, so the difference between pulses is reduced, possibly leading to false positive perfusion defects. Air-filled micro-bubbles are optimal for this technique because of rapid dissipation of the gas, allowing closely spaced pulses.
There are several high MI techniques, some developed for the moving myocard.

Triggered harmonic imaging
Intermittent high power imaging can improve imaging, with the best opacification obtained using intermittent harmonic imaging. During intermittent high power imaging, high energy ultrasound is transmitted at specified intermittent intervals, triggered to the ECG (e.g. 1 of 4 cardiac cycles). The time between destructive pulses allows the micro-bubbles to replenish the myocardium. With each destructive pulse, high amplitude backscatter rich in harmonics is returned to the transducer, enabling static images of myocardial perfusion.

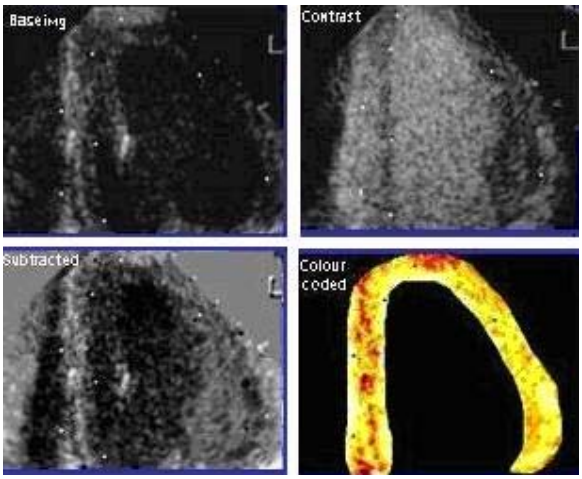


Fig. 2 From top left to bottom right: Base imaging, no bubbles; contrast, bubbles; subtracted, echo's subtracted; gray-scale, color coded.

Pulse Inversion Doppler
Another grey scale high MI technique is pulse-inversion imaging whereby two synchronized beams of pulses impinging onto the myocardium. The second pulse sent is a mirror image of the first (i.e. 180° phase shift). The scanner processes the echo's of the two types of pulses by adding them together. When the bubbles generate a linear echo, the addition of one pulse to the other should cancel out to zero and no signal is generated. However, micro-bubbles produce non-linear echo signals at high MI and the summation of returning pulses will not equal zero.
Using this technique, processing can theoretically be limited only to signals generated by bubbles and not by other objects. However, tissue motion artefacts are a major limitation, as movement of tissue also creates non-linear signals.

More specific applications can be found in ref. 5.

Physical advantages of CEU
- Ultrasonic molecular imaging is safer than molecular imaging modalities such as radionuclide imaging.
- Since microbubbles can generate such strong signals, a lower intravenous dosage is possible; micrograms compared to milligrams for other molecular imaging modalities such as MRI contrast agents.

Physical disadvantages of CEU
- Ultrasound produces more heat as *f* increases, so *f* must be carefully monitored.
- Equipment settings are subjected to safety indexes (see Ultrasound). Increasing ultrasound energy increases image quality, but microbubbles can be destructed, resulting in microvasculature ruptures and hemolysis.
- Low targeted microbubble adhesion efficiency, which means a small fraction of injected microbubbles bind to the area of interest.

Literature

1 De S. On the oscillations of bubbles in body fluids. J Acoust Soc. Amer, 1987, 81, 56-567.
2 Postma, M., Bouakaz A., Versluis M. and de Jong, N. IEEE T Ultrason Ferr, 2005, 52, in press.
3 http://en.wikipedia.org/wiki/Contrast_enhanced_ultrasound
4 http://e-collection.ethbib.ethz.ch/ecol-pool/diss/fulltext/eth15572.pdf
5 http://www.cardiovascularultrasound.com/content/2/1/15

Doppler echocardiography

Principle

The echocardiogram is the most common application of echography imaging and the Doppler velocity measurement (see Doppler principle). It allows assessment of cardiac valve function, left-right shunt (e.g. open oval foramen), leaking of blood through the valves (valvular regurgitation), and calculation of the cardiac output.
The echo-image is depicted in black and white, and the direction and velocity of the blood are depicted in red for approaching the probe and blue for removing away. The more blue or red the higher the velocity. See Echography and Contrast enhanced ultrasound for further explanation and examples of images in figures.

Application

Contrast echocardiography (CE, e.g. for detection of the right heart Doppler signals, intra-cardiac shunts). It uses intravenously administered micro-bubbles to traverse the myocardial microcirculation in order to outline myocardial viability and perfusion, see Contrast enhanced ultrasound.

Echography

Principle

Echography, also called medical (ultra)sonography is an ultrasound-based diagnostic imaging technique.
Echography uses a probe containing acoustic transducers (generally of piezo crystals, see piezoelectricity) to send sound into a material (here tissue). Whenever a sound wave encounters a material with a different acoustical impedance, part of the sound wave is *reflected* (see Waves) which the probe detects as an *echo*, see Fig. 1.

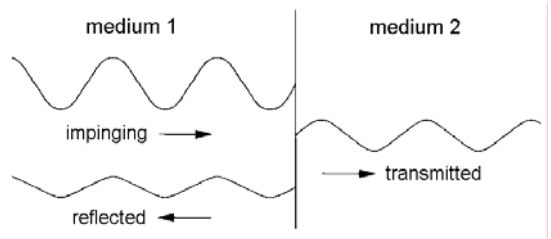


Fig. 1

The time it takes for the echo to travel back to the probe is measured and used to calculate the depth of the tissue interface causing the echo. The greater the difference between acoustic impedances, the larger the echo is. The reflection coefficient R is:

$$R = \frac{A_r}{A_i} = \left| \frac{Z_1 - Z_2}{Z_1 + Z_2} \right| \qquad (1)$$

with A_r and A_i the amplitudes of reflected and impinging wave and Z_1 and Z_2 the acoustic impedences of medium 1 and 2 respectively. When these media are respectively water (Z=152.000 rayl) and air (Z=400 rayl) then R=0.999474, which is equivalent to a transmission loss of 20log(1-R)=65.6 dB. When medium 1 is air and 2 is water, the same holds (see (1)). Consequently: *with a large ratio of the both impedances, the reflection is large*.
The above consideration does not take in account scatter from the object, which diminishes the reflectance and disturbs imaging. Taking water as substitute for blood, R of blood-muscle interface is only 0.034 (see for values Acoustic impedance), which asks for highly sofisticated hardware and software to obtain a good image (noise reduction). When bone is involved, R is some 2-20 times higher.
A water-based gel ensures good acoustic coupling between skin and the ultrasound scan head.
A 2D-image can be obtained by a probe with many transducers and a 3-D images can be constructed with a specialized probe.
From the amount of energy in each echo, the difference in acoustic impedance can be calculated and a color is then assigned accordingly.

The echographic modes
In the A-mode the strength, i.e. **a**mplitude, of the reflected wave is indicated on the vertical axis and time at the horizontal one, with time zero at the origin.
In the B-mode the strength is indicated by the **b**rightness (grayscale) of a dot. With the B-scan, along the vertical axis the penetration depth is indicated. The beam of the ultrasound changes slightly its angle of incidence every time a new sound pulse is emitted. In this way a kind of section of the anatomical object is obtained. However, the less depth, so the closer to the sound source, the more compressed is the image in the horizontal direction (so parallel to the surface). Such a scan is made many times a second and in this way a kind of movie is made. This is very helpful for moving structures, especially the heart (Fig. 2).

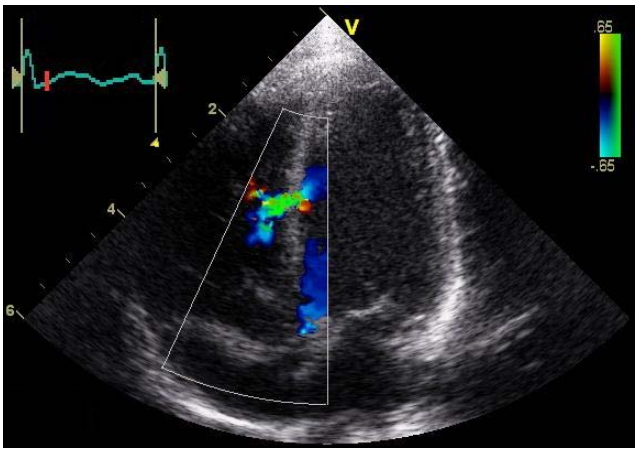


Fig. 2 Abnormal echocardiogram showing a mid-muscular ventricular septum defect in a new-born child. The red line in the ECG mark the time instant that the image was captured. Colors indicate blood velocity measured with the combined Doppler apparatus.

In the *M-mode*, **m**ovement is visualized by displaying time along the horizontal axis and an image is made for a single beam direction. When for instance this beam impinges on a mitralis valve, the image shows opening and closing of the valve (Fig. 3).

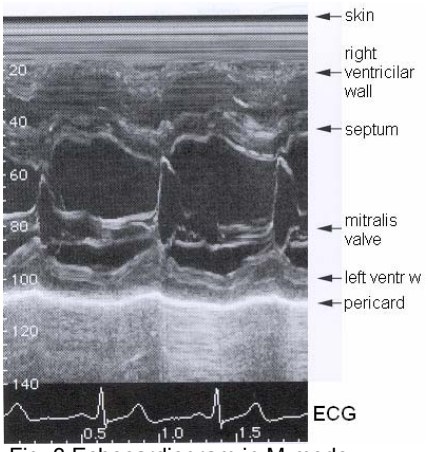


Fig. 3 Echocardiogram in M-mode

Technical strengths

- It images soft tissues very well and is particularly useful for delineating the interfaces between solid and fluid-filled spaces.
- It renders "live" images.
- It shows the structure as well as functional aspects.
- Widely available and comparatively flexible.

Small, cheap, easily carried scanners (bedside) available.

Mechanical weaknesses

- Classical ultrasound devices have trouble penetrating bone but current research on ultrasound bone imaging will make it possible with dedicated devices in the future.
- Performs very poorly when there is a gas between the scan head and the organ of interest, due to the extreme differences in acoustical impedance.
- Even in the absence of bone or air, the depth penetration of ultrasound is limited, making it difficult to image structures that are far removed from the body surface, especially in obese patients.
- The method is operator-dependent. A high level of skill and experience is needed.

Applications

Echography is widely utilized, often with a hand-held probe. It is especially practiced in cardiology, gynecology and obstetrics, urology (kidney, prostate), vascular medicine, gastroenterology (also liver), endocrinology and ophthalmology.

More info

Limitations
The spatial resolution in the axial direction (the depth) is directly related to the wavelength λ of a pure ultrasound frequency. In the lateral direction the resolution is determined by the width (the aperture angle) of the beam due to divergence.
Further, there is an ambiguity in depth position. This occurs when the time lapse between sending and receiving a wave is larger than the period time t_{per}, The reflected wave of the objects at all distances n times 0.5λ are superimposed at the reflected pulse of the object itself. Mathematically: depth = $(0.5t_{recieve}/t_{per}, - \text{integer}\{0.5t_{recieve}/t_{per}, -\})c + n\lambda$. This problem is solved by sending short pulses and adjusting the pulse interval time such that any reflection from boundaries to at least the depth of interest are arriving within the pulse interval time. Since a pulse comprises many frequencies (see Signal Analysis and Fourier) the received signal needs some complicated computation (deconvolution) to reconstruct the echo-image.

Doppler echography

There exist several types of echography, nowadays combined with employ the Doppler effect (see Doppler principle): the Doppler echography. This technique assess whether structures (usually blood) are moving towards or away from the probe, and its relative velocity. By calculating the frequency shift of a particular sample volume, for example a jet of blood flow over a heart valve, its velocity and direction can be determined and visualized. This is particularly useful in cardiovascular studies and vascular examinations of other organs (e.g. liver portal system). The Doppler signal is often presented audibly using stereo speakers: this produces a very distinctive, although synthetic, sound. Doppler echography can by distinguished in several modifications. The most common ones are here discussed.

The duplex (Doppler) scanner
The duplex scanner detects in a selected part of the image the moving blood by using the Doppler effect. The scanner calculates the actual velocity of the blood provided the angle between the direction of the ultrasonic beam and the direction of movement is known. The operator therefore aligns a marker along the direction of flow in the blood vessel and positions a cursor at the height of the peak systolic blood velocity.
In the common and superficial femoral arteries, the waveform normally has a forward component followed by a reverse component and a second smaller forward component. This is called a triphasic waveform because of the three phases. More distally in the superficial femoral artery, the second forward component may be absent, giving a biphasic waveform with two phases.

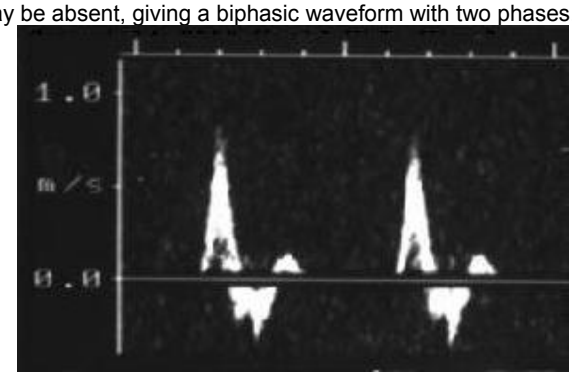


Fig. 4 Duplex scan of superficial femoral artery with a triphasic blood velocity waveform Horizontal axis: time (large ticks 1 s); vertical axis velocity.

The frequency shift is normally in the audio range (due to the ”ultra” frequencies), so most duplex scanners send the signal to a pair of audio speakers, and this enables the operator to hear the signal in addition to seeing the display (as in Doppler echocardiography).

Color Doppler scanners
Color Doppler (CD) scanners detect and display moving structures by superimposing color onto the grey-scale image. Color is superimposed wherever the scanner detects a moving structure, usually blood (Fig. 2). The hue of the color shows the direction and magnitude of the blood velocity. In this image, red and yellow indicate flow away from the probe, with dark red representing low velocities and

orange and yellow indicating higher velocities. Flow towards the probe is indicated in blue and green, with green indicating higher velocities. The hue can therefore be used to identify sites where the artery becomes narrower and the blood has to move faster to achieve the same volume flow rate. When the blood velocity exceeds the limit of the color scale, aliasing occurs (high velocity in one direction interprets as lower velocity in the other, wrong, direction). Color Doppler can also be used to display venous blood flow.
CD and Duplex sonography are often combined to Duplex/CD sonography, especially for assessing stenoses.

Power Doppler (PD)
Duplex/CD sonography is not an effective technique when the artery under study is almost perpendicular to the ultrasonic beam or by other poor conditions as bowel gas, breathing movements and obesity. Power Doppler (PD) has improved diagnostic capabilities of vascular Duplex/CD sonography, mainly because it is independent from the angle of insonation and has more sensitivity. PD generates an intravascular color map reflecting the integrated power in the Doppler signal, which essentially depends on the density of red blood cells within the sample volume. However, due to its intrinsic limitations, PD cannot replace conventional sonographic techniques and especially CD. So, PD is used as an adjunctive tool in vascular sonography.

Tissue Doppler Imaging (TDI)
Tissue Doppler Imaging (TDI) measures and displays peak velocities high temporal resolution (ca. 8 ms) at any point of the ventricular wall during the cardiac cycle. The mean velocities can be calculated with time velocity maps and displayed as color encoded velocity maps in either an M-mode or a two-dimensional format. Indeed, since all the points within the ventricular walls are velocity-encoded in real-time, the color-coded display should provide a huge amount of information which could form the basis for the application of accurate, reproducible quantitative evaluation.

Literature
http://en.wikipedia.org/wiki/Ultrasonography
http://www.worldwidewounds.com/2000/sept/Michael-Lunt/Doppler-Imaging.html#Colour%20Doppler%20scanners

Optoacoustic imaging

Principle

Central to optoacoustic imaging is the optoacoustic effect whereby pulsed laser energy is absorbed by a medium causing a local temperature increase followed by the generation of pressure transients (acoustic waves).
In optoacoustic imaging, short laser pulses irradiate sound scattering (see Waves) biological tissue and adiabatically (see Compression and expansion) heat absorbing structures, such as blood vessels, to generate ultrasonic pressure transients by means of the thermo elastic effect. These acoustic transients propagate to the tissue surface and are recorded with transducers (ultrasound or electromagnetic) to reconstruct high contrast images of the initial absorbed energy distribution exactly resembling the absorbing structures. They can be recorded using high frequency pressure sensors (piezoelectric or optical). The speed of sound in tissue (~1500 m/s) allows for the time resolved detection of these pressure waves and the determination of depth from where these pressure waves originated. By using an array of sensors the temporal delay of these incoming pressure wave fronts can be combined into an ultrasound image. The optoacoustic technique combines the accuracy of spectroscopy with the depth resolution of ultrasound.

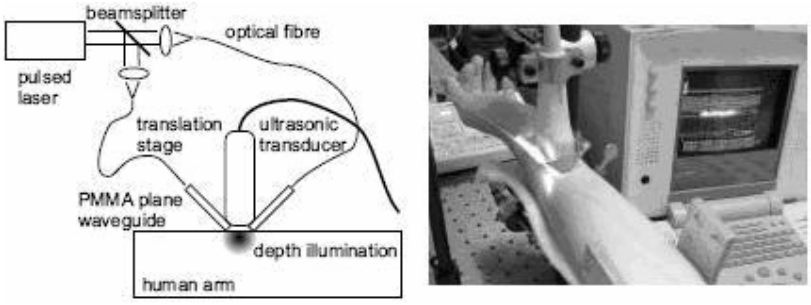


Fig. 1

Application

The *in vivo* optoacoustic images acquired from human finger, arm and legs show high contrast detailed blood vessel structures, which are hard to see on the corresponding ultrasound echography images.
The optoacoustic imaging and ultrasound techniques extract complementary information, which strongly suggests a combination (overlay) of the two techniques in a single device.
There are acoustic imaging devices in development for breast cancer detection, also utilizing the change in the optical properties of blood in respect to oxygen saturation and the strong optical contrast between hemoglobin and surrounding tissue.

More Info

Two real-time optoacoustic imaging systems have been developed recently:
(1) Laser excitation combined with classical medical ultrasound system for comparison of the two complementary techniques. In medical applications real-time imaging avoids motion artifacts (heartbeat and breath), facilitates imaging procedure and allows instant diagnosis.
(2) Optical systems detecting the Schlieren effect and equipped with an acoustic lens system for 3D imaging (Fig. 1). Schlieren are optical inhomogeneities in transparent material, often liquid, mostly not visible to the human eye. They are applied to visualize flow, based on the deflection of light by a refractive index (see Light) gradient (resulting e.g. from temperature of salt gradients). The index gradient is directly related to flow density gradient.) This method is still experimentally.
Both methods can be combined with ultrasound echography.

In the 2nd optical system, the Schlieren transducer images the pressure transient in a fluid filled cuvette below the tissue with an ns-flash lamp and reconstructs images on a computer for visualization.
The first optical system uses an acoustic lens to directly reconstruct a 3D image of the original pressure distribution. This copied pressure image is optically dark field imaged at two angles to provide a stereo image pair of the original absorbing structures.
Both optical systems operate at real-time frame rates of 10-20 Hz and provide high resolutions up to 30-100 µm.

