Chapter 2

Optics and Optical Devices

All optical spectroscopy instruments are optical devices in that they use light sources, manipulate the light and measure the light. Optics and optical devices have a long history going back to 17th century when the wave and corpuscular light theories were developed by two famous scientists Christian Huygens and Isaac Newton. Nowadays optics is a well developed branch of natural sciences with numerous subtopics, application fields and wide range of instruments and tools available commercially. Giving its importance for understanding the principles of the optical spectroscopy instruments this Chapter will discuss a few general topics, such as interference and interferometers, diffraction and diffraction resolution limits, monochromators, and calculation of optical systems in geometrical optics approximation. However, this is rather fragmentary selection of optics subjects and readers are advised to refer to general optics text books for more complete study of the subject.[1]

The photon, being a quantum object, has a controversy of wave and particle presentations. Also there are unified theories, it is common to use wave theory to discuss interference or diffraction properties of the light, and to present photons as particles for ray tracing or to study their interactions with matter. Accordingly, the wave presentation of light will be discussed at first, following by its application to interference and diffraction. In the last section we shall switch to geometrical optics to discuss calculations of beam tracing in optical systems.

2.1 Waves

2.1.1 Wave equation

In a simple one dimension case (1D) the wave equation is

$$\frac{\partial^2 f}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = 0 \tag{2.1}$$

where $f = f(x,t)$ is a function of coordinate x and time t, and c is a constant. For example a string vibration can be described by the wave equation, then $f(x,t)$ can be the string

[1] The author used a book by Robert Guenther as a reference [1], though there are many other excellent text books on modern optics.

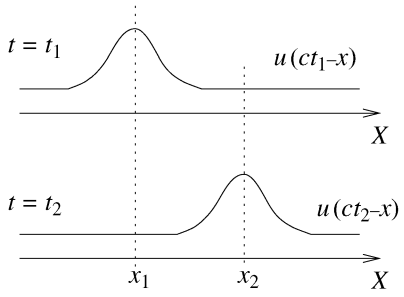

Figure 2.1: Propagation of a pulse along the string (1D wave). The amplitude and the shape of the pulse do not change as the pulse propagates along the string.

displacement at position x. A general solution of the equation is

$$f = u(ct - x) + v(ct + x) \tag{2.2}$$

Where u and v are any functions of a single parameter. These functions present two waves propagating in opposite directions: the wave $u(ct-x)$ propagates in direction of increasing x, and the wave $v(ct+x)$ in decreasing x.

An illustration of a pulse propagating along the string is presented in Fig. 2.1. Let us assume that at time $t = t_1$ the shape of the pulse is given by a pulse-like function $u(y)$ which has a single maximum at y_0. Naturally, in our case the argument of the function u is $y = ct_1 - x$, that is $u = u(ct_1 - x)$. At fixed time $t = t_1$, function $u(ct_1 - x)$ depends only on x. The position of the maximum, x_1, is given by a simple relation $y_0 = ct_1 - x_1$, i. e. at time t_1 the coordinate of the maximum is $x_1 = ct_1 - y_0$. At time $t = t_2$ the shape of the pulse is determined by the same pulse-like function u, although now it reads as $u(ct_2 + x)$. Thus, at time t_2 the maximum of the pulse is at point $x_2 = ct_2 - y_0$.[2] The displacement of maximum $\Delta x = x_2 - x_1$ in time interval $\Delta t = t_2 - t_1$ is $\Delta x = ct_1 - y_0 - (ct_2 - y_0) = c\Delta t$, and the velocity of the pulse propagation is $c = \frac{\Delta x}{\Delta t}$, so the constant c in eq. (2.1) is the wave velocity.

In three dimensional (3D) case, the wave equation is, for any scalar field or potential component,

$$\nabla^2 U - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0 \tag{2.3}$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator and $U = U(x, y, z, t) = U(\vec{r}, t)$ is a function of coordinates and time. This equation describes acoustic waves, for example. The

[2] One can notice that $y_0 = ct_1 - x_1 = ct_2 - x_2$, which means that $ct - x$ is invariant of eq. (2.1).

electric filed is the vector field, for which the wave equation in free space is[3]

$$\nabla^2 \vec{E} - \epsilon_o \mu_o \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \tag{2.4}$$

and the velocity of the electromagnetic waves (e. g. light) in free space is $c = \sqrt{\frac{1}{\epsilon_o \mu_o}}$, where ϵ_0 and μ_0 are permeability and permitivity of vacuum, respectively. In dielectric medium the velocity is $c = \sqrt{\frac{1}{\epsilon \epsilon_o \mu \mu_o}}$.

Solution of 3D wave eq. (2.3) is not as straightforward as for 1D case, since there is infinite number of propagation directions. Usually a concept of the wave front is used to solve eq. (2.3). However, the problem can be simplified considering harmonic waves.

