We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Chapter

Properties and Behaviour of Waves

Ivo Čáp, Klára Čápová, Milan Smetana and Štefan Borik

1. Introduction

Waves have specific properties, which do not depend on their physical substance. It means they are common for all kinds of waves as electromagnetic, mechanical, etc. It deals mainly with such typical demonstrations as interference, diffraction, and wave displaying connected with them.

2. Wave function

The wave function is a function of space coordinates and time. It describes the wave quantity in a proper place and time, for example, $u(r, t)$—an acoustic displacement of a mechanical wave or $E(r, t)$ and $H(r, t)$—intensities of electrical and magnetic components of an EM wave. Any wave propagates in space. At any point, we can define the velocity of propagation $c = c\boldsymbol{n^0}$, where $\boldsymbol{n}^0$ is the unit vector in the wave propagation direction.

The general form of the wave function can be expressed as

$$\mathbf{u}(\mathbf{r}, t) = \mathbf{f}\ (\mathbf{r} \pm \mathbf{c}t), \tag{1}$$

where the sign "−" stands for a wave propagating in $\boldsymbol{n}^0$ direction a "+" in the opposite direction.

2.1 Wave polarisation

The wave quantities are vector quantities, which have their direction relating to the direction of the wave propagation. If this direction is defined, we speak about a polarised wave. For example, an acoustic wave generated by a converter or EM wave generated by an antenna is polarised. One of the polarizations is a ***longitudinal polarisation***—vector of the wave quantity is parallel to the wave propagation direction, for example, sound in the air, sound or ultrasound in a liquid, ultrasound in soft tissue, etc. ***Transverse polarisation*** means direction of the wave quantity perpendicular to the propagation direction, for example, EM wave in free space, wave on an elastic string. Transverse polarisation can be ***a linear*** one if the direction of the wave quantity does not vary along with the wave, a ***circular*** one if the vector of the transverse wave quantity follows a circle or an ***elliptical*** one if the transverse wave quantity follows an ellipse. Circular or elliptical polarisation should be taken as a superposition of two harmonic linearly polarised transverse waves propagating together with mutually perpendicular polarisation and phase shift of $\pi/2$ rad. If the amplitudes of both waves are the same, the resulting polarisation is circular. In the case of different amplitudes, the polarisation is elliptical.

There are also elliptically polarised waves in a plane parallel to the propagation direction. It can be considered a combination of the transversely polarised and longitudinally polarised waves. One example is the surface wave on the water. If we observe the movement of a tiny body floating on the surface of the water with the surface wave, we can see that the body oscillates in a vertical direction, but it also performs an oscillating motion in the longitudinal direction. It follows an ellipse.

If the wave is generated by many individual uncoordinated sources of polarised waves, the polarisation of the resulting wave cannot be unambiguously determined, and such a wave is called a non-polarised one. Bulk acoustic waves in gases and liquids are always longitudinally polarised because in these media the transversally polarised waves cannot propagate due to the fluidity of the medium (this does not apply to the surface waves in the liquid). If the EM wave is generated by a temperature source (hot fibre, infrared radiation of the body surface, starlight, etc.), the individual sources are single atoms of the substance that emit their waves accidentally. Each elementary wave is transversally polarised, and all polarisation directions are present equally. The resulting wave is therefore non-polarised. Typical sources of such non-polarised light are bulbs, gas discharge lamps, LEDs, etc.

Some applications require the use of polarised radiation. In this case, a source of polarised radiation, for example, LASER, can be used. Another possibility is to use a polarizer (polarising filter), a tool that suppresses (filters) one polarisation of non-polarised radiation. Polarizers use anisotropic crystals, lattice structures or reflective elements. Periodic structures (dense optical gratings) are most often used as light polarizers. Polarizers are used on glasses, cameras, or microscope lenses, polarising filters are part of LCDs (digital watches, laptops, mobile phone screens, etc.). The polarisation of light also occurs when reflected from shiny surfaces.

Reflection and transmission of light at the plane interface of two media express Fresnel's relations

$$I_{ro} = I_{rd} \left[\frac{n_2 \cos \alpha - n_1 \cos \beta}{n_2 \cos \alpha + n_1 \cos \beta} \right]^2 = I_{rd} \left[\frac{\tan (\alpha - \beta)}{\tan (\alpha + \beta)} \right]^2 \tag{2}$$

$$I_{ko} = I_{kd} \left[\frac{n_1 \cos \alpha - n_2 \cos \beta}{n_1 \cos \alpha + n_2 \cos \beta} \right]^2 = I_{kd} \left[\frac{\sin (\alpha - \beta)}{\sin (\alpha + \beta)} \right]^2 \tag{3}$$

$$I_{rp} = I_{rd} \frac{n_1}{n_2} \left[\frac{2 n_1 \cos \alpha}{n_2 \cos \alpha + n_1 \cos \beta} \right]^2 = I_{rd} \frac{n_1}{n_2} \left[\frac{2 \cos \alpha \sin \beta}{\sin (\alpha + \beta) \cos (\alpha - \beta)} \right]^2 \tag{4}$$

$$I_{kp} = I_{kd} \frac{n_1}{n_2} \left[\frac{2 n_1 \cos \alpha}{n_1 \cos \alpha + n_2 \cos \beta} \right]^2 = I_{kd} \frac{n_1}{n_2} \left[\frac{2 \sin \beta \cos \alpha}{\sin (\alpha + \beta)} \right]^2, \tag{5}$$

where index r denotes a wave polarised in the plane of incidence, index k a wave polarised perpendicularly to the plane of incidence (parallel to the plane of the interface), index d the incident wave, index "o" the reflected wave and index "p" the passing wave. The angle α is the angle of incidence (equal to the angle of reflection) and β the angle of refraction to which Snell's relation $n_1 \sin \alpha = n_2 \sin \beta$ applies.

As we can see from the relation (2), for $\alpha + \beta = \pi/2$ rad, $\tan(\alpha + \beta) \to \infty$ and $I_{ro} = 0$. It means the EM wave with this polarisation is not reflected from the interface when this condition is met. From the condition $\alpha + \beta = \pi/2$ rad we get the relation

$$\frac{\sin \alpha}{\sin \beta} = \frac{\sin \alpha}{\sin \left(\frac{\pi}{2} - \alpha \right)} = \frac{\sin \alpha}{\cos \alpha} = \tan \alpha = \frac{n_2}{n_1} = \tan \alpha_B, \tag{6}$$

where α_B is Brewster's angle. When light falls on the interface at the angle of incidence α_B, only the 'k' wave polarised perpendicularly to the plane of incidence is reflected

$$I_{ko} = I_{kd} \left[\sin^2 \alpha_B - \cos^2 \alpha_B \right]^2 = I_{kd} \left(\frac{n_2^2 - n_1^2}{n_2^2 + n_1^2} \right)^2. \tag{7}$$

A wave polarised parallel to the plane of incidence is not reflected.

If the *non-polarised* EM wave hits the interface at the Brewster angle of incidence, the *reflected* light is polarised perpendicularly to the plane of incidence. If proper, this polarised wave can be suppressed by the polarising filter, for example, light reflections from shiny surfaces when shooting. Similarly, unpolarized sunlight is partially polarised when scattered in the atmosphere and so the blue brightness of the sky can be partially suppressed by a polarising filter to emphasise clouds.

The incident wave with r-polarisation is reflected with the same phase as incident one when $\alpha < \alpha_B$ and with the opposite phase (shifted by π rad) when $\alpha > \alpha_B$. The wave with k-polarisation is always reflected with the opposite phase.

Light affects charged particles in biological tissue. Apart from the thermal effect, EM waves also have a direct effect on cellular structures, it can support or suppress some cellular processes. The beneficial effects of light utilise *phototherapy*. They are particularly evident when using polarised light. Polarised light has stimulation and regenerative effects and a beneficial effect on the immune system overall. It also speeds up the healing of wounds after surgery. Lasers or special lamps are used for this reason. Polarised light can also be better focused. It utilises, for example, *laser scalpel* for very fine operations in ophthalmology and neurosurgery.

2.2 Waveform coherence

In an ideal harmonic wave, the phase is defined at each place and time by the wave function $e^{j\,(\mathbf{k}\cdot\mathbf{r}-\omega t)}$. In actual cases, however, this phase relationship is maintained only at a certain distance l along the propagation direction. This distance is called the *length of coherence*. The concept of *coherence* expresses the fact that it is still the same wave described by the given wave function, in which there is a constant phase difference between the two points of the space in which the wave propagates. In case of harmonic wave source (powered by harmonic voltage source)— transmitting antenna, piezoelectric transducer, acoustic loudspeaker, etc. the coherence lasts for the duration Δt_c of the duration of the harmonic excitation signal, it is coherence length $l_c = c\Delta t_c$. The coherence length is related to the spectral width Δf of the waveform $l_c \approx c/(\pi \Delta f)$.

In the case of light sources, the coherence length is limited by the fact that light is generated by individual atoms of a substance. Their photon emission is more or less coordinated. In the case of temperature sources (filament lamps, gas discharge lamps, LEDs), coordination restricts to a very small space, and thus the length of coherence is very small. The large coherence length is achieved with sources that use optical resonator—lasers. Approximate values of length of coherence are filament lamp ~ 1 μm, mercury lamp ~ 10 μm, LED ~ 100 μm, semiconductor laser about 1 m, gas He-Ne laser up to 100 m, fibre laser up to 100 km.

Apart from the coherence of the waves, we also define the coherence of wave sources. The wave sources are coherent in the case of their constant phase relationship. They are then the sources of coherent waves. At the same time, there is a constant phase difference between the waves generated by such sources. For example, two loudspeakers connected to a common AC generator, or two antennas connected to a common transmitter are such coherent sources. Another example of coherent sources is an ultrasonic transducer used in ultrasonography. It is assembled from many segments powered by the same electric generator. Waves generated by two independent sources cannot fulfil the conditions of coherence. The

slight difference in the frequency will earn a considerable phase mismatch. To ensure the coherence of two wave beams, they must be coupled to each other, so that the phase difference of the source signals is constant. Coherent sources can be created by letting a harmonic wave fall on two or more slits in a shielding grating. Single slits act as coherent sources for the space behind the grating (e.g., an optical grid). A similar effect is also achieved when the harmonic wave hits a suitable mirror assembly (e.g., a segmented reflector).

2.3 Plane, cylindrical, and spherical waves in lossless medium

For simplicity, consider waves in a lossless, homogeneous, isotropic, and linear medium. In the free space (outside the source), the wave equation has a form

$$\Delta \mathbf{u}(\mathbf{r},t) - \frac{1}{c^2} \frac{\partial^2 \mathbf{u}(\mathbf{r},t)}{\partial t^2} = 0, \tag{8}$$

where u is the wave quantity and c the velocity of the wave propagation, irrespective of the physical substance of the wave (mechanical or EM).

The solution is generally complicated and usually obtained by numerical methods. There are several software tools for modelling wave fields. The concrete solution of the equation depends on initial and boundary conditions. The initial conditions are determined by the time dependence of the source quantity, the boundary conditions result from the geometrical arrangement of both the source and the objects influencing the propagation of the wave.

In the following section, we will show some simple cases of wave propagation from sources with significant symmetry.

2.3.1 Plane wave

If the source is a sufficiently large and planar one that generates a wave with the same amplitude and phase over its entire surface, the wave quantities depend only on the coordinate z in the direction perpendicular to the plane of the source, $u(z,t)$. The Laplace operator Δ is reduced only to the second derivative according to the variable z

$$\frac{\partial^2 \mathbf{u}(z,t)}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \mathbf{u}(z,t)}{\partial t^2} = 0. \tag{9}$$

The solution of the equation has a shape of the wave function

$$\mathbf{u}(z,t) = \mathbf{f}\left(t \pm \frac{z}{c}\right), \tag{10}$$

as shown for mechanical or electromagnetic waves. A wave depending on only one spatial variable is called a *planar wave*. Its wavefronts are planes perpendicular to the direction of propagation. The quantity c is the velocity of the wave propagation, the sign "−" indicates a wave that propagates in the z-direction and "+" in the opposite direction, in other words the *forward wave* and the *backward wave*. The vector function $\mathbf{f}$ describes the waveform (impulse, harmonic, etc.).

In the case of the plane wave in a lossless medium, the shape of the wave propagating along the z-axis remains unchanged—it is not distorted. On this principle, for example, delay lines through which the time shift of the signal is realised.

A plane wave is only an idealised model of a real wave. In practical cases, the wave has a more complex spatial distribution, since the source surface is never infinitely large, which is a prerequisite for the formation of a plane wave.

2.3.2 Cylindrical wave

Simple cases of spatial two- and three-dimensional waves include cylindrical and spherical waves.

The cylindrical wave is generated by a very long rectilinear line source that generates a wave of equal amplitude and phase along its entire length. In this case, the wave quantities do not depend on the longitudinal coordinate z in the source direction and depend only on the transverse ones. In this case, it is proper to use cylindrical coordinates ρ (distance from the source axis) and φ (angle in a plane perpendicular to the source axis). Moreover, if the source is isotropic, it generates a wave equally in all directions φ, reducing the dependence of the wave quantities only on the coordinate ρ. The Laplace operator, and thus the wave equation, has then a shape

$$\left(\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho}\right)\mathbf{u}(\rho,t) - \frac{1}{c^2}\frac{\partial^2\mathbf{u}(\rho,t)}{\partial t^2} = 0. \tag{11}$$

The solution of this equation depends on the specific time dependence of the wave function given by the time dependence of the source quantity. Due to the harmonic time dependence of the excitation, the solution is the Bessel functions, which for longer distances converge towards the shape

$$\mathbf{u}(\rho,t) = \mathbf{u}_0\sqrt{\frac{\rho_0}{\rho}}\ \sin\left[k\left(\rho - ct\right) + \psi\right]. \tag{12}$$

The direction of the vector $\boldsymbol{u}_0$ determines the polarisation direction of the wave (transverse or longitudinal), ρ_0 is the radius of the source cylinder.

As you can see, the displacement $\boldsymbol{u}$ amplitude decreases with distance ρ from the axis with $\frac{1}{\sqrt{\rho}}$ function. Since the wave intensity I is proportional to the square of the wave quantity, it decreases disproportionately with the distance from the source, $I \sim \frac{1}{\rho}$. This conclusion can also be reached by the idea that the generated power P, when propagating the wave perpendicularly to the source, is spread over an increasing cylindrical surface $S = 2\pi\rho l$, and hence the intensity $I = P/S \sim 1/\rho$.

A cylindrical electromagnetic wave occurs, for example, around the line antenna or in the transverse direction within the coaxial line or other cylindrically symmetrical structures.

2.3.3 Spherical wave

We consider a point (spherical) isotropic source that generates waves to the surrounding homogeneous and isotropic medium. Due to the source point symmetry, the generated field has also the same point symmetry. Preferably, the spherical (spherical) coordinates r, ϑ, φ is used to describe the field. If the field is isotropic (independent of the coordinates ϑ and φ), the wave equation takes the form

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)\mathbf{u}(r,t) - \frac{1}{c^2}\frac{\partial^2\mathbf{u}(r,t)}{\partial t^2} = 0. \tag{13}$$

The solution to this equation has the form

$$\mathbf{u}(r,t) = \frac{\mathbf{u}_0}{r} f(r \pm ct) \text{ for } r > r_0, \tag{14}$$

where r_0 is the radius of the surface of the spherical source and $f(r - ct)$ is the non-attenuated component of the wave function, representing the wave propagating in the radial direction at the speed c. As a result, the wave amplitude decreases with a distance from the centre of symmetry proportional to $1/r$.

In the case of a harmonic wave, the solution is the wave function

$$\mathbf{u}(r,t) = \mathbf{u}_0 \ \frac{r_0}{r} \ e^{j(\omega t \pm \mathbf{k} \cdot \mathbf{r})}, \tag{15}$$

where $\mathbf{u}_0$ is the value of the wave quantity on the surface of the source ball with radius r_0.

Note: The correctness of solution (14) or (15) can be approved by direct substitution into the Eq. (13).

The wave intensity is proportional to the square of the wave quantity and therefore decreases with the square of the distance from the source $I \sim 1/r^2$. This dependence is also obtained from the energy considerations—if the source generates power P, it decomposes into a spherical surface with an area $S = 4\pi r^2$, and thus $I = P/S \sim 1/r^2$.

From a long-distance l (much greater than the source dimensions), each source appears to be a point source, and thus the field has the character of a point source field, it means the intensity decreases with distance according to the function $1/r^2$.

In the case of an anisotropic source, which emits different intensities in different directions (e.g., acoustic loudspeaker, directional radio transmitter), the dependence on distance $1/r^2$ remains valid, but there occurs dependence of the wave quantities on the coordinates ϑ and φ.

If we observe the wave at a long-distance r from the source and on a surface dimension of which is much less than r, the wavefronts appear to be planar (curvature of the surface is not observed in a small area). The wave thus appears a plane one in this confined space. The plane wave model, therefore, can be used in this case.

Example 1 Solar constant

The Sun emits an essential part of its energy in the form of EM radiation with a wide spectrum of frequencies. The total irradiated power gives us Stefan-Boltzmann's law for the radiation of a black body

$$P = I_0 S_0 = \sigma T_S^4 4\pi R_S^2,$$

where $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4} is the Stefan-Boltzmann's constant, $T_S \approx 5778$ K temperature of the Sun surface, $R_S = 6.96 \times 10^8$ m radius of the Sun.

Numerical results are $P \approx 3.8 \times 10^{26}$ W, $I_0 \approx 63$ MW m^{-2}.

There is radiation intensity at the distance of the Earth's orbit around the Sun

$$I = \frac{P}{S} = \frac{\sigma T_S^4 4\pi R_S^2}{4\pi d_{ZS}^2} = \sigma T_S^4 \left(\frac{R_S}{d_{ZS}}\right)^2,$$

where $d_{ZS} = 1.50 \times 10^{11}$ m is the distance Earth-Sun.

After substitution, we get $I = 1.36$ kW m^{-2}, which is the so-called solar constant.

For each square meter perpendicular to the direction of radiation, this power (partially reduced by the atmosphere), which heats the Earth's surface, evaporates

water and causes rain and water to supply rivers, drives atmospheric currents and thus wind, provides energy to living organisms, etc. The energy of these renewable sources, as well as direct energy, is used to generate electricity for human needs.

2.3.4 Waves propagation in waveguides

The wave may propagate in an open space or a structured transmission medium. A special case is represented by waveguides; spatial boundaries that guide the transmission of waves, and hence signal and energy of waves. The simplest case represents a homogeneous waveguide having the same properties along the entire direction of wave propagation. A classic case is a coaxial power line (coaxial cable), which is commonly used to transmit mainly a high-frequency signal. The coaxial cable has an inner conductor and an outer coaxial conductor, the space between the conductors being filled with a dielectric. Between the conductors, there is a radial electric field with an intensity E perpendicular to the conductor axis and a magnetic field with an intensity H characterised by circle induction lines in a plane perpendicular to the conductor axis. The Poynting vector thus has a wave propagation direction in the direction of the line axis. The advantage is that the EM wave remains enclosed within the cable in the dielectric so that it is not lost by dissipation into the environment, and there is no interference with the electromagnetic field of the line because the external coaxial line field is zero.

However, the waveguide does not need an inner conductor and the form of a tube with a rectangular or circular cross-section is sufficient. A waveguide performs its function if the wave is reflected from its walls without loss and remains within. It uses a full reflection of the waves at the interface of the internal and external media. The complete reflection of the EM wave occurs on a perfectly conductive wall, in which case it is the metal waveguide of the EM wave. Total reflection also occurs at the interface of dielectric media if the propagation speed c_1 in the internal medium is less than the propagation speed c_2 in the external medium, and the angle α of incidence on the interface is greater than the cut-off angle α_m of total reflection, it means $\alpha > \alpha_m = \arcsin(c_1/c_2)$, for example, optical fibres. Acoustic waveguides work in the same way.

Example 2 Conductive waveguide EM wave

A simple idea can be obtained from the EM waveguide according to **Figure 1**.

We consider two parallel planar conductive walls with a mutual distance a. Between the walls, we send out a plane EM wave at the angle α of incidence on the wall. Suppose that the wave is polarised parallel to the wall; the vector E is parallel to the wall and the vector H to the wall angle α. The plane wave propagation velocity is c. A wave vector k of magnitude $k = \omega/c$ and of an angle α with a normal line can be divided into the longitudinal component $k_{\parallel} = k \sin \alpha$ and the transverse

Figure 1.
Propagation of the EM wave in the waveguide.

one $k_\perp = k \cos \alpha$. Since at the conductive interface $E = 0$, there must be wave nodes on the walls; $k_\perp a = n\pi$, where $n = 1, 2, \ldots$ indicates the waveguide mode number. From there we have $ka \cos \alpha_n = n\pi$ and after substitution

$$\cos \alpha_n = n \frac{c}{2fa}, \text{ and therefore } v_\parallel = c \sin \alpha_n = c \sqrt{1 - \left(\frac{n}{2}\right)^2 \left(\frac{c}{fa}\right)^2}.$$

The condition of spreading of the n-th mode

$$f > n \frac{c}{2a} = f_{nm} \text{ and simultaneously } f < (n+1) \frac{c}{2a} = f_{(n+1)m}$$

(if $f > f_{(n+1)m}$ the mode n changes into $n + 1$).

The phase velocity of the waveform v_f in the direction of the waveguide can be seen on the upper wall. If the wavelength is λ, the distance between the points with the phase difference of 2π on the wall is $\lambda_f = \lambda / \sin \alpha$

$$\lambda_f = \frac{\lambda}{\sqrt{1 - \left(\frac{f_{nm}}{f}\right)^2}} \text{ and } v_f = \lambda_f f = \frac{c}{\sqrt{1 - \left(\frac{f_{nm}}{f}\right)^2}} > c. \quad (16)$$

The group velocity of the wave progression in the z-direction is $v_g = (\lambda \sin \alpha) f$. It means

$$\lambda_g = \lambda \sqrt{1 - \left(\frac{f_{nm}}{f}\right)^2} \text{ and } v_g = \lambda_g f = c \sqrt{1 - \left(\frac{f_{nm}}{f}\right)^2} < c. \quad (17)$$

The velocity v_g is the velocity at which the energy, and hence the signal, propagates in the waveguide.

If the waveguide cross-section is rectangular (a, b), the same considerations apply to both transverse directions. For an EM wave that is polarised perpendicularly to the longitudinal axis (TE mode), the boundary condition applies to the wave nodes on both pairs of opposing walls. In the case of the general orientation of vector E, the condition applies

$$v_\parallel = c \sin \alpha_n = c \sqrt{1 - \left(\frac{n}{2}\right)^2 \left(\frac{c}{fa}\right)^2 - \left(\frac{m}{2}\right)^2 \left(\frac{c}{fb}\right)^2}, \quad (18)$$

where n, m are integers indicating the wave mode TE_{nm}.

The condition for the transmitted wave frequency and the cut-off frequency is

$$f > \sqrt{n^2 \left(\frac{c}{2a}\right)^2 + m^2 \left(\frac{c}{2b}\right)^2} = f_{nm}. \quad (19)$$

The higher the mode numbers, the greater the geometric dispersion of the waves (as opposed to the material dispersion); it means the dependence of phase velocity on the frequency and thus distortion of the transmitted signal. Modes with low mode numbers TE_{10}, TE_{01}, TE_{11} are therefore used (both numbers n, m cannot be zero—the wave would not exist). The number of modes is limited by dimensions, or frequency using the condition for the propagation of the given mode $f > f_{nm}$.