Medical ultrasound is limited by low acoustic contrast, which particularly deteriorates or inhibits imaging of smaller structures in near skin regions. The two systems combining laser excitation and commercial ultrasound are provide high optical contrast and sub-millimeter acoustical spatial resolution for *in vivo* biomedical imaging. Variation of the laser wavelength allows spectroscopy and functional imaging of blood oxygenation level based on oxygen dependent Hb absorption spectra.

Literature

After http://e-collection.ethbib.ethz.ch/ecol-pool/diss/fulltext/eth15572.pdf
"http://en.wikipedia.org/wiki/Optoacoustic_imaging"

Phonocardiography

Principle

For auscultation of the heart, i.e. listening to the sounds generated by the beating heart, the classical acoustic stethoscope and more and more frequently the electronic stethoscope is used (see Stethoscope). Performed with the latter is phonocardiography. It is a method not solely to record the sounds produced by the heart valves, but also to visualize them on a PC and to analyze them with a computer algorithm in order to discover dysfunction of a valve.

Heart sounds
In healthy adults, there are two normal heart sounds often described as a *lub* and a *dub* (or *dup*), that occur in sequence with each heart beat. These are the first heart sound (S1) and second heart sound (S2).
S1 is caused by the sudden block of reverse blood flow due to closure of the atrioventricular valves, mitral and tricuspid, at the beginning of ventricular contraction, or systole.
S2 is caused by the sudden block of reversing blood flow due to closure of the aortic valve and pulmonary valve at the end of ventricular systole, i.e beginning of ventricular diastole.
S1 and S2 can by split in various components. The extra sounds S3 and S4 are rarely audible. In addition, murmurs, clicks and rubs can be audible. They are especially of importance for diagnostics.

Analysis
The analysis is generally a spectral analysis (Fast Fourier Transform) of the whole signal or of one of the sounds. Very suitable is a frequency (vertical) versus time plot with the amplitude (or power) depicted in a color scale, just as done in sonograms (see Fig.2 of Sound Spectra).

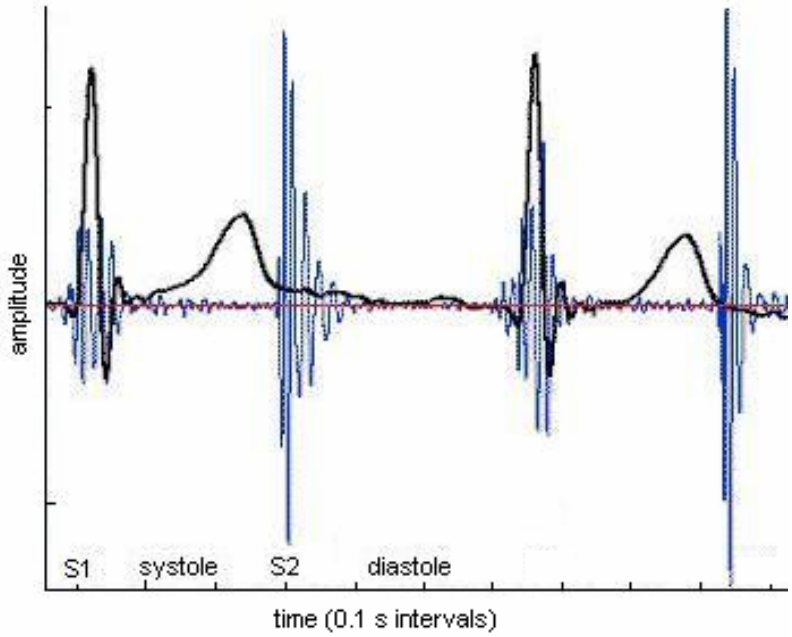


Fig. 1 Phonocardiogram (blue) with superimposed ECG (black) of nearly two heart cycles. Note the correspondence between S1 and QR, and S2 and end of T wave.

Application

Routinely by physicians of many disciplines and cardiologists.
With phonocardiography, pressure and sound manifestations of the pulse wave spreading in the cardiovascular system can be recorded. In addition to standard diagnostic application of phonocardiography can be used to measure pulse wave velocity. Then two signals are needed at minimum for an estimation of the pulse wave velocity (by cross-correlation). These signals have to be measured simultaneously from different points on a human body.
Stethoscopes are also used by mechanics to isolate sounds of a particular moving engine part for diagnosis.

More Info

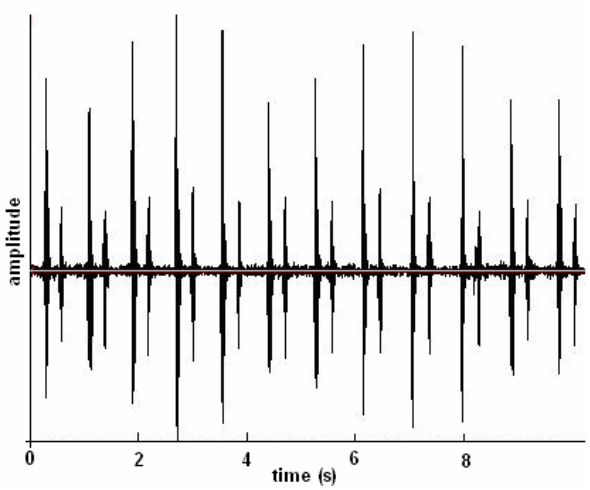


Fig. 2 Amplitude variation due to respiration typical for the healthy heart.

For digital signal processing of phonocardiograms see the review Crit Rev Biomed Eng. 1995; 23, 63-219 and also Proceedings of the 3rd IEEE International Symposium on phonocardiography, 14-17 Dec. 2003, 491-493.

Sound and Acoustics

Basic Principles

In any elastic medium, gas, liquid or solid, an acoustic wave can be described as mechanical compressions and rarefactions. To be propagated, in liquids and solids, the elasticity guarantees some compressibility, which is necessary to evoke and propagate sound pressure. The wave comprises a pressure variation within the medium and as well as a to-and-fro oscillation of the component particles. The oscillations of the particles can be expressed as displacement in time (and space). The simplest case is a standing (not traveling) sinusoidal wave. Such a wave occurs when a tout string of a violin is plucked to make a pure tone. However, this is a transversal wave. Longitudinal standing waves are produced by organ pipes. With a standing wave, the sound pressure within a single sound cycle is maximal when the displacement of particles of the medium is maximal. At that instant, the particles move towards each other and 0.5λ further away from each other, causing a compression and rarefaction respectively.
In formula, the standing wave is described as $y = y_0 \sin(2\pi f t)$ with $y_0 = \sin(2\pi x'/\lambda)$ with f the frequency, t time, λ the wavelength, x' the distance from the place of maximal excursion (the crest). However, most waves are traveling or propagating and then the equation comprises a single sine wave which is a function of t and the distance from the sound source x:

$$y(x,t) = y_0 \sin(\omega\,(t - x/c)) \qquad (1)$$

with *y(x,t)* the displacement of particles from the equilibrium position, y_0 the maximal amplitude of the particle displacement, ω the angular frequency (=2πf), c the sound propagation velocity. x/c is the time that the wave needs to travel the path x.
When the particle oscillations are in the direction of propagation, the wave is longitudinal. When the direction is perpendicular on the propagation direction it is a transversal wave, like a taut string of a violin. Equation (1) holds for both types. However, generally sound is a longitudinal wave phenomenon (in the so called far-field, see More Info). Fig. 1 illustrates the propagation of a longitudinal sound wave.

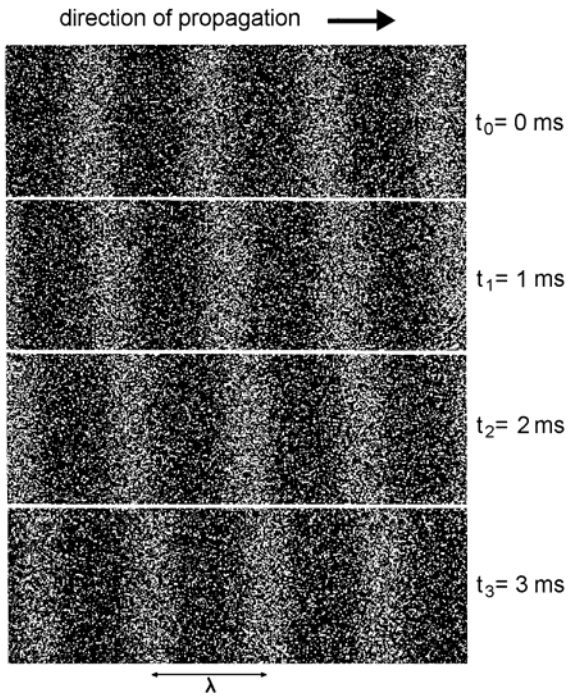


Fig 1 Longitudinal sound wave visualized at 4 time instants. The density of the white specks represents the particle density in the medium and so the sound pressure. The propagation direction is arbitrarily chosen to the right. The distance between to compression peaks is the wavelength λ. Shifting a peak over the distance of one wavelength λ takes about 6 ms. So, the frequency is ca. 167 Hz.

The oscillatory displacements of the particles is denoted with the vector **ξ** (vectors are denoted in bold), its time derivative, the particle velocity with v. The particle motion around an equilibrium position causes the pressure increase and decrease. This sound pressure p is superimposed at the ambient (the atmospheric) pressure. Sound pressures are generally much smaller than atmospheric pressure. Even at the human pain limit some 5000 smaller (Table 1 Hearing and Audiometry). With a longitudinal displacement wave the pressure wave is also longitudinal (see Fig. 1).
With a sound wave through a medium, the particle displacement is the physical movement of a particle (e.g. a molecule of air). Far from the source the direction of the oscillations is the same as that of the sound wave. Close to the source the direction of the oscillations is dependent on the type of sound source (see **More Info**, near and far field). For a free propagating wave with a plane wave front, the sound pressure (the scalar p) and the particle velocity v are mutually related such that $p = \rho c v$, with ρ being the (characteristic) density of the medium. The acoustic impedance Z is $Z = \rho c$, in $N \cdot s \cdot m^{-3}$ (rayl, from Rayleigh). v is the acoustic analogue of electric current, and p the analogue of voltage. So, Z is also p/v and consequently, pv or p^2/Z gives the intensity J (in W/cm^2) of the sound. When the wave acts on surface A we obtain the power J·A (W).
Notice that in a gas the particle motion is random and many 100 m/s (see Gas laws), whereas the motion due to sound is coherent for all particles and in the nm to mm range. This motion is superimposed on the random motion.

Sound pressure
The amplitude or strength is expressed as the effective amplitude (= $0.5\sqrt{2}$ of the maximal amplitude of a sinusoidal pressure wave) of the sound pressure. SPL or sound level L_p is a logarithmic measure of the energy of a particular sound relative to a reference sound source.

$$L_p = 20\log(p_1 p_2) = 10\log((p_1^2/p_2^2)) \text{ dB SPL} \quad (2)$$

where p_0 is the reference sound pressure and p_1 is the sound pressure being measured.

Application

The use of sound in medical practice is enormous. In much classical application the physician is the listener to evaluate the sound. The sound can be evoked by him (palpation) or is generated by the subject (patient) and listened by an instrument (stethoscope etc.). Many modern applications generate sound as a probe and record the sound evoked by the probe. Examples are found in echography, audiology, cardiology etc.

More Info

Sound intensity
The sound intensity, *I*, (acoustic intensity, W/m^2) is defined as the sound power P_{ac} per unit area A. It is also equal to P_{eff}^2/Z and to pv. The usual context is the measurement of sound intensity (W/m) in the air at a listener's location.

$$I = \frac{1}{T} \cdot \int_0^T p(t) \cdot v(t) \cdot dt \quad (3)$$

For a spherical sound source, the intensity as a function of distance r is:

$$I_r = P_{ac}/A = P_{ac}/(4\pi r^2) \quad (4)$$

The sound intensity *I* in W/m^2 of a plane progressive wave is:

$$L_J = 10\log(J_1/J_0) \text{ dB SPL} \quad (5)$$

where J_1 and J_0 are the intensities. If J_0 is the standard reference sound intensity, where $J_0 = 10^{-12}$ W/m^2. Then instead of "dB" we use "dB SIL". (SIL = sound intensity level).
Our ears as sensors cannot convert sound intensities and powers. In most conditions for the sense of hearing, the auditory system only utilizes the sound pressure as input signal (see Hearing and Audiometry).
With all these parameters defined, now particle displacement can be represented in terms of many combinations of the other parameters.

$$\xi = \frac{v}{2 \cdot \pi \cdot f} = \frac{p}{Z \cdot \omega} = \frac{1}{\omega}\sqrt{J/Z} \quad (6)$$

Sound velocity

Sound's propagation speed c (from Latin *celeritas*, velocity) depends on the type of the medium (a gas, fluid or solid) through which it propagates, and the ambient temperature and pressure. See Sound impedance for the sound velocity of some (biomaterials).

Near-field and far-field

The near- and far-field are concepts to indicate whether an object (or listener) is close or remote from a sound source in relation to the emitted frequency (see Kalmijn 1988). The border between near-field and far-field is defined as $\lambda/2\pi$. The behavior of pressure and displacement is different in the near- and far-field, the actual reason why this distinction is made.

For hearing in the air with air borne sound, only the far-field is of importance and sound is perceived by the sound pressure received by the ear drum. Within the near field particle displacement also plays a role. Now, sound is also transmitted by particle displacement and perceived via conduction by the tissues of the head (with bone conduction dominating). Listening with headphones is based on a transmission mixture of pressure and displacement, with the ear drum and the head-tissues as receiver.

Sound sources

The easiest sound source to deal with mathematically is the monopole source, a pulsating sphere that produces spherical waves. At distances from the source that are large with respect to λ, the amplitudes of v and p decrease linearly with the distance from the sound source (R), as does the time integral ξ of v (i.e., the particle displacement, and the acceleration, respectively). It is clear that intensity of the sound decreases with R^2. For $R << \lambda$, p decreases linearly but ξ decreases with the R^2. The phase difference between p and v depends upon R. When $R >> \lambda$, p and v are in phase with one another. However, for $R < \lambda$, p leads v for at least 45°, up to a maximum of 90°. For distances approximately $\lambda/2\pi$, there is a gradual transition from the near- to the far-field. In addition to the near- and far-field effects of λ, there is also an effect of the frequency f (where $f=c/\lambda$). Under the condition that v at the interface of the pulsating source is the same for all frequencies, p is proportional to f, irrespective of R. In the near-field, displacement is proportional to 1/f, but in the far-field displacement is independent of f.

Many (biological) sound sources are dipoles (vibrating spheres) rather than monopoles. Dipole sources produce complicated sound fields that are composed of a radial and a tangential component. The tangential component, in contrast to the radial, is very small in the far-field. All frequency effects for the radial component are a factor of f (in Hz) times stronger in a dipole field than in monopole field.

In the far-field of a dipole source, distance effects are the same as for a monopole source, aside from the effect of the angle between the direction of oscillation of the dipole and the direction of observation (multiplication by a cosine function). The near-field of a dipole is very complicated. A more complete discussion of the fields of monopole and dipole sources can be found in Kalmijn (1988).

Kalmijn, A. J. (1988). Hydrodynamic and Acoustic Field Detection. In *Sensory Biology of Aquatic Animals* (ed. J. Atema, R. R. Fay, A. N. Popper and W. N. Tavolga), pp. 83-130. New York, Berlin, Heidelberg, London, Paris, Tokyo: Springer.

Sound Spectra

Basic Principles

Types of sounds
Sound is defined as mechanical oscillations perceivable for the human ear, generally from 16 to 16.000 Hz.
Sounds that are sine waves with fixed frequency and amplitude are perceived as pure tones. While sound waves are usually visualized as sine waves, sound waves can have arbitrary shapes and frequency content. In fact, most sounds and so the waves are composed of many sine waves of different frequencies, according to the principle of Fourier (see Signal Analysis and Fourier). Waveforms commonly used to approximate harmonic sounds in nature include saw-tooth waves, square waves and triangle waves. The frequency of these periodic sounds is called the fundamental frequency f_1 and the higher frequencies in the sound are the overtones or harmonics. A harmonic with frequency f_k is an integer multiple (k) of the frequency f_1, so $f_k = k f_1$. Table 1 presents a number of sounds, indicated by its range of composing frequencies or by f_1. Tones produced by a music instrument are generally periodic (not with drums), but speech is a-periodic.

Table 1

type of sound	frequency (Hz)
central c of piano (C^1) :	262
'high c' (c^3)	1046
range of concertgrand	27-4186
range of singing-voice (bass)	82-330
range of singing-voice (soprano)	262-1046
speech fundamental (man)	163
speech fundamental (woman)	262

Instruments can play a nearly (pure) tone, but mostly notes have many harmonics, accounting for the timbre. Noises (strictly speaking) are irregular and disordered vibrations including all possible frequencies. They are a-periodic, i.e. the wave shape does not repeat in time.

Spectra
With Fourier analysis (see Signal Analysis and Fourier) the spectrum of the signal can be calculated. Spectral analysis yields the amplitude spectrum (amplitude versus frequency) and the phase spectrum (phase versus frequency). They present the frequencies of which the signal is composed. With the amplitudes and phases of all harmonics the signal can be composed uniquely (Fourier synthesis). In the analysis of sounds, generally only the amplitude spectrum is calculated.
The emitted spectra are limited by the technology of the emitting apparatus. For instance, low frequencies (< 100 z) are hard to produce by loudspeakers, such that they are not contaminated by distorting higher harmonics. Another limitation is that the emitted spectrum is filtered by the medium in between generator and receiver. In air, high frequencies are strongly diminished (a train listened at a long distance produces a dull rumble). Finally, the receiver should by capable to sense all emitted frequency and, moreover, with the same sensitivity.
Fig. 1 presents two sounds as a function of time, together with their spectra, evoked by singing a Dutch aa (top) and a Dutch oo (bottom).

The periodicity of the signal is characteristic for vowels. The fundamentals of vowels are the same but the higher harmonics quite different. The fundamental can be adjusted by changing the tension of the vocal cords. This can be done within about two octaves (factor 2x2). The length of the vocal cords determines the lowest fundamental. The harmonics are determined by the mouth opening and by the shape of the mouth cavity, which acts as a resonator. Their spectrum is called the formant. It is specific for the vowel. Consonants are generated in a similar way but the contain more noise, produced by pressing air though a constriction. Their wave shapes are not periodic. This hampers a straightforward Fourier analysis. To get round this problem, the analysis is made in a short span of time and this time window glides over the signal, producing a spectrum at each time sample point. In this way a phonogram can be constructed with the frequency versus time. The amplitude of each frequency is depicted by a gray-scale. Fig. 2 presents the phonogram of two subjects, which pronounced the same Dutch sentence. Comparison of both panels shows that the speech-phonogram is highly subjects specific.

Applications

Speech analysis is widely used in clinical speech-training examinations, in the technology of hearing aids, cochlear implants etc. Such technologies will develop fast, since hearing-loss and disorders will become eudimic due to hair cell damage (too frequently too long exposure to too high sound levels).