2.1.2 Harmonic waves

Harmonic waves are practically important for spectroscopy applications, since the emission and absorption usually occur in relatively narrow spectrum range. The photon energy determines the frequency of its electromagnetic wave, thus, electro-dynamically, photons are essentially harmonic electro-magnetic oscillations. Another term used in optics to denote harmonic electro-magnetic oscillations is a monochromatic wave.

Harmonic oscillations are given by functions $\sin(\omega t)$ or $\cos(\omega t)$, where $\omega = 2\pi\nu$, and ν is the oscillation frequency and ω is the circular frequency.[4] Another useful and widely used mathematical notation for harmonic oscillations is (Euler formula)

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t) \tag{2.5}$$

For example, a harmonic wave in 1D case can be presented as

$$f = f_o e^{i(\omega t - \kappa x)} \tag{2.6}$$

where the argument of the function was changed to be dimensionless as required by sine, cosine or exponential functions. It is also convenient to use $\omega t - \kappa x$ as the argument in mathematical presentation of harmonic waves since it shows the frequency (ω) of oscillations. In terms of eq. (2.6) the velocity of the wave is $c = \frac{\omega}{\kappa}$, and parameter κ is called (circular) wave number. In space the period of wave, or the wavelength, is $\lambda = \frac{2\pi c}{\omega}$, or $\lambda = \frac{2\pi}{\kappa}$.[5]

2.1.3 Plane waves

Extending the equation of monochromatic wave to 3D case, one can rewrite an equation for plane waves

$$U = U_0 e^{i(\omega t - \vec{\kappa}\vec{r})} \tag{2.7}$$

[3] The electromagnetic waves have two components: electric and magnetic fields. However, as follows from Maxwell equations, these two components are tightly related with each other, and only one of them is needed to describe completely the electromagnetic wave. By convenience the electric component will be used here.

[4] The circular frequency is convenient and preferred notation here as it gives shorter form of equations.

[5] Conversions from the circular to linear frequency and wave number are $\omega = 2\pi\nu$ and $k = 2\pi\kappa$, respectively. An equivalent presentation of harmonic wave is $f = f_0 e^{i2\pi(\nu t - kx)}$, with the wave velocity $c = \frac{\nu}{k}$ and wavelength $\lambda = k^{-1}$, respectively.

Where $\vec{r} = \vec{r}(x, y, z)$ is the vector from the origin of the coordinate system to point with coordinates (x, y, z), and $\vec{\kappa}$ is the wave vector. In isotropic dielectric medium the wave vector, $\vec{\kappa}$, determines the wave propagation direction and its absolute value is $|\vec{\kappa}| = \kappa = \frac{\omega}{c}$, or it is equal to the wave number of the 1D case considered above. One can select the coordinate system so that, e. g., axis Z is directed along the vector $\vec{\kappa}$, then the projections of the vector are $\kappa_x = 0$, $\kappa_y = 0$ and $\kappa_z = |\vec{\kappa}| = \kappa$. Thus, the product $\vec{\kappa}\vec{r} = \kappa_x x + \kappa_y y + \kappa_z z = \kappa_x x = \kappa z$, and eq. (2.7) can be rewritten as $U = U_0 e^{i(\omega t - \kappa z)}$. In other words, by proper selection of the coordinate system, the 3D plane waves can be reduced to 1D waves.

The wave given by eq. (2.7) has infinite wave front and its amplitude, U_o, is a constant in the whole space. This is not very useful (practical) model, usually we like to know how do waves change when propagating through different media, e. g. optical system, lens for instance. Then it is reasonable to limit the size of the wave, i. e. the value U_0 can not be a constant. This can be done by rewriting eq. (2.7) as

$$U = U(\vec{r}) e^{i(\omega t - \vec{\kappa}\vec{r})} \tag{2.8}$$

where $U(\vec{r})$ is a slow function of coordinates (compared to the wavelength) and is called wave amplitude. Substituting eq. (2.8) into eq. (2.3) one can obtain equation for the wave amplitude, $U(\vec{r})$, also known as Helmholtz equation

$$(\nabla^2 + \kappa^2) U = 0 \tag{2.9}$$

This equation is only valid when the function $U(\vec{r})$ is much slower than function $e^{i(\vec{\kappa}\vec{r})}$. Then one can consider only wave amplitude distribution over the space but can omit oscillating part, $e^{i(\omega t - \vec{\kappa}\vec{r})}$.[6]

In optics, we usually can neglect a change of the wave amplitude, $U(\vec{r})$, at distances compatible with the wavelength, λ. In addition oscillations of the electromagnetic field at optical frequencies are much quicker than the time resolution of measuring instruments. Therefore, experimentally available value is power averaged over a space region which is much greater than the wavelength and in a time interval much longer than the wave period.