Analogously, in the waveguide can be excited a wave that has a magnetic intensity H perpendicular to the longitudinal axis, it means transverse magnetic mode

TM. The boundary conditions are similar as for the TE mode. The waveguide can transmit TM_{nm} modes, where n, m = 1, 2, 3, ... The lowest is the TM_{11} mode.

Waveguide modes also propagate in waveguides with other cross-sections, most often circular. Similarly, like for rectangular, there exist TE_{nm} and TM_{nm} modes with similar transmission properties.

The same principle is used by optical fibres—waveguides for optical waves. Instead of a perfectly reflecting conductive wall, a total wave reflection from the dielectric interface with different refractive indices (fibre core n_1 and envelope $n_2 < n_1$) is used. Different filament modes also propagate in the fibre, like metallic waveguides.

If the cylindrical metal waveguide has an inner conductor (coaxial line), the EM wave has only a transverse character, that is, a TEM mode, which is characterised by signal propagation without distortion due to geometric dispersion. However, in the coaxial line dielectric, there are heat losses due to dielectric imperfection, and therefore the coaxial line is not used to transmit very high EM wave power when the dielectric becomes overheated. Coaxial lines are particularly advantageous for signal transmission.

2.4 Transmission of information utilising waves

The harmonic wave propagating in a medium does not transmit the information itself, but it is only an information carrier. To transmit information, the wave must be modulated by an appropriate information signal, Wyatt [1].

As an example, consider transmitting a data pulse. It can be decomposed into harmonic components using the Fourier integral. For a rectangular pulse with time length τ and carrier angular frequency ω_0 resp. frequency $f_0 = \omega_0/(2\pi)$,

$$\mathbf{u}_0(t) = U_0 e^{j\omega_0 t} \text{ for } -\frac{\tau}{2} \leq t \leq \frac{\tau}{2} \tag{20}$$

is a complex Fourier image of the pulse function

$$\mathbf{A}(\omega) = U_0 \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{j\omega_0 t} e^{-j\omega t} dt = \frac{2U_0}{\omega - \omega_0} \sin \frac{(\omega - \omega_0)\tau}{2} = U_0 \tau \frac{\sin \frac{(\omega - \omega_0)\tau}{2}}{\frac{(\omega - \omega_0)\tau}{2}}. \tag{21}$$

This is a $\sin \alpha/\alpha$ function with a main frequency band $\Delta f = f - f_0 = 1/\tau$. To transmit a pulse with a length τ, this frequency band Δf must be transferred.

The propagation of individual harmonic components can be expressed by the wave function

$$\mathbf{u}_\omega(z,t) = \mathbf{A}(\omega) e^{j(\omega t - kz)}. \tag{22}$$

If the pulse is modulated upon a base wave with an angular frequency ω_0 and a wavenumber k_0, the current angular frequency, and the wave number can be expressed as $\omega = \omega_0 + \Delta\omega$ and $k = k_0 + \Delta k$. Thus, we get relation for the individual harmonic components

$$\mathbf{u}_\omega(z,t) = \left(U_0 \tau \frac{\sin \frac{\Delta\omega\tau}{2}}{\frac{\Delta\omega\tau}{2}} e^{j(\Delta\omega t - \Delta k z)} \right) e^{j(\omega_0 t - k_0 z)}. \tag{23}$$

The carrier wave with angular frequency ω_0 propagates with phase velocity

$$c_f = \frac{\omega_0}{k_0}. \tag{24}$$

The modulation envelope is represented by the parenthesis, and thus the modulated signal propagates as a wave with parameters $\Delta\omega$ and Δk at with the velocity $c_{\text{sign}} = \Delta\omega/\Delta k$. Thus, we define a group velocity c_g for the close vicinity $d\omega$ of the basic angular frequency ω_0 as

$$c_g = \left(\frac{d\omega}{dk}\right)_{\omega_0}, \tag{25}$$

which characterises the rate of signal transmission in a medium.

If we have $c_f \neq c_g$, there occurs a signal distortion. In the mentioned case of propagation of the pulse signal, the pulse height decreases, and its width increases along the propagation path.

Note 1: In addition to the material dispersion mentioned above, a geometric dispersion, associated with the geometry of the waveguide, affects information transmission in waveguides.

Note 2: If the dispersion relationship (25) is linear, i.e., c_f is a constant independent on the angular frequency, then $c_g = c_f$ and the medium is non-dispersive. To minimise signal distortion, the frequency bands, in which the dispersion is minimal, are used for transmission, e.g., in optical fibres.

2.5 Wave modulation

In the previous paragraph, we showed that waves can transmit information. For example, the sound of our voice is a superposition of single waves with frequencies from tens of Hz to several kHz. It is similar in case of the tones of musical instruments or EM waves generated by lightning during a storm.

Sometimes we need to utilise proper conditions for the propagation of waves in some range of frequency or to transmit several parallel information channels by the wave. In such a case, we use the so-called carrier wave of the required frequency ω_0 and modulate the transmitted information onto this wave. The modulation is most often amplitude or frequency one. Phase modulation is used only in special cases, for example, in MRI signal analysis. The modulation utilises control of the respective wave parameter (amplitude, frequency, or phase) by the information signal.

Suppose the signal given by its time dependence

$$\mathbf{s}(t) = \sum_{n=-N}^{N} \mathbf{S}_n e^{jn\Omega t}, \tag{26}$$

where $\mathbf{S}_{mn}$ are phasors and $-N\Omega, \ldots -\Omega, 0, \Omega, 2\Omega, \ldots, N\Omega$ angular frequencies of spectral components of the signal. The frequency range is $\Delta\Omega = 2N\Omega$.

2.5.1 Amplitude modulation: AM

The wave with amplitude modulation (AM) represents the relationship

$$\mathbf{u}(x,t) = [U_0 + \mathbf{s}(t)]\, e^{j(\omega_0 t - k_0 x)} = U_0 e^{j(\omega_0 t - k_0 x)} + \sum_{-N}^{N} \mathbf{S}_n e^{j[(\omega_0 + n\Omega)t - k_n x]}, \tag{27}$$

where it applies to phasors $\mathbf{S}_n = \mathbf{S}_{-n}^*$ of harmonic components. From here, we can see that the modulated wave contains frequency components from $(\omega_0 - N\Omega)$

to $(\omega_0 + N\Omega)$. The frequency bandwidth is thus twice the cut-off frequency $N\Omega$ of the modulation signal.

If two independent information channels are to be transmitted, their carrier frequencies ω_{01} and ω_{02} must be separated from each other by more than $N_1\Omega_1 + N_2\Omega_2$, so that the signals do not interfere with one another.

The wave amplitude changes with modulation. Since the frequency component with the carrier frequency ω_0 contains no information, it is often suppressed to reduce the energy of the transmitted (and generated) waves.

AM is used to transmit radio broadcasts on long, medium, and short waves. The officially agreed bandwidth of the radio station is 9 kHz so that the maximum frequency of the modulation signal spectrum is 4.5 kHz. This is enough for clear speech transmission but not enough for high-quality music transmission.

A special case is pulse amplitude modulation. It is used mainly for digital signal transmission. Pulsed AM is used, for example, when sampling an analogue signal. The bandwidth of the EM wave required to transmit pulses with a length τ is $\Delta f \sim 1/\tau$. For 1 ns pulse transmission, a channel width of the order of 1 GHz is required. Since $\Delta f \ll f_0$ is required, microwaves are used for high-frequency data transmission, for example, transmission by satellite with frequencies of the order of 100 GHz, or terrestrial fibre-optic links with IR wavelength of about 1.6 μm ($\sim$150 THz).

The advantage of amplitude modulation is the relative simplicity of modulators and demodulators. The main disadvantage is the liability to interferences and a higher signal-to-noise ratio.

2.5.2 Frequency modulation: FM

In VHF and UHF radio and TV channels with a carrier frequency of about 100 MHz, there is enough space for a wider band (for radio 40 kHz, for TV 7–8 MHz) that provides high-quality audio and video transmission. The fast transfer is also required for the transmission of large data files in telemedicine (e.g., CT and MRI images, online video transfer of the medical operation, etc.).

Concerning transmission better security (suppression of signal error and interference) and reduction of the signal to noise ratio, frequency modulation of the waves is used. Frequency modulation consists of the control of the frequency of the wave by the modulation signal. The amplitude and thus the power remains constant (as opposed to AM when the power fluctuates and is thus more susceptible to interference).

The frequency modulated wave can be described by function

$$\mathbf{u}(x,t) = U_0 e^{j\,[(\omega_0 + m\,u(t))\,t - k_\omega x]} = U_0 \ \exp \ j\left[\left(\omega_0 + m \sum_{-N}^{N} \mathbf{S}_n e^{j\,n\,\Omega t}\right) - k_\omega x\right],$$

(28)

which can be expressed as a series of Bessel functions, where the spectral components of the wave are the same as in the case of AM.

Pulse FM used in data transmission, consists of a change of frequency at the time pulse duration. Since this change is limited in time, the same condition as AM applies to bandwidth $\Delta f \sim 1/\tau$.

2.6 Material dispersion of the wave

The wave velocity depends on the parameters of the medium, depending on the frequency of the wave. This phenomenon is called material dispersion. As shown in

the previous paragraph, $c_f \neq c_g$ applies to the dispersive medium, that is, the velocity of information propagation is less than the phase velocity of the wave. The material dispersion is most often manifested in light waves. The dispersion prism is used in spectrometers to decompose white light into individual monochromatic components. The dispersion of light in water causes the formation of a rainbow due to the internal reflection of the light in the raindrops. The material dispersion also contributes to the distortion of infrared waves transmitted by optical fibres.

The material dispersion appears significantly at a large relative bandwidth $\Delta f/f_0$ of the transmitted wave. Therefore, for a given bandwidth Δf, the maximum carrier frequency f_0 is chosen.

The material dispersion in the case of the transmission of (pulse) data information in the optical fibres is reduced by selecting a frequency domain with minimal material dispersion, for example, in glass fibres around $\lambda \sim 1.6\ \mu m$.

3. Ray optics

For the wave at a great distance from the source, that is, at the distance much greater than the dimensions of the source (so-called far Fraunhofer region), the ray representation of the wave can be used.

A ray is a line along which a wave passes. In addition to the rays, there are defined wavefronts. They represent the surfaces of the constant phase, in the case of a harmonic wave, and the surfaces that the wave reaches over some time, in the case of pulse or other waves. The wavefronts and rays are orthogonal to each other. At any point, the beams are perpendicular to the wavefronts.

Wave propagation obeys two basic principles, the Fermat principle, and the Huygens principle, which follow from the basic wave equations.

3.1 Fermat's principle

Fermat's principle says that the wave propagates from point A to point B along a spatial curve at which the time t_{AB} required to overcome the path l_{AB} is minimal

$$t_{AB} = \int_{l_{AB}} \frac{dl}{c} - \text{minimal}, \quad \text{or} \quad \delta\left(\int_{l_{AB}} \frac{dl}{c}\right) = 0, \tag{29}$$

where the symbol δ represents a variation of the respective functional (expression in parentheses).

For EM waves, phase velocity c can be expressed using refractive index n

$$t_{AB} = \frac{1}{c_0} \int_{l_{AB}} n\, dl - \text{minimal}, \quad \text{or} \quad \delta \int_{l_{AB}} n\, dl = \delta l_{opt} = 0,$$

where $l_{opt} = \int_{l_{AB}} n\, dl$ is *optical path length*.

Fermat's principle then says that EM waves propagate from point A to point B along the shortest optical path.

Example 3 Reflection and refraction of waves on the plane boundary of two homogeneous media

It follows from the Fermat's principle that in a homogeneous medium, where c is constant, the shortest wave path is the abscissa. Hence the notion that the waves
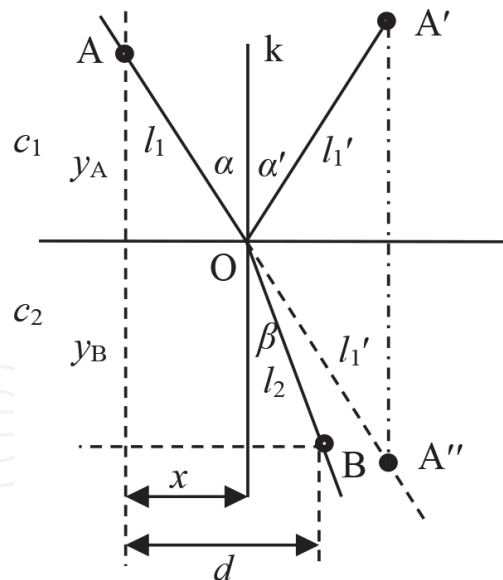
Figure 2.
Reflection and transition of wave through an interface.

propagate straightforwardly. However, this applies only to a homogeneous medium. In a non-homogeneous medium, the rays are curved.

At **Figure 2** there is a ray going out from point A and reflected to point A′ and refracted to point B.

The curvature of the optical beam in a non-homogeneous medium utilises, e.g., GRIN lens, see Section 6.2.2.

The ray OA′ has the same length as its mirror image OA″. As the ray travels in the same medium, the shortest wave path is equal to the shortest geometric path whose length corresponds to the length of the abscissa between points A and A″. The point of reflection O lies on the line AA″so that the angles $\alpha' = \alpha$. This is the law of reflection, which states that the ray incident, reflected, and the line perpendicular to the interface lie at one plane perpendicular to the interface and the angle of incidence α is equal to the angle of reflection α'.

Consider now the transition to the second environment from point A to point B. We denote the distances x and d, as shown. We are looking for a minimum of expression

$$t_{AB} = \frac{l_1}{c_1} + \frac{l_2}{c_2}, \text{where } l_1 = \sqrt{x^2 + y_A^2}, \text{and } l_2 = \sqrt{(d-x)^2 + y_B^2}. \quad (30)$$

We express the relation for t_{AB} as a function of the variable x, which indicates the position of the point O

$$t_{AB} = \frac{\sqrt{x^2 + y_A^2}}{c_1} + \frac{\sqrt{(d-x)^2 + y_B^2}}{c_2}.$$

We obtain the minimum t_{AB} using the zero value of the derivative according to the variable x:

$$\frac{dt_{AB}}{dx} = \frac{2x}{2c_1\sqrt{x^2 + y_A^2}} - \frac{2(d-x)}{2c_2\sqrt{(d-x)^2 + y_B^2}} = \frac{\sin\alpha}{c_1} - \frac{\sin\beta}{c_2} = 0,$$

from where we get Snell's refraction law

$$\frac{\sin \beta}{\sin \alpha} = \frac{c_2}{c_1}, \text{ or for light } \frac{\sin \beta}{\sin \alpha} = \frac{n_1}{n_2}. \tag{31}$$

Example 4 Total reflection

We express the relation for the angles of incidence and refraction in the form

$$\sin \beta = \frac{c_2}{c_1} \ \sin \alpha \ \leq \ 1.$$

From this inequality, we get the condition of the refraction of the beam

$$\sin \alpha \ \leq \ \frac{c_1}{c_2}, \text{ or } \ \sin \alpha \ \leq \ \frac{n_2}{n_1}.$$

For $c_1 > c_2$, or $n_2 > n_1$, there is no restriction on the angle α of incidence. If $c_1 < c_2$, or $n_2 < n_1$ there exists a limit angle α_m given by the relation

$$\sin \alpha_m = \frac{c_1}{c_2}, \text{ or } \ \sin \alpha_m = \frac{n_2}{n_1}, \tag{32}$$

where the refraction of the ray occurs when the condition $\alpha < \alpha_m$ is fulfilled.

For the angles of incidence $\alpha > \alpha_m$, the refraction cannot occur, and therefore the incident wave is only reflected and does not pass into the second medium—this phenomenon calls total reflection.

The total reflection phenomenon is used, for example, in optical fibres, or generally in waveguides. If the refractive index n of the optical fibre is greater than the refractive index n_0 of the surrounding medium, $n > n_0$, and the light enters the fibre at an angle $\alpha > \alpha_m$ relative to the normal to the cylindrical surface of the fibre, it spreads inside the fibre without losing energy by irradiation into the surroundings.

In biomedical applications, optical fibres are used in laser lithotripsy, laser scalpel, endoscopy, and the like.

Example 5 Optical fibre

Many applications use waves led in optical fibre. These are mainly signal transmission (e.g., optical computer networks), illumination of inaccessible areas (e.g., an internal organ in the body using an endoscope), the transmission of radiation power (e.g., in laser lithotripsy). The essence of the optical fibre transmission is that the wave is completely reflected from the fibre walls and cannot escape from the fibre. This creates an optical waveguide. The basic condition is that the wave must strike the fibre wall at angle of incidence $\alpha > \alpha_m$, greater than the total reflection one. The transition of the wave along the fibre is in **Figure 3**.

The beam of parallel rays is centred by the lens S on the inlet surface of the cylindrical filament, the angle of incidence on the filament front being $\beta < \beta_0$. From
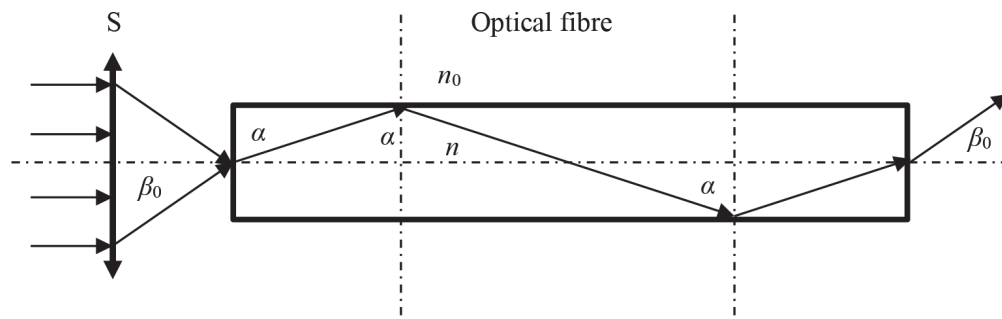
Figure 3.
Transition of light along with an optical fibre.

the surrounding medium with refractive index n_0 the light refracts into the fibre environment with refractive index $n > n_0$ at the refracting angle $\pi/2 - \alpha$, for which the refractive law $\sin(\pi/2 - \alpha) = \cos \alpha = n_0/n \sin \beta$. The angle α is the angle of incidence on the fibre wall and must be $\alpha > \alpha_m$, where $\sin \alpha_m = n_0/n$. Thus, we obtain a condition for the cut-off angle β_m: $\cos \alpha_m = \frac{n_0}{n} \sin \beta_m$, from where

$$\sin \beta_m = \frac{n}{n_0} \cos \alpha_m = \frac{n}{n_0} \sqrt{1 - \sin^2 \alpha_m} = \sqrt{\left(\frac{n}{n_0}\right)^2 - 1}. \qquad (33)$$

If $\beta_0 < \beta_m$, the entire beam enters the fibre and travels through the filament along its axis. At the other end, it gets out at the same angle β.

The wave proceeds along the fibre axis at a phase velocity $v_f = \frac{c}{\sin \alpha}$ and group velocity $v_g = c \sin \alpha$, which depends on the angle α, or β. As a result, rays incident at different angles interferes with each other, thereby suppressing some directions and strengthening some others. Thus, only some modes are effectively propagated in the fibre, like the metallic waveguide, see Example 2. The interference of modes distorts the transmitted signal. A more detailed analysis suggests that the number of modes depends on the ratio of fibre diameter to wavelength. The single-mode fibre that is mostly used to transmit information (e.g., data transmission over a computer network) has a diameter of $(6–7)\,\lambda$, that is, for IR radiation with $\lambda = 1.3\ \mu m$, the core diameter of the filament is $(8.5–9.5)\ \mu m$. In the case of energy transfer only, thicker multimode fibres with a diameter of up to 0.1 mm or more can be used, depending on the application.

The optical cable consists of many fibres. In endoscopes, an optical cable is used to transfer the image so that the image produced by the lens of the objective is projected onto a bundle of optical fibres, each transmitting 1 pixel of the image. The cable consisting of 100,000 fibres with a diameter of 5 μm (for a light wavelength of about 500 nm) has a diameter of approximately 1.5 mm. After leaving the cable, the light strikes the detector (CCD chip).

3.2 Huygens and Huygens-Fresnel principle

The propagation of waves in space is described by the *Huygens principle*. Huygens had the following idea:

If the wave reaches a certain wavefront, each point of the wavefront becomes a point source for the next part of the space. The following wavefront is then the contour surface of the wavefronts of these elementary spherical waves.

In this way, we can construct one wavefront after another and gradually depict the whole wave field. The procedure is shown in **Figure 4**. In the figure, three wavefronts correspond to times t_0, $t_0 + \Delta t$, and $t_0 + 2\Delta t$. Between the wavefronts are drawn elementary spherical waves originated in individual points of the previous wavefront.

In the figure, the elementary wavefronts do not have the same radius, which relates to different velocities of wave propagation in different places of the non-homogeneous medium. The rays are orthogonal to the wavefronts. We can see that in a non-homogeneous medium, the rays decline towards parts of the space with the lower velocity of propagation, lower right part in the figure.