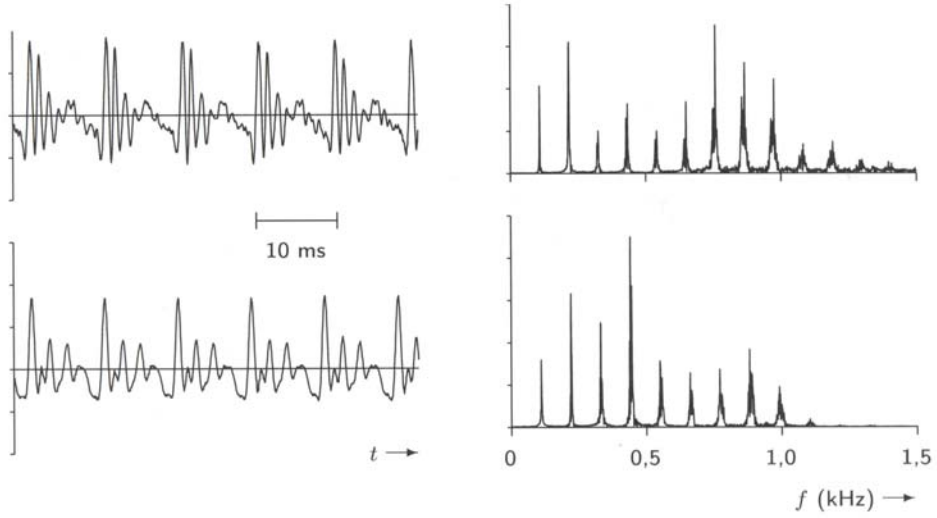


Fig. 1 Wave shapes (left) and amplitude spectra (right) of two vowels.

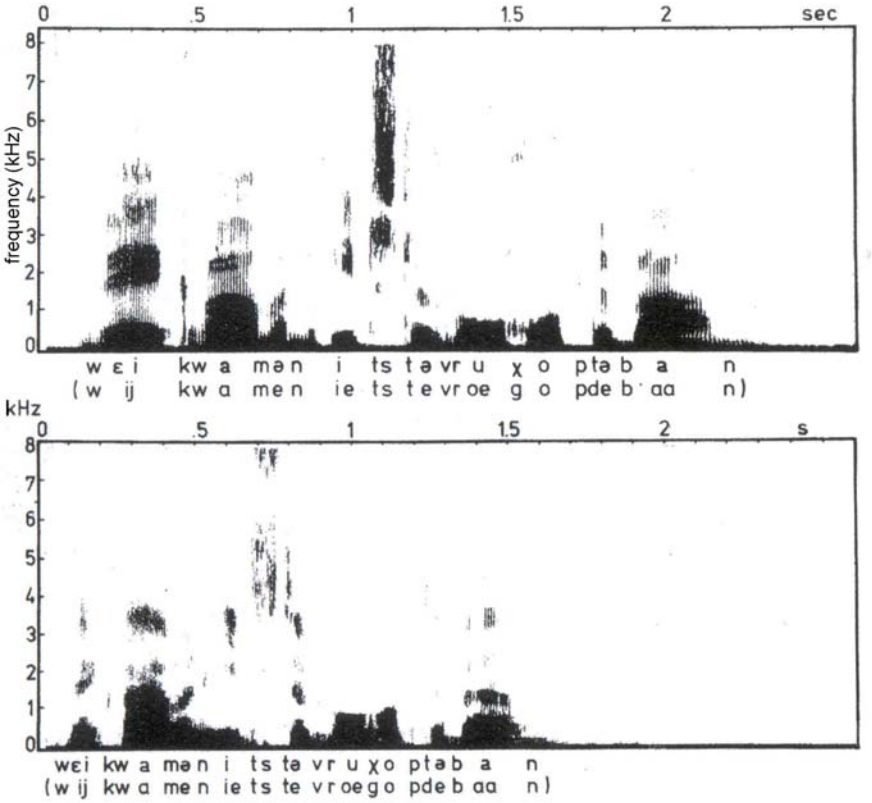


Fig. 2 Phonogram of two different subjects pronouncing the same sentence.

Other applications are artificial speech understanding/interpretation and producing artificial speech. Sound analysis is applied in many medical equipment and apparatus, e.g. in echography.

Stethoscope

Principle

The classical *acoustic stethoscope* operates on the transmission of sound from the two sided chest piece (a diaphragm and a bell) via air-filled hollow tubes to the listener's ears. If the diaphragm is placed on the skin, body sounds vibrate the diaphragm, creating acoustic pressure waves (see Sound and Acoustics) which travel up the tubing to the listener's ears. If the bell is placed on the patient, the vibrations of the skin directly produce acoustic pressure waves. The *bell* transmits *low* frequency sounds, while the *diaphragm* transmits *higher* frequency sounds. The problem with acoustic stethoscope is that the sound level is extremely low.
For children small size stethoscopes (also electronic) have been developed and another familiar stethoscope is the trumpet-shaped fetal stethoscope.

Electronic stethoscopes overcome the low sound levels by amplification. Unlike acoustic stethoscopes, which are all based on the same physics, transducers in electronic stethoscopes vary widely. Microphones and accelerometers behave poorly due to ambient noise interference. Piezoelectric crystals (see Piezoelectricity) and a capacitive sensor attached to the diaphragm are suitable. With the latter, the diaphragm responds to sound waves identically to a conventional acoustic stethoscope, with changes in an electric field replacing changes in air pressure. This preserves the sound of an acoustic stethoscope with the benefits of amplification.
A more recent development is the electronic phonendoscope, an endoscopic stethoscope.
Ambient noise filtering is nowadays available in electronic stethoscopes.
The analysis is generally a spectral analysis (Fast Fourier Transform) of the whole signal or of one of the sounds.

Application

Application comprises routine physical examination by physicians of many disciplines and specifically by cardiologists.
For the examination of the heart sounds, the electronic stethoscope is very suitable with digital analysis advancing. Most helpful for evaluation of the heart sounds (see Phonocardiography) is a frequency versus time plot with the amplitude (or power) depicted in a gray or color scale, just as done in sonograms (see Fig.2 of Sound Spectra).

Ultrasound

Principle

Ultrasound comprises frequencies > 20 kHz, the upper limit of human hearing. Above this limit the phase of spike firing in the axons of the acoustic nerve are not longer related to the phase of the impinging sound frequency. In other words, spike firing, if any, is random in time. Many mammals such as dogs can hear the most lower range of ultrasound. Echolocating bats and sea mammals (dolphins) have audiograms (see Hearing and Audiometry) up to 150 kHz and occasionally up to 200 KHz. Ultrasound, that goes up to ca. a GHz (Acoustic microscopes), has many industrial and medical applications. Typical ultrasound frequencies *f* are 1 to 10 MHz. A higher frequency results in a higher resolution of the image, just as holds for the frequency (color) of the light in light microscopy. In ophthalmology with its fine structures, 5-20 MHz is applied, and in obstetrics, cardiology and urology 2-5 MHz. In the same way as with Doppler applications, ultrasound can by emitted continuously or pulsed.

	f	half-distance
	(MHz)	mm
material		
water	1.0	14000
muscle	1.0	0.27
fat	0.8	0.69
brain	1.0	0.32
bone	0.6	0.095
bone	0.8	0.034
bone	1.2	0.021
bone	1.6	0.011
bone	1.8	0.008
bone	2.25	0.006
bone	3.5	0.0045

Table 1 half-distances for various tissues

Producing an ultrasound wave In medical echography (or ultrasonography) a sound wave is produced by creating continuous or short pulses of sound from an array of piezoelectric transducers (see Piezoelectricity). Wiring and transducers are encased in a probe. The electrical pulses vibrate the piezo crystals to create a series of sound pulses from each, which in turn produce together a single focused arc-shaped sound wave from the sum of all the individual emitted pulses. To make sure the sound is transmitted efficiently into the body (a form of impedance matching, see Acoustic impedance) between probe and skin a thin layer of a special gel is administrated. The sound wave, which is able to penetrate bodily fluids, but not (hardly) solids, bounces off the solid object and returns to the transducer. This rebound is an echo.
A completely different method is to produce ultrasound with a laser (ref. 1), see Optoacoustic imaging.
Receiving the echo's The piezo-transducer is in addition to sending also capable to receive the echo. Just in reverse, the echo-produced vibrations give rise to electrical pulses in the probe. These signals are analyzed in a PC and transformed into a digital image.

Dangers of ultrasound can basically not be excluded since ultrasound is energy and energy is always partly transformed to heat. Heat may damage tissue directly or may give rise to bubble formation (boiling). Bubbles may result in micro-traumata (as with decompression illness). Damage can be limited by reducing *f*. Since the tissue is in the far-field (see Sound and acoustics) of the source, which can be approximated by an acoustic dipole, the sound energy increases with f^3 (for a monopole the increase is proportional with *f*). Therefore, f is limited. This prevents tissue damage by heating. There is another reason to limit *f*: The absorption (this is heat dissipation) increases with f. In air, a tenfold rise in f increases the dissipation by a factor of about 25 and in water about a factor of 12. Consequently, the penetration depth strongly decreases with *f*. From Table 1 it can be seen that this is progressively with *f*. The maximal beam-intensity I_{max}, the mean beam intensity I_{mean} and the total administrated dose D all have their maximal allowable values. With constant I, $D= I_{max}t$, with t the application time. With pulsed ultrasound, pulse period t_{per} and pulse duration t_{pulse} the dose is $(t_{per}/t_{pulse})I_{max}t$. Fig. 1 shows the I_{max} versus t safety relation. The protocols of the application (Echography) are such that damage is actually impossible and (systematic) damage has not been reported in literature.

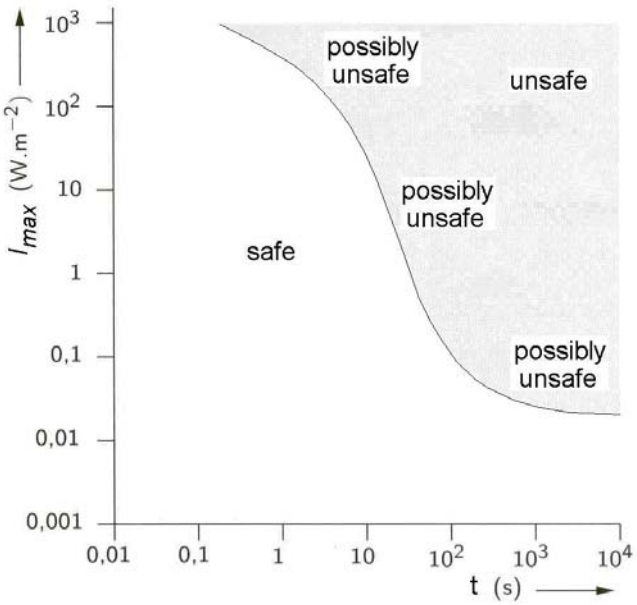


Fig. Principle of safety norms of ultrasound with power/area (=intensity/(area·time) on the vertical axis.

Application

Echography, also called (ultra)sonography, can visualize muscle and soft tissue, making them useful for scanning the organs, also of the unborn child. The use of microbubble contrast media to improve ultrasound signal backscatter is known as Contrast enhanced ultrasound (CEU) as e.g. applied in echocardiography and recent applications in molecular imaging and drug delivery.
Diagnostic echography scanners operate at 2-13MHz. More powerful ultrasound sources are used in non-imaging applications, such as to generate local heating in biological tissue (physical therapy). Focused ultrasound sources may be used to break up kidney stones or for cataract treatment, cancer treatment and cleaning teeth. The latter applications rely directly on the generated mechanical vibrations.

A well-know application is the ultrasonic cleaner (*f* is 20-40 kHz). In medicine (and other fields) they are used for cleaning of surgical and dental instruments, lenses and other optical parts. The cleaner generate millions of microbubbles in the liquid of the cleaner The main mechanism for the cleaning action is actually the energy released from the collapse of the cavitation of the microbubbles. Ultrasound when applied in specific configurations can produce exotic phenomena such as sonoluminesence (the emission of short bursts of light from imploding bubbles in a liquid when excited by ultrasound frequencies).

More Info

The safety index
In practice, for safety of echo-apparatus' the thermal index (TI) and the mechanical index (MI) is used. TI expresses the warming of the tissue, with different measures for soft tissue, bone and cranial bone. In general, the thermal index (TI) is defined by the relationship:

$$TI = W_p/W_{deg},$$

where W_p is the power of the impinging beam and W_{deg} is the estimated power necessary to raise the target tissue 1°C. As an absolute maximum TI <4 is considered for 15 min sessions.
MI expresses the risk of cavitational effects of bubbles. (Cavitation is the phenomenon where small cavities of partial vacuum form in fluid, then rapidly collapse, producing a sharp sound. Cavitation occurs in pumps, propellers etc.). MI is defined as the quotient of the peak negative pressure P^- and the square root of the ultrasound frequency, i.e.:

$$MI = P^-/\sqrt{f_0} \qquad \text{(in MPa MHz}^{-1/2}\text{)}.$$

MI<1.9 is safe. (4 MPa is also used. This is more conservative for echography in fine structures). When exposed to *high* MI (> 0.6, i.e. the MI often used for standard imaging) the bubbles oscillate wildly and

burst. Upon destruction (cavitation), micro-bubbles produce a brief, high amplitude signal, with good resolution. Caviation, in principle, can result in microtraumata, but in specific applications cavitation is pursued and without risk of damage, as in some versions of Contrast enhanced ultrasound . When the bubbles are filed with air, after bursting the air is rapidly dissolved in the liquid phase.
The safety limits holds also for the Doppler mode (see Echography).

Constructing the image
The computer must determine three things from each electrical impulse received: 1.) sorting out the impulse of the tens of transducers; 2.) determining the impulse amplitudes. 3.) determining the time interval between sending and receiving the impulses. Now the image in grey scale can be constructed. Transforming the electrical signal into a digital image can be best explained by using a blank column-row sheet (so a spreadsheet) as an analogy. The wire receiving the impulse determines the 'Column' and the time that it took to receive the impulse determines the 'Row', and the amplitude of the impulse determines the color that the spreadsheet cell should change too (white for a strong pulse, black for a weak pulse, and varying shades of gray for everything in between.)

Literature
1. http://e-collection.ethbib.ethz.ch/ecol-pool/diss/fulltext/eth15572.pdf

Hearing and Audiometry

Audiology

Principles

Audiometry is the testing of hearing ability. Typically, audiometric tests determine a subject's hearing levels, but may also measure the ability to discriminate between different sound intensities, recognize pitch, or distinguish speech from background noise. The tympanogram (see Tympanometry), the acoustic reflex or Stapedius reflex and Otoacoustic emissions may also be measured. Results of audiometric tests are used to diagnose hearing loss or diseases of the ear.

Sound pressure level and hearing level
Although pressure is measured in Pascals (Pa; see Sound and Acoustics), its amplitude is generally referred to as *s*ound *p*ressure *l*evel L_p and measured in dB, abbreviated as dB SPL. 1 dB SPL ≡ 20 μPa, the reference sound pressure p_0, which is the standard human threshold at 1000 Hz. L_p is defined as the logarithmic ratio of the energy. This is the same as the log-ratio of the squared sound pressures of the sound p_1 and the reference p_0. In formula:

$$L_p = 20\log(p_1 p_0) = 10\log((p_1^2/p_0^2)) \text{ dB SPL}$$

In underwateracoustics a reference level of 1 μPa is mostly used.
Table 1 presents values of some basic characteristics of various sounds.

Table 1 Typical sounds with their characteristics

sound	distance	sound level	pressure P_{eff}	particle velocity	displacement eardrum
	m	dB SPL	μPa	μm/s	μm
1000 Hz threshold		0	20	0.05	$0.1 \cdot 10^{-6}$
leave rustle	5	10	63	0.16	$0.3 \cdot 10^{-6}$
whispering	2	20	200	0.5	$1 \cdot 10^{-6}$
traffic	107	60	$20 \cdot 10^{3}$	50	$0.1 \cdot 10^{-3}$
normal speech	2	65	$36 \cdot 10^{3}$	89	$0.2 \cdot 10^{-3}$
pneumatic hammer	2	90	$0.63 \cdot 10^{6}$	$1.6 \cdot 10^{3}$	$3.1 \cdot 10^{-3}$
airplane	50	100	$2 \cdot 10^{6}$	$5 \cdot 10^{3}$	0.01
pain limit		120	$20 \cdot 10^{6}$	$50 \cdot 10^{3}$	0.1

For particle velocity (p = ρcv)see Sound and Acoustics

The ear drum als pressure-to-displacement transducer
Table 1 shows that at the 1000 Hz threshold particle velocity and eardrum displacement is extremely small, smaller than the diameter of a H_2O molecule. From here a lot of amplification is needed to reach a hearing sensation. The acoustic impedance (Z=ρc, with ρ the specific density of the medium and c the sound speed, see Acoustic impedance) of water, and so of the watery tissues in the cochlea, is about 4000 times that of water. Therefore, air born sound hardly penetrates water (loss ca. 70 dB). The mismatch in impedance is solved by the eardrum, which comes in mechanical vibration due to the sound pressure. In most conditions, the auditory system only utilizes the sound pressure and not particle displacement (as in bone-conduction) as input signal.
The bones in the middle ear amplify the eardrum displacements and transmit them to the oval window, which has a much smaller area (some 40 times). The oscillations of the oval window evoke a kind of displacement wave in the scala media.

Perceived sound strength: phon
Since the perceived loudness correlates roughly logarithmically to sound pressure, a new measure is introduced, the *phon*. This is a psychophysical measure, a unit of perceived loudness. At 1 kHz, 1 phon is 1 dB above the nominal threshold of hearing, the sound pressure level SPL of 20 μPa. By definition two sine waves that have equal phons are equally loud. An *equal-loudness contour*, also called a loudness level contour or a Fletcher-Munson curve, is a measure in dB SPL versus frequency for which a listener perceives a constant loudness. The loudness levels of each curve is expressed in phon.

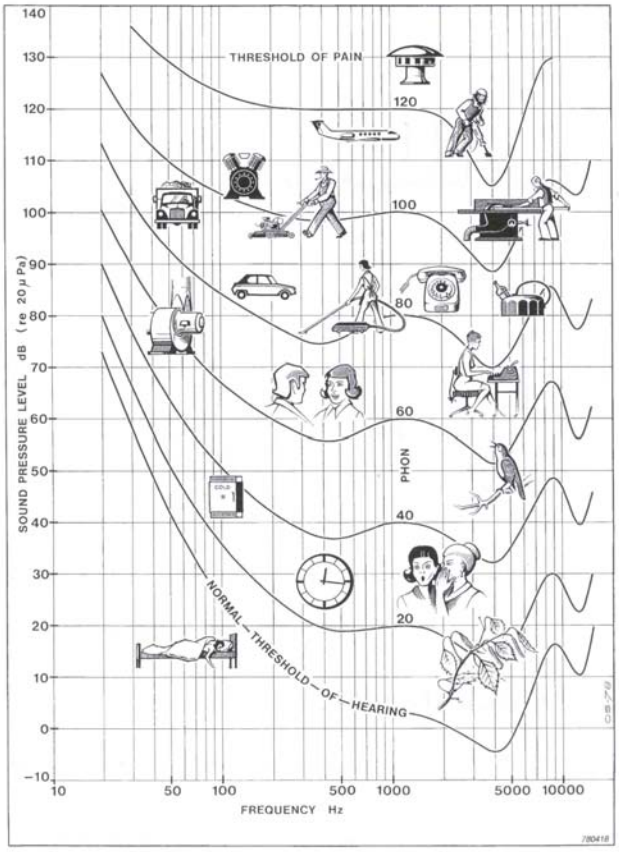


Fig. 1 Isophones of human hearing

Applications

In case of hearing loss the clinician likes to know where in the pathway from outer ear to cortex there is a disorder. To find this out there is a multitude of tests, psychoacoustic tests and measurements of the performance of the 'hardware' of the auditory system, from tympanum to cortex.

Psychophysical tests

An audiogram is a graphical representation of how well sound frequencies, i.e. pure tones, are perceive. It is a normalized threshold curve (the most lower curve of Fig. 1), converted from dBSPL to dBHL, with HL hearing level. Normal hearing is between -10dBHL and 15dBHL, with 0dB from 250Hz to 8kHz the 'average' normal hearing. Audiograms are generally measured with headphones. This gives a slightly different results as compared in the free field (performed in an anachroic, i.e. an echo-free room). Sometimes the audiogram is not recorded with pure tones but with small band noise (e.g. 1/3 octave noise) or with a pure tone masked by background noise.

In addition, all kind of audiological tests can by performed, for instance focused on just audible frequency or intensity differences etc, time resolution (gap detection), whether or not with a noise mask (presented just before the probe signal, during the probe signal or after it (backward masking), amplitude and frequency modulation, speech perception (repeating one-syllable words by the patient), directional hearing, also with masking by noise (cocktail party effect) etc. This test may have a more experimental rather than clinical character. A bone conduction hearing test is administered to figure out whether the hearing loss is conductive (caused in the outer or middle ear) or sensorineural (cochlear). In this test, a vibrator is placed behind the ear or on the forehead. The vibrator emits frequencies, and the person being tested hears tones or beeps just like in the test with earphones.

A routine audiological examination comprises a pure tone audiogram, often a audiogram with a noise mask and tests for speech perception.

Measurements of the 'hardware' of the auditory system

A tympanogram (see Tympanometry) is performed to examine the function of the tympanum and the middle ear. The acoustic reflex test (Stapedius Reflex) measures the reflexive contraction of the stapedius muscle, which is important in protecting the ear from loud noises.

Otoacoustic Emission

Basic principle

An otoacoustic emission (OAE), first discovered by David Kemp in 1979, is a sound which is generated from within the inner ear by a number of different cellular mechanisms. OAEs disappears with inner ear damaged, so OAEs are often used as a measure of inner ear health. There are two types of otoacoustic emissions: Spontaneous Otoacoustic Emissions (SOAEs) and Evoked Otoacoustic Emissions (EOAEs). OAE is considered to be related to the amplification function of the cochlea. OAEs occur when the cochlear amplifier is too strong. Possibly outer hair cells enhance cochlear sensitivity and frequency selectivity and provide the energy. The outer hair cells have few afferent fibers but receive extensive efferent innervation, whose activation decreases cochlear sensitivity and frequency discrimination.

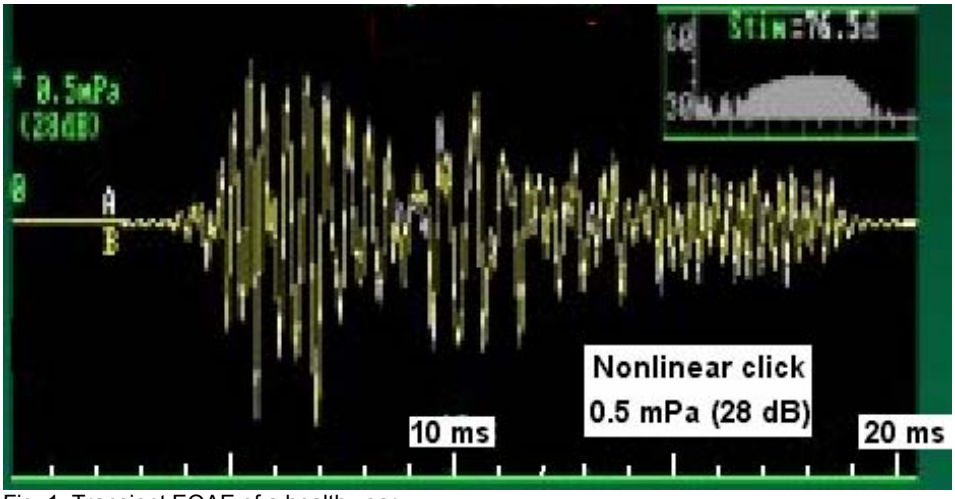


Fig. 1 Transient EOAE of a healthy ear

Applications

Middle and inner ear disorders, and retrocochlear disorders, tinnitus.

More Info

EOAEs are currently evoked using two different methodologies. Transient EOAEs (TEOAE or TrEOAE) are evoked using a click stimulus that repeat at about 20 instances per second. 98% of the healthy ears have TEOAEs (0.5-4 kHz), but above 60 year only 35%. The evoked response from this type of stimulus covers the frequency range up to around 4 kHz. Transient potentials were the original methodology of examination used by Kemp. Distortion Product OAEs (DPOAE) are evoked using a pair of tones with particular intensity (usually either 65 - 55 dB or 65 for both) and frequency ratio (F2/F1). The evoked response from these stimuli occurs at a third frequency. This distortion product frequency is calculated based on the original F1 and F2. Inner ear damage diminishes the distortion product.

Last decade, OAEs became increasingly important in the clinical audiological practice. It is still uncertain if there is an interaction, e.g. at the level of the olivocochlear bundle, between the ears regarding OAEs. In general, contralateral stimulation does not provoke EOAEs. The frequency resolution, e.g. for speech, depends on very fast modulation of the incoming signal. Due to the neural distance, this modulation would lag behind if otoacoustic emissions in one ear would effect the opposite one.

Literature

"http://en.wikipedia.org/wiki/Otoacoustic_emission"
Probst R, Lonsbury-Martin BL, Martin GK. A review of otoacoustic emissions. J Acoust Soc Am. 1991, 89: 2027-67.

Stapedius reflex

Basic principle

While tympanometry measures the change of the compliance caused by changing pressure in the outer auditory canal, an acoustic reflex, or contraction of the stapedial and tensor tympani muscles, occurs under normal conditions when a loud acoustic stimulus (e.g. loud speech) is presented to the auditory system. The stapedius pulls the stapes (stirrup) of the middle ear away from the oval window and the tensor tympani muscle pulls the malleus (hammer) away from ear drum. This contractions cause a stiffening of the ossicular chain, which decreases the compliance (volume change over pressure) of the middle ear system. As in tympanometry, a probe tone is used to measure this change in compliance. Regardless of whether the acoustic stimulus is active in the left, in the right, or in both ears, the Stapedius reflex is always binaural, i.e. it occurs in both ears at the same time. The examination can be performed ipsilateral (probe and sensor at the same ear) or contralateral (different ears; Fig. 1).

Fig. 1 Ipsilateral and contralateral measurement of the Stapedius reflex.

The reflex measurement is expressed as the air pressure value where the compliance peak occurred during the tympanometric test. Stimulus tones of varying intensities at 500, 1000, 2000 or 4000 Hz are presented as short bursts. If the applied stimulus causes a reflex, the impedance measuring instrument registers a decrease in compliance in the “probe ear” which indicates a Stapedius reflex. If a change greater than say 0.05 ml in the pressure/volume diagram (see Fig. 2 of Tympanometry) is detected a reflex is considered present for the applied test level and frequency. Because this means an extremely small compliance change, any movement of the probe during the test may produce an artifact (false response). The smallest test level (in dBHL, dB hearing level) evoking the reflex is the reflex threshold. The Stapedius reflex is activated in normal-hearing adults with sound pressure levels between 70 and 105 dBHL.

Application

The reflex is measured when tympanometry is also done. The acoustic reflex is affected by middle ear status as well as the amount of hearing loss.
Middle ear pressure should be equivalent to ambient air pressure (0 daPa (decaPa) difference). Minor shifts of the peak compliance to the negative may occur when the patient is congested and they are rarely to the positive side. Negative pressure greater than -150 daPa generally indicates further examination. A perforation in the tympanic membrane will cause a high ear canal volume measurement because the instrument then measures the volume of the entire middle ear space included.
An extremely soft tympanic membrane or an ossicular chain discontinuity will yield a very high peak compliance in the presence of normal middle ear pressure. Ear canal volume will be normal and the reflex may be absent.
A fixation of the ossicular chain, as in otosclerosis, will produce a tympanogram with very low compliance in the presence of normal middle ear air pressure. A resolving case or beginning case may produce a reduced peak in the presence of severe negative middle ear pressure. The ear canal volume is normal and the reflex is either absent or at an elevated level.
Eustachian tube dysfunction in the absence of fluid will show a normal compliance curve, but it will be displayed to the negative side of the tympanogram. Ear canal volume will be normal and the reflex may be present, depending on the degree of involvement.

Literature
S.B. Smyth www.maico-diagnostics.com/eprise/ main/Maico/Products/Files/MI24/Tymp.Guide.pdf -

Tympanometry

Basic principle

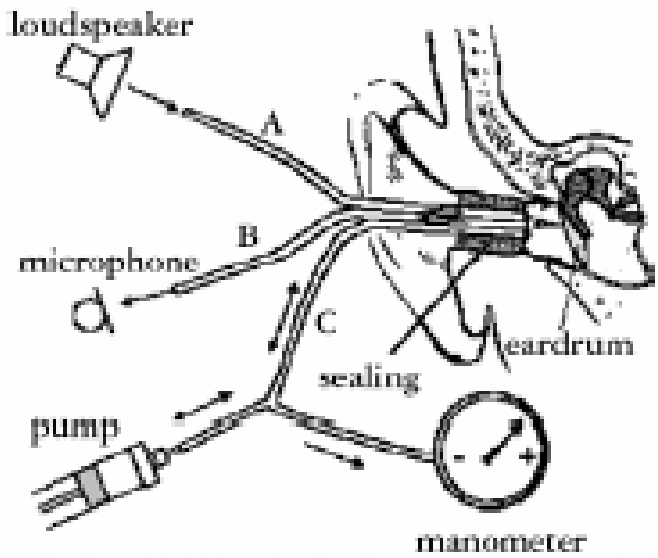


Figure 1
Principle of the impedance measurement

Fig. 1

Tympanometry is a measurement of the acoustic impedance of the middle ear. If the eardrum is hit by a sound, part of the sound is absorbed and sent via the middle ear to the inner ear while the other part of the sound is reflected (see Utrasound). The stiffer the eardrum, the less sound reaches the inner ear. With fluid in the middle ear, the drum behaves very stiff (remind the rebound of a tennis ball against a stone wall and again the net.)
Inside the probe of the tympanometer are three small tubes. One contains a small loudspeaker which emits a low frequency sound (Fig. 1: A). Another tube (B) is connected to a microphone. The third tube (C) contains a manometer and a pump which produce positive ad negative pressure. The probe, covered by a soft tip, is inserted airtight nearby the eardrum. Pressure is rapidly swept from positive to negative (50-200 daPa/s (deca-Pascals)). The highest compliance (the lowest impedance) is normally reached with an inner ear air pressure corresponding to the outside pressure. When performing tympanometry, the pressure in the canal varied continuously from +400 daPa to –200 daPa. Impedance is lowest (maximal compliance) when pressure in the canal equals pressure within the middle ear. Fig. 1 shows a tympanogram. Maximal compliance is at the peak and measured as an equivalent volume of air (in mL). The sharpness of the peak is expressed as its width at half of peak and measured in daPa. The highest compliance is reached with normal pressure. With positive and negative pressure the eardrum stiffens and the compliance decreases. The position of the peak on the horizontal axis and on the vertical axis provides diagnostic information regarding the function of the middle ear system.

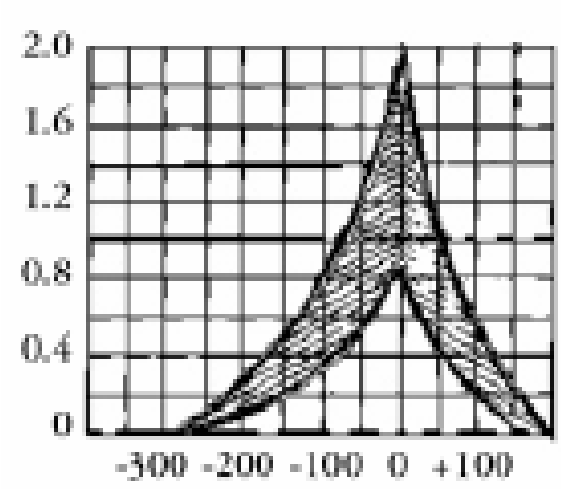


Fig. 2 Tympanogram with normal curve area hatched. Along the horizontal axis is pressure (daPa) and along the vertical axis the equivalent volume (mL).

Application

Tympanometry is the most utilized clinical test in audiology (see Hearing and audiometry). The test requires a clean ear canal. There are three types of tympanograms:
1. Type A. The peak compliance occurs at or near atmospheric pressure indicating normal pressure within the middle ear.
2. Type B. No sharp peak, little or no variation in impedance over a wide sweep range, usually secondary to non-compressible fluid within the middle ear space (otitis media, middle ear squeeze, or tympanic membrane perforation, the latter two often due to diving).
3. Type C. Peak compliance is significantly below zero, indicating negative pressure (sub-atmospheric) within the middle ear space. This finding is suggestive of Eustachian tube dysfunction or middle ear fluid.
The information derived from the tympanogram provides the physician with the additional information regarding the patient's middle ear function (to document or rule out the presence of otitis media, tympanic membrane perforation or Eustachian tube dysfunction).
A compliance peak within the range 0.2-2.0 ml (children and adults) indicates normal mobility of the middle ear system.

More Info

More correctly, the compliance (the reciprocal of elastance, see Lung gas transport 2, pressure, volume and flow) is the slope of the tympanogram, e.g. measured in mL/daPa, which is maximal near the peak (see Fig. 2). However, in audiological practice, for convenience the difference referred to the healthy tympanogram is expressed as height of the peak in mL.

Literature

S.B. Smyth www.maico-diagnostics.com/eprise/ main/Maico/Products/Files/MI24/Tymp.Guide.pdf

Vestibular mechanics

Basic Principle

The vestibular system, or *balance system*, is the sensory system that provides the dominant input about our movement and orientation in space. It is situated in the *vestibulum* in the inner ear (Fig. 1). As our movements consist of rotations and translations, the vestibular system comprises two subsystems: the semicircular canals, which indicate rotational movements; and the *otoliths*, which are sensitive for linear accelerations in order to perceive changes of linear motion of the head or body. The vestibular system sends signals primarily to the neural structures that control the eye movements, and to the muscles that keep us upright. The projections to the former provide the anatomical basis of the vestibulo-ocular reflex (VOR), which is required for clear vision; and the projections to the muscles that control our posture are necessary to keep us upright.

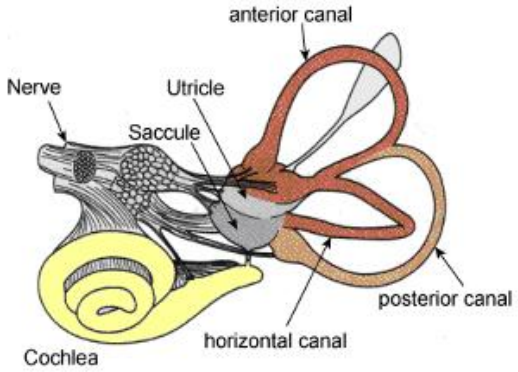


Fig. 1 Human labyrinth, from the left ear. The light blue pouch (upper right) is the endolymphatic sac, and contains only fluid.

Semicircular canals

Since we perceive the environment within three spatial dimensions, accordingly, our vestibular system contains three semicircular canals in each labyrinth. They are approximately orthogonal to each other, and are called horizontal (or lateral), anterior (or superior), and posterior (or inferior) canal.

The canals are arranged in such a way that each canal on the left side has an almost parallel counterpart on the right side. Each of these three pairs works in a push-pull fashion: when one canal is stimulated, its corresponding partner on the other side is inhibited, and vice versa.

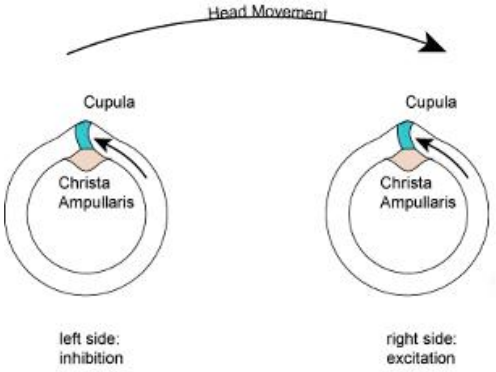


Fig. 2: Push-pull system of the semicircular canals, for a horizontal head movement to the right.

This push-pull system allows us to sense all directions of rotation: while the right horizontal canal gets stimulated during head rotations to the right (Fig 2), the left horizontal canal gets stimulated (and thus predominantly signals) by head rotations to the left.

Otoliths

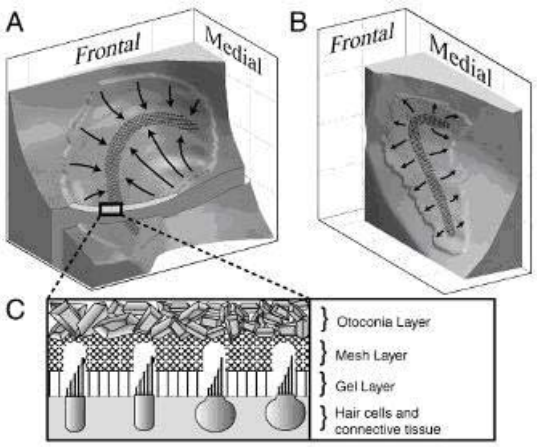


Fig. 3 Otoliths, left side. A) the *utricle*, and B) the *saccule*. C) Cross-section through the utricle: the *Mesh Layer* is fairly stiff, while the underlying *Gel Layer* is more viscous. When the *Hair cells* are bent in the directions indicated by the arrows in A) and B) they get excited, while a deflection in the opposite direction inhibits them. The upper part of the otoconia layer is embedded in the endolymph

While the semicircular canals respond to rotations, the otoliths sense linear accelerations. We have two on each side, one called *Utricle*, the other *Saccule*. Fig. 3C shows a cross section through an otolith: the otoconia crystals in the *Otoconia Layer* (Fig. 3, top layer) rest on a viscous gel layer, and are heavier than their surroundings. Therefore, during linear acceleration of the head, they get displaced. In this way the hair cells are deflected (Fig. 3, bottom layer) and thus produces a sensory signal. Most of the utricular signals elicit eye movements, while the majority of the saccular signals projects to muscles that control our posture. While the interpretation of the rotation signals from the semicircular canals is straightforward, the interpretation of otolith signals is more difficult: since gravity is equivalent to a constant linear acceleration, we somehow have to distinguish otolith signals that are caused by linear movements from such that are caused by gravity. We can do that quite well, but the neural mechanisms underlying this separation are not yet fully understood.

For a more extensive basic explanation see Dickmans website Vestibular Primer, for the more specialized physiology Highstein et al. (2004).