Fresnel extended Huygens' idea to add a quantitative dimension. The wave quantity at a given point P of the following wave is a superposition of individual waves of elementary sources in the points M of the previous wavefront S, for illustration, see **Figure 5**.
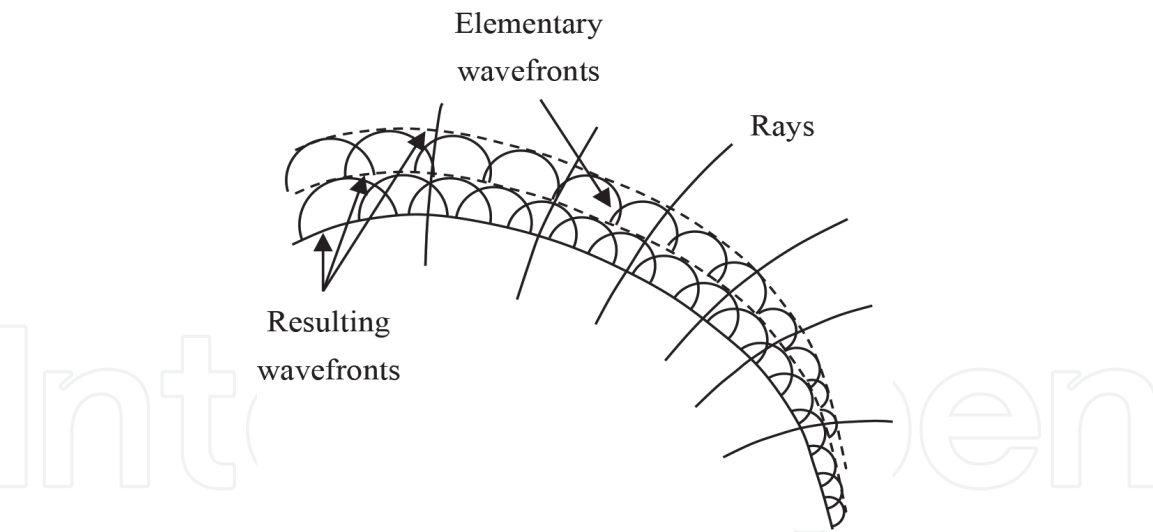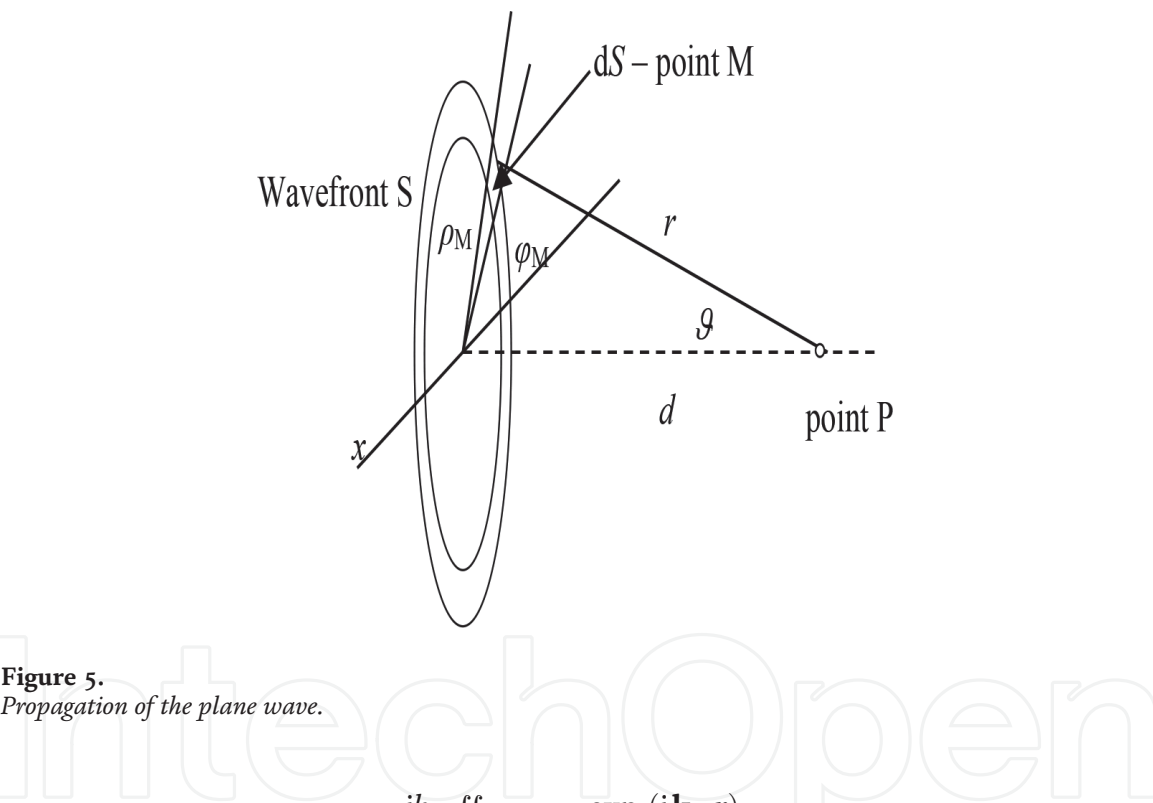
Figure 4.
Illustration of the Huygens principle.

Figure 5.
Propagation of the plane wave.

$$\mathbf{u}(\mathrm{P}) = -\ \frac{jk}{2\pi}\ \iint\limits_{S} \mathbf{u}_0(\mathrm{M})\ \frac{\exp\,(j\mathbf{k}\cdot\mathbf{r})}{r}\ \cos\vartheta\ dS, \qquad (34)$$

where $\boldsymbol{k}$ is the wave vector, $\boldsymbol{r} = \boldsymbol{r}_\mathrm{P} - \boldsymbol{r}_\mathrm{M}$ the position vector of the point P originating from the point M and ϑ the angle between the vector $\boldsymbol{r}$ and the normal to the wavefront at point M. The constant before the integral was formed by matching this approach to the proper solution of the wave equation, see the following Example 6.

Example 6 Plane wave and the Huygens-Fresnel (H-F) principle

As an illustration of the application of the H-F principle, let us give an example whose result we know. Consider a plane wavefront of a plane wave as a source (points M) and examine the further propagation of the wave—to the point P at a distance d from the wavefront, **Figure 5**. We divide the wavefront into elementary annular rings with radius ρ_M and width $d\rho_\mathrm{M}$, and then into angular segments $d\varphi_\mathrm{M}$. Source elementary area $dS = d\rho_\mathrm{M}(\rho_\mathrm{M}d\varphi_\mathrm{M})$, and $\cos\vartheta = d/r$.

The H-F integral then has an expression

$$\mathbf{u}(P) = -\frac{jk}{2\pi}\,\mathbf{u}_0 \int_0^{2\pi}\!\!\int_0^\infty \frac{\exp{(jkr)}}{r}\,\frac{d}{r}\,\rho_M \mathrm{d}\rho_M\;\mathrm{d}\varphi_M = -jk\,\mathbf{u}_0\,d \int_0^\infty \frac{\exp{(jkr)}}{r^2}\rho_M\,\mathrm{d}\rho_M.$$

Since $r^2 = d^2 + \rho_M^2$, by differentiation, we get $r\mathrm{d}r = \rho_M\mathrm{d}\rho_M$ and adjust the integral to form

$$\mathbf{u}(P) = -jk\,\mathbf{u}_0\,d \int_d^\infty \frac{\exp{(jkr)}}{r}\,\mathrm{d}r = -jk\,\mathbf{u}_0\,d \int_{kd}^\infty \left(\frac{\cos kr}{kr} + j\,\frac{\sin kr}{kr}\right)\mathrm{d}(kr). \quad (35)$$

If $kd > 2\pi$, that is, in the distance $d > \lambda$, we can use the approximation of functions *integral sine* and *integral cosine*

$$\int_x^\infty \frac{\sin x}{x}\mathrm{d}x \approx \frac{\cos x}{x} \quad \text{and} \quad \int_x^\infty \frac{\cos x}{x}\,\mathrm{d}x \approx -\frac{\sin x}{x}$$

and we adjust the H-F integral to this shape

$$\mathbf{u}(P) = -jk\,\mathbf{u}_0\,d\,\left(-\frac{\sin kd}{kd} + j\frac{\cos kd}{kd}\right) = \mathbf{u}_0\,e^{jkd}, \quad (36)$$

which is a complex function of the plane shifted by d.

From the initial planar wavefront, we get the following parallel planar wavefront, which corresponds to the propagation of the plane wave in space.

This example has shown that (34) correctly describes the wave propagation and the construction of the next wavefront, which is more than λ away from the previous one.

The Huygens-Fresnel principle will be used in the next section to explain the diffraction phenomenon.

4. Wave interference

If several waves propagate in a linear medium, they compose—*linearly interfere* with each other. The resulting waveform is the vector sum of the waveforms of the individual waves $\mathbf{u} = \sum_{i=1}^{n}\mathbf{u}_i$, where n is the number of interfering waves. We obtain significant results if the individual waves are coherent with each other, that is, have a defined phase relation expressed by their wave functions. Interference may result in amplification of the wave (*constructive interference*) or mutual suppression (*destructive interference*).

4.1 Constructive and destructive interference

Let us consider two plane harmonic waves with the same polarisation, and the same frequency, which propagates in the same direction and the phase difference φ between them. The resulting wave is

$$\mathbf{u}(z,t) = \mathbf{u}_1(z,t) + \mathbf{u}_2(z,t) = \mathbf{u}_1\,e^{j(\omega t - kz)} + \mathbf{u}_2\,e^{j(\omega t - kz + \varphi)}. \quad (37)$$

We set the relationship to shape

$$\mathbf{u} = \mathbf{u}_1 \; e^{j(\omega t - kz)} \left[1 + \frac{u_2}{u_1} \; e^{j\varphi} \right] = \mathbf{u}_1 \; e^{j(\omega t - kz)} \left[\left(1 + \frac{u_2}{u_1} \; \cos\varphi \right) + j \; \sin\varphi \right]$$

The resulting wave has the wave properties of the original waves, the amplitude

$$u = u_1 \sqrt{ \left(1 + \frac{u_2}{u_1} \; \cos\varphi \right)^2 + \sin^2\varphi } \tag{38}$$

and phase shift for the wave $\boldsymbol{u}_1$

$$\psi = \arctan \frac{\sin\varphi}{1 + \frac{u_2}{u_1} \; \cos\varphi}. \tag{39}$$

If the waves have the same amplitude $u_1 = u_2$, we get the amplitude and phase shift of the resulting wave

$$u = u_1 \sqrt{ (1 + \cos\varphi)^2 + \sin^2\varphi } = u_1 \sqrt{ 2(1 + \cos\varphi) } = 2u_1 \left| \cos\frac{\varphi}{2} \right|, \tag{40}$$

and

$$\psi = \arctan \frac{\sin\varphi}{1 + \cos\varphi}.$$

Under constructive interference $\varphi = 2n\pi$, it means, the shift by $\Delta z = n\lambda$, where $n = 0, 1, 2, \ldots$ (even multiple π or half-wave). Then we obtain

$$u = 2u_1, \text{ and } \psi = 0.$$

The resulting wave has twice the amplitude and the same phase as the original waves.

Under destructive interference $\varphi = (2n + 1)\pi$, it means, $\Delta z = (2n + 1)\lambda/2$ (odd multiple π or half-wave). Then we obtain

$$u = 0, \text{ and } \psi \text{—has no sense.}$$

In the first case, the wave is amplified to double the amplitude, in the second one the waves are mutually suppressed.

Example 7 Reflection of waves from thin film

A typical interference phenomenon is the reflection of the wave from a thin layer of substance. Consider the perpendicular impact of the wave from the medium with propagation velocity c_1 on the layer with the propagation velocity c_2. Let us suppose wave impedances $Z_1 < Z_2 < Z_3$, for example, light falling from the air onto the reflective layer on the glasses.

The incident wave is reflected from the first interface backward with the opposite phase, that is, $\Delta\varphi_1 = \pi$ rad and the layer penetrates without phase change. From the second interface, the wave is reflected again with a phase change $\Delta\varphi_2 = \pi$ rad. After passing through the layer forward and backward, a phase shift $\Delta\varphi_3 = k2d$ occurs at the wave reflected from the second interface, where d is the layer thickness, and $k = 2\pi/\lambda = (2\pi/\lambda_1)(c_1/c_2)$. This wave then passes through the first interface out of the layer without changing the phase and interferes with the first reflected wave.

The interference is constructive when $\Delta\varphi = \Delta\varphi_1 - (\Delta\varphi_2 + \Delta\varphi_3) = 2n\pi$, from where we get the condition for layer thickness

$$d_k = n \ \frac{\lambda_1}{2} \frac{c_2}{c_1} = n \ \frac{\lambda_2}{2}. \tag{41}$$

If the layer has a thickness d_k, it reflects the maximum extent, and it means that wave penetrates the layer and therefore enters the third medium to a minimum extent. This principle is used in the production of reflective layers, for example, UV interference filters on lenses of glasses or optical instruments.

As we can see from the result, constructive reflection depends on the wavelength. If white light falls on the layer, only a certain colour is reflected. If the layer thickness changes, the colour of the reflected light changes as well. This is used, for example, for measuring the thickness of thin films. If there is a thin oil layer on a surface of the water, which does not have the same thickness everywhere, different colour patterns are formed because of interference reflection. We can see it on water puddles. The effect is most noticeable when $n = 1$.

Destructive interference occurs when $\Delta\varphi = (2n + 1)\pi$, from where we get layer thickness

$$d_d = (2n + 1)\frac{\lambda_1}{4}\frac{c_2}{c_1} = (2n + 1)\frac{\lambda_2}{4}. \tag{42}$$

A layer of a thickness dd reflects minimally, and thus maximally transmits. This is used to form anti-reflective layers, which we use to get the maximum of the incident light. They are used, for example, for glasses to increase the brightness of the observed objects, for telescope lenses, etc.

As we can see, the layer is anti-reflective only for certain wavelengths, but for other wavelengths, it should be reflective. The thickness of the layer can be set to wavelength according to our demand to support or suppress the light transmission. Since the wavelength of yellow light is about twice the wavelength of ultraviolet one, the layer can be reflective for the UV light and at the same time transparent for a yellow one.

4.2 Wave beats

In another case, consider two harmonic waves with the same direction of polarisation and the same amplitude, which propagate in the same direction and whose angular frequencies differ by a small difference $\Delta\omega = \omega_1 - \omega_2$. By folding both waves, we get a wave field

$$\mathbf{u}(z,t) = u \ \left\{ e^{j \ \left[\left(\omega_0 + \frac{\Delta\omega}{2} \right)t - \left(k_0 + \frac{\Delta k}{2} \right) z \right]} + e^{j \ \left[\left(\omega_0 - \frac{\Delta\omega}{2} \right)t - \left(k_0 - \frac{\Delta k}{2} \right) z \right]} \right\}.$$

We adjust the resulting wave function to shape

$$\mathbf{u}(z,t) = \left[2 \ u \cos \ \left(\frac{\Delta\omega}{2} t - \frac{\Delta k}{2} z \right) \right] \ e^{j \ (\omega_0 t - k_0 z)}. \tag{43}$$

The composition of waves results in a wave that represents a carrier wave with a mean angular frequency of ω_0 and a corresponding wavenumber k_0, whose phase velocity $c_f = \omega_0/k_0$, with the modulation envelope described in square brackets. A modulation envelope is a wave with a frequency of $\Delta\omega/2$ and a wavenumber of $\Delta k/2$ that proceeds with a group velocity of $c_g = \Delta\omega/\Delta k$.

Wave intensity is

$$I \propto |\mathbf{u}(x,t)|^2 = 4\, u^2 \cos^2\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x\right) = 2\, u^2\, [1 + \cos\,(\Delta\omega t - \Delta kx)].$$

(44)

The wave intensity I represents a wave with an angular frequency $\Delta\omega$ that travels in space at c_g. At a given point in the wave field, the intensity varies with an angular frequency $\Delta\omega = |\omega_1 - \omega_2|$ between zero and a maximum value proportional to u^2. This variation of the wave intensity is called *wave beats*.

Note: Beats are used, e.g., when tuning musical instruments. If the tone frequency approaches the tuning fork frequency, the beats gradually disappear. Similarly, in the "octave" music interval where $\omega_2 = 2\omega_1$, the angular frequency of the beats is $\Delta\omega = \omega_1$ and thus beats fuse with a wave of lower frequency. Wave beats are therefore not observed. This serves to check the correct tuning of the octave interval.

Example 8 Stereo sound reproduction

The preferred way of listening to recorded music is stereophonic reproduction. The record is scanned by two microphones located at an appropriate distance. The recording is then played back from two parallel speakers whose distance is d.

The listener P moves along a line parallel to the speaker apertures at a distance r, **Figure 6**. The distance of P from the centre is indicated by x.

Suppose that the electric current through the loudspeakers has the same frequency and phase. Both speakers thus generate coherent waves with the same amplitude and phase. The waves from both speakers are composed at the listener position P

$$\mathbf{u}(P) = -\frac{jk}{2\pi}u_0 S\left(\frac{e^{-jkr_1}}{r_1}\cos\vartheta_1 + \frac{e^{-jkr_2}}{r_2}\cos\vartheta_2\right),$$

(45)

where $\cos\vartheta_{1,2} = r/r_{1,2}$.

The sound intensity in the point P of the listener is

$$I(P) = I_0\frac{S^2 r^2}{\lambda^2 r_1^4}\left[1 + \left(\frac{r_1}{r_2}\right)^4 + 2\,\frac{r_1^2}{r_2^2}\,\cos k\,(r_2 - r_1)\right],$$

(46)

where $r_{1,2} = \sqrt{r^2 + \left(x \mp \frac{d}{2}\right)^2}$.

The dependence of the relative sound intensity in front of a pair of speakers is shown in the graph at **Figure 7**. For the clearness of the illustration, the plots for two different frequencies are drawn. We can see that in some places the sound of some frequency is not heard. Thus, the spectral composition of the music you listen to depends significantly on the position P of the listener. It results in the change of
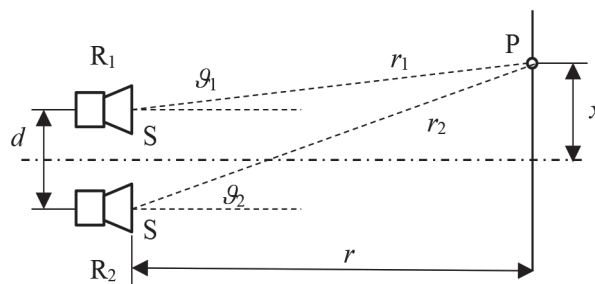
Figure 6.
Interference of sound of two loudspeakers.
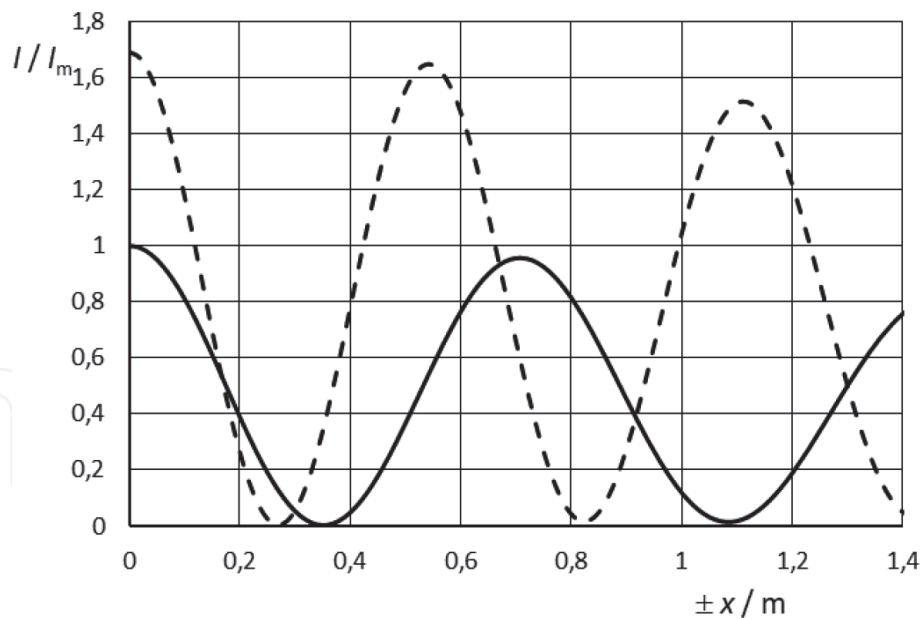
Figure 7.
Result of two speakers' interference for f₁ = 1.0 kHz (solid line) and 1.3 kHz (dashed line) relative to the maximum intensity.

tone colour, and thus the whole harmony of the music. The only place where we do not hear this distortion is on the axis of the system $x = 0$. It is advisable, for a quality experience of recorded music, to sit in the axis of a pair of speakers.

Suppression of these negative phenomena in two-source stereo is solved by a set of more speakers, for example, a quadraphonic system.

4.3 Standing waves and resonators

Consider two harmonic plane waves with the same polarisation and frequency that propagate in opposite directions. The resulting wave is the superposition of both waves

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 = u_1 e^{j(\omega t - kz)} + u_2 e^{j(\omega t + kz)}.$$

The term can be broken down into two parts

$$\mathbf{u} = (u_1 - u_2) \, e^{j(\omega t - kz)} + u_2 \left[e^{j(\omega t - kz)} + e^{j(\omega t + kz)} \right]$$
$$= (u_1 - u_2) \, e^{j(\omega t - kz)} + 2u_2 \, \cos(kz) \, e^{j\omega t}. \tag{47}$$

The resulting wave has two different components. The first one is a moving wave that propagates in the z-direction and has an amplitude equal to the difference of amplitudes of both waves. The second component represents oscillations with frequency ω and amplitude $u(z) = 2u_2 \cos(kz)$ dependent on the coordinate z. It is a *standing wave*.

If the amplitudes of the two waves propagating against each other are the same, a pure standing wave occurs

$$\mathbf{u} = 2u_1 \cos(kz) \, e^{j\omega t}. \tag{48}$$

Points where $kz_u = n\pi$, resp. $z_u = n(\lambda/2)$, where $n = 0, 1, 2, \dots$ is $u_u = 0$. The amplitude of the oscillations is zero and these points are called *nodes of the standing wave*.

Conversely, points where $kz_k = (2n + 1)\pi/2$, resp. $z_k = (2n + 1)\lambda/4$, the maximum amplitude $u_k = 2u_1$. These points are called *antinodes of the standing wave*.

Standing wave or partially standing wave arises when the wave is reflected from the interface of two media by the composition of both direct and reflected waves.

The relation (22) shows that the maximum displacement is $u_1 + u_2$ and the minimum displacement $|u_1 - u_2|$.

The quantity

$$SWR = \frac{u_1 + u_2}{|u_1 - u_2|} \tag{49}$$

defines the standing wave ratio—*SWR*. *SWR* = 1 is a purely moving wave, *SWR* → ∞ is a purely standing wave. *SWR* is mainly used to evaluate signal reflections on lines or the reflectance of various acoustic structures.

The principle of standing wave formation is used in resonators. One example of a resonator is the reflection of the wave from the thin film described above. The resonator may be used as a wave amplifier. Multiple wave reflections, inside the resonator, can cause its amplification, like resonance in a serial RLC electrical circuit, wherein the resonance state the voltage at the capacitor is Q times a voltage of the source. Q is the quality factor (in the low-loss circuit Q ≫ 1). Similarly, in the wave resonator, the intensity of the wave in the resonator is many times greater than the intensity of the coming wave, resp. of the wave generated near the node. A typical example is creating a sound on a piano string—the hammer gently strikes the string near the knot (point of string fixation). The string sounds only with tones corresponding to the resonance condition (basic and higher harmonic frequencies). It is like the air column of the wind musical instrument, where the tongue vibrates the air near the knot, and the entire column sounds with multiple intensities.

Example 9 Standing wave on the string—vocal cords

The vibration of vocal cords whose frequency determines the tone of the voice can be modelled with a simple model of standing wave on the string. The string (the musical instrument) is fixed at the endpoints, which cannot move. They represent, therefore, standing wave nodes. For the standing wave then applies

$$l = n \frac{\lambda}{2} = n \frac{c}{2f} = \frac{n}{2f} \sqrt{\frac{F}{\mu}}, \text{and } f_n = \left(\frac{1}{2l} \sqrt{\frac{F}{\mu}}\right) n = nf_1 \tag{50}$$

On the string, there arise standing waves with discrete frequency spectrum f_n. The term in parentheses represents the base frequency f_1, which determines the base pitch of the tone. Other higher harmonics change the colour of the tone (different contents of higher harmonics cause different colours of the same tone at different musical instruments). Wave velocity on the string depends on the force F tensioning the string and the linear weight $\mu = m/l$ of the string (see chapter Mechanical waves). We can tune the resonance frequency, and therefore the pitch of the tone, changing values of F, μ and l.