Application

Vestibulo-ocular reflex (VOR)
The vestibular system needs to be fast: for clear vision, head movements need to be compensated almost immediately. Otherwise our vision corresponds to a photograph taken with a shaky hand. To achieve clear vision, signals from the semicircular canals are sent as directly as possible to the eye muscles. This direct connection involves only three neurons, and is correspondingly called *Three-neuron-arc* (Fig. 4). Using these direct connections, eye movements lag the head movements by less than 10 ms, one of the fastest reflexes in the human body. The automatic generation of eye movements from movements of the head is the *VOR*.

This reflex, combined with the push-pull principle described above, forms the physiological basis of the *Rapid head impulse test* or *Halmagyi-Curthoys-test*: when the function of the right balance system reduced by a disease or by an accident, *quick* head movements to the right cannot be sensed properly any more. As a consequence, no compensatory eye movements are generated, and the patient cannot fixate a point in space during this rapid head movement.

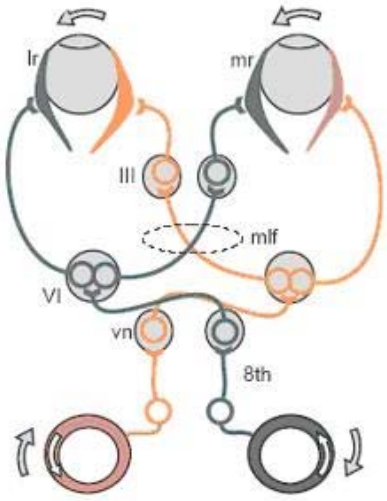


Fig. 4 Three-neuron arc, during a head movement to the right. *8th* facial nerve, from the peripheral vestibular sensors to *vn*, the vestibular nuclei in the brainstem. *VI* abducens nucleus. The medial lateral fascicle (*mlf*) projects from the abducens nucleus to *III*, the oculomotor nucleus. The left lateral rectus muscle *lr* and the right medial rectus muscle *mr* get contracted, turning the eyes to the left. The blue objects are excited, the red ones inhibited. (From SensesWeb, Vilis, an educational website concerning al sensory systems).

Pathologies

Diseases of the vestibular system can take different forms, and usually induce vertigo and instability, often accompanied by nausea. In addition, the function of the vestibular system can be affected by tumors on the cochleo-vestibular nerve (vestibular schwannoma), an infarct in the brain stem or in cortical regions related to the processing of vestibular signals, and cerebellar atrophy. Less severe, is vertigo caused by the intake of large amounts of alcohol.

Benign Paroxysmal Positional Vertigo (BPPV) is probably caused by pieces that have broken off from the otoliths, and have slipped into one of the semicircular canals. In most cases it is the posterior canal that is affected. In certain head positions, these particles push on the cupula of the canal affected, which leads to dizziness and vertigo. This problem occurs rather frequently, often after hits to the head or after long bed rest. BPPV manifest itself by vertigo attacks which repeatable appear when the head is brought into a specific orientation. In most cases BPPV can be eliminated (for the patient in an almost miraculous way) by lying down, bringing the head in the right orientation, and sitting up quickly.

See for a more extensive description of clinical applications e.g. Brandt (2003).

More Info

Mechanics of the semicircular canals

The mechanics of the semicircular canals can be described by the Navier-Stokes equations. By applying its solution (Grant, 1995), the transfer function (see System Analysis) of the semicircular canals can be found.

If we designate the deflection of the cupula with θ, and the head rotational velocity with ω, the cupula deflection in Laplace notation is approximately:

$$\frac{\theta}{\omega}(s) = \frac{\alpha s}{(s+1/\tau_1)\cdot(s+1/\tau_2)}$$

with α a proportionality factor, and *s* corresponding to the frequency (s=iω). It appears to be a band-pass filter with a first order high frequency cut off and a first order low frequency cut off. For humans, the time constants τ_1 and τ_2 are approximately 3 ms and 21 s, respectively, or in the frequency domain 41 Hz (a rotation with 93^0/ms) and 0.0077 Hz (17^0/s).

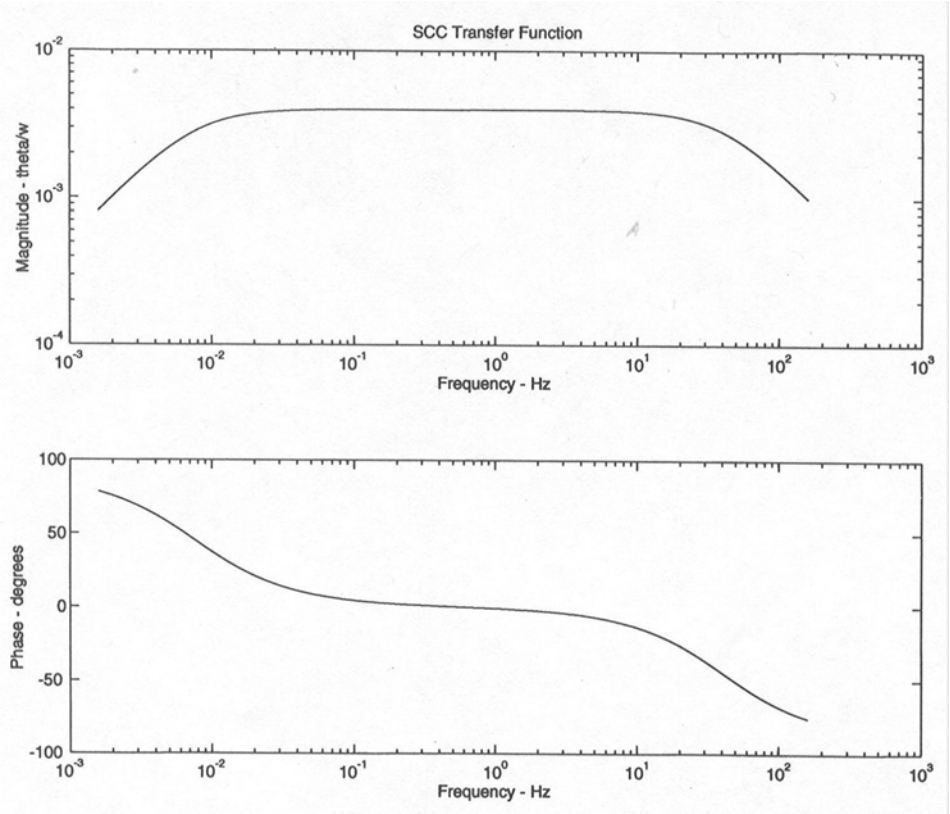


Fig. 5 Frequency response of semicircular canals for the transfer function of mean angular displacement of endolymph fluid θ as a function of angular head velocity ш(from Grant, 1995).

So, in the frequency domain it has a span of some 3.5 decades. As a result, for typical rotational head movements, which cover the frequency range of 0.1 Hz and 10 Hz, the deflection of the cupula is approximately proportional to the head-rotational velocity. This is very useful, since the velocity of the eyes must be opposite to the velocity of the head in order to have clear vision.

Mechanics of the otoliths

When the acceleration of the otoliths is plotted versus frequency, than they behave as an overdamped high-pass second order system. Using the acceleration and not velocity or displacement is obvious, since the otolith system behaves as an accelerometer. This means that any accelaration with a high frequency content (see Fourier analysis) is well perceived whereas very slow accelarations are not perceived, as can easily be confirmed from daily live experience.
The mechanics is dominated by the elasticity and viscosity of the gel(and mesh) layer. For the derivation of the transfer function, again the Navier-Stokes equations are basic, and in addition the motion equations are of importance. The derivations, which are more complicated than for the semicircular canals can be found in Grant (1995). For a moderate elasticity and a rather high viscosity ratio (gel over endolymph), the lower high cut off frequency is at ca. 3 Hz and the upper one 3000 Hz. Higher elasticity results in less damping (with closer cut off frequencies) and the same holds for a lower viscosity ratio.

References

Dickman D. Vestibular Primer *http://vestibular.wustl.edu/vestibular.html*.
Grant W., Vestibular mechanics. In: The biomedical engineering handbook, Bronzino (ed), CRC Press, Boca Raton, pp 517-527, 1995.
Highstein S.M., Fay R.R. and Popper A.N. (eds), The vestibular system, Springer-Verlag, Berlin, 2004.
Brandt, T. Vertigo : its multisensory syndromes, Springer-Verlag, Berlin, 2003.
Vilis T. SensesWeb *http://www.med.uwo.ca/physiology/courses/sensesweb/*.

Electricity and Bioelectricity

ECG: augmented limb leads

Principle

Unipolar leads and augmented limb leads
Basically a unipolar lead is an exploring electrode placed on a chosen site linked with an indifferent or reference electrode with a very small potential.
To obtain a common reference with 'zero potential', two limb electrodes and an electrode placed at the sternum (called central terminal, CT, also Wilson's central terminal) are connected through 5 kΩ resistances to form the indifferent electrode, connected to the negative terminal of the ECG machine. The potential of the active electrode, connected to one the three limbs, creates one of the leads aVR, aVL, or aVF. The other two electrodes of the standard leads are connected with CT. These are the *augmented limb leads*. So, the only difference with the bipolar leads I, II, and III is the choice of the common reference. As with the standard leads, the right leg is never connected for a lead, but it can be used to attach an earth electrode (see EEG: 12-lead ECG).

The augmented limb leads can be presented as vectors with a certain angle in some plane and time varying amplitude.

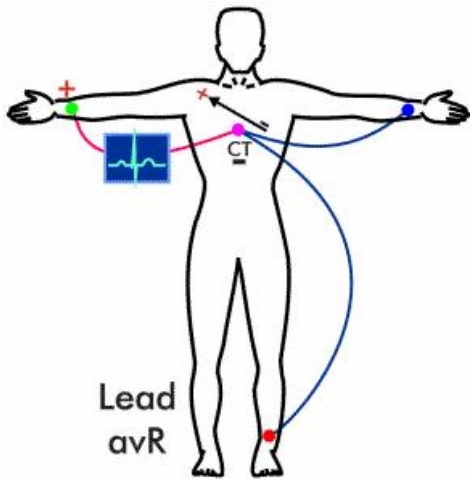


Fig. 1 Recording of aVR. Since the vector is the source, the current which creates the voltage, flows in the body from negative to positive (as in a battery). The vector point to -150 degrees (up are negative angles, just as in the hexaxial reference system (see ECG: Hexaxial reference system).

The three augmented leads are:
Lead aVR or "augmented vector right" has the positive electrode (green in Fig. 1) on the right arm.
Lead aVL or "augmented vector left" has the positive (blue) electrode on the left arm.
Lead aVF or "augmented vector foot" has the positive (red) electrode on the left leg.

The augmented limb leads aVR, aVL, and aVF are amplified (augmented) compared to the signal when a unipolar limb lead is made with CT as the reference electrode. Such a lead is rather small to be useful. Together with leads I, II, and III, the augmented limb leads aVR, aVL, and aVF form the basis of the hexaxial reference system, which is used to calculate the heart's electrical axis in the *frontal plane*. See for an example of all six leads of a healthy subject Fig. 2 of EEG: 12-lead ECG.

Application

This three leads are frequently used in the clinic, for instance for intensive care and for routine non-cardial surgery, but then V5 is used in addition.

More info

When the limb electrodes are connected one by one with CT the leads VR (right arm), VL (left arm) and VF (left leg) are obtained. aVR is equal to:

$$aVR = VR - (VL + VF)/2 \quad (1)$$

since (left arm) and VF (left leg) are connected, so the mean of both signals results. CT is the common reference and therefore it is not in the equation. Eithoven's triangle says that VR + VL + VF = 0 or VR = – (VL + VF). Substitution in (1) yields that aVR = 3VR/2. For the other two, aVL and aVF the same holds. So, recording the unipolar leads VR, VL and VF is a disadvantage. Therefore, the leads I, II and III are recorded together with the augmented leads.

ECG: basic electrocardiography

Principle

An electrocardiogram or ECG is a graphic representation of the electrical activity of the heart over time produced by an electrocardiograph. Understanding the various waves and normal vectors of depolarization and repolarization yields important diagnostic information.

Calibration
A typical electrocardiograph (or PC monitor) runs mostly at a paper speed of 25 mm/s. With a paper ECG, the finest division is a block of 1 mm^2 and 25 mm/s represents 1 mm/40 ms. A diagnostic quality 12-lead ECG is calibrated at 10 mm/mV.

Filter selection
Modern ECG monitors offer multiple filters for signal processing with a *monitor mode* and *diagnostic mode*. In monitor mode, the low frequency filter (also called the *high-pass filter*: signals above the threshold or cut-off frequency are allowed to pass) is set at either 0.5 Hz or 1 Hz. The high frequency filter (the *low-pass filter*) is set at 40 Hz (see Linear filters). The *high-pass filter* reduces wandering of the baseline and the low pass filter reduces high frequency noise and 50 or 60 Hz power line hum. (Hum can also be suppressed by a selective band pass filter, a T-filter). In diagnostic mode, the high pass filter is set at 0.05 Hz, allowing accurate ST segments. The high low pass filter is set to 40, 100, or 150 Hz. Consequently, the monitor mode ECG display is more filtered than diagnostic mode.

Waves and intervals of the ECG

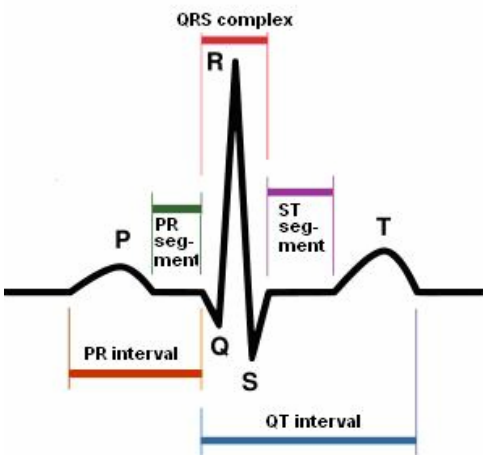


Fig. 1 Schematic representation of normal ECG

The baseline voltage of the electrocardiogram is known as the *isoelectric line*. A typical ECG tracing of a normal heartbeat (or cardiac cycle) consists of a P wave, a QRS complex and a T wave. A small *U wave* is normally visible in 50 to 75% of ECGs. For a detailed description the reader is referred to the textbooks of clinical physiology.

Working principle of electrodes

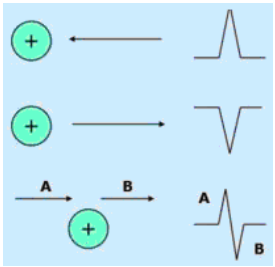


Fig. 2 Relationship between positive electrodes (green circles) and the propagating depolarization wavefronts (at the right).

An ECG is obtained by measuring electrical potentials between various points of the body using a biomedical instrumentation amplifier. A *lead*, records the electrical signals of the heart from a particular combination of recording electrodes which are placed at specific points on the body.
When a depolarization wavefront (or electrical vector) moves toward and away a positive electrode, it creates a *positive* and *negative* deflection in the corresponding lead respectively as depicted in Fig. 1.
When a depolarization wavefront (or electrical vector) moves perpendicular to a positive electrode, it creates an *equiphasic* (or *isoelectric*) complex. It will be positive as the depolarization wavefront (or mean electrical vector) approaches (A), and then become negative as it passes by (B).

Standard ECG leads

Fig. 3 Lead II

The basic three bipolar limb leads Leads I, II (Fig. 3) and III are the so-called bipolar *limb leads* because electrodes are attached to the arms and legs forming the *Einthoven's triangle*. All three electrodes are 'active' and there is no reference electrode. Therefore the leads are bipolar. Einthoven, who discovered the ECG, placed legs and arms in buckets of salt water (Fig. 4). They are the first three leads of the modern 12 lead-ECG (see ECG: 12-lead ECG).

Fig. 4 ECG as done by Willem Einthoven based on electromechanical technology with a wire galvanometer as sensor.

The bipolar (standard) leads
The electrodes are attached as follows:
• *lead I* = left arm positive minus right arm negative (LA–RA)
• *lead II* = left leg positive minus right arm negative (LL–RA).
• *lead III* = left leg positive minus left arm negative (LL–LA).

Application

- It is the gold standard for the evaluation of cardiac arrhythmias
- It guides therapy and risk stratification for patients with suspected acute myocardial infarction.
- It helps detect electrolyte disturbances (e.g. hyperkalemia and hypokalemia)
- It allows for the detection of conduction abnormalities (e.g. right and left bundle branch block)
- It is used as a screening tool for ischemic heart disease during a cardiac stress test
- It is occasionally helpful with non-cardiac diseases (e.g. pulmonary embolism or hypothermia)

However, the electrocardiogram does not assess the contractility of the heart.

More info

Lead II should be equal to the sum of leads I and III, so I + III = II. This is called *Einthoven's Law*. It is written this way (instead of I + II + III = 0) because Einthoven reversed the polarity of lead II in Einthoven's triangle. Then, QRS complexes are upright.

The position from which the heart is viewed by each of these leads is shown in Figure 5.

ECG: body surface mapping

Principle

Body surface mapping of the electric activity of the heart refers to the use of many recording sites (>64) arranged on the body so that isopotential surfaces could be computed and analyzed over time. Extensive computer algorithms, applied in experimental cardiologic research, calculate these isopotential surfaces.

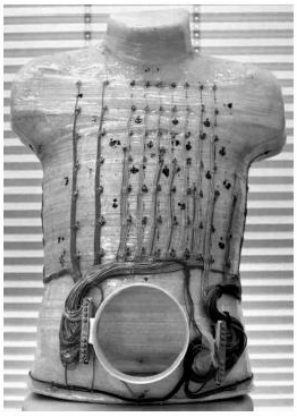


Fig. 1 Front view of a torso phantom with 138 electrodes (63 on the left anterior chest wall).

The main objective of body surface potential maps (BSPMs) is to calculate with complicated mathematics the underlying electric sources approximated by equivalent current dipoles (ECDs). There are many solutions of this calculation, in terms of number of dipoles, their time-varying positions, directions and magnitudes. This problem is called the *backward problem*. In contrast, the forward problem (calculating the potential distribution on the surface of the torso) has an unique solution. The theory of the forward and backward problem holds for any electrical source in the body and is, in addition to the heart, most advanced developed for brain activity (see Magnetoenceplalography (MEG)). The backward problem is piece wise solved. For instance, by focusing to a specific temporal window and restricting the area of location, one of the dipoles can be found. The solution can suppose electric isotropy of the heart and the surrounding torso tissues. A refinement is reached by supposing anisotropic heart models (Fig. 2). The myocardial anisotropy can be determined by a heterogeneous 3D matrix from segmented magnetic resonance images.
The most relevant dipoles are located in the septum, apical area, left ventricular wall or right ventricular wall. The solution of the apical source is most dependent on the anisotropy of the torso tissues.

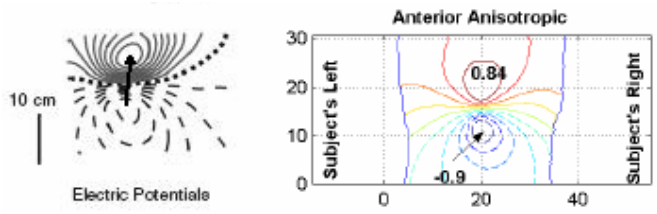


Fig. 2 Left Isocontour lines (0.5 mV resolution). Solid lines positivity, dashed lines negativity, zero line is dotted. The arrow indicates the ECD calculated with a BEM model. Right Contour plot of the BSPM in frontal view for a *z*-oriented (vertical) apical ECD (FEM). Potentials in mV and distances in cm.