The vocal cords are the muscle bundles between which there is a gap. As the air flows through the gap, the vocal cord muscles tremble, the fundamental frequency of oscillations being dependent on the thickness of the vocal cords, and the force that strains the vocal cords. Children and women have vocal cords thinner, so they have a higher voice frequency. In men in adolescence, the vocal cords coarsen (mutation), and the voice becomes deeper. You can control the pitch of a tone by changing the vocal cord tension, so you can intonate when singing (intonation—pitch control).

Example 10 Standing wave in a tube—ear canal

The outer ear canal is an acoustic resonator that increases the sensitivity of hearing at the middle frequency of audible sound (about 3 kHz). The incident longitudinally polarised wave is reflected on the eardrum and interferes with the wave passing directly. This creates a standing wave and resonantly amplifies the sound.

A simple model idea is provided by the description of the standing wave in the air column in the canal. If the tube is closed at the end, there is a standing wave node at the end, there is an anti-node in the open mouth. The length of the ear canal thus represents a quarter of the wavelength.

$$l = \frac{\lambda}{4} = \frac{c}{4f}, \text{ resp. } f = \frac{c}{4l}. \tag{51}$$

Substituting the values of $c \approx 330$ m·s^{-1} and the length of the ear canal $l \approx 2.5$ cm, we obtain a resonant frequency $f \approx 3.3$ kHz. It is known from audiometric measurements that the maximum sensitivity of the auditory organ is around this frequency.

Example 11 Ultrasonic transducer

To generate ultrasound, for example, in sonography or lithotripsy, piezoelectric transducers are used. The ultrasonic transducer is a plate of an anisotropic piezoelectric material with suitable orientation and electrodes on its surface. When the AC voltage is applied, the mechanical stress in the plate changes alternately, and thus the deformation (plate thickness slightly alternately increases and decreases) generates oscillations that represent a standing wave in the plate. If the wave impedance of the outer medium is less than the impedance of the plate, there occurs reflection of the wave on the surfaces of the plate and the standing wave arises in the plate. Thus, resonance occurs when the plate thickness is approximately equal to half the wavelength. For values used in sonography, for example, $f = 5.0$ MHz, and ultrasonic velocity in ceramic piezoelectric $c \approx 3.4$ km s^{-1}, we obtain plate resonance thickness $d \approx \lambda/2 = c/(2f) \approx 0.34$ mm.

Example 12 Infrasound in the room

The danger for a human is represented by vibrations with frequencies below the audible limit of $f = 15$ Hz, referred to as infrasound. The walls of rooms, especially concrete structures, vibrate due to the transmission of vibrations from the ground or the operation of machines in the building. The oscillations of the subsoil, unless we consider an earthquake, come mostly from traffic. Wall oscillations generate acoustic waves in the room, which can be amplified by wave resonance. High impedance concrete walls represent almost ideal reflective surfaces and are nodes of rising standing waves. The basic resonant frequency corresponds to the condition $l = \lambda/2$.

For a room with a length of $l = 10$ m and a sound velocity $c = 330$ m s^{-1}, the base frequency of the standing wave is $f_1 = c/2l \approx 16.5$ Hz. The maximum oscillation is in the centre of the room. If in the office working place, for example, in the middle of the room, acoustic vibrations can harm the health of a person. To reduce the risk of low-frequency vibrations, it is possible to treat the walls with an anti-reflective cover.

5. Diffraction

Wave diffraction is a phenomenon that is sometimes referred to as wave bending and, in simplified terms, can be characterised as wave propagation beyond an obstacle. When interpreting the phenomenon, we will use the Huygens-Fresnel principle.

5.1 Radiation of the planar source

Planar sources are often used to generate waves, for example, piezoelectric plates for generating ultrasound, reflectors of sources of EM radiation, or funnel antennas of microwaves. The planar source can also be an aperture in a screen, which is hit by a plane wave—the aperture becomes a planar source for additional space.

5.1.1 Long rectangular strip source

Thin strip transducers, especially ultrasonic ones, are often used. Such a source may be also a slot that is hit by a plane wave, for example, the light.

Using the Huygens-Fresnel principle, the strip of width a is divided into elementary sources, elementary strips of width dx, **Figure 8**.

The interference of cylindrical waves of elementary sources is expressed as

$$\mathbf{u}\,(\mathrm{P}) = -\,\frac{\mathrm{j}}{\sqrt{\lambda}}\,u_0\,(\mathrm{M})\int\limits_{-a/2}^{a/2}\frac{\exp\,(\mathrm{j}\mathbf{k}\cdot\mathbf{r})}{\sqrt{r}}\,\cos\,\varphi\;dx. \tag{52}$$

Consider a wavefield at a distance $r \gg a$ (Fraunhofer region—distant field) for which we get vectors r and r_0 to be parallel and $\cos\varphi$ constant.

The wave function argument is expressed

$$\mathrm{j}\boldsymbol{k}\cdot\boldsymbol{r} = \mathrm{j}\boldsymbol{k}\cdot\boldsymbol{r}_0 - \mathrm{j}kx\,\sin\,\varphi$$

and after calculation of the integral, we obtain

$$\mathbf{u}\,(\mathrm{P}) = -\,\frac{\mathrm{j}a}{\sqrt{\lambda r}}\,u_0\,\cos\,\varphi\;\mathrm{e}^{\mathrm{j}\mathbf{k}\cdot\mathbf{r}_0}\,\frac{\sin\left(\frac{ka}{2}\,\sin\,\varphi\right)}{\left(\frac{ka}{2}\,\sin\,\varphi\right)}. \tag{53}$$

The wave propagates from the elementary source as a cylindrical one in all directions, but the wave is not isotropic and has a distinct directional pattern depending on the angle φ under which we observe the radiation.

The wave intensity is

$$Z\,\omega^2u^2\,(\mathrm{P}) = \frac{a^2}{\lambda r}\,Z\,\omega^2u_0^2\,\cos^2\varphi\quad\mathrm{F}(\alpha),\,\mathrm{resp.}\quad I = I_0\,\frac{a^2}{\lambda r}\,\cos^2\varphi\;\mathrm{F}(\alpha), \tag{54}$$
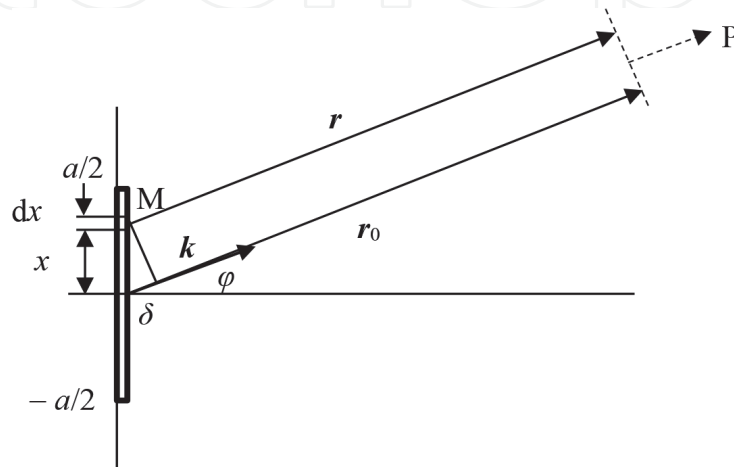
Figure 8.
Diffraction on the strip.
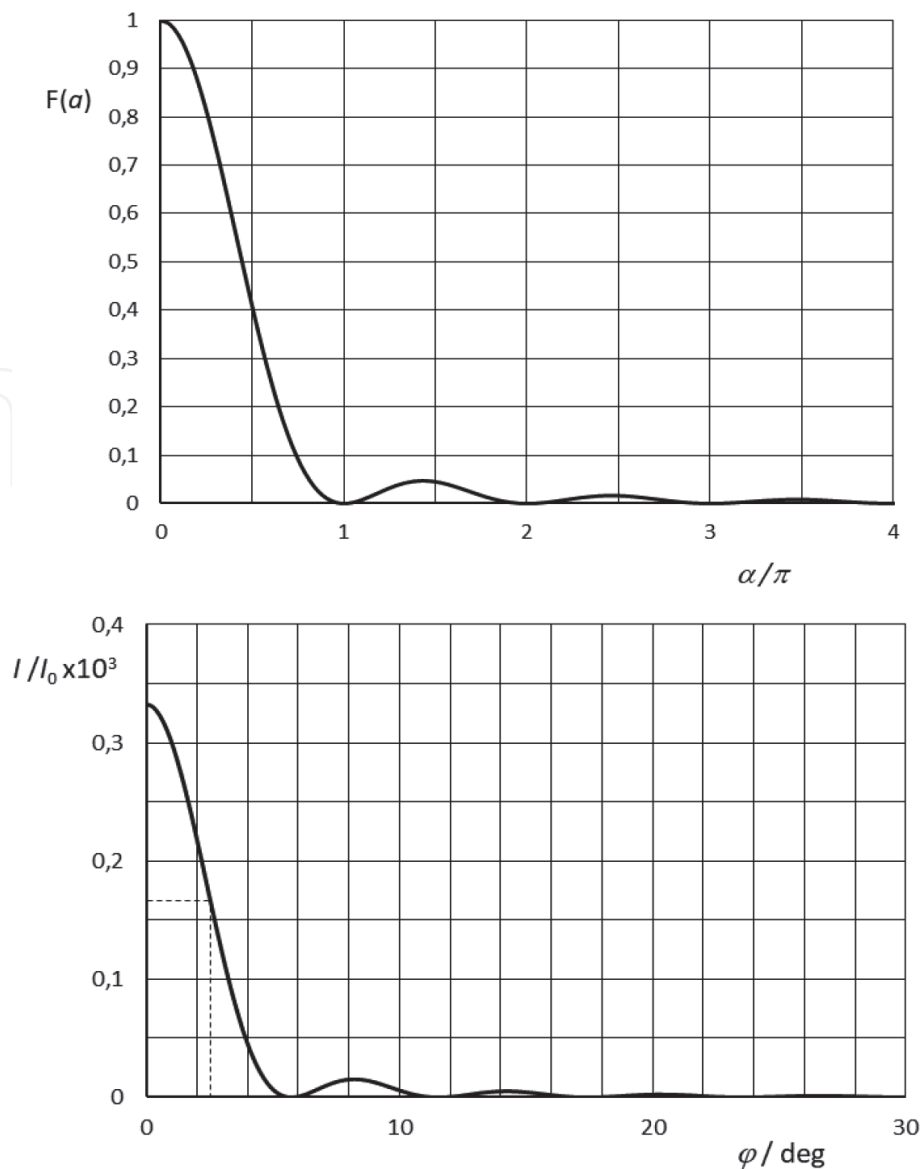
Figure 9.
Graph of the function F(α) = (sin α/α)² versus α/π (upper), and directional radiation characteristic of the strip source (λ = 500 nm, a/λ = 10, r = 15 cm)—lower one.

where $F(\alpha) = \left(\frac{\sin \alpha}{\alpha}\right)^2$, and $\alpha = \frac{ka}{2} \sin \varphi = \pi \frac{a}{\lambda} \sin \varphi$.

In the graphs in **Figure 9**, we see the diffraction function $F(\alpha)$, and the directional radiation characteristic I/I_0, from which we can see to which angles φ the source emits the radiation. The main lobe is significant. Its width is determined at the level of intensity decrease to 50% of the maximum value. The characteristic width in the figure is $\Delta\varphi \approx 5°$. While the cos φ function changes only slowly and monotonically from the unit value in a straight line perpendicular to the surface of the plate, to zero in a direction parallel to the plate. The profile of the diffraction pattern is mainly described by the function $F(\alpha)$.

The maxima of the function $F(\alpha)$ determine the directions in which the radiation has a local maximum. As we see in the picture, the maximum of the 0th order is the biggest one, others are significantly lower. The maximum of the 1st order, for $\alpha/\pi \approx 1.4$, is only 5% of the main maximum.

The value $\alpha/\pi = 1$ (minimum of function) corresponds to the geometric angle $\varphi_{\min 1}$, for which we have

$$\alpha = \frac{ka}{2} \sin \varphi_{\min 1} = \pi, \text{ or } \sin \varphi_{\min 1} = \frac{\lambda}{a}. \tag{55}$$
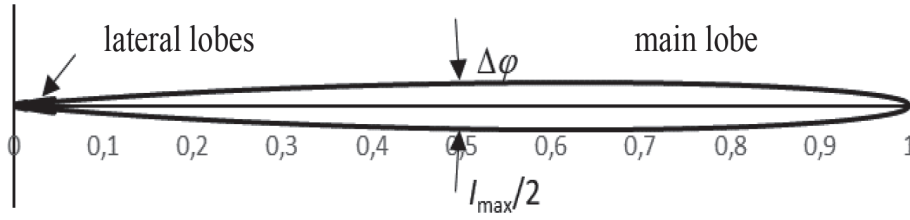
Figure 10.
Radiation characteristic in the polar coordinates for a/λ = 10.

In **Figure 10**, the radiation characteristic of a strip with $a/\lambda = 10$ is plotted in the polar coordinates I/I_0 versus φ.

The width $\Delta\varphi$ of the main lobe is determined at the level of intensity decrease -3 dB (50%). If $\cos^2\varphi \approx 1$ for narrow characteristics $\Delta\varphi \ll 1$, the decrease is determined by the function $F(\alpha)$. For the intensity decrease to 50%, applies the equation

$$\left(\frac{\sin\alpha}{\alpha}\right)^2 = \left(\frac{\sin\left(\pi\frac{a}{\lambda}\sin\varphi\right)}{\pi\frac{a}{\lambda}\sin\varphi}\right)^2 = \frac{1}{2}, \text{ resp. } \sin x = \frac{x}{\sqrt{2}}. \tag{56}$$

It results in the numerical solution $x \approx 1.39$.
The width of the main lobe is then

$$\Delta\varphi = 2\arcsin\left(0.44\ \frac{\lambda}{a}\right). \tag{57}$$

For the characteristic at **Figure 10** (λ/a = 1/10) we get $\Delta\varphi \approx 5.0°$.
For a wide strip with $a \gg \lambda$, we can use an approximate relation

$$\Delta\varphi \approx 0.88\frac{\lambda}{a}. \tag{58}$$

If we want to achieve a narrow characteristic, for example, in sonographic imaging, we must choose a small ratio λ/a. For f = 5 MHz, and c = 1500 m s^{-1} is λ = 300 μm. To generate a beam with divergence $\Delta\varphi$ = 1°, a plate with a width $a \approx 1.5$ cm is required. This dimension is too large. Therefore, a narrowing of the radiation characteristic we obtain using a set of parallel strips (lattice) as described in the following paragraph.

The above-described diffraction occurs at a distance $r \gg a$. In the near, the so-called Fresnel region, the wave advances as a beam with a transverse profile corresponding to the shape of the transducer. It proceeds in a direction perpendicular to the transducer, **Figure 11**.
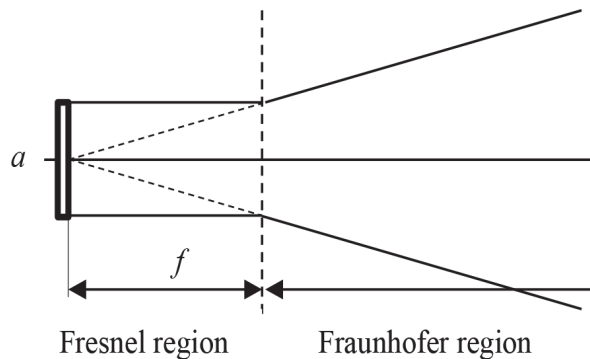
Figure 11.
Near and distant region of the transducer.

Only from a distance f occurs the divergence of the wave due to diffraction. We have

$$\frac{a}{2f} = \tan \varphi_m = \frac{\sin \varphi_m}{\sqrt{1 - \sin^2 \varphi_m}} = \frac{0.44 \frac{\lambda}{a}}{\sqrt{1 - \left(0.44 \frac{\lambda}{a}\right)^2}},$$

from where

$$\frac{f}{\lambda} = \frac{1}{2} \frac{a}{\lambda} \sqrt{\left(\frac{1}{0.44} \times \frac{a}{\lambda}\right)^2 - 1}. \tag{59}$$

For $\lambda = 300$ μm and $a = 1.5$ cm is $f/\lambda \approx 2.8 \times 10^3$ and $f \approx 84$ cm. In this case, which corresponds to the ultrasound frequency of sonography, the length of the Fresnel region f is greater than the depth of the organs of the body. The lateral resolution of 1.5 cm is unacceptably high for imaging. It is, therefore, necessary to narrow the wave beam. For a thin strip, $a = \lambda = 0.3$ mm $f/\lambda \approx 1.0$ and hence $f \approx 0.3$ mm. However, the beam divergence in the distant region is $\Delta \varphi \approx 52°$, which is too large to display. This problem is solved by a structured grid transducer, see Chapter 3.

5.1.2 Rectangular and circular planar source

To illustrate the diffraction effect, there are shown, in **Figure 12**, diffraction patterns formed on the projection screen after irradiation of a rectangular aperture (left image) and a circular aperture (right image) with a plane wave of the laser beam. The series of bright points on the horizontal axis (left) corresponds to the diffraction maxima according to **Figure 9**. The rectangular source can be taken as the combination of two mutually perpendicular strips. Each side induces diffraction in a direction perpendicular to the respective side with parameter α according to (54) to side a or equally β to side b. The resulting diffraction pattern of the rectangular (or square) source is shown in **Figure 12** left. The length of the side of the rectangular source can be calculated by measuring the distance of the maximums on the screen.

In the case of a circular source, the Huygens-Fresnel integral has a form

$$\mathbf{u}(P) = -\frac{jk}{2\pi} \iint\limits_S u_0(M) \frac{\exp\left[jk\left(r - \frac{x}{r}\xi - \frac{y}{r}\eta\right)\right]}{r} \cos \vartheta \; d\xi \, d\eta, \tag{60}$$
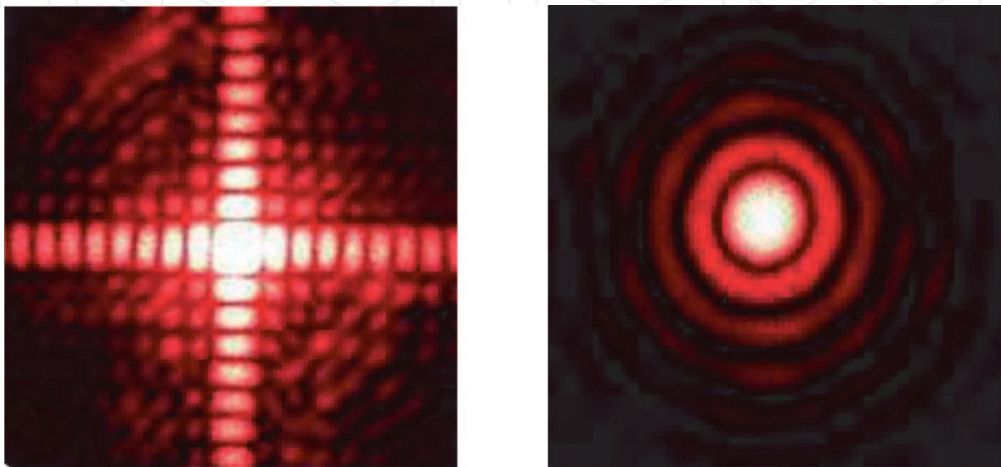
Figure 12.
Diffraction patterns of the square and circular sources of light.

where x, y are coordinates of the point P on the screen, ξ and η coordinates of the point M on the surface of the source, r is the mutual distance of points M and P.

Considering $r \gg R$, where R is the radius of the circular source, the distance r can be considered independent of the position of the point M of the source and hence taken from the integral. Since the system is axially symmetrical, it is preferable to use polar coordinates

$$x = r \sin \vartheta \ \cos \alpha, \qquad y = r \sin \vartheta \ \sin \alpha, \qquad \xi = \rho \ \sin \beta, \qquad \eta = \rho \ \cos \beta,$$

and the surface element of the source $\mathrm{d}S = \mathrm{d}\xi \ \mathrm{d}\eta = \rho \ \mathrm{d}\beta \ \mathrm{d}\rho$.

$$\mathbf{u}(P) = -\ \frac{\mathrm{j}k}{2\pi}u_0 \ \frac{\mathrm{e}^{\mathrm{j}kr}}{r} \ \cos \vartheta \int_0^R \int_0^{2\pi} \mathrm{e}^{-\mathrm{j}k\rho \sin \vartheta \ \cos (\beta - \alpha)} \ \rho \, \mathrm{d}\beta \ \mathrm{d}\rho. \tag{61}$$

By solving this integral we get

$$\mathbf{u}(P) = -\ \frac{\mathrm{j}k}{2\pi}u_0 \ \frac{\mathrm{e}^{\mathrm{j}kr}}{r} \ \cos \vartheta \ \ 2\pi \int_0^R J_0(k\rho \sin \vartheta) \ \mathrm{d}\rho,$$

where $J_0(x)$ is the Bessel function of the 0th order. After its integration we have

$$\mathbf{u}(P) = -\ \mathrm{j} \ \frac{2\pi R^2}{\lambda}u_0 \ \frac{\mathrm{e}^{\mathrm{j}kr}}{r} \ \cos \vartheta \ \frac{J_1(kR \sin \vartheta)}{kR \sin \vartheta}, \tag{62}$$

where $J_1(x)$ is the Bessel function of the 1st order. We get the wave intensity as a function of the angle ϑ relating to the axis of the transducer

$$I = I_0 \left(\frac{2\pi R^2 \cos \vartheta}{\lambda r}\right)^2 F(\gamma), \tag{63}$$

where $F(\gamma) = \left[\frac{J_1(\gamma)}{\gamma}\right]^2$, and $\gamma = kR \sin \vartheta = \frac{2\pi}{\lambda}R \ \sin \vartheta$.

In **Figure 13** (right) we see the irradiation characteristics. At 50% of the maximum, the width of the main lobe is $\Delta\vartheta \approx 2.9°$. In this case, the diffraction pattern forms circular traces, **Figure 12** (right).

The central bright circle is called the Airi disk. It is surrounded by a circle of zero intensity $F(\gamma)$ and thus $J_1(\gamma) = 0$ for $\gamma \approx 3.832$, which corresponds to the angular radius
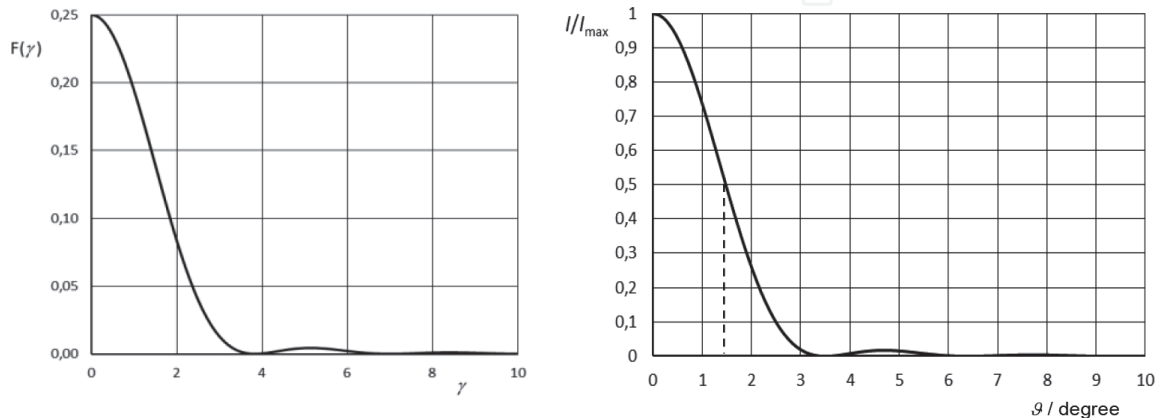
Figure 13.
Plot of F(γ) versus parameter γ—left, and relative intensity versus angle ϑ for λ = 500 nm, R/λ = 10—right.

$$\vartheta_{1\ min} = \arcsin\ \left(\frac{1.22\ \lambda}{2R}\right). \tag{64}$$

For the values in **Figure 13** is $\vartheta_{1\ min} \approx 3.5°$.

The results indicate that the angular diameter of the Airi disk and thus the scattering angle of the radiated wave decreases as the diameter of the source increases.

Again, the result is valid in the far Fraunhofer region $r \gg R$. In the near Fresnel region, the beam has a diameter of the transducer.

5.2 Structured plane sources

5.2.1 System of parallel strip sources with the same phase

Consider a simple transducer structure consisting of a grid of N parallel equal strips (according to). As in the previous paragraph, we observe the angular dependence of the radiation intensity. The distance between the centres of the strips is d and their width a. The resulting wave is a superposition of the waves emitted by the individual strips, see (25), using a phase shift resulting from the unequal ray length from the individual strips

$$\mathbf{u}(P) = \sum_{n=1}^{N} \mathbf{u}_n = -\ \frac{j\,a}{\sqrt{\lambda r}}\ u_0\ \cos\varphi\ e^{j\mathbf{k}\cdot\mathbf{r}_0}\frac{\sin\alpha}{\alpha}\sum_{n=1}^{N}\ e^{-jknd\,\sin\varphi}. \tag{65}$$

If we express the sum of the geometric progression, we get

$$\mathbf{u}(P) = -\ \frac{j\,a}{\sqrt{\lambda r}}\ u_0\ \cos\varphi\ e^{jk\ r}e^{-j\ (N+1)\frac{kd}{2}\sin\varphi}\frac{\sin\alpha}{\alpha}\ \frac{\sin\left(N\frac{kd}{2}\sin\varphi\right)}{\sin\left(\frac{kd}{2}\sin\varphi\right)} \tag{66}$$

and radiation intensity

$$I = I_0\ \frac{a^2}{\lambda r}\ \cos^2\varphi\ \left(\frac{\sin\alpha}{\alpha}\right)^2\ \left(\frac{\sin N\beta}{\sin\beta}\right)^2 = I_{max}\ F(\alpha)\ G(N,\beta), \tag{67}$$

where $\alpha = \frac{ka}{2}\sin\varphi$ and $\beta = \frac{kd}{2}\sin\varphi$.

The function $F(\alpha)$ describes the diffraction on a single strip and the function $G(N,\beta)$ is the so-called lattice diffraction function. The width of the strip a shall be chosen so that its radiation characteristic is sufficiently wide. The lattice function $G(N,\beta)$ is periodic with period π. For values $\beta_n = n\pi$, where $n = 0, 1, 2, ...$ the function acquires the maximum value of $G(N, \beta_n) = N^2$. For the angle φ_n of the diffraction maximum, we get the relation

$$\sin\ \phi_n = n\frac{\lambda}{d}. \tag{68}$$

For the width of this maximum, we have $\Delta\beta = \pi/N$, and when converted to the angular width

$$\Delta\varphi = 2\ \arcsin\ \left(\frac{1}{N}\ \frac{\lambda}{2d}\right). \tag{69}$$

For small values of angular width, we get approximate relation

$$\Delta\phi \approx \frac{1}{N}\frac{\lambda}{d}. \tag{70}$$

The relation (68) shows that the original beam of incident waves is divided into individual diffraction beams deviated from the original direction by angles φ_n. The angle φ_n depends on the wavelength of the wave.

Example 13 Diffraction grating spectroscope

A diffraction grating is used to separate the wave components of different wavelengths from one another. If white light falls on the grating, it is broken into individual colour components. The first diffraction maximum is used for the decomposition of light (λ = 400–600 nm). For λ = 400 nm, we choose the ratio $\lambda/d \approx 0.57$, so that $2\lambda/d > 0$ and thus the second maximum does not arise. The grating period is d = 700 nm and strip separation a = 350 nm. The range of diffraction angles is from φ_{1f} (400 nm-violet) = 35° to φ_{1r} (700 nm-red) = 59°.

For grating with N = 5000 strips, that is, the total width of the structure D = 3.5 mm, for λ = 520 nm (green), the diffraction maximum width $\Delta\varphi \approx 1.5 \times 10^{-4}$ rad.

The diffraction grating can distinguish wavelengths with the difference $\delta\lambda = \lambda_1 - \lambda_2$, for which the difference in angles of the maxima is $\delta\varphi \sim \Delta\varphi$. From the relation (32) we get for the 1st order maximum for λ = 500 nm (green) φ_{1g} = 44°.

$$\delta(\sin\varphi_1) = \cos\varphi_1\ \delta\varphi = \frac{\lambda}{d}\frac{\delta\lambda}{\lambda} = \sin\varphi_1\frac{\delta\lambda}{\lambda},$$

from where $\dfrac{\delta\lambda}{\lambda} = \dfrac{\Delta\varphi}{\tan\varphi_1}$.

For the given values $\delta\lambda/\lambda \approx 1.6 \times 10^{-4}$ and wavelengths resolution $\delta\lambda \approx 0.1$ nm.

In the light spectrum of the sodium lamp that emits yellow light, there is an emission double line (doublet) with wavelengths λ_1 = 588.9 nm and λ_2 = 589.6 nm. The difference $\delta\lambda$ = 0.7 nm, and therefore the doublet is distinguished by the diffraction grating (*note: we do not distinguish them with the naked eyes*). Analysis of the radiation spectra is called spectroscopy. Based on the spectrum analysis, we can investigate the properties of the radiation source, for example, chemical composition. Thus, we can analyse the composition of stars using spectral analysis of the light coming from them. The spectrum of radiation is also modified as it passes through the material. In this way, we can analyse the properties of the material, for example, to determine the chemical composition of human body fluids (cerebrospinal fluid, blood, plasma, etc.). In haematology, the blood fat content before blood donation is determined by spectral analysis.

Example 14 Diffraction grating monochromator

In some applications, monochromatic radiation (single wavelength) is required. In such a case, the light with the desired wavelength can be obtained using its selection from intense white light, for example, of halogen lamp, by a diffraction grating monochromator.

The principle is the same as in the previous example. The white light is dispersed by the diffraction grating into individual colour components, and the desired wavelength is selected using a slit-shaped aperture. In addition to the transmission grating, a reflective one is often used as well. It consists of parallel strips with a reflective surface that is illuminated by white light. The individual strips thus represent, in reflected light, a set of parallel wave sources, which interfere with each other as in the case of a transmission grating.

The arrangement of the monochromator is in **Figure 14**. The white light source and the aperture are fixed. The reflective grating is deflected so that light with the desired wavelength λ passes through the aperture gap. The deflection of the grating is usually done by a micrometre screw calibrated at wavelengths.
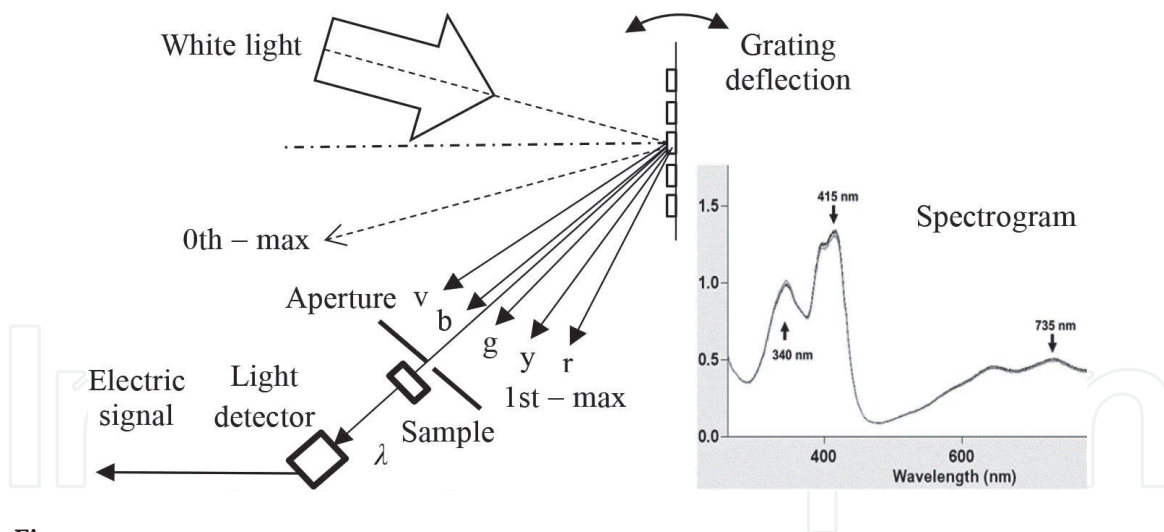
Figure 14.
Monochromator with reflection grating.

In spectrometers, the sample to be examined is placed behind the aperture. Its absorbance is determined from the signal of the detector versus wavelength (deflection of the grating). The result is recorded in a spectrogram, a picture on the right, from which it is possible to determine the presence of certain substances in the sample, for example, the haemoglobin spectrum or the fat content in the blood.

From the results (69), resp. (70), transducers with multiple periodic structures have a significantly narrower radiation pattern than a single strip source. If the aim is to concentrate the wave power to a small scattering angle $\Delta\varphi$, it is required that no non-zero order diffraction maxima occur, in order not to dissipate the wave energy. This is achieved when $\lambda/d > 1$. If λ/d is approximately equal to one, the width of the radiation beam is approximately $1/N$ in radians.

Example 15 Sonographic probe

A structured transducer consisting of more parallel piezoelectric strips is used as the ultrasonographic probe. In the case of $N = 64$ strips with a width of $a = 0.1$ mm and a spacing $d = 0.2$ mm at 5 MHz and ultrasonic speed $c = 1500$ m·s^{-1} ($\lambda = 0.3$ mm and $\lambda/d = 1.5 > 1$) radiation characteristic width $\Delta\varphi \approx 1.3°$.

Such a structured transducer has a sufficiently high angular resolution.

5.2.2 Electronic deviation of the radiated wave

Many applications require changing the direction of radiation, for example, radars, or ultrasonography. This can be achieved by a mechanical deviation of the antenna, for example, swinging, or rotating radar antennas, or USG probes with a mechanical deviation of the ultrasonic transducer.

The controlled deviation of the radiation characteristic can also be achieved electronically. In the previous case of structured multistrip transducer, all strips generated their waves with the same phase on their surface. If, however, the individual strips are excited, so that the phase shift ψ of the excitation of adjacent strips is constant, the resulting wave is formed by superposition of the waves of the individual strips, **Figure 15** on the left, and we have

$$\mathbf{u}(z,t) = \sum_1^N \mathbf{u}_n = \sum_1^N u_{\mathrm{m}n} \; e^{j \; (\omega t - k r_0 + k n d \sin \phi + n \psi)} \frac{\sin \alpha}{\alpha} =$$
$$= u_{\mathrm{m}} \; e^{j \; (\omega t - k r_0)} \; e^{j(k d \sin \phi + \psi)\frac{N-1}{2}} \frac{\sin \alpha}{\alpha} \frac{\sin \; N\gamma}{\sin \; \gamma} ,$$

(71)

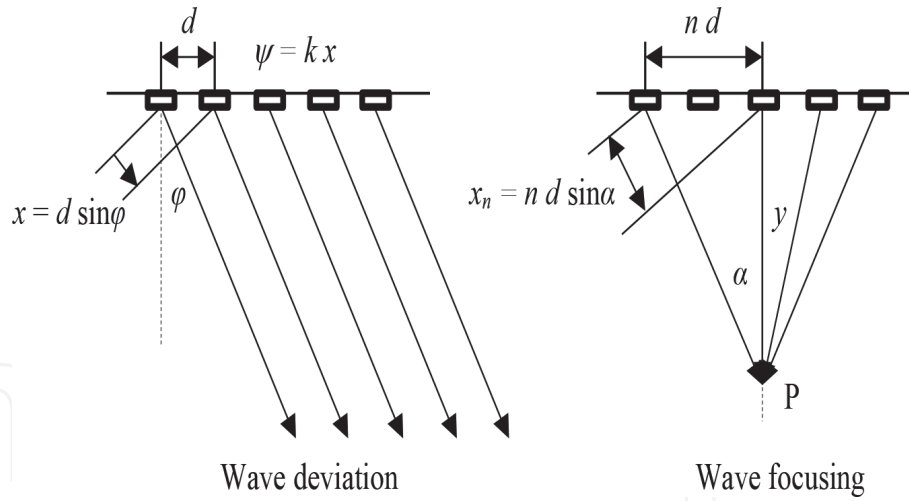where $\gamma = \frac{kd}{2} \sin \phi + \frac{\psi}{2}$.

Figure 15.
Electronic deviation and focusing of wave.

The main maxima correspond to $\gamma = 0, \pm\pi, \pm 2\pi, \ldots$, from we get

$$\sin\phi = \frac{\lambda}{d}\left(n - \frac{\psi}{2\pi}\right), n = 0, \pm 1, \pm 2, \ldots$$

If $\lambda/d \gg 1$ and $\psi/2\pi \ll 1$, the relation has only one real solution for $n = 0$. The deviation angle is then

$$\varphi = -\arcsin\frac{\lambda}{d}\frac{\psi}{2\pi} \approx -\frac{\psi}{kd}. \tag{72}$$

Due to varying the phase shift ψ between the exciting signal of the strips, we can electronically deviate the radiated wave by an angle φ, **Figure 15** on the left. Thus, by periodic changing the phase difference ψ using the electrical supply of the strips, the generated wave continuously deviates. This method of electronic deviation of the ultrasonic beam is used in sectoral ultrasonography. To create a typical sectoral image of the internal organs, the elder probes used mechanical beam deviation. It had, however, some disadvantages. The probes for electronic deviation have no moving parts, and therefore the beam deflection can be faster than in the case of the mechanical ones.

5.2.3 Electronic focusing of the radiated wave

To achieve the necessary resolution of the ultrasonographic image of the tissue at the desired depth, it is necessary to focus the wave to that depth. The elder method used an acoustic lens on the surface of the ultrasonic transducer. The current method consists of focusing the wave electronically. It is achieved by a suitable phase shift of the individual waves radiated by the strips of the transducer, **Figure 15** on the right.

The individual waves from the strips are concentrated in the focus P when they encounter in the same phase, and constructive interference occurs.

The path of the n-th individual wave is longer than the distance from the middle wave by $x = nd \sin\alpha$, which corresponds to the phase shift ψ_n, by which its exciting signal needs to be back phase-shifted,

$$\psi_n = kx_n = nkd \sin \alpha = \frac{2\pi}{\lambda} nd \sqrt{(nd)^2 + y^2}. \tag{73}$$

If the electrical exciting signals of the individual strips are programmed so that their phase decreases gradually according to (73), the resulting wave concentrates in the focus P at a depth y from the transducer. Changing the value of the y parameter in the phase shift calculation can change the focusing depth of the wave.

Different generated beam profiling can be achieved through digital processing of the exciting signal. The digital control of phase shifts of the individual elements of the transducer using a computer is considerably simpler than the analogue solution.

6. Wave imaging

Waves, both mechanical or electromagnetic, occur in a wide range of phenomena and applications. One of the technical and natural phenomena is the transmission of information. We can see objects by vision or perceive sound through hearing. The basis of human perception of the waves is processing many signals provided to the brain by biological sensors, which are light-sensitive cells on the retina of the eye or cells of the cochlea sensitive to vibrations in the inner ear. Paired sensor organs (a pair of eyes, a pair of ears) provide the brain with information to create a stereo effect. In space, we orient ourselves by sight and hearing. The essence of sensory perception is the detection of the spatial–temporal distribution of the respective wave, which is created by its source and carries information about the properties of the source. By detecting this field, we infer the configuration of the waveform and create such an image of the source (optical, acoustic, etc.). The observed image is formed in our brain. The projection of an object into the structure of brain cells represents a certain transformation by which the object is transmitted to the brain cell's structure using neural signals. From the created image in the brain, we infer the position, arrangement, and other observed properties of the object. But the image in the brain only shows the perceived waves field. Between the object and the human sensors, the perceived field may be modified by many obstacles. It means, the image created in the brain need not match the object faithfully, for example, if we look into a crooked mirror the image created in the brain does not match the real object (we do not believe our eyes). In the perception of vision, the waves emanating from the observed object are modified before their impact on the eye retina cells by the eye lens and cornea. On the retina, the incident wave creates a two-dimensional image of a three-dimensional object. For example, kilometres large object is displayed as a centimetre image on the retina, and then it is sensed by sensory cells.

The brain has a memory, and it stores received images. Human memory is subjective so that various methods of wave-field recording are used for objective image preservation. If not only the amplitude but also the phase information is recorded, we get a complete record—a hologram. But two-dimensional recording only amplitude information is much simpler and widely used, for example, an ordinary photo. A suitable combination of amplitude records can create a spatial impression, for example, an acoustic stereo effect using headphones with a separate signal for the right and left ear, or optical by creating separate two-dimensional images for the right and left eye (e.g., using stereo glasses). However, this stereo effect is incomplete because it provides a spatial view from only one direction. Full digital image recording, tomography, is allowed by computer memory which can store a series of two-dimensional images of thin slices (*Greek: tomos—slice*) of the object with different values z of the third spatial dimension.

Wave imaging follows three basic goals. The first is to make the observed object visible or audible to the human senses, for example, observation of infrared, X-ray, ultrasound, or infrasound and ultrasound, or other waves "invisible" to humans. The second one is to allow observation of very distant, very large or very small objects, observation of very fast or very slow events, etc. The third goal is to record images for both analysis and archiving purposes. In addition to artistic recordings (paintings, sculptures), there are recordings on storage media, which may be chemical emulsions (classical photography), mechanical, magnetic or optical sound recordings (turntables, magnetic tape, magnetic disc, optical disc) or digital recording in computer memory. Records can be temporary (short term) or permanent (long term).

Short-term recording (computer display, microscope or telescope image, computer RAM recording, etc.) serves mainly to analyse the current structure and properties of the observed object, for example, observation of tissue by a microscope, observation of internal organs by an endoscope, etc. The long-term records are mainly used for archiving, for example, photo albums, X-rays images, etc. A modern and very economical way of archiving is archiving in digital form on various storage media. Digital archiving saves considerable space in healthcare facilities instead of storing printed reports or photographic images. Also, archived digital records allow rapid search and operative communication using communication networks.

The basis of imaging is to create a readable image, either for the observer's eye or for the respective recording medium. Wave imaging, except for volumetric holography, is two-dimensional (eye retina, film, camera CMOS or CCD chip, tissue tomography slice). The wave imaging aims to display the object as accurately as possible on the surface of the recording medium. For this purpose, serve the individual elements of the display systems, such as mirrors and lenses, or different complex systems, for example, optical, and acoustic microscopes, telescopes, photographic objectives.

6.1 Elements of display systems

6.1.1 Mirrors

One group of display system elements are mirrors. The mirror may be planar or curved. In **Figure 16** are several types of used mirrors.

(a) The planar mirror creates an image virtual, straight, and right-left inverted (that is why persons know themselves differently from a mirror and a photograph). When properly positioned, a planar mirror allows looking around the corner (known application is a periscope). (b) The 3D corner reflector consists of three mutually perpendicular planar mirrors forming a "corner". At the present 2D figure, only a pair of mirrors is shown. The ray of wave incident at any angle on one mirror is reflected from the other one exactly in the opposite direction. The principle of the 3D corner reflector is the same. It is used, for example, in roadside reflecting pieces, or in the safety reflecting elements, where the light of a vehicle headlamp is always reflected towards the driver. (c) Curved concave mirror (usually spherical or parabolic in more demanding applications). The beam of rays parallel to the optical axis incidents in the mirror and is reflected the focal point F located at the centre of the distance between the top of the mirror and the curvature centre. It is used in medicine, for example, as an ORL mirror that the doctor puts on the head. It concentrates the light from the lamp on the examined place, usually an ear. The doctor looks at the place through the aperture A at the top of the mirror. In advance, the mirror shields the disturbing direct light falling
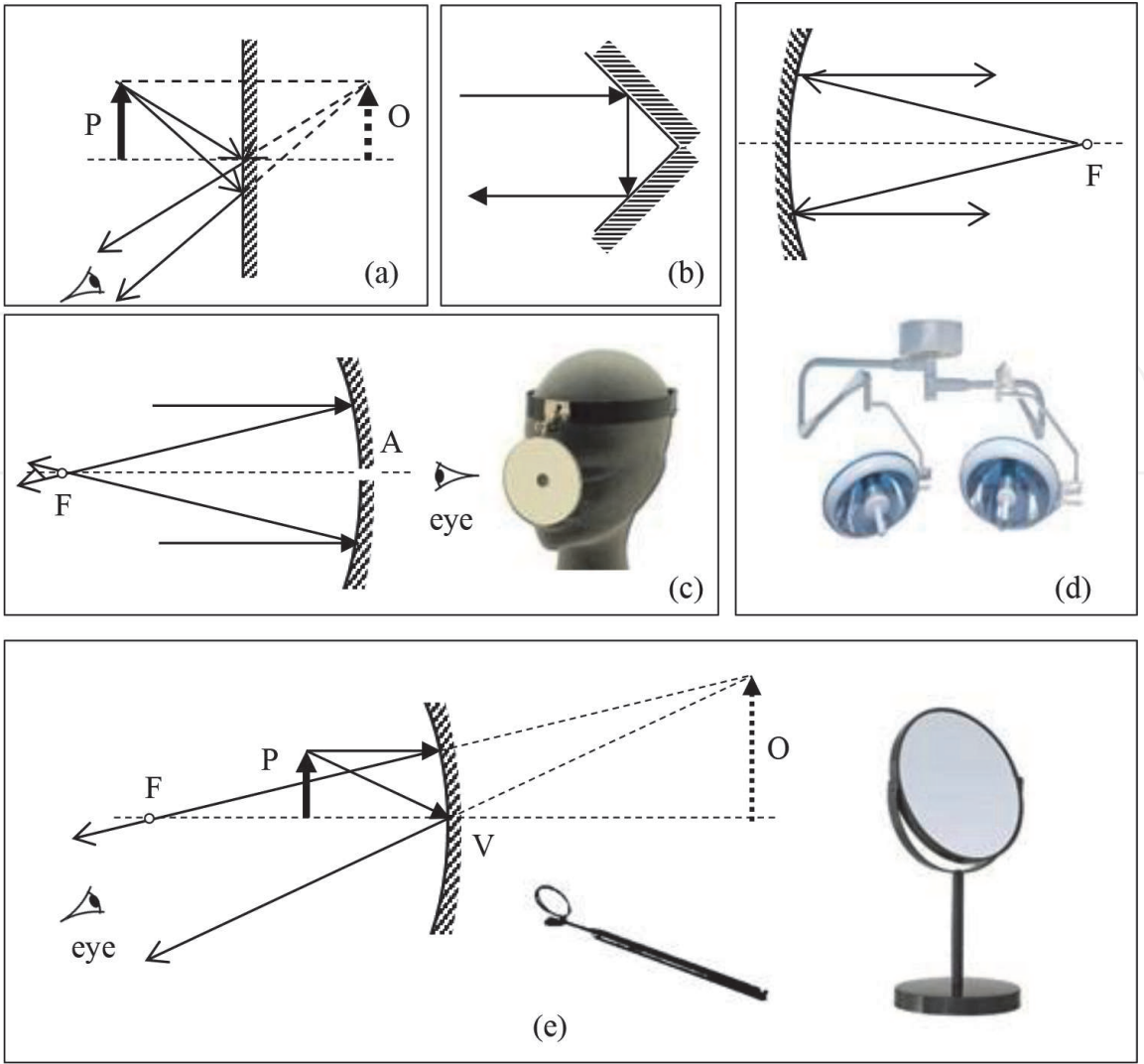
Figure 16.
Different types of mirrors.

into the doctor's eye, thus increasing the contrast of the object under investigation. If a detector is placed in the focus of the concave mirror, the energy of radiation striking the entire surface of the mirror (parabolic antennas) is concentrated there. The larger the mirror radius, the stronger the signal is detected. This principle is used, for example, at antennas for satellite reception or microwave radio transmission of a digital signal. Similarly, the principle also uses a parabolic antenna to receive mechanical waves of distant acoustic sources. (d) If a point source of the wave is placed in focal point F, the wave, when reflected from the concave mirror (reflector), forms a parallel beam of rays. It is used, for example, when illuminating the operating field during surgery or dental procedures. By changing the position of the source on the axis, a slightly diverging or converging beam of radiation is achieved. (e) The curved concave mirror is also used to display the object. If the object P is placed between the focus F and the top V of the mirror, a direct and enlarged virtual image is produced, with a greater distance from the mirror. This is used, for example, as dental mirrors for the diagnosis of teeth or cosmetic mirrors. These principles are common to any wave, regardless of its physical nature, whether mechanical (sound, ultrasound) or electromagnetic (light, infrared radiation, radio waves, etc.).