More Info

The current and potential distributions in the torso can be computed using ECDs decomposed in the orthogonal x, y, z direction, or with a 3D algorithm. Solution can be based on the *finite element method* (FEM) or *boundary element method* (BEM). FEM is a volume-discretisation method of numeric calculus applied to a set of equations in a 3D-matrix (the elements). BEM is a numerical method of solving linear equations formulated in the *boundary integral* form. Conceptually, it works by constructing a "mesh" over the surface (with a small surface/volume ratio to be effective). The mesh is often a set non-uniform triangles (*triangular segmentation*), depending on the local curvature of the surface.
Often BSPM is combined with magnetocardiography. Together they give better solutions of the ECDs.

ECG: hexaxial reference system

Principle

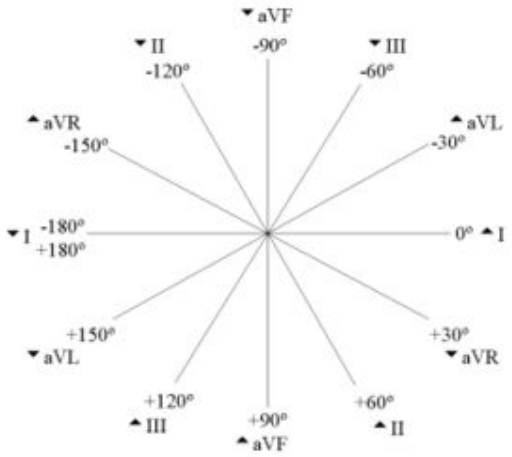


Fig. 1 Hexaxial reference system. Notice that negative is up. Triangle up denotes that with normal polarity the main peak of the QRS complex is positive. This holds for all 6 leads from 0 ^{o} to 150^{o} except AVR.

The hexaxial reference system is a diagram that is used to determine the heart's electrical axis in the frontal plane. The heart's *electrical axis* refers to the general direction of the heart's depolarization wavefront (or *mean electrical vector*) by using the polarity of the QRS complex in leads I, II, and III in the frontal plane. It is usually oriented in a right shoulder to left leg direction, which corresponds to the right inferior quadrant of the hexaxial reference system, although a slightly broader range, -30^{o} to +90^{o} is considered to be normal.

The diagram is based on the first six leads (I, II, III, aVR, aVL, and aVF) of the 12-lead ECG (see ECG: 12-lead ECG). To determine the heart's electrical axis, first the most isoelectric (or equiphasic) lead should be located on a diagnostic quality ECG with proper lead placement. Then the corresponding spoke on the hexaxial reference system should be found. The perpendicular spoke will point to the heart's (principal) electrical axis. To determine which of the two opposite located numerical value (in degrees) should be used, use the polarity of the perpendicular lead on the ECG. For example, if the most isoelectric lead is aVL, the perpendicular lead, being lead II has an angle of +60^{o}.

Application

The reference system is used for diagnosis of cordial disorders.
Normal and deviating directions are classified as:

- Normal axis: -30^{o} to +90^{o}
- Left axis deviation: -30^{o} to -90^{o}, may indicate left anterior fascicular block or Q waves from inferior myocardial infraction.
- Right axis deviation: +90^{o} to +180^{o} may indicate left posterior fascicular block, Q waves from high lateral myocardial infarction, or a right ventricular strain pattern.
- Extreme axis deviation: -90^{o} to -180^{o}

ECG: vectorcardiography

Principle

Vectorcardiography is the registration, by formation of a loop display on a PC, of direction and size (vector) of the moment-to-moment electric activity of the heart during one complete cycle. The electric activity is generated by one of the major event of the myogenic activity, the auricular depolarization (P waves), the ventricular depolarization (QRS complex) and waves of ventricular repolarization (T waves). The source of the activity are supposed to be a single *electric dipole* which position, direction and magnitude changes during the heart cycle. The position describes a single loop during one heart cycle. Because the resultant traces were all loops of variable shapes, the traces were referred to as P loop, QRS loop and T loop. The 3D-vector loops are represented by projections upon the frontal plane. Fig. 1 illustrates how the QRS vector is constructed in the frontal plane at the instant of the R peak.

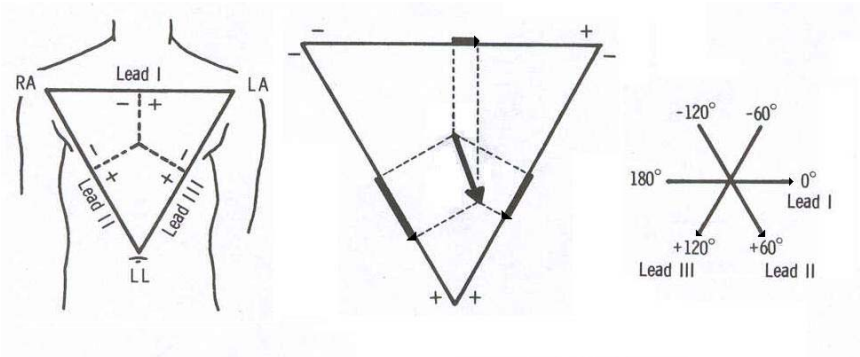


Fig. 1 Left: Configuration of Einthoven's leads. Middle: vector representation. Right: the 3 leads in the hexaxial presentation (see ECG: hexaxial reference system).

Application

Its application is mainly in the field of experimental cardiology. Clinical applications are limited. On-Line vectorcardiography has been applied during coronary angioplasty. Monitoring ST (segment) vector magnitude and QRS vector difference by vectorcardiography is used for identifying myocardial ischaemia during carotid endarterectomy.

More Info

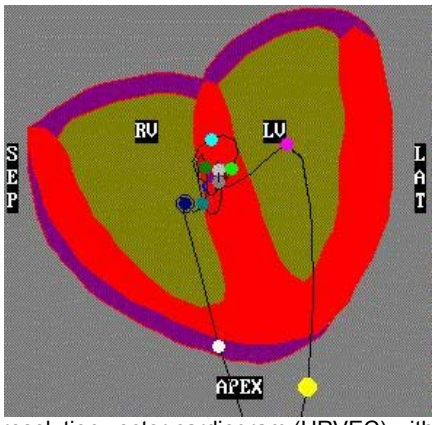


Fig. 2 High resolution vector cardiogram (HRVEC) with many loops.

The nomenclature and symbolic representation of the spatially oriented loops is standardized. The various letters denote the various vectors. One of the systems is:
q: the spatial direction of the vector of the QRS loop having the greatest magnitude,
p: the unit vector perpendicular to QRS and to the 'QRS plane' (the plane containing the QRS loop),
t: the spatial direction of the vector of the T loop having the greatest magnitude.

ECG: 12-lead ECG

Principle

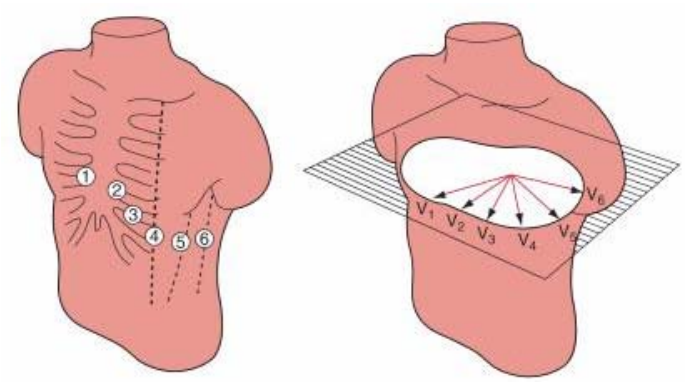


Fig. 1 Placement and vector representation of the precordial leads.

Unipolar chest leads, the precordial leads

When unipolar leads are recorded from the chest wall, the exploring electrode is connected to the positive pole of the ECG and the negative to the central terminal CT of Wilson (see ECG: augmented limb leads) at the sternum. By convention, the following sites are normally selected (Fig. 1):
• *V1*, the fourth intercostal space just to the right of the sternum
• *V2*, the fourth intercostal space just to the left of the sternum
• *V3*, midway between V2 and V4
• *V4*, the fifth intercostal space in the midclavicular line
• *V5*, the left anterior axillary line at the same horizontal level as V4
• *V6*, the left midaxillary line at the same horizontal level as V4.

Since the precordial leads V1, V2, V3 (the right precordial leads) and V4, V5, and V6 (left precordial leads) are placed directly on the chest, close to the heart, they do not require augmentation. CT is used as reference (negative input of the ECG machine), and consequently these leads are considered to be unipolar. The precordial leads view the heart's electrical activity in the so-called horizontal plane, in which the electrical Z-axis is located.
The QRS complex should be negative in lead V1 and positive in lead V6. The QRS complex should show a gradual transition from negative to positive between leads V2 and V4.
The precordial leads are always recorded together with the three basic Einthoven leads I, II and III (see ECG: basic electrocardiography) and the augmented leads (see ECG: augmented limb leads). Fig.2 presents them all 12 for a healthy subject.

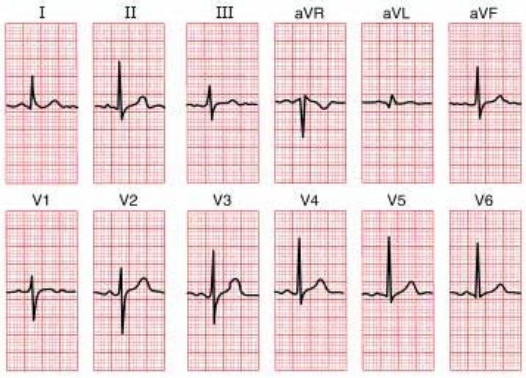


Fig.2. Normal 12-lead ECG

Ground electrode
In modern four-lead (the augmented leads and V5) and twelve-lead ECGs, an additional electrode is the ground electrode (usually green). This electrode is placed on the right leg by convention, although in theory it can be placed anywhere.

With a three-lead ECG, when one dipole is viewed, the remaining lead becomes the ground lead by default. Notice, that the ground lead is not the reference electrode (generally CT, see also ECG: augmented limb leads). The ground electrode suppresses 50 or 60 Hz hum of the mains.

Application

The equiphasic (or isolectric or biphasic, see bottom configuration Fig. 2 of ECG: basic electrocardiography) of lead is referred to as the transition lead. When the transition occurs earlier than lead V3, it is referred to as an *early transition*. When it occurs later than lead V3, it is referred to as a *late transition*. There should also be a gradual increase in the amplitude of the R wave between leads V1 and V4. This is known as *R wave progression*. Poor R wave progression is a nonspecific finding. It can be caused by conduction abnormalities, myocardial infarction, cardiomyopathy, and other pathological conditions.

More info

The twelve leads, each recording the activity from a different perspective, which also correlates to the area of identifying acute coronary injury, are classified as follows.
The *inferior leads* (leads II, III and aVF) look at electrical activity from the vantage point of the inferior or diaphragmatic wall of the left ventricle.
The *lateral leads* (I, aVL, V_5 and V_6) look at the electrical activity from the vantage point of the lateral wall of left ventricle. Because the positive electrode for leads I and aVL are located on the left shoulder, leads I and aVL are sometimes referred to as the high lateral leads. Because the positive electrodes for leads V5 and V6 are on the patient's chest, they are sometimes referred to as the low lateral leads.
The *septal leads*, V_1 and V_2 look at electrical activity from the vantage point of the septal wall of the left ventricle. They are often grouped together with the anterior leads.
The *anterior leads*, V_3 and V_4 look at electrical activity from the vantage point of the anterior wall of the left ventricle.
In addition, any two precordial leads that are next to one another are considered to be contiguous. In other words, even though V4 is an anterior lead and V5 is a lateral lead, they are contiguous because they are next to one another.
Lead aVR offers no specific view of the left ventricle, but views the endocardial wall from the right shoulder.

The modern ECG machine is completely integrated with an analog front end, a 12- to 16-bit analog-to-digital (A/D) converter, a computational microprocessor, and dedicated input-output (I/O) processors. These systems compute the 12 lead signals and analyze them with a set of rules. Fig. 3 shows the ECG of a heartbeat and the types of measurements that might be made on each of the component waves of the ECG and used for classifying each beat type and the subsequent cardiac rhythm.

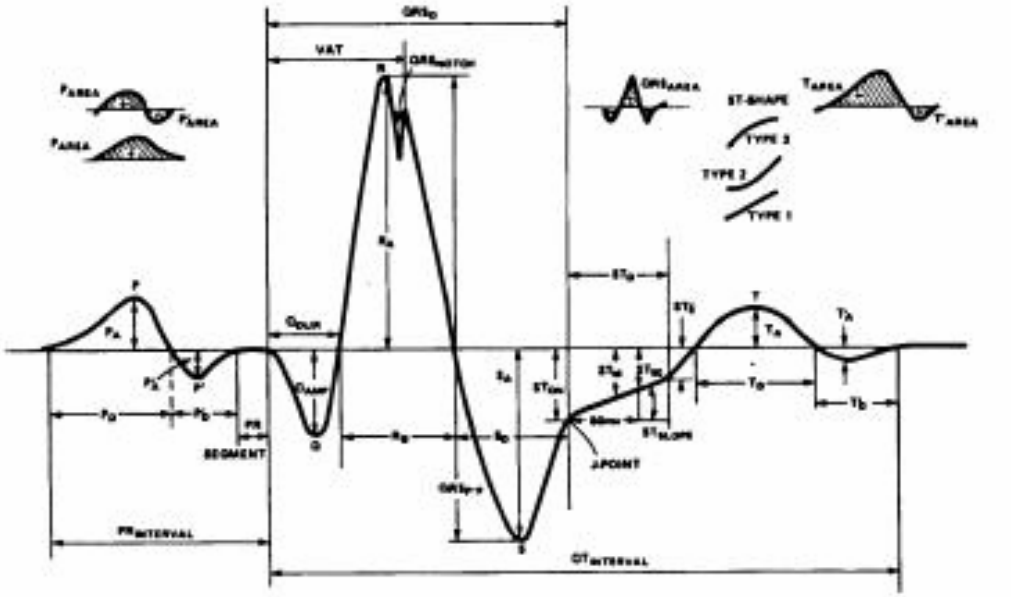


Fig. 3 The numerous ECG measurements that can be made with computer-based algorithms, with for instance artificial neural networks, wavelet analysis, component analysis. The measurements are primarily durations, amplitudes, and areas.

Electroencephalography

Basic Principles

Electroencephalography is the neurophysiologic measurement of the electrical activity of the brain by recording from electrodes placed to the scalp, or in special cases on the cortex (sub-dural). The resulting traces are known as an electroencephalogram (EEG).

The recording is obtained by placing electrodes on the scalp, usually after preparing the scalp area by light abrasion and application of a conductive gel to reduce impedance. Each electrode is connected to an input of a amplifier, which amplifies the voltage (typically 1,000–100,000 times, or 60–100 dB of voltage gain), and then displays it on a screen or inputs it to a computer. The amplitude of the spontaneous ongoing EEG is about 100 µV when measured on the scalp, and about 1-2 mV when measured on the surface of the brain. The amplitude of evoked potentials and pre-motor activity is up to some 15 µV. In general averaging is necessary to elucidate these signals from the spontaneous EEG.

Methods
The electrode-amplifier relationships are typically arranged in one of three ways:
Common reference derivation
One terminal of each amplifier is connected to the same electrode, and all other electrodes are measured relative to this single point. It is typical to use a reference electrode placed somewhere along the scalp midline, or a reference that links one or both earlobe electrodes.
Average reference derivation
The outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each amplifier.
Bipolar derivation
The electrodes are connected in series to an equal number of amplifiers. For example, amplifier 1 measures the difference between electrodes A and B, amplifier 2 measures the difference between B and C, and so on.
This distinction has become void with the advent of computerized or *paperless* EEGs, which record all electrodes against an arbitrary reference and will calculate the above montages post hoc.
The choice of the reference is crucial for the resulting spontaneous or evoked EG and the activity maps. EEG has several limitations. Scalp electrodes are not sensitive enough to pick out individual action potentials, the electric way signaling in the brain, resulting in releasing inhibitory, excitatory or modulatory neurotransmitters. Instead, the EEG picks up synchronization of neurons, which produces a greater voltage than the firing of an individual neuron. Secondly, EEG has limited anatomical specificity when compared with other functional brain imaging techniques such as functional MRI. With the use of a large number of electrodes EEG brain activity maps can be constructed (see Fig. 1) and the source of the activity can be estimated (see Magnetoencephalography).

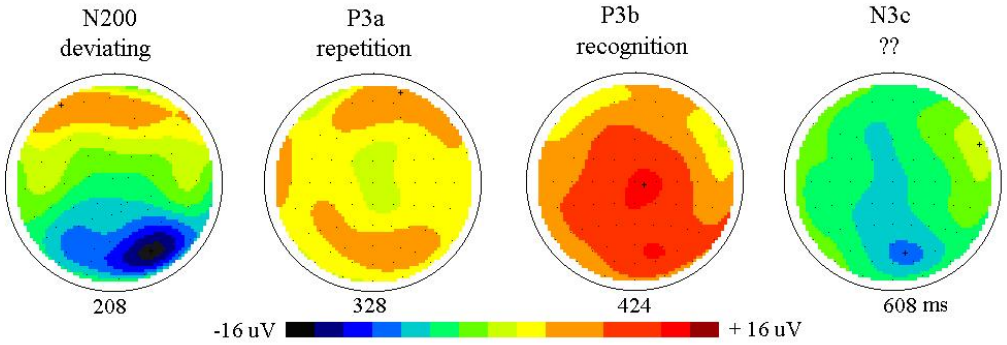


Fig. 1 Maps of the response components a 68 year old subject to the presentation of a rare course visual pattern within a continuous series of a fine pattern stimulus. The component names and their supposed functions are indicated at the top of the map. Nc3 is possibly related to anticipation, expectancy and action to a given stimulus. The numbers below the maps are the times after the start of the stimulus at which the maps are constructed from the responses of 64 channels.

EEG has several strong sides as a tool of exploring the brain activity. As other methods for researching brain activity have time resolution between seconds and minutes, the EEG has a resolution down to sub-millisecond. The brain is thought to work through its electric activity. EEG is in addition to the highly specialized MEG method the only method to measure it directly. Other methods for exploring functions in the brain do rely on blood flow or metabolism which may be decoupled from the brain electric activity.

Newer research typically combines EEG or Magnetoencephalography with MRI, SPECT or PET to get high temporal and spatial resolution.