Note: In addition to parabolic mirrors, which focus a beam of parallel rays into the focus of a parabola and vice versa, there are also elliptical mirrors (rotating ellipsoid-shaped mirror surfaces) that concentrate rays radiated from a point source located in one

focus into the second one. There are various architectural structures, such as halls, with vaulted walls and ceilings, which consist of ellipsoidal areas. There exist so-called "corners of lovers". If one person is in one corner (focus) and quietly speaks in a noisy room, the other person in the other corner (focus) can hear what the first person is, while other people outside the place hear nothing.

6.1.2 Lenses

Unlike mirrors, which use reflection, lenses use a transition through the material of the lens with a refractive index different from the refractive index of their surroundings. The basic types of lenses are shown in **Figure 17** with the indicated passage of the light rays through the lenses.

Lenses in the figure have refractive index $n > n_0$, where n_0 is the refractive index of their surroundings.

(a) The converging (positive) lens concentrates a beam of rays parallel to the axis into the image focus F′ and forms a diverging beam behind the focus like a point source located in the focus. The intensity of the wave in the focus is greater, the larger the area of the lens from which it gathers incident rays. Like an optical lens, an acoustic lens works by concentrating an incident ultrasonic or sound wave into the focal plane. Acoustic lenses have been used in ultrasonography to focus the ultrasonic beam. Today they are replaced by digital focusing. (b) If we place a point source in the object focus F of the lens, for example, halogen bulb, the lens forms a parallel beam of rays similar to a spherical mirror (while in the case of a mirror the source shields part of the reflected beam, this does not occur when the lens is used). It is used in projection devices as a condenser. (c) The figure shows the display of points lying outside the lens axis. If the object is located at a distance $a > f$, where f is
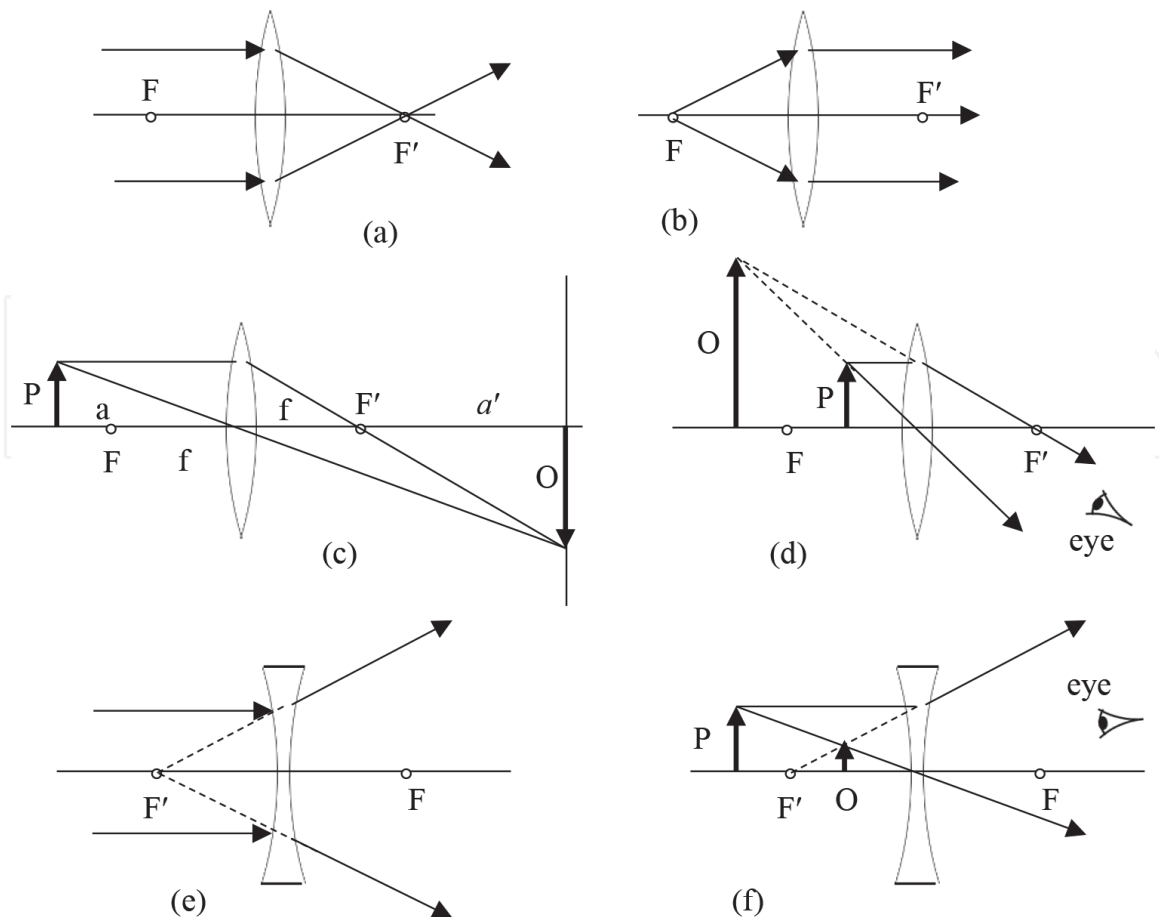
Figure 17.
Converging and diverging lens (P-object, O-image).

the focal length of the lens (inverse value $D = 1/f$ is the optical power and is given in dioptres, denoted D), a real and inverted image occurs in the image plane. It can be recorded with a suitable storage medium. This is used by photographic cameras, which focus incident rays on the detection surface (film, CCD, or CMOS chips) in the image plane, where creates an image of the object. The lens of the eye also concentrates the light incident on the pupil to a point on the retina of the eye and creates there an image of the observed object. We see from the figure that the image distance needs to be adjusted to the proper position to sharpen the image, for example, by moving the lens (focusing on the camera) or changing the optical power of the lens (eye accommodation). (d) If the object is between the object focus F and the lens, a direct, apparent, and magnified image occurs at a greater distance and can be observed with the eye. This is the principle of a magnifying glass for observing small objects. (e) The diverging (negative) lens scatters the incident beam as if the rays were coming from a point source located in the image focus F'. Thus, for example, create a wave field of the point source from the laser beam of parallel rays. (f) Displaying by a diverging lens is indicated in the figure. The object placed in front of the lens creates a virtual, direct, and reduced image that can be directly observed with the eye.

Lenses (glasses, contact lenses) are used to correct long-sightedness (positive lenses) or short-sightedness (negative lenses) if the eye lens is not capable of the necessary accommodation.

Note: If the inequality $n < n_0$ applies, the function of the lenses will be reversed (e.g., an air bubble in the glass or the water). The positive lens acts as a negative one and vice versa.

6.2 Imaging systems

Various imaging systems are assembled from the individual optical elements to display the object for observing it by a human eye with sufficient resolution or an image recording by a proper device (photograph, file on a memory disk). Imaging systems allow observing very distant objects (binoculars—telescopes) or very small objects (microscopes). The displaying system can be supplemented, for example, by diffraction grating for spectral analysis of the incident radiation. See also Splinter [2].

6.2.1 Photographic camera and projector

The simplest optical system is a photo camera, projector, or the human eye. The principle of the camera is in **Figure 17** on the left. The device objective can be a single lens or a set of lenses. Due to the dispersion of light in the lens material and correction of this undesirable phenomenon, complex objectives composed of several lenses (converging and diverging) of different materials are created for cameras to correct these defects. The camera objective at all is a converging system that concentrates the rays from the object on a recording medium (photographic film or plate or detecting sensor—CCD or CMOS chip) to create an image. As shown in the section on lenses, a distant object creates an image in the focal plane of the lens. The focal length f is indicated on the objective. For conventional cameras with a width of 35 mm recording field is $f \approx 50$ mm. If we want to enlarge the displayed field, we use a wide-angle objective with a focal length of about 28 mm.

To display distant objects a telephoto lens with a focal length of up to 400 mm is used, which has a small viewing angle but displays very distant objects on the recording medium. Optical devices also use lenses with variable focal lengths, zooms, which allow using both a wide-angle and a telephoto lens when retuning. The zoom can be optical (optical zoom) when the focal length of the imaging

system changes, or electronic (digital zoom), when the selection of image points on the recording medium is electronically changed. By selecting the material of the elementary parts of the optical system and the detecting medium, it is possible to realise cameras for displaying visible light (VL)—photo cameras, or cameras for displaying infrared radiation (IR)—thermographic cameras. The IR camera works on the same principle as a VL camera, only the elements of the optical system must be transparent for infrared radiation and opaque for visible light (e.g., monocrystalline germanium), and the detector must be sensitive to IR radiation. CMOS detectors have the best features (sensitivity, noise) for it. IR cameras are used in thermographic diagnostics, whether technical or medical.

Projector, **Figure 18** on the right, is used to project film slides or digital recordings on a screen. The current modern tool is a digital data projector. The projector contains an intense light source (usually a halogen lamp). The diverging beam of the point source is changed to a collinear beam using a condenser (positive lens). The light hits a film or digital recording and projects it onto the screen using an objective. Projectors are used to present movies, images, or digital recordings from a variety of storage media.

6.2.2 GRIN lens

In an optically or acoustically inhomogeneous medium, the rays of the waves are curved. It is used in various technical applications. The optical inhomogeneity of the medium is used, for example, in optical fibres or GRIN lenses.

Example 16 Curvature of the acoustic beam in the surface layer of water in the sea

In the sea, the chemical composition and pressure of the saltwater change with depth. This affects the velocity of sound propagation in water. At steady conditions, the velocity of sound in saltwater decreases approximately linearly to a depth of about h_1 = 500 m and then increases again with the same gradient. At the surface, the velocity of sound in water is c_0 = 1500 m s^{-1} and at depth h_1 = 500 m is c_1 = 1480 m s^{-1}. Velocity gradient k = dc/dh = 4.0 $\times$ 10^{-2} (m s^{-1}) m^{-1}.

We show that the sound beams are curved and have the shape of a circular arc, **Figure 19**. The figure on the right shows the elementary section of the beam with the radius of curvature R in a thin layer with a thickness dh. When the wave velocity c changes, the angle α to the normal to the water level changes as well. According to Snell's law of refraction, we have

$$\frac{\sin \alpha}{c} = \frac{\sin \alpha_1}{c_1} = \text{constant, or after differentiation} \tag{74}$$

$$-\frac{\sin \alpha}{c^2} \mathrm{d}c + \frac{1}{c} \cos \alpha \ \mathrm{d}\alpha = 0.$$
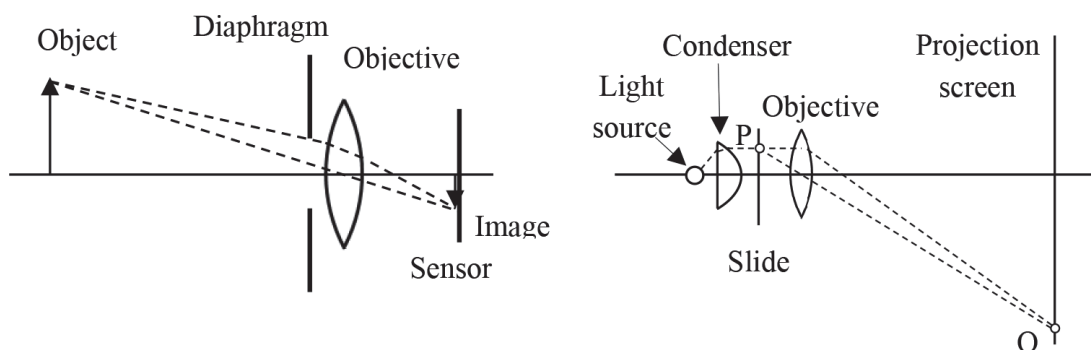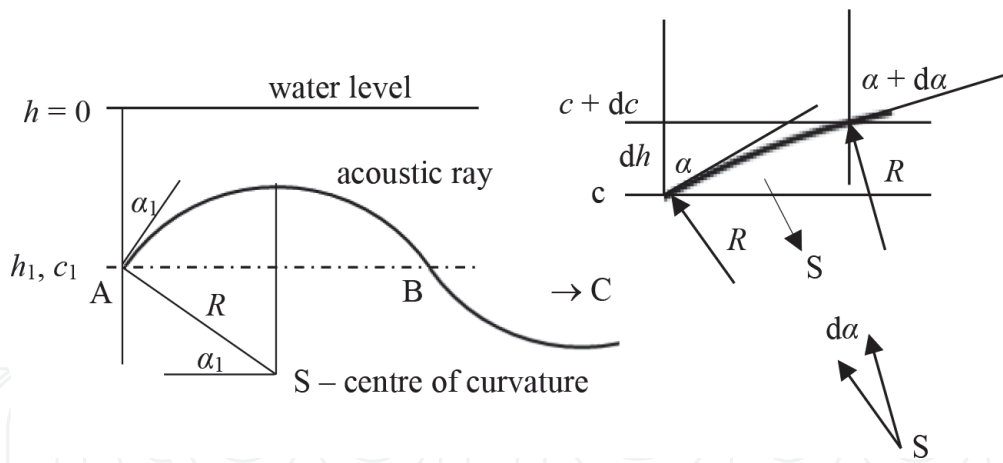
Figure 18.
Camera (left) and projector (right).

Figure 19.
Curvature of the acoustic ray in inhomogeneous sea water.

The length of the elementary section of the ray can be expressed by $R\,\mathrm{d}\alpha$ or $\mathrm{d}h/\cos\alpha$. After being replaced by refraction law, we get

$$\frac{1}{R} = \frac{\sin\alpha}{c}\frac{\mathrm{d}c}{\mathrm{d}h} = \frac{k}{c_1}\sin\alpha_1, \text{ which is constant.} \tag{75}$$

The beam is, therefore, a curve with a constant curvature, that is, circular arc. If an acoustic wave is emitted from a source at point A, it propagates along a circular ray to point B at the same depth h_1. The horizontal distance x_{AB} is

$$x_{AB} = \frac{2c_1}{k\tan\alpha_1}, \tag{76}$$

then it continues along a next circular arc to point C at the depth h_1, etc. It represents a waveguide through which the wave propagates. However, this applies only in the plane perpendicular to the water level. In other directions, this circular curvature does not apply, because in them the velocity gradient is zero.

Note: We observe a similar phenomenon with light in a medium with an inhomogeneous distribution of the refractive index. Like in the example, light propagates in a gradient optical fibre whose refractive index is greatest at the centre of the fibre and decreases towards its surface. If light enters the fibre with a sufficiently small angle α1, the light beam is curved and does not leave the fibre, i.e., propagates in the fibre without attenuation associated with radiation.

Elements of optical systems with a modulated refractive index, are mainly used in micro-optical applications. As mentioned, the basic element with a transversally modulated refractive index is an optical fibre.

Another element, especially of endoscopic imaging systems, is the GRIN lens (Gradient Refraction Index). The principle of focusing is indicated in **Figure 20**, like Example 16. In the middle is a sample of a GRIN lens with its typical dimensions. On the right is a sample of the endoscope tip.

GRIN lenses are particularly advantageous in that they have very small dimensions and do not require any specially shaped surface like conventional lenses. By suitable profiling of the refractive index, a special shaping of the optical beam can be achieved. **Figure 20** indicates the use of GRIN technology to implement a converging lens. This application is mainly used as an end optical element of endoscopes.
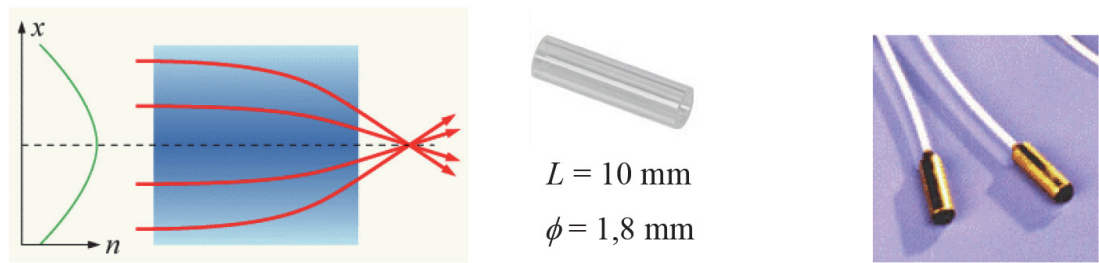
$L = 10$ mm

$\phi = 1,8$ mm

Figure 20.
GRIN lens.

6.2.3 Endoscopy

The endoscope is an optical imaging device that allows investigating internal organs inaccessible to direct observation. The endoscope is a cable that is inserted into the examined area. It contains illumination of the investigated area and a camera for its displaying. Typical applications are gastroscopy (endoscope inserted into the oesophagus), colonoscope (endoscope applied through the rectum to the large intestine), and the like. In the endoscope, the GRIN lens has two different functions. They are used as collimators to illuminate the field of view. A laser beam propagates through the optical fibre, and the GRIN lens extends it into a diverging light beam. The second function is in getting the image, where it has the same function as the camera objective.

The GRIN lens displays the subject in the image plane, from which it is captured. Two techniques are used. Most endoscopes are fiberscopes that introduce images into the optical fibre bundle (about 100,000 fibres with a diameter of several µm) in the image plane, and the fibres guide the light of the individual pixels of the image to a detection device at the outer end of the endoscope. The second option is to use a CCD sensor that transforms the image in the image plane into an electrical signal. Fiber-scopes have a smaller head and a higher signal-to-noise ratio. They are mainly used in laparoscopic operations and endoscopy of fine structures (in angiography, neurology, etc.). The CCD chip has larger dimensions and a lower quality signal, but it allows an electric connection with an output. It is used, for example, in endoscopic capsules, **Figure 21**. The capsule with a diameter of about 10 mm contains an optical system with a GRIN lens and LED illumination around the circumference (see the front view of the capsule). The image is captured using a CCD, and the electric signal is

A picture of the small intestine
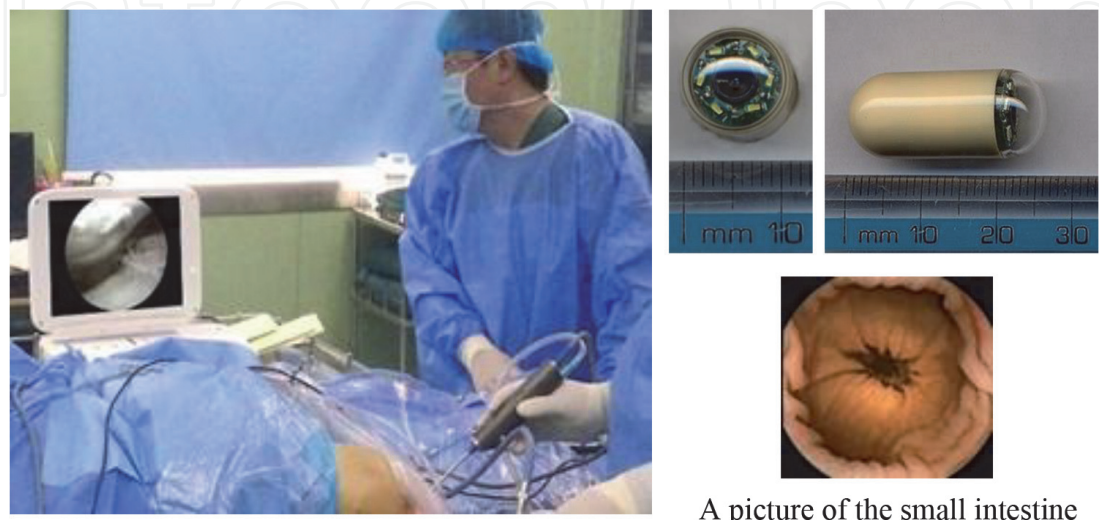
Figure 21.
Endoscopy (on the left a classic endoscope during laparoscopic surgery, on the right a wireless endoscopic probe and an image of the small intestine created by this probe).

transmitted via Bluetooth to an external sensor on the body surface. The capsule has an internal power supply. It is used to investigate the small intestine, which cannot be reached by a standard endoscopic probe, gastroscopy of the stomach and duodenum, or colonoscopy of the large intestine. After swallowing, the probe moves through the digestive tract and transmits images of the intestinal wall.

6.2.4 Optical system of the telescope

A telescope is an optical system that is mainly used to image distant objects or to expand or compress a wave beam.

The basic set of the telescope consists of two lenses, where the image focus F_1' of the first lens (objective) is identical with the object focus F_2 of the second lens (eyepiece). Lens telescope has an objective converging lens. The first Galileo telescope used an eyepiece the diverging lens, **Figure 22a**. The focal length f_1 of the objective was greater than the focal length f_2 of the eyepiece.

Then the angle φ_1 of view of the distant object is smaller than the view angle φ_2 at which we observe the object through the telescope. We see, there occurs an angular magnification. A small field of view expands, and so it is possible to observe details that cannot be seen by the naked eyes. As can be seen from the image, Galileo's telescope creates a direct image. When observing, for example, stars, or in compression of the wave beam, the inversion of the image does not matter, and a Kepler telescope with an eyepiece converging lens can be used, **Figure 22b**. This telescope has better optical properties but is structurally longer and produces an inverted image. An increase in the angular magnifying is in **Figure 23c**.