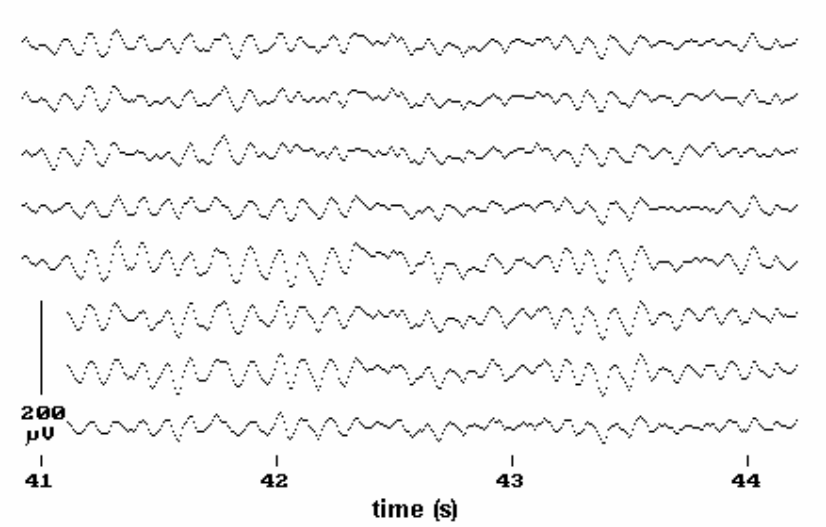


Fig. 2 Three seconds of spontaneous EEG recorded, shown for eight of the 64 channels. The sine-wave like signals represent the alpha activity. In the spectrum this activity gives rise to the so-called alpha-peak.

Wave types
Historically four major types of continuous (spontaneous) rhythmic sinus-like EEG waves are recognized: delta (up to 4 Hz), theta (4-8 Hz), alpha (8-13 Hz; Fig. 2) and beta (13-40 Hz). An alpha-like normal variant called mu is sometimes seen over the motor cortex (central scalp) and attenuates with movement, or rather with the intention to move. The sensorimotor rhythm (SMR) is a middle frequency (about 12-16Hz) associated with physical stillness and body presence. Gamma is the frequency range above 40 Hz (approximately 30-80 Hz to be precise).

Application

This device is used to assess brain damage, caused by tumors, CVA's and trauma's. It also applied for supporting the diagnosis of neurological disorders like dementia, etc. and psychiatric disorders. Neuroscientists and biological psychiatrists use EEGs to study the function of the brain by recording the spontaneous and evoked EEG during controlled behavior of human volunteers and animals in lab experiments. Theories to explain sleep often rely on EEG patterns recorded during sleep sessions. In addition, the procedure is routinely used clinically to assist in the diagnosis of epilepsy. In some jurisdictions it is used to assess brain death.
The EEG has a history of tens years and so there is a mass of journal publications and handbooks about EEG theory and applications.

Reference
Regan D. Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. New York: Elsevier, 1989.

Lorentz force

The Lorentz force is the force exerted on a charged particle in an electromagnetic field. The particle will experience a force due to electric field of *q*E, and due to the magnetic field *q*v × B. Combined they give the Lorentz force equation (or law):

$F = q(E + v \times B)$, (1)

where
F is the force (in N)
E is the electric field (in V/m)
B is the magnetic field (in Webers/m^2, or equivalently, Tesla's)
q is the electric charge of the particle (in C, Couloms)
v is the instantaneous velocity of the particle (in m/s)
and × is the cross product (see below).

Thus a positively charged particle will be accelerated in the *same* linear orientation as the *E* field, but will curve perpendicularly to the *B* field according to the right-hand rule ("kurketrekker regel").
The cross product is a vector operation in a three-dimensional Euclidian space (with generally Cartesian coordinates, so with orthogonal axes). It is also known as the vector product or outer product. It differs from the dot product (also called the inner or scalar product, i.e. in a one-dimensional space the common multiplication) in that it results in a vector rather than in a scalar. Its main use lies in the fact that the cross product of two vectors is orthogonal to both of them. Fig. 1 illustrates the vector product.

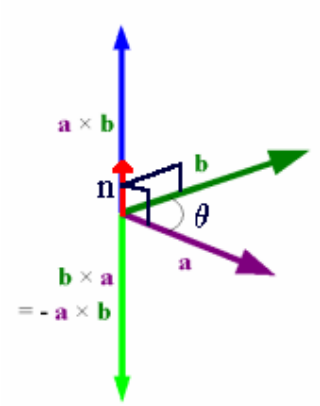


Fig. 1 Cross product of vectors a and b, with θ the measure of the angle between **a** and **b** (0° ≤ θ ≤ 180°), on the plane defined by the span of the vectors. **n** is the unit vector perpendicular to both **a** and **b** (a vector of size 1 with the dimension of the both vectors taken together).

Equivalently to equation (1), we can express the Lorentz force law in terms of the electric charge density ρ and current density J as;

$$F = \int_V (\rho E + J \times B) \cdot dV \,. \qquad (2)$$

Applications

The Lorentz force is a principle exploited in many physical devices including those also applied for medical research and in the clinic. A simple application is the electrophoretic trough and complicated one the mass spectrometer (see Mass spectrography) and the cyclotron and other circular path particle accelerators (necessary for PET, see Positron emission tomography).

Magnetoencephalography (MEG)

Basic Principles

A magnetoencephalograph (Fig. 1) is the device to measure the magnetic field produced by the coherent electrical activity of assemblies of neurons in the brain.

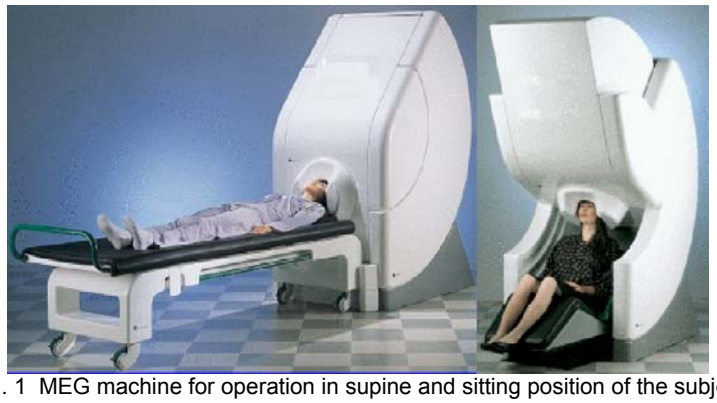


Fig. 1 MEG machine for operation in supine and sitting position of the subject

The magnetic signals themselves derive from currents flowing in the dendrites of neurons. The laws of electromagnetism govern the parameters of the field (see http://en.wikipedia.org/wiki/Magnetic_field). Given the direction of the current, the right-hand rule applies for the direction of the field. The fields are measured just outside the scalp, using extremely sensitive magnetic sensors (mostly gradiometers), including a so-called SQUID, (Super conducting Quantum Interference Device; see More info), mounted in a kind of helmet (Fig. 2), which is filled with liquid helium to provide super conducting.

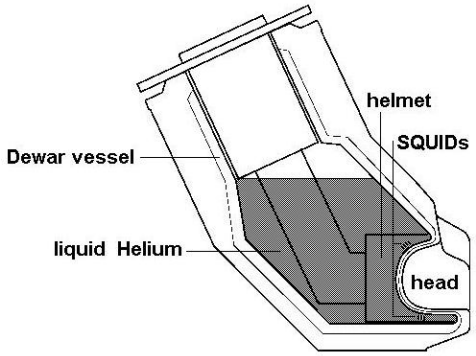


Fig. 2 Basic principle of operation of MEG machine

Because the magnetic signals emitted by the brain are on the order of a few femtotesla (1 fT = 10^{-15} T), the measurements are performed in a magnetically shielding room to exclude interference with external magnetic signals, including the 10^8 stronger earth's magnetic field (see Fig. 3). The room reduces the high-frequency noise, while noise cancellation algorithms reduce low-frequency signals. The signals of additional sensors mounted far from the helmet reduce low frequency spatial noise.

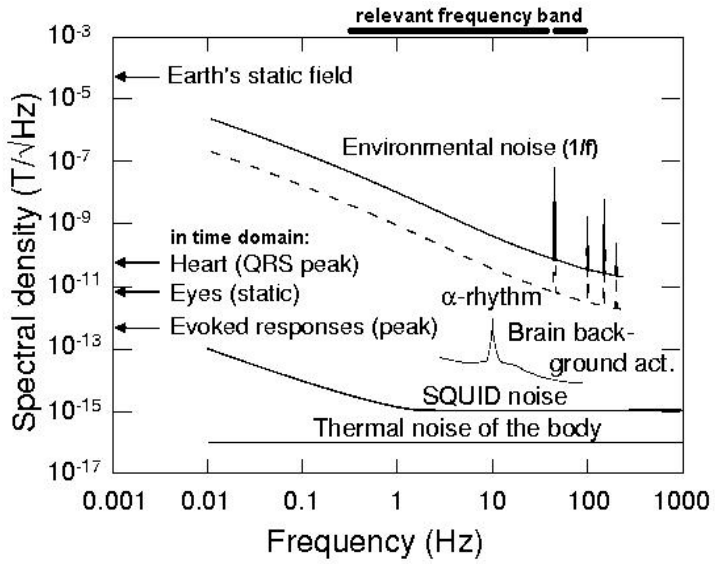


Fig. 3 Amplitudes of MEG signals and noise sources

Modern whole-head systems have roughly 300 channels, and have a noise floor of around 5-7 fT (> 1 Hz. The ongoing spontaneously evoked magnetic field of the brain is typically around 100 to 1000 fT, while sensory responses and pre-motor signals are at least ten times smaller and often close or lower than the noise floor. These signals can only be made visible by applying signal-averaging techniques. Fig. 4 (top) gives an example of an analysis, a MEG map constructed 162 ms after applying a visual stimulus.
MEG is a relatively new technique with good spatial and a temporal resolution, thus complementing other brain activity measurement techniques such as electroencephalography (EEG), positron emission tomography (PET), and functional Magnetic Resonance Imaging (fMRI). The primary technical difficulty with MEG is its high sensitivity for interfering signals evoked outside the brain by motion of charged objects (e.g. the eyeball, heart, limbs) and metal implants, and its vulnerability for non-brain generated signals (ECG, EMG, loudspeaker).

Applications

Detecting and localizing epileptiform spiking activity in patients with epilepsy, and in localizing eloquent cortex for surgical planning in patients with brain tumors. (see More info). Localization of the source, source modeling, is performed with complicated mathematics by solving the 'inverse-problem', which is subject of intensive research (see **More info**). Diagnostics of cognitive and behavior disorders are often performed in combination with EEG or fMRI.
In the Netherlands the VUMC and UMC St Radboud have MEG machines. These machines are as expensive as fMRI machines and for its operation a highly specialized technical staff is necessary. Specialized commercially MEG machines are developed for examining the heart and the unborn human fetus.

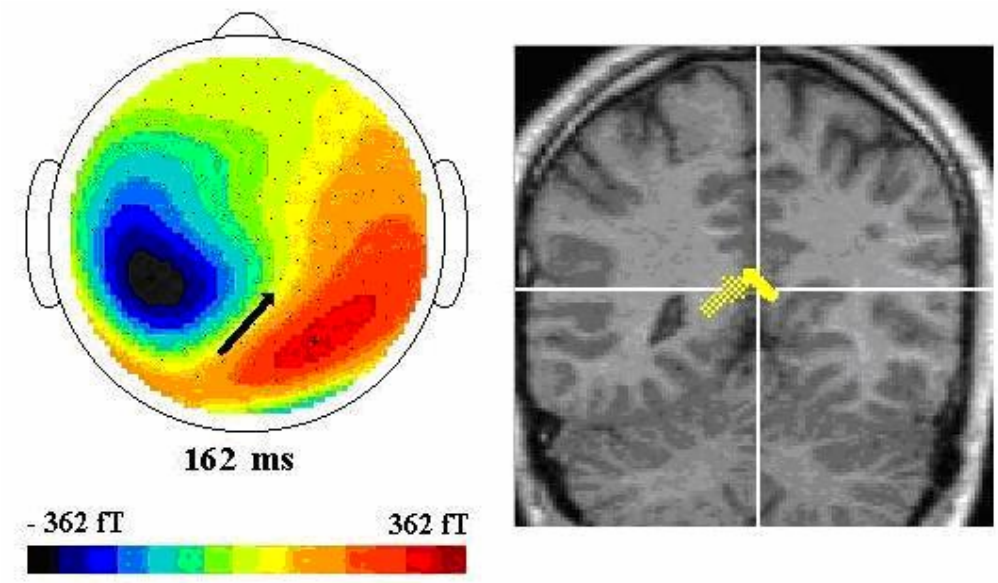


Fig. 4 Field and source analyses On top, the map is constructed 162 ms after applying a visual stimulus. The ECD is presented by the black arrow. An advanced analysis (bottom) of the same field, now with the MRI image, shows that this single ECD is comprised of two ECDs closely together. (Data provided by the author.)

More info

Localizing the source of activity

With mathematical physics it is possible to localize the source of electric brain activity. To obtain the source(s) the so-called *inverse problem* should be solved, a procedure of complicated numerical mathematics. A set of equations (the model) with numerically known constants is applied to an input-data set comprised of the signals of the sensors (the channels) and the anatomical data of the head (often provided by a MRI image), in order to estimate the source(s), the output of the model. This estimate is not unique, since there are many solutions. The solution, called an equivalent current dipole (ECD) is only unique if the system is noise-free, if there is only one source active and if the head is approximated by a magnetically homogeneous sphere. In reality these conditions certainly do not hold. In practice the simplest approach is mostly to consider a pair of ECDs, one in each hemisphere, and a spherical head. The best accuracy of localization is 2-3 mm, about two times better than with EEG. An advanced approach is to use MRI images of the head and supposing a distributed source. Now, the density (a cloud of sources of equal strength) of hypothetical sources is calculated and visualized similarly as fMRI and SPECT analyses.
Inverse problems are typically ill posed, as opposed to the well-posed problems The forward problem is well-posed: for a given source and system parameters, the magnetic field can be calculated and there is only one unique solution.

SQUID

There are various types of SQUID, but all have at least one super conducting Josephson junction (Fig. 5).

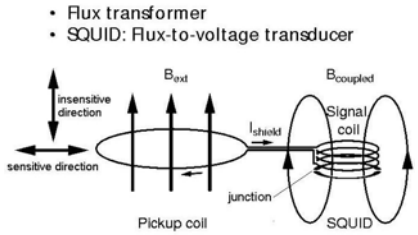


Fig. 5 Principle of operation of the magnetic sensor, in this case a magnetometer. B is the magnetic field density.

The basic principle of operation is to cancel the current though the junction with an adjustable current though the signal coil (Fig. 5). The strength and direction of the compensating current is the primary sensor signal. In MEG, often gradiometers are applied. They are characterized with two compensation coils with opposite winding.
SQUIDs are the most sensitive devices to measure extremely tiny magnetic fields with noise levels as low as 3 $fT \cdot Hz^{-½}$. Another application is the scanning SQUID microscope.

References

Sato S, et al. Principles of magnetoencephalography. J Clin Neurophysiol. 1991;8:144-56. Review.
Faugeras O, et al.Variational, geometric, and statistical methods for modeling brain anatomy and function. Neuroimage. 2004;23 Suppl 1:S46-55. Review.
Turner R, Jones T. Techniques for imaging neuroscience. Br Med Bull. 2003;65:3-20. Review.

Piezoelectricity

Basic principle

Piezoelectricity is the ability of certain quartz analogue crystals or ceramics to *generate a voltage* in response to applied mechanical stress. The word is derived from the Geek *piezein*, which means to squeeze or press. The piezoelectric effect is reversible in that piezoelectric crystals when subjected to an externally applied voltage, can change shape by a small amount. The deformation, about 0.1% of the original dimension, is of the order of nm, but nevertheless finds numerous applications.
In a piezoelectric crystal, the positive and negative electric charges are symmetrically separated, such that the crystal overall is electrically neutral. Each of these sites of charge forms an assembly of aligned electric dipoles. When a mechanical stress is applied, the symmetry is disturbed, and the charge asymmetry generates a voltage across the material. For example, a 1 cm cube of quartz with 2000 Newton of correctly applied force upon it, can produce a voltage of 12,500 V.
Piezoelectric materials also show the opposite effect, called converse piezoelectricity, where the application of an electrical field creates mechanical deformation in the crystal. The crystal is maximally excited by a sine wave signal at the *resonance frequency* of the crystal. (See for a technical description of the piezoelectric effect http://en.wikipedia.org/wiki/Piezoelectric.)

Applications

As very high voltages correspond to only tiny changes in the width of the crystal, this width can be changed with better-than-µm precision, making piezo crystals the most important tool for positioning objects (e.g. by stepper motors) with extreme accuracy.
The applications of piezoelectricity in medicine, science, technology and daily life (e.g. electric lighter) is enormous. Examples are the production and detection of sound (e.g ultrasound in medicine, loudspeakers, microphones in medical (e.g. audiology) and daily live applications), ultra-fine focusing of optical assemblies (atomic force microscope and scanning tunneling microscope) employ converse piezoelectricity to keep the sensing needle close to the probe), fine-tuning a laser's frequency, generation of high voltages, electronic frequency generation, microbalances, laser mirror alignment, acousto-optic modulator, a device that vibrates a mirror to give the reflected beam a Doppler shift. This is useful for fine-tuning a laser's frequency.

More info

Ultrasonic transducers Piezoelectric materials are used as utrasonic transducers for imaging applications (e.g. medical imaging,) and high power applications e.g. in medical treatment and sonochemistry, e.i. the effect of sound waves on chemical systems, e.g in sonoluminesence (the emission of short bursts of light from imploding bubbles in a liquid when excited by ultrasound) and sonic cavitation. For imaging applications, the transducer can act as both a sensor and an actuator.
Piezoelectric motors Types of piezoelectric motor include the well-known travelling-wave motor used for auto-focus in cameras, inchworm motors (single cell electrophysiology) for linear motion, and rectangular four-quadrant motors with high power density (2.5 W/cm³) and speed ranging from 10 nm/s to 800 mm/s. All these motors work on the same principle. Driven by dual orthogonal vibration modes with a phase shift of 90°, the contact point between two surfaces vibrates in an elliptical path, producing a frictional force between the surfaces. Usually, one surface is fixed causing the other to move. In most piezoelectric motors the piezoelectric crystal is excited at the resonance frequency of the crystal.
Quartz clocks employ a tuning fork made from quartz that uses a combination of both direct and converse piezoelectricity to generate a regularly timed series of electrical pulses that is used to mark time. The quartz has a precisely defined natural frequency of oscillating and this is used to stabilize the frequency of a periodic voltage applied to the crystal. The same principle is critical in all radio transmitters and receivers, and in *computers* where it creates a *clock pulse*. Both of these usually use a frequency multiplier to reach the MHz and ranges.

Vision

Ophthalmic Corrections

Principle

Optical corrections with lenses (spectacles or contact lenses) have primarily the aim to compensate for an unallowable deficit (hyperopia) or surplus (myopia) of the refractive power of the eye. Advanced measures like intraocular lenses and cornea reshaping are beyond our scope. With correction the image is no longer blurred, apparently increases the acuity. Actually acuity is not changed (see Visual acuity).

Application

Hyperopia
Hyperopia may be corrected with positive (convex) lenses. Mild to moderate hyperopia can be compensated through accommodation and may not need correction (or may not even be noticed). However, the perennial accommodation is a strain on the eye and may cause headache.
Myopia
Myopia causes problems already at 0.5D. Correction is with negative (concave) lenses.
Presbyopia
Against presbyopia, a lack of accommodation due to aging, some means of changing the correction according to distance is needed. This may be achieved using:

- Separate spectacles for long distance (weak positive or negative lens) and for reading (positive lens).
- Bifocal lenses, where most of the lens surface has a power suitable for vision at long distance, but a region in its lower part has a different power, for reading etc. There are also trifocal lenses.
- Progressive lenses, which use the same idea as bifocals, but there is a smooth grading between the upper and lower part of the lens, see Fig. 1.