The disadvantage of lens binoculars is the loss of part of the power of the incident radiation by reflection at the lens-air interfaces. This is solved by mirror telescopes, in which the objective lens is replaced by a hollow mirror (Newton's or Cassegrain's telescope). The intensity of the image depends on the aperture of the lens $S_1 = \dfrac{\pi d_1^2}{4}$, which catches the power $P_1 = S_1 I_1$. If we do not consider losses, this power is then radiated by an eyepiece with area $S_2 = \dfrac{\pi d_2^2}{4}$, **Figure 22d**.
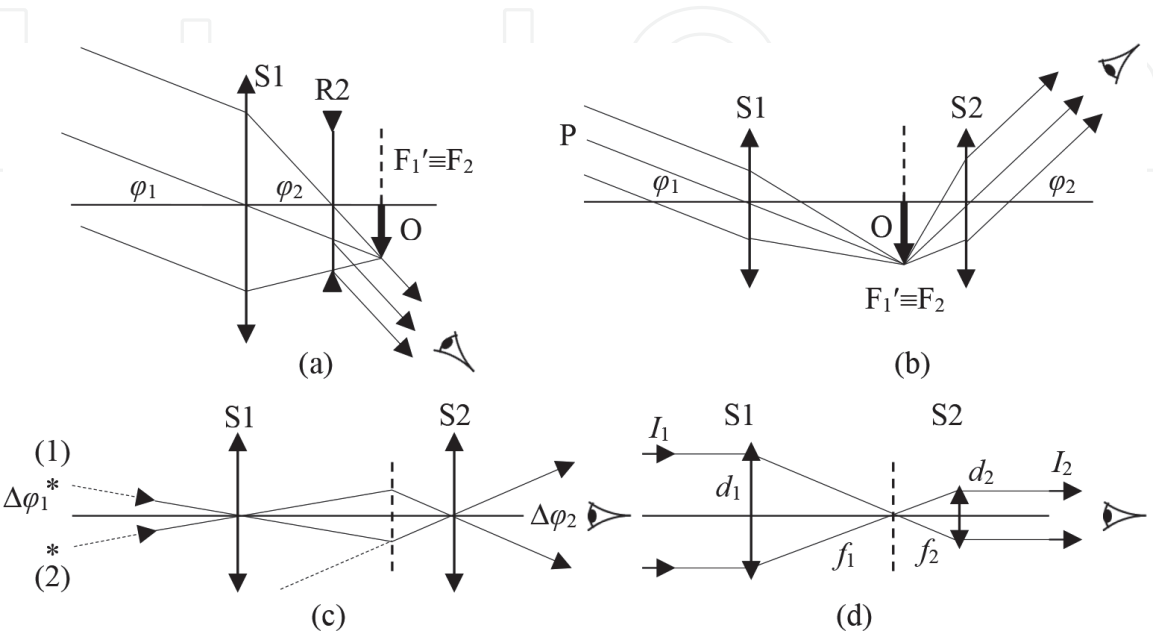

Figure 22.
Display the telescope function.

The intensity of the compressed beam

$$I_2 = \frac{P_1}{S_2} = I_1 \ \frac{d_1^2}{d_2^2} = I_1 \left(\frac{f_1}{f_2} \right)^2.$$

For d_2 = 8 mm (the eyepiece diameter corresponds to the size of the pupil of the eye) and d_1 = 50 mm (ordinary hunting telescope) is I_2/I_1 = 40, that is, increases the intensity of the incident light 40 times. In the dark, details visible through the binocular are invisible to the naked eyes. In mirror telescopes, apertures with a diameter of several meters (the largest astronomical telescopes) can be realised, and thus observe distant objects on the edge of the Universe.

Binoculars are also used together with a recording device mounted behind the eyepiece instead of the eye, or with a Doppler analyser, which allows determining the speed of movement of the observed object. Such a device uses, for example, police to measure speed and record road vehicles. The camera (photo or video) with a telephoto objective also performs a similar function.

6.2.5 Optical system of the microscope

Unlike the telescope, the microscope is used to image very small and near objects. We will explain the principle of the optical microscope, but the same principle uses infrared or acoustic one.

The microscope, **Figure 23**, consists of two converging lenses S1 and S2. There is an optical interval Δ between the image focus F_1' of the objective S1 and the object focus F_2 of the eyepiece S2. Subject P is in front of the focus F_1 of the objective. The lens S1 forms a real image O, and the focal plane of the eyepiece (F_2) is adjusted into this image plane. The eyepiece then forms a beam of parallel rays, which form an angle of view φ_2 to the optical axis. From the conventional distance of view l_k = 25 cm, the same object can be seen at a viewing angle φ_1. The angular magnification of the microscope is defined as $z = \varphi_2/\varphi_1 \gg 1$. A red blood cell with size d = 7.2 μm is observed from the distance l_k at an angle $\varphi_1 \approx 0.1'$ (angular minutes). The resolution of the eye is $1'$, that is, the blood cell cannot be seen with the naked eyes. Using a microscope with magnification z = 100, the angle of view is $\varphi_2 \approx 10'$, that is, the shape of the blood cell is visible under a microscope. Optical microscopes have a magnification z up to 2000. Despite the high magnification, objects of the order of less than the wavelength of light cannot be observed due to diffraction. An increase in the cut-off resolution is achieved by shortening the wavelength (minimum wavelength of light is approximately
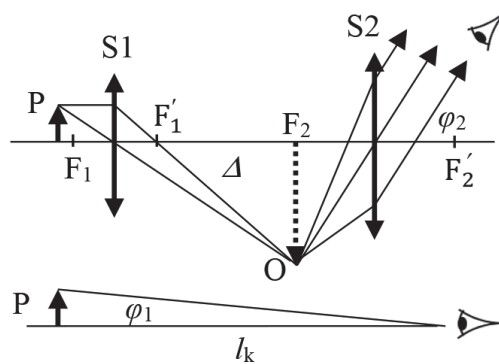

Figure 23.
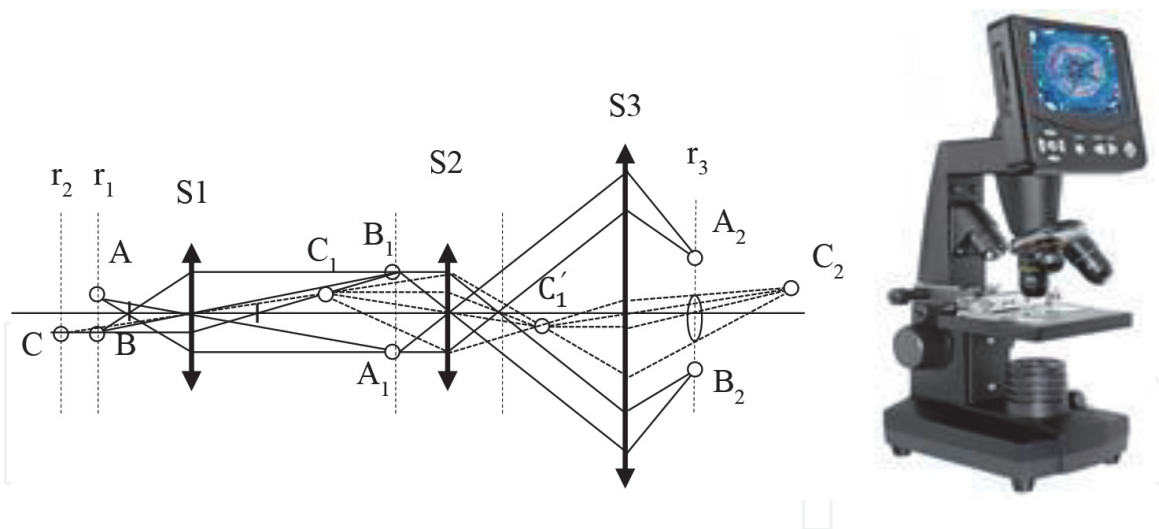*Principle of microscope imaging.*

Figure 24.
Imaging by a microscope and depth of field.

350 nm). Extremely high resolution is achieved using an electron microscope, which achieves a resolution of up to 10^{-9} m, that is, allows to display viruses (15–300 nm), macromolecules, and the like. Top electron microscopes achieve magnification up to 10^6 and resolution up to 0.5 nm. Although the electron microscope image is in principle like that of a scanning optical microscope, the function of the apparatus, which uses an electron beam instead of an optical one, is considerably more complicated.

Conventional microscopes are used to image only thin planar (2D) samples. The image of the volume (3D) sample is blurred. **Figure 24** is the construction of images of objects that lie in different places and object planes r_1 and r_2. Let us first consider two objects A, B in one object plane r_1 the microscope is focused on. If we observe the course of the characteristic rays, objective S1 creates real images A_1, B_1 in the focal plane of the eyepiece. Behind the eyepiece, beams of parallel rays at different angles concerning the axis correspond to each of the objects. There is another imaging lens behind the eyepiece, either the lens of the eye in the case of visual observation or the camera lens if the image is recorded on a recording medium. In modern digital microscopes, the recording medium is a CCD chip. We can see that the lens S3 focuses the rays sharply to the points A_2, B_2 in the only image plane r3.

Even in this simple figure, we see that the distance A_2B_2 is greater than that of AB. The image is then projected on the LCD screen, or the digital recording is transferred to a computer. The advantage of a digital microscope is that it allows the image to be saved to disk, archived, or further processed.

Consider now point C, which is in another plane r_2 of the 3D object. If we observe the course of the characteristic rays, we see that the objective S1 creates a real image C_1, which is displayed by the eyepiece to a point C_1' at the intersection of the converging rays. The lens S3 then creates an image C_2 behind the detection plane. The beam pointing to point C_2 covers the trace indicated by a little ellipse in the detection plane. Object C thus does not create a sharp point image in the detection plane, but a wider track. Points that are in an object plane other than r_1 thus cover the resulting image in the detection plane with such surface tracks and blur the image.

Thus, a conventional microscope is not suitable for imaging 3D objects. However, it is suitable for imaging thin planar slices or liquid samples placed between two microscope slides, for example, in the investigations of various histological findings, blood particles, movement of sperm in the semen, the occurrence of chromosomes in cells in genetic tests, etc. Conventional optical microscopes are used in precision surgery, ophthalmology, neurology, dentistry, and the like.

6.2.6 Confocal microscope

A digital confocal microscope (CLSM—Confocal Laser Scanning Microscopy) is used to observe 3D objects. The scheme of the microscope is in **Figure 25a**. There is displayed only point A of the object plane. The overall image of a given sample plane is obtained by scanning the sample in two x, y directions perpendicular to the system axis. The confocal microscope displays only the points of the object plane on which the microscope is focused. Points from other planes of the object are suppressed and do not blur the resulting image (like a conventional microscope). In **Figure 25** are, for comparison, images of a defect on the skin surface: (b) from a conventional microscope and (c) from a confocal microscope focused on different planes of the object. It can be seen from the images that the confocal microscope images are significantly sharper than the image (b). By moving the sample in the z-direction, we get images of different planes of the sample layer by layer. By combining these images in a computer, a 3D image of the object can be created. It is an optical tomography. An example of a 3D image of cancer cells is shown in **Figure 25d**. A confocal microscope is standard equipment in the pathology department and the biochemical laboratory.

The principle of imaging is in **Figure 25a**. The laser beam of parallel beams is focused by the lens L1 on a small aperture of the screen S1, thus creating a point source of coherent light. The diverging beam is reflected by the semi-transparent mirror SM onto the lens L2. It focuses the rays to point A in the object plane. At point A, the light is reflected or scattered, and the rays going back from this point of the object incident on the lens L2 are returned to the semi-transparent mirror SM. Some of the light reflects towards the laser, but now we are interested in the light passing through the semi-transparent mirror.

At the same distance from the SM as the screen S1, there is a screen S2 with a small hole in the straight direction. Since the beams from point A are focused on the centre of its aperture, only the light from point A passes through the aperture to the lens L3 of the recording detector. The magnitude of the detector signal is directly
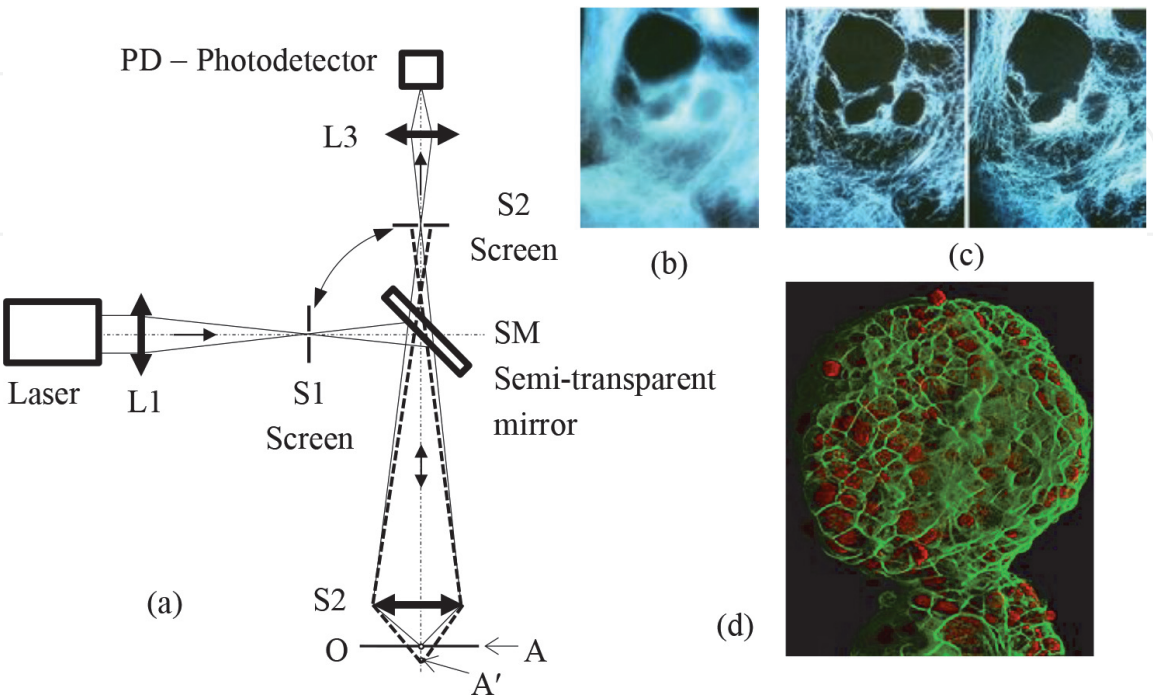

Figure 25.
Confocal microscope.

proportional to the reflectivity of point A of the object. By moving the sample (scanning), the detector signal is changed, which is digitised and recorded.

The laser light incident on the sample illuminates also points outside point A. The light emanating from point A′ at another depth also hits the lens L2 but forms a converging beam (dashed in the figure) that is not focused on the small aperture S2. It does not, therefore, pass through the aperture. The light emanating from a depth other than the depth of point A thus does not affect the detector signal and, therefore, cannot blur the image of the O plane. Similarly, the points in the plane O sideways from point A form an emanating beam, which is deflected by direction and thus also does not pass through the aperture S2. The detector signal thus corresponds only to point A of the displayed sample and its very close surroundings. The created digital image is, therefore, very sharp.

6.2.7 Microscope with phase contrast

The subjects of observation are often samples, especially in biochemistry and biology, which contain transparent objects like cells, small microorganisms, etc. A standard microscope evaluates the intensity (amplitude) of light, and this is not affected by transparent objects. Such objects are very faintly visible in the sample image, **Figure 26a**. However, the investigated objects have a different refractive index than the surrounding medium. As light passes through the sample, a different phase shift of the light beam occurs. If such a beam can interfere with a suitably arranged reference beam, the light may be increased or reduced depending on whether the interference is constructive or destructive. Thus, the phase shift caused by the different refractive index of the object is converted into amplitude information visible in the resulting image. In this way, originally invisible objects become visible—phase contrast is achieved, **Figure 26b**. The phase-contrast microscopes belong to the fundamental equipment of laboratories working with biological samples (pathology, biochemistry, etc.)

6.2.8 Fluorescence microscope

The fluorescence microscope is used to image organic and inorganic components of samples that are characterised by fluorescence. It uses the excitation of the fluorescent components of the object by ultraviolet radiation and detects the visible light that the substances emit when relaxed. The construction of the fluorescence microscope is like that of a conventional one. It is only extended by a source of UV radiation, **Figure 27**. The light from the UV lamp passes through the filter F1. UV
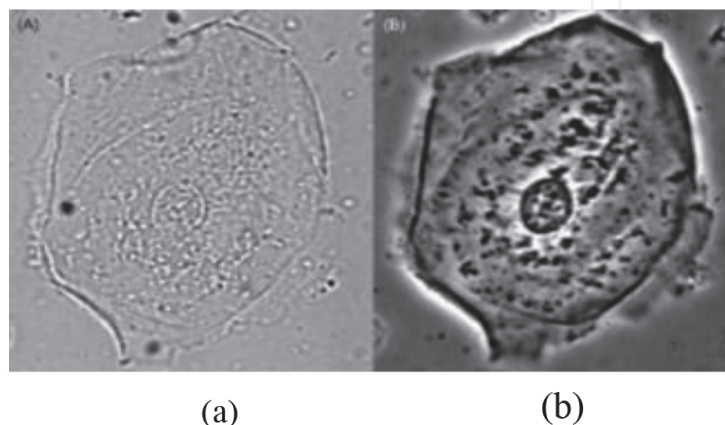

(a)   (b)

Figure 26.
Image of a transparent sample created by (a) amplitude sensitive microscope, (b) phase-contrast microscope.
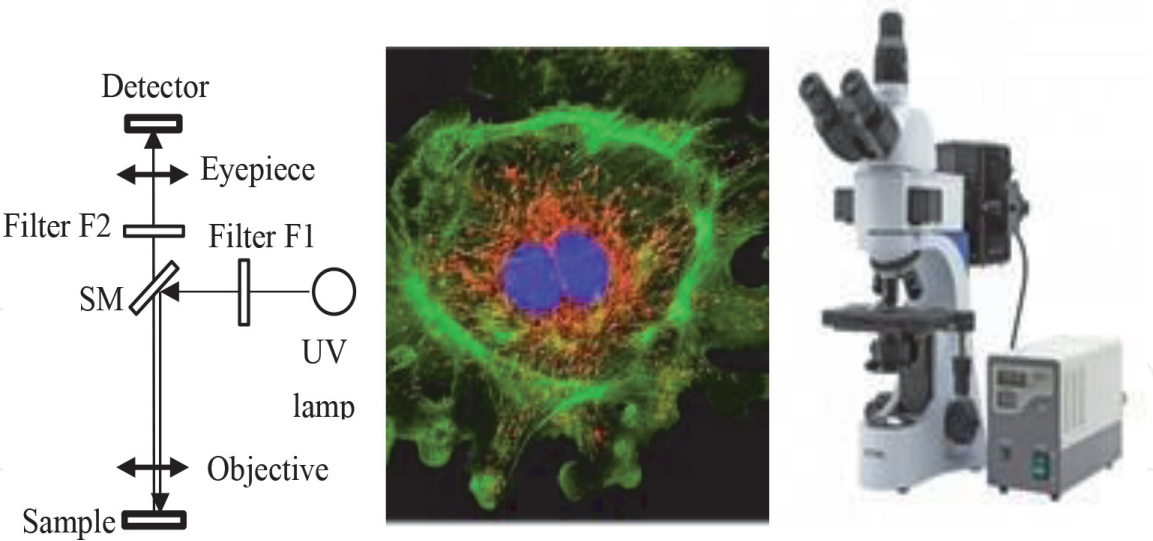
Figure 27.
Fluorescence microscope.

light with the selected wavelength proceeds to the semipermeable mirror SM and is reflected towards the sample. It induces fluorescence in the sample. The emitted visible light enters the microscope objective and passes through the SM to the optical filter F2, the eyepiece, and the detector. Filter F2 suppresses reflected UV radiation. The image is observed directly with eyes or captured with a digital camera. The parts of the sample containing a fluorescent substance are displayed markedly in the resulting image. Since different molecules or cellular components have different excitation energy, it is possible to change the display of the individual fluorescent components by retuning the F1 filter.

In **Figure 27** right is an apparatus for direct observation or with digital output and display of the observed sample on the LCD. The figure also shows an image of a cell with structures that emit light of different wavelengths (colours). In some cases, the primary fluorescence of certain molecules (e.g., some proteins—amino acids) is used. More often objects are "coloured" with fluorescent dyes, which penetrate biological structures and make them visible. In this way, the processes taking place in them can also be observed (for example, nucleic acids—DNA, RNA—are visualised). Fluorescence microscopes are mainly used in biological research laboratories.

6.3 Resolution of the wave imaging

If we use waves to image a certain structure, we are interested in what details can be observed, that is, resolution. Suppose we observe a small circular detail with radius r on the surface of the sample. The light reflected from the circle produces a characteristic diffraction pattern consisting of a zero maximum and higher-order diffraction maxima, see Section 5.1.2. For the reconstruction corresponding at least approximately to the observed formation, we must capture at least the main lobe of the spectrum, which contains the most significant spectral components, **Figure 13**. Therefore, if we want to distinguish the detail, we must capture a beam of the rays of the main lobe up to the first minimum, so-called *Airi disk*, which falls on the lens of the detection device (eye, microscope, camera lens), **Figure 28**.

For the first minimum, the relation (30) applies. If the lens is at a focal length f from the detail and R is the radius of the lens, the cut-off angle is $\varphi = \arctan(R/f)$. The requirement for the circle resolution (display resolution) is $\varphi \approx \varphi_{1\,min}$. If we consider the maximum value $\varphi \approx 45°$, that is, $R = f$, we get the resolution limit of a circular detail with diameter d the condition $d = 2R \approx 1.22\lambda$.
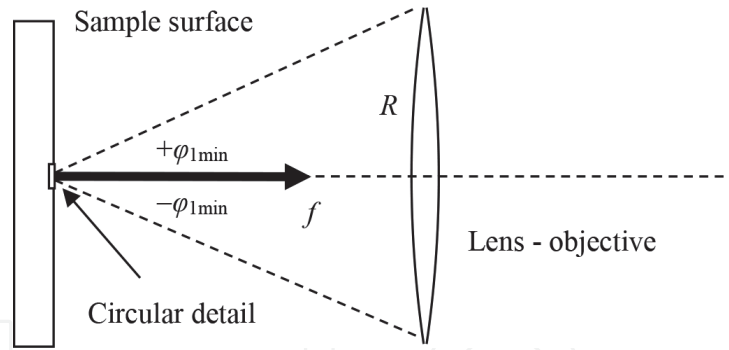
Figure 28.
Reconstruction of the diffraction pattern.

The main information of the zero-order lobe is in its width at the level of -3 dB (50%). The contours of the image of the circular detail will no longer be sharp but can still be distinguished. For $I/I_m = 0.5$, relation (29) and **Figure 13**, we get the value of the parameter $\gamma \approx 1.16$. Using (29) and estimating $\varphi = 45°$ we get the resolution limit of the circular detail $d_{min} \approx 0.37\lambda$.

It follows this analysis that wave imaging allows observing objects with minimum dimensions of about one-half of the wavelength of the used wave. With an optical microscope, details with dimensions of up to about 200 nm can be observed in blue light.