Progressive lenses are adequate but not everybody can get used to having sharp sight only in a horizontal band across the field of vision. They require very precise fitting.

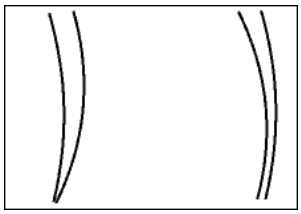


Fig. 1 Progressive lenses The left lens is intended for an emmetrope, the right one for a myope.

Astigmatism
Against astigmatism (i.e. different power of the eye for different sectors of the visual field), astigmatic lenses with the opposite orientation are needed, eventually as a progressive lens.
Contact lenses
Contact lenses, mostly chosen for cosmetic reasons, cause less distortion in the outer parts of the field of vision than spectacles do (since they are closer to the main plane of the eye optics).
The most important subdivision of contact lenses is that of soft (commonest) and hard lenses.
Hyperopia and myopia are corrected according to the same principles as with spectacles.
Presbyopia and astigmatism are more problematical, because contact lenses tend to rotate. Several devices can prevent this rotation mechanically; they are successful in some people and not in others.
There are several other ways to correct presbyopia with contact lenses, none of them entirely satisfactory:

- Different lenses on each eye - you use one eye for distant sight and the other for near sight. Stereoscopic vision and judgment of distance are impaired, but it works fairly well in practice.
- The centre of the lens has the power for distant sight and the periphery of the lens the power for near sight. The idea is that when you look downwards, the lower eyelid prevents the lens from following, so that you look through its peripheral part. This works only with hard lenses (soft ones stick too strongly to the cornea).
- The centre of the lens has the power for near sight and the periphery of the lens the power for distant sight (i.e. the opposite of the above). Here the idea is that when you accommodate, the pupil constricts, and the eye only receives light which has passed through the centre of the lens. Since this reflex weakens with age, the method is not too reliable.

Ophthalmoscopy

Principle

The ophthalmoscope, originally invented by the famous Herman von Helmholtz, is an instrument used in determining the health of the retina, the vitreous humor and other interior structures of the eye. It consisting essentially of a mirror that reflects light from a light source into the eye. Through a central hole in the mirror the physician can examine the patient's eye. Fig. 1 presents the principle with a the light pathway from the patient's retina to its image on the retina of the physician. b present a part of the pathway of illumination of the patent's retina. The physician actually watches the image (with height h', see c) at a distance of 250 mm from his eye.

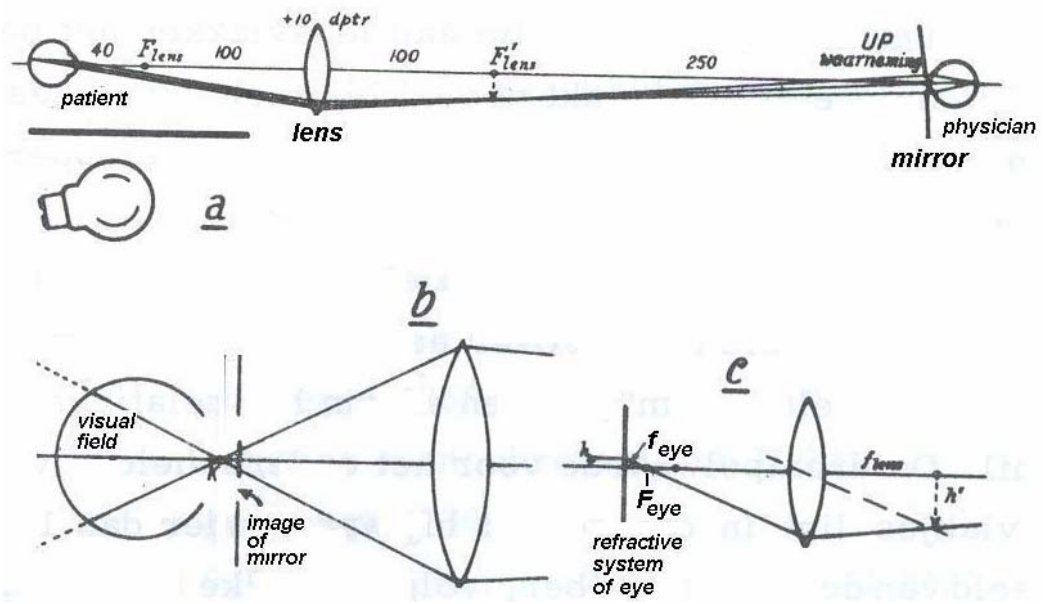


Fig. 1 Ophthalmoscope with image (a and c) and illumination (b) light pathways. All distances in mm.

Application

In patients with headaches, the finding of swollen optic discs, or papilledema, on ophthalmoscopy is a key sign, as this indicates raised intracranial pressure which could have several causes. Cupped optic discs are seen in glaucoma.
In patients with diabetes, regular ophthalmoscopic eye examinations (once every 6 months to 1 year) is mandatory to screen for diabetic retinopathy as visual loss due to diabetes can be prevented by retinal laser treatment if retinopathy is spotted early.
In arterial hypertension, hypertensive changes of the retina closely mimic those in the brain, and may predict cerebrovascular accidents (strokes).

More info

There are several methods of types of ophthalmoscopy each performed with a different version of the opthalmoscope.

Direct ophthalmoscope
Direct ophthalmoscopy using a slit-lamp and negative auxiliary lenses can provide a very high level of magnification-even greater than that of the monocular hand held direct ophthalmoscope. Stereopsis is provided to a greater degree than all other examination techniques.

Binocular indirect ophthalmoscope
Binocular indirect ophthalmoscopy is a technique used to evaluate the entire ocular fundus. It provides for stereoscopic, wide-angled, high-resolution views of the entire fundus and overlying vitreous. Its

optical principles and illumination options allow for visualization of the fundus regardless of high ametropia, hazy ocular media, or central opacities.

Monocular indirect ophthalmoscope
Monocular indirect ophthalmoscopy combines the advantages of increased field of view (indirect ophthalmoscopy) with erect real imaging (direct ophthalmoscopy). By collecting and redirecting peripheral fundus-reflected illumination rays, which cannot be accomplished with the direct ophthalmoscope, the indirect ophthalmoscope extends the observer's field of view approximately four to five times. An internal lens system then re-inverts the initially inverted image to a real erect one, which is then magnified. This image is focusable using the focusing lever/eyepiece lever.

Laser scanning ophthalmoscopy
This is a method of examination of the eye. It uses the technique of confocal laser scanning microscopy for diagnostic imaging of retina or cornea of the human eye. It is helpful in the diagnosis of glaucoma, macular degeneration, and other retina disorders. It has been combined with adaptive optics technology to provide sharper images of the retina. The principle of illumination and imaging is illustrated in Fig. 2 and 3.

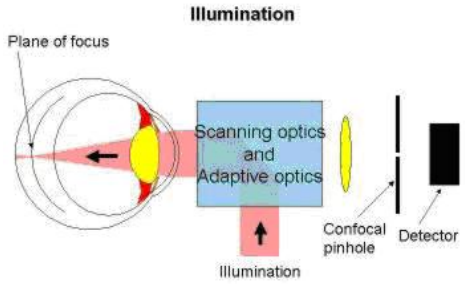


Fig. 2 Illumination The illumination beam uses the eye's optics to focus the illumination light to a point on the retina. The scanning optics move the focused spot across the retina in a raster pattern.

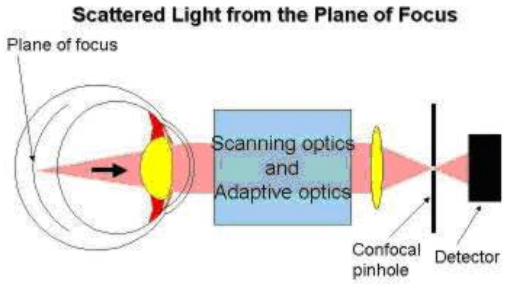


Fig. 3 Detection The confocal pinhole is conjugate to the plane of focus on the retina, so light that scatters from the plane of focus passes the through the pinhole and is detected by the photomultiplier.

Optometry

Principle

Optometry is the health care profession concerned with examination, diagnosis, and treatment of the eyes and related structures with the objective to correct vision using lenses and other optical aids (see also Retinoscopy and Ophthalmic Corrections).

Fig. 1 An optical refractor, also called a phoropter in use.

More info

An optometrist (or opticians or optometric physicians) is a care practitioner for most vision and ocular health concerns, including, but not limited to, fitting and prescribing glasses and contact lenses, diagnosing and treating (excluding surgery) muscular abnormalities, treating minor ocular injuries, diagnosing and treating diseases such as glaucoma and diagnosing others such as diabetic retinopathy.

Retinoscopy

Principle

Retinoscopy is a technique to obtain an objective measurement of the refractive condition of a patient's eye (see also Optics of the eye and Ophthalmic Corrections). The examiner uses a retinoscope to shine light into the patient's eye and observes the reflection off the patient's retina. While moving the streak or spot of light across the pupil the examiner observes the relative movement of the reflex. Next, he uses a phoropter (a large an instrument to measure an individual's refractive error in order to determine the eyeglass prescription, see Optometry) or manually places lenses over the eye to "neutralize" the reflex.

Application

Retinoscopy is especially useful in prescribing corrective lenses for patients who are unable to undergo a subjective refraction that requires a judgment and response from the patient. It is also used to evaluate accommodative ability of the eye and detect latent hyperopia.

More info

Retinoscopy works on a principle called Foucault's principle. Basically it indicates that the examiner should simulate the infinity to obtain the correct refractive power. Hence a power corresponding to the working distance is subtracted from the gross retinoscope value.
Static retinoscopy is performed when the patient does not accommodate. Dynamic retinoscopy is performed when the patient has active accommodation from viewing a near target.

Visual acuity

Principle

Visual acuity (VA) is acuteness or clearness of vision, especially form vision, which is dependent on the sharpness of the retinal focus within the eye, the sensitivity of the nervous elements, and the interpretative faculty of the brain. It is also dependent on the stimulus and lightning conditions. Clinically, VA is a quantitative measure of the ability to identify black symbols on a white background at a standardised distance as the size of the symbols is varied. It represents the smallest size that can be reliably identified.
If the eye watches two objects, for example two black dots, it will see them under a certain angle. Obviously, if this angle gets too small, the eye cannot distinguish them from each other. The smallest angle that still allows the eye to see that there are two dots is the *resolution angle*. VA is defined as the 1/resolution angle (in minutes of arc. One minute, 1', of arc is $1/60^{o}$). The "typical" human resolution angle is 1'. The VA-test is the most common clinical test of visual function since it easily detects a resolution impairment in normal daily vision.

Application

Traditional the Snellen chart (Fig. 1) is used for VA testing. This chart (and any other VA-chart) should be used under certain specified conditions of lighting etc. at a specific distance. (For more details about the method and other details of measurement see e.g. the VA chapter of Wikipedia.)

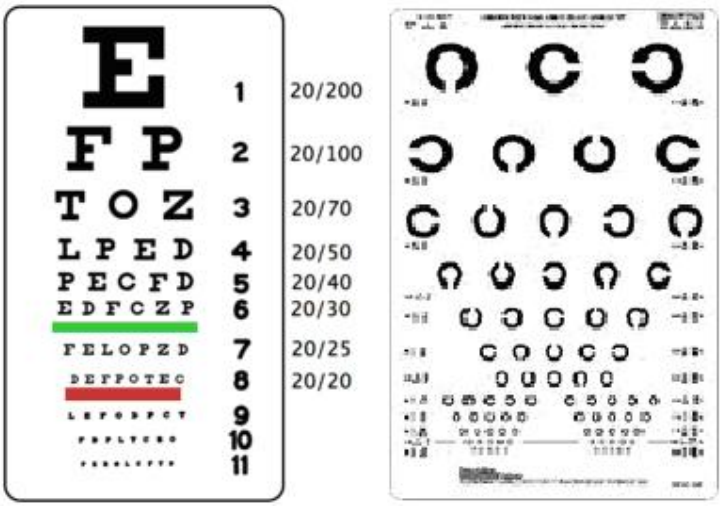


Fig. 1 Snellen (left) and Landolt C chart

Another frequently used and better chart is the Landolt C chart. The broken ring optotype is made with a "C" like figure in a 5 x 5 grid that subtends 5' (standard size) and an opening measuring 1'. The subjects have to indicate the orientation of the gap (left, right, up, down, see Fig. 2). The advantage is that all C's of the same size are equally recognisable. This chart is preferred for laboratory experiments and it can also be used for illiterates.
Other charts are the Sloan's chart and the Bailey&Lovie chart. The Lea chart (figures like apple, house, etc.) and Tumbling E chart are also useful for illiterate people.

More Info

The eye media (tear film, cornea, anterior chamber, pupil, lens, vitreous, and finally the retina) affect the quality of the image. The retinal pigment epithelium is responsible for, among other things, absorbing light that crosses the retina to prevent backward scatter.
To resolve fine details, the eye's optical system has to project a focused image on the fovea. The fovea has the highest density of cone photoreceptors and no rods, thus having the highest resolution and best colour vision (but no night vision). VA and colour vision, both based on the same receptor cells, are different but unrelated physiologic functions as also holds for other function such as reduced contrast, or inability to track fast moving objects.
Along the visual pathway, after the passing the lateral geniculate body, a relay station, the visual cortex, the posterior (occipital) part of the cortex, is the first major centre for visual processing. The central 10°

of the visual field (approximately the extension of the macula) is represented by at least 60% of the visual cortex. Much of these neurones are believed to be involved directly into VA processing.

Table 1 Visual acuity scales

Foot	Metre	Decimal	LogMAR
20/200	6/60	0.10	1.0
20/160	6/48	0.13	0.9
20/125	6/37	0.16	0.8
20/100	6/30	0.20	0.7
20/80	6/24	0.25	0.6
20/63	6/18	0.32	0.5
20/50	6/15	0.40	0.4
20/40	6/12	0.50	0.3
20/32	6/9	0.63	0.2
20/25	6/7	0.80	0.1
20/20	6/6	1.00	0.0
20/16	6/4.8	1.25	-0.1
20/12.5	6/3.75	1.60	-0.2
20/10	6/3	2.00	-0.3

VA is expressed as a vulgar fraction (m/n) or as a decimal number (see Table 1 for conversions). Using the foot, VA is expressed relative to 20/20, the normal value. Otherwise, using the metre, the equivalent is expressed relative to 6/6. In the decimal system, the acuity is defined as the reciprocal value of the size of the gap (in arc minutes) of the smallest Landolt C that can be reliably identified. A value of 1.0 is equal to 20/20.
LogMAR is another commonly used scale, which is expressed as the logarithm of the resolution angle. It is a conversion to a linear scale. Positive values indicate vision loss, while negative values denote normal or better VA. This scale is mainly used in basic research.

With normal eyesight (VA is 6/6) one can see can see detail from 6 m away and with VA is 6/12 one can see detail from 6 m away as a subjects with VA is 6/6 sees details from 12 m away. So with 6/12 acuity is halved.
In humans, the maximum acuity of a healthy, emmetropic eye (and even ametropic eyes with correctors) is approximately 6/5 to 6/3.6 and6/6 is considered as the lower limit of normal. Maximum VA without visual aids (such as binoculars) is around 6/3. In case of hyperopia or myopia, the eye should be corrected for the viewing distance. Some birds, such as hawks, have acuity about 5 times maximal human VA.

VA is typically measured monocularly rather than binocularly. In some cases, binocular VA will be measured. Usually binocular VA is slightly better than monocular VA.
"Dynamic VA" is VA for a moving object.
VA can also be measured as a function of eccentricity in the various directions. It rapidly falls off with eccentricity.

References
http://www.neuro.uu.se/fysiologi/gu/nbb/lectures/VisAcuity.html

Optics of the eye

Principle

There are a number of refracting surfaces in the eye, with the important ones the anterior surface of the cornea, and the lens.
Of these the cornea, because of the large difference in refractive index between air (1.0) and corneal tissue (1.37) is the more powerful, with a typical power of about 40 dioptre (dptr). The lens in the relaxed (not accommodated) state has a power of about 17 dptr. Accommodation may increase this, by about 14 dptr in children, less with increasing age.

These several components may, by approximation, be thought of as equivalent to a single ideal lens. The resulting simplification is the *reduced eye* (see Fig. 1).
Location and strength of the ideal lens vary in literature but the following is a good approximation. Its principal plane is situated just behind the iris and so its distance to the retina (the axis length) is 20 mm, and its power is 50 dptr.

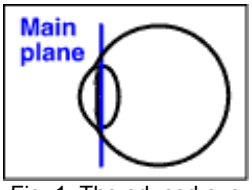


Fig. 1 The educed eye

Note that 50 dptr is less than the sum of the powers of cornea and lens. They are not close enough together for their powers to be additive. In the normal, resting reduced eye, somewhat arbitrarily, 36 dptr are due to the cornea and 14 dptr to the lens.
Both the axis length and the principal plane power are abstractions. However, the quantity power - 1/axis length (in m), the "net power" of the reduced eye, is concrete and measurable. The eye will see a sharp image when the image is focused on the retina, i.e. when the image distance is equal to the axis length. Replacing axis length by image distance and applying the lens formula (see Light: the ideal lens), under the condition of a sharply seen object, we find that the net power (in dptr, m^{-1}) is:
$\varphi = 1/f = 1/d_{object}$
The difference $1/d_{object}$ – 1/axis length directly gives the correction. Negativity means myopia and a negative lens equal to the absolute difference. With this correction the myopic eye will see clearly at infinite distance (the Foucault's principle, see Retinoscopy).

More info

More precisely the eye should by presented as a system with two principal planes (see The Ideal Lens), as illustrated in Fig. 2. It appears that the two principal planes are only 0.3 mm from each other. This is caused by the small differences between most eye media and the small refractive contribution of the lens. Therefore, for most applications the reduced eye with one principal plane is adequate.

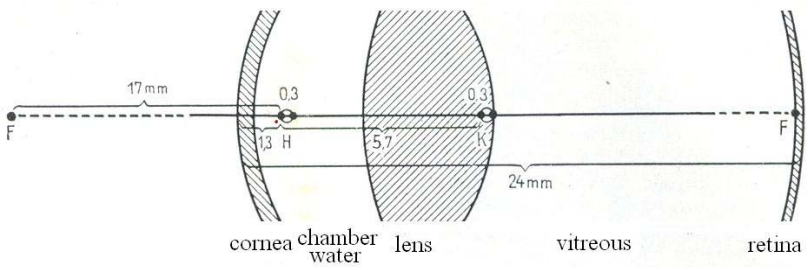


Fig. 2 The reduced eye with two principal planes.

The power of the reduced eye of Fig. 2 is:
$\varphi = 1/f = 1/0.017 = 58.8$ dptr or $\varphi = 1.33/(0.024 - 0.0013 - 0.0003) = 59.4$ dptr,
where 1.33 is the mean refractive index of the eye.

Index