If we observe a sample with naked eyes, the radius of the pupil is approximately $R = 2$ mm and the distance of comfortable observation $f = 25$ cm, which corresponds to an angle $\varphi \approx 8 \times 10^{-3}$ rad. At the wavelength $\lambda = 500$ nm, the resolution is $d \approx 46$ $\lambda \approx 0.02$ mm. A mesh having a fineness of about 0.05 mm can be seen with the naked eye. If we observe a finer structure, we cannot capture the diffraction maxima of small details with the eye lens, and thus we see only a homogeneous surface without any structure.

The resolution increases if the object can be observed from a closer distance so that the if angle φ is as large as possible. A magnifying glass or a microscope is used to increase the viewing angle of the observed object. Microscope objectives are therefore designed to meet the condition of maximum aperture, that is, $R \approx f$.

In edge cases, blue light is preferable to red one. X-rays provide greater resolution than light, but the device is technically more complicated. A significant rise in resolution provides electron beam imaging (electron microscopy), in which the electron wavelength λ can be set by the accelerating voltage U

$$\lambda = \frac{h}{\sqrt{2meU}}, \tag{77}$$

where h is the Planck constant, m and e are the mass and charge of the electron. For the voltage $U = 100$ kV we get the wavelength $\lambda \approx 4 \times 10^{-12}$ m, which is two orders of magnitude less than the size of individual atoms or interatomic distances in substances.

Bacteria, cells, larger viruses, etc., can be observed with an optical microscope. The electron microscope can image such small objects as macromolecules (e.g., DNA), proteins, small viruses, etc. The mentioned resolution rule also applies to ultrasound. USG imaging uses ultrasound with frequencies of 2–10 MHz. For fine structures, for example, eyes or nerves, higher frequencies > 10 MHz are used.

6.4 Focusing a wave beam

If the source is a strip, the wavefield described by the relation (26) arises. For a circular source (e.g., a circular hole or a laser output hole), the relevant relation (29) for the spatial distribution of the wavefield is similar, with Bessel functions instead of harmonic ones. We can see from the shape of the resulting relation for the strip source that the relation (25) for the angular dependence of the radiation characteristic corresponds to a Fourier image of a rectangular pulse with length a. If we do a reverse Fourier transform of this image, we get a reconstructed image that has the same structure as the source, that is, the strip with the width a. When reconstructing an impulse from its Fourier image, it is necessary to use at least the main part of the zero-lobe from the diffraction image according to the Eq. (27). The reverse transformation can be realised by a lens, cylindrical in the case of a strip, and spherical in the case of a circular source (**Figure 29**).

A lens with a circumference diameter D and a focal length f has a beam convergence angle $\Delta\varphi$ given by the relation when focusing a plane-wave beam

$$\tan\frac{\Delta\varphi}{2} = \frac{D}{2f}. \tag{78}$$

Using Eq. (27) we get a condition

$$\frac{\Delta\varphi}{2} = \arctan\frac{D}{2f} = \arcsin\left(0.44\frac{\lambda}{a}\right) \tag{79}$$

and after adjustment

$$a = 0.44\lambda\sqrt{1+\left(\frac{2f}{D}\right)^2}. \tag{80}$$

The $2f/D$ ratio acquires a real minimum value of about one, then the minimum width of the focused track is the $a_{\min} \approx 0.6\lambda$. As the $2f/D$ ratio increases (i.e., the angle $\Delta\varphi$ decreases), the width of the focus track increases. In the case of a circular lens, the results are similar. In the optimal case, the wave beam can be focused on a track with a dimension equal to approximately half the wavelength. This phenomenon occurs regardless of the wavelength of the wave or its nature (mechanical or electromagnetic).

This fact must be respected, for example, in gentle operations with a laser scalpel in neurology, eye surgery, etc. We get a smaller focused track using blue light, in confront with a red light or IR radiation. The condition of the best focus applies to coherent waves, for example, laser. When using incoherent radiation, the possibility of focusing the beam is always worse (e.g., sunlight or LED).
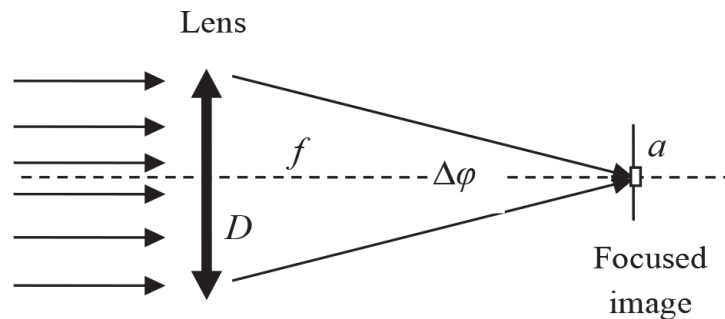

Figure 29.
Focusing a wave beam.

7. Non-linear phenomena

7.1 Non-linear wave interactions

In a linear medium, wave interference occurs, which means the waves add at the place of their common occurrence, but the waves themselves do not affect each other. However, if the medium is non-linear, the wave is deformed. It leads to arising higher harmonic components, and due to a non-linear interaction to different combination frequencies. These combination components may propagate in different directions relating to the direction of the original waves. These interactions are most often described by the quantum model. For example, the acoustic wave modulates the optical refractive index of the medium, and the electromagnetic wave thus passes through the medium with a periodic change of the propagation parameter, like through an optical grating. It causes a diffraction pattern, that is, the main lobe and lateral ones. Due to the movement of the grating at the speed of the acoustic wave, there arises also a shift of the EM wave frequency by the frequency of the acoustic wave.

Example 17 Acoustic modulator of the laser beam

In some information transmission by light, for example, optical fibre, it is necessary to modulate the light beam frequency. One way is an acoustic-optical modulator, which uses the nonlinearity of the modulation crystal, **Figure 30**. The carrier of the signal is a frequency-modulated acoustic wave, which causes a temporal change of the refractive index of the crystal corresponding to a transient change of the signal. A laser beam enters this crystal perpendicularly to the direction of propagation of the acoustic wave and is diffracted. The modulated first-order diffraction beam is used, which is angularly separated from the 0-order beam.

The principle of the deflector (modulator) is easily explained using the quantum concept of waves. EM waves represent photons with energy $E = \hbar\omega$ and momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$, and acoustic waves phonons with energy $E_{\mathrm{ph}} = \hbar\Omega$ and momentum $\boldsymbol{p}_{\mathrm{ph}} = \hbar\boldsymbol{K}$.

In the interaction, where the photon captures the phonon, the law of conservation of energy and momentum is fulfilled.

$$\hbar\omega_0 + \hbar\Omega = \hbar\omega$$
$$\hbar\mathbf{k}_0 + \hbar\mathbf{K} = \hbar\mathbf{k} \tag{81}$$

If we consider only a very small change in frequency $\Omega \ll \omega_0$ is $p \approx p_0$ and according to **Figure 30** is

$$\omega = \omega_0 + \Omega \tag{82}$$

and $2k \sin \frac{\varphi}{2} = K$, or $\varphi = 2 \arcsin\left(\frac{\lambda}{2\Lambda}\right)$, for $\lambda \ll \Lambda$ we have $\varphi \approx \frac{\lambda}{\Lambda}$.
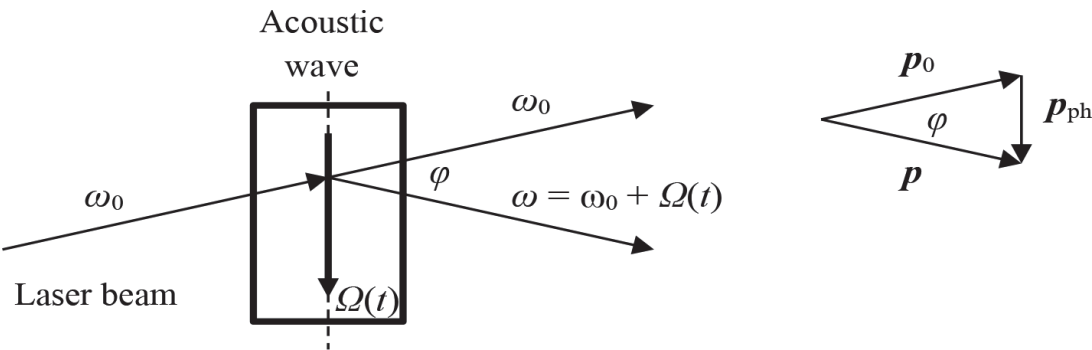

Figure 30.
Modulator of the laser beam.

The deflected beam is frequency modulated by an acoustic signal and angularly separated from the unmodulated 0-th order beam.

Example 18 Two-photon excitation microscopy

Another example of the nonlinear interaction of a laser beam with a substance is fluorescence microscopy used in biochemistry, which uses the principle of two-photon excitation of atoms. Conventional fluorescence microscopy—see Section 6.2.8, utilises excitation by high energy photons of UV radiation, which have a shorter depth of penetration into the sample due to the higher frequency and which produce a relatively strong white background of the image.

Two-photon excitation uses the excitation of molecules by a pair of low-energy photons. If we observe the fluorescence of the most common fluorophores (molecules exhibiting fluorescence) in the optical range of 400–500 nm, IR laser radiation of 800–1000 nm is used for excitation. If an electron captures a photon in the ground state, it is excited to a higher energy level. The electrons from this excitation level relax back to the ground state, emitting photons in the IR band, that is, optically invisible.

However, if it is excited again to the next higher energy level before relaxation, due to the relaxation from this second level photons of the visible light are re-emitted—fluorescence occurs. Each fluorescent molecule has its characteristic excitation energy, so that specific fluorophore can be discovered. The probability of double capture of excitation photons is very small and increases with an increasing energy density of radiation. The conditions for double excitation are only in the focus of the focused laser beam. In such a way, the location of the fluorophore can be precisely addressed. The method of scanning individual layers of the sample is used for the display. It permits creating a 3D image of the structure.

7.2 Photo-acoustic imaging

Other modern diagnostic tools use a nonlinear photoacoustic (PA) effect, for example, Xu [3], Beard [4]. As mentioned in Chapter 3, the ultrasound can be excited thermally (by thermal pulse) using a powerful pulse laser. The laser focuses at a location on or near the surface of the object and emits a very short (nanosecond) pulse of radiation. The absorbed energy causes a local increase in temperature, and thus a mechanical expansion of the material. This thermal deformation causes a mechanical shock wave to propagate from this point to all sides. If a series of pulses is applied, a series of shock waves is generated which has a harmonic ultrasonic component with a frequency equal to the repetition rate of the pulses. Since the pulse length is significantly shorter than the repetition period, it is sufficient to return the temperature to the original site temperature before the next pulse arrives. The amplitude of the generated mechanical wave depends on the amount of energy absorbed, and this is directly proportional to the absorption coefficient (absorbance) of the substance at a given location. The generated mechanical (ultrasonic) waves are detected by ultrasonic sensors. In this way, the absorbance distribution of the sample, which is related to the tissue structure, can be mapped using a photoacoustic phenomenon. For example, the blood has a large absorbance for a red light, and therefore the bloodstream and small blood capillaries are displayed very well. The method has a very good resolution due to the possibility of focusing the light beam on an area with dimensions of units of μm. Another advantage is that, unlike light, the attenuation of mechanical waves is relatively small and the generated wave shocks propagate well through an investigated sample. Some specific imaging devices work on the principle of the photoacoustic phenomenon. Some of them are listed below.

7.2.1 Photoacoustic microscopy

The first method that uses the principle of the photoacoustic phenomenon is photoacoustic microscopy (PAM), for example, Wang [5]. The laser beam and the detection acoustic transducer are focused on the examined point of the sample. The energy of the laser beam is thus concentrated to one point and a shock wave with a repetition frequency of 500 MHz, is detected by an acoustic transducer focused at the same point. PAM is a scanning method. The sample is shifted in the transverse directions x, y, and the acoustic transducer signal is recorded. It is also possible to adjust the optical and acoustic lens holder perpendicularly to the sample surface and to image the sample structure at different depths below the surface.

PAM allows displaying a depth up 3–4 mm. By composing images from different depths, a 3D image can be created. **Figure 31** is the basic principle of PAM. The box with an optical lens and a focused acoustic transducer contains water as a transmission medium for the acoustic wave. The box is acoustically bound to the displayed object using a binding gel. Point A is the common focus and displayed point. The picture shows four scans from different depths and a 3D reconstruction.
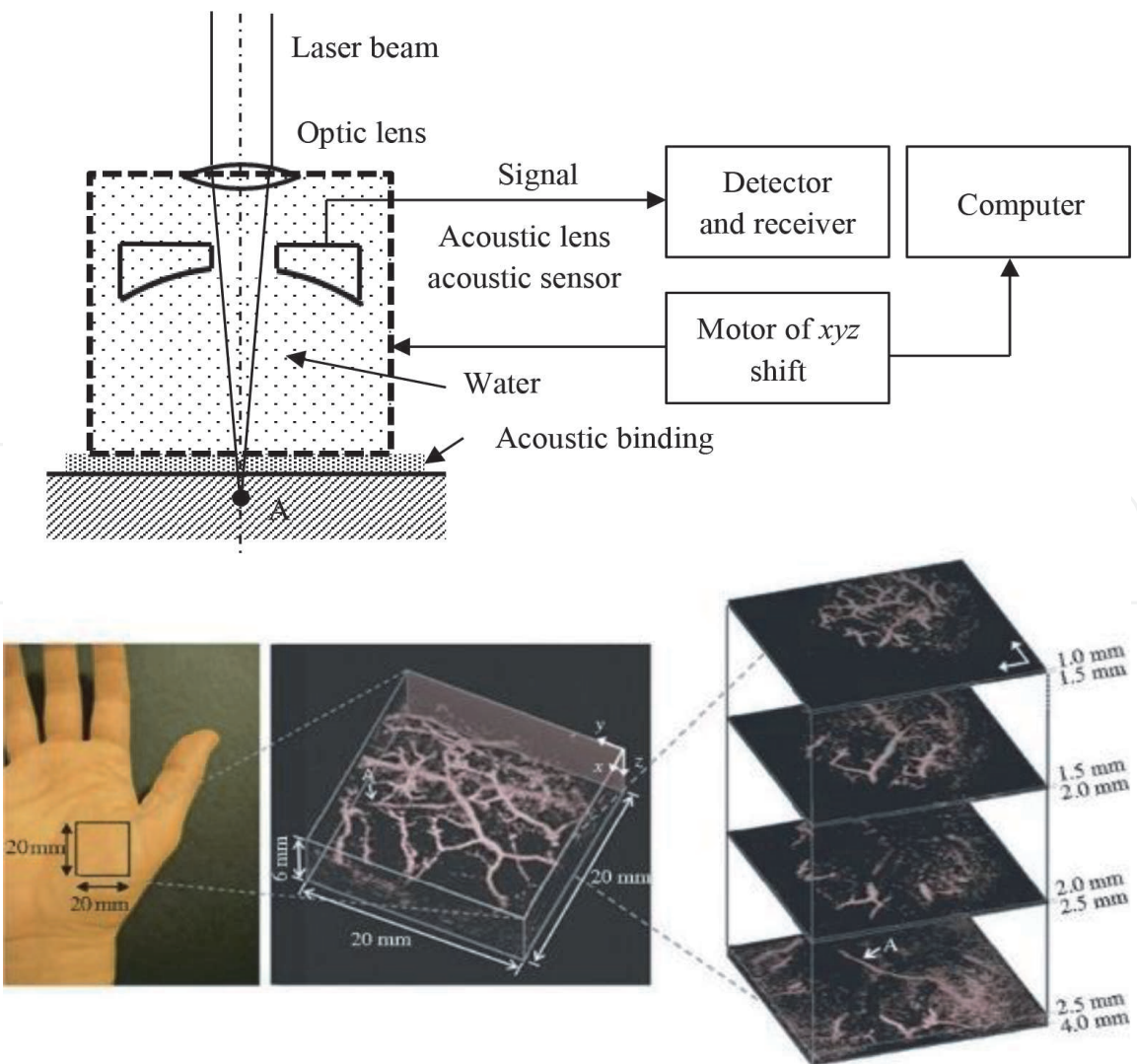

Figure 31.
Principal of photoacoustic imaging—PAM.

7.2.2 Photoacoustic computing tomography

The second principle of information processing of a photo-excited sample is photoacoustic computed tomography (PACT). Unlike PAM, the light pulse of the power laser illuminates the entire examined surface of the sample. As a result, all points with higher absorbance in the sample become a source of acoustic waves. The sample is surrounded by a system, for example, 512 small acoustic detectors, arranged on a hemisphere to obtain a 3D image, or on a circular ring around the sample to create a 2D image. The signals of all detectors are scanned and fed to a computer. They are converted by a complex integral transformation into a spatial or planar image of the distribution of points with different absorbances. In this way, different tissues in the surface layer of the sample are displayed. PACT can display the structure up to a depth of 50 mm, in the case of using a set of detectors arranged on a hemispherical surface, with a resolution of 0.5 mm at a pulse repetition frequency of 5 MHz.

In **Figure 32** on the left is a 3D image of the tissue perfusion of the top layer of the skin. Different levels of the signal are distinguished in colour, which corresponds to different absorbances of the cells. Such an image allows, for example, detection of skin tumour (melanoma)—the picture on the right.

PAM and PACT provide similar resolution as ultrasonography (USG). Against USG, they have a much smaller depth of view, which is limited by the attenuation of light in the tissue but allows seeing the information that USG does not provide. This is due to another mechanism of ultrasound generation. While different acoustic impedances of tissues are applied in USG, different absorbance of tissues is decisive for PA methods. Photoacoustic imaging methods are not yet widely used but are the subject of intensive research.

7.2.3 Photoacoustic spectroscopy

Another method that uses the photoacoustic phenomenon is photoacoustic spectroscopy (PAS), for example, West [6]. It uses different spectral compositions of absorbance of different substances in the tissue.

Figure 33 is the spectrum of various substances in the blood: melanin, Hb-haemoglobin, Mb-myoglobin, and others. The difference in the absorbance of HbO_2 and HbR (oxidised and reduced haemoglobin) at a light wavelength of about 680 nm can be used to distinguish oxygenated blood from deoxygenated blood. In this way, it is possible to determine the degree of oxygenation of the blood—*oximetry*.
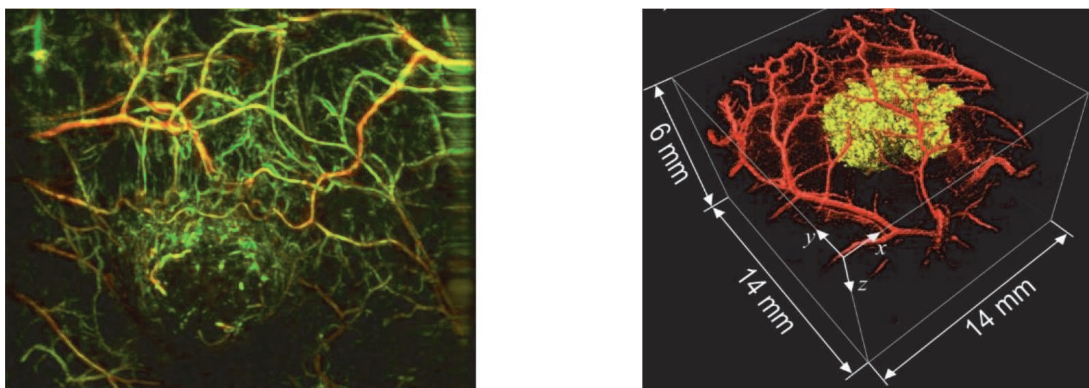

Figure 32.
PACT—on the left image of the surface layer of the skin, right view of melanoma cells (source: https://www.charite.de/en/service/press_reports/artikel/detail/photoakustische_bildgebung_ermoeglicht_tiefen_blick_ins_gewebe/).
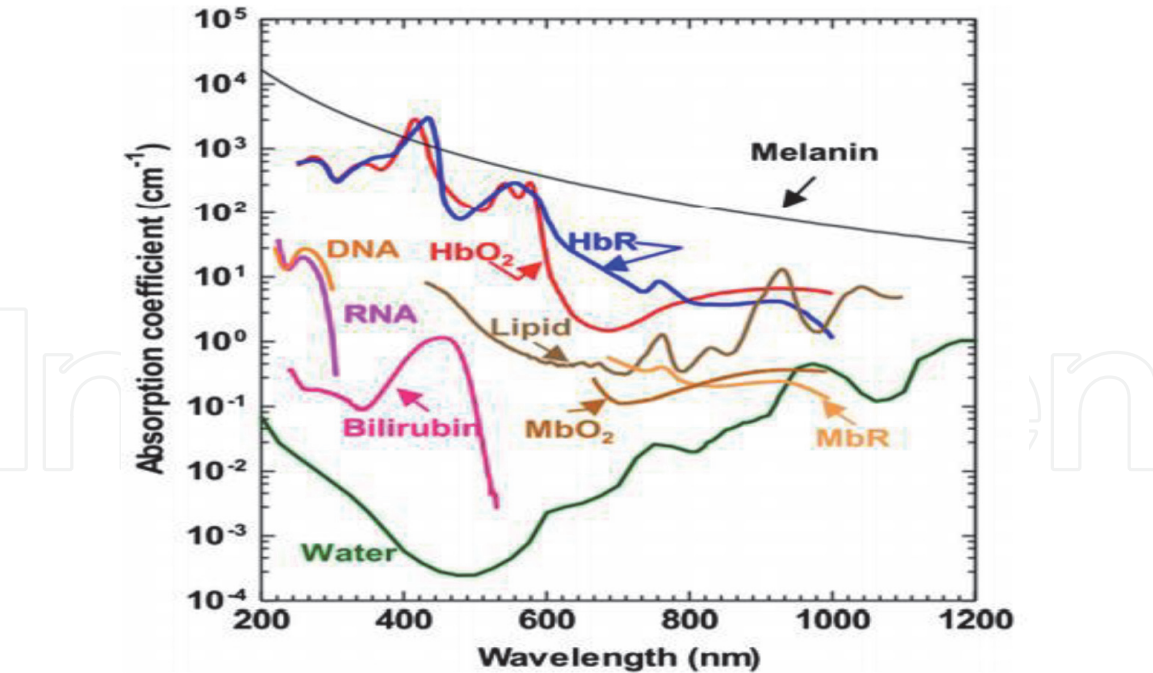
Figure 33.
The spectrum of absorbance of different substances in the blood.

Author details

Ivo Čáp*, Klára Čápová, Milan Smetana and Štefan Borik
Faculty of Electrical Engineering and Information Technology, University of Žilina, Slovakia

*Address all correspondence to: ivo.cap@fel.uniza.sk

IntechOpen

References

[1] Wyatt CL. Electro-Optical System Design—For Information Processing. New York: McGraw-Hill; 1991. ISBN 978-0-0707-2184-5

[2] Splinter R, Hooper BA. An Introduction to Biomedical Optics. Boca Raton: CRC Press; 2007. ISBN 978-0-7503-0938-7

[3] Xu M, Wang LV. Photoacoustic imaging in biomedicine. Review of Scientific Instruments. 2006;**77**(4): 041101

[4] Beard P. Biomedical photoacoustic imaging. Interface Focus. 2011;**1**(4): 602-631

[5] Wang LV. Multiscale photoacoustic microscopy and computed tomography. Nature Photonics. 2009;**3**(9):503-509

[6] West GA et al. Photoacoustic spectroscopy. Review of Scientific Instruments. 1983;**54**:797-817