The energy flow of the electromagnetic field is given by Poynting vector $\vec{S} = \vec{E} \times \vec{H}$. The light intensity[7] is the time average of Poynting vector $I = \left\langle |\vec{S}| \right\rangle$, and it is proportional to the square of the electric field amplitude for the electromagnetic wave, $I \propto E^2_{ampl}$.[8] The intensity is the parameter which is available experimentally and commonly used to measure the light at different points of optical systems. Therefore, in calculations of the electric field we shall finally look for light intensity or equivalent measure describing the electromagnetic wave.

[6] It is also important to notice that the amplitude in eq. (2.9) does not depend on time. Therefore Helmholtz equation describes stationary wave flow.

[7] Here the intensity is power density. However, the term light intensity is ill defined in itself and can be used to refer to different forms of the light power characteristics.

[8] In dielectric medium $I = \left\langle |\vec{S}| \right\rangle = \frac{cc}{2} E^2_{ampl}$.

2.2 Interference

In this section the interference of plane monochromatic waves will be discussed. This means in particular that the wave front is assumed large enough to neglect its distortions at distances characteristic for the interference phenomena. Therefore the wave presentation of eq. (2.7) will be used.

Interference is mutual influence of two or more waves producing certain characteristic phenomena. In the case of electromagnetic waves (e. g. light) the mutual influence is superposition of the electric fields produced by different waves (or by difference sources of the waves). Let us consider two sources of the electromagnetic waves. If one source produces electric field $\vec{E_1}$ at a point $\vec{r}$ and another source produces field $\vec{E_2}$ at the same point, then the total field at this point is $\vec{E} = \vec{E_1} + \vec{E_2}$. Let us further assume that the waves are flat and have the same frequency ω, so that the corresponding wave vectors, $\vec{\kappa}_1$ and $\vec{\kappa}_2$, have the same length, i. e. $|\vec{\kappa}_1| = |\vec{\kappa}_2|$. Then, the electric field of the first wave is $\vec{E_1}(\vec{r},t) = \vec{E_1}e^{i(\omega t - \vec{\kappa}_1\vec{r})}$ and of the second is $\vec{E_2}(\vec{r},t) = \vec{E_2}e^{i(\omega t - \vec{\kappa}_2\vec{r})}$, respectively.[9]

It is important to note that the orientations of the pairs of the vectors $\vec{E_1}$ and $\vec{\kappa}_1$, and $\vec{E_2}$ and $\vec{\kappa}_2$ cannot be arbitrary. In dielectric media the electromagnetic waves are propagating in the direction perpendicular to the plane formed by electric and magnetic field vectors, thus $\vec{E}\cdot\vec{\kappa} = 0$.

For further simplification let us assume that the vectors $\vec{E_1}$ and $\vec{E_2}$ have the same orientation, which means that $\vec{E_1}$ and $\vec{E_2}$ are perpendicular to the plane formed by the vectors $\vec{\kappa}_1$ and $\vec{\kappa}_2$. Therefore, the vector sum can be replaced by the sum of scalar values $\left|\vec{E_1}\right| + \left|\vec{E_2}\right|$. Then, the total field created by these two waves is

$$E(\vec{r},t) = E_1 e^{i(\omega t - \vec{\kappa}_1\vec{r})} + E_2 e^{i(\omega t - \vec{\kappa}_2\vec{r})} \tag{2.10}$$

The experimentally available parameter is light intensity, which is proportional to the square of the amplitude of the electric field oscillations $I \sim E_0^2$, where E_0 is the amplitude of the oscillations,[10] and can be calculated as $I = EE^*$, where E^* is the complex conjugated number of E[11]

$$\begin{aligned} I &= E\cdot E^* = \left(E_1 e^{i(\omega t - \vec{\kappa}_1\vec{r})} + E_2 e^{i(\omega t - \vec{\kappa}_2\vec{r})}\right)\left(E_1 e^{-i(\omega t - \vec{\kappa}_1\vec{r})} + E_2 e^{-i(\omega t - \vec{\kappa}_2\vec{r})}\right) \\ &= E_1^2 + E_2^2 + E_1E_2\left(e^{i(\vec{\kappa}_1 - \vec{\kappa}_2)\vec{r}} + e^{-i(\vec{\kappa}_1 - \vec{\kappa}_2)\vec{r}}\right) \\ &= E_1^2 + E_2^2 + 2E_1E_2\cos\left(\left(\vec{\kappa}_1 - \vec{\kappa}_2\right)\vec{r}\right) \end{aligned} \tag{2.11}$$

The only variable term of I is $2E_1E_2cos\left(\left(\vec{\kappa}_1 - \vec{\kappa}_2\right)\vec{r}\right)$. It gives sinusoidal variation of the light intensity in direction $\vec{\kappa}_1 - \vec{\kappa}_2$. The period of the modulation in this direction is $\frac{2\pi}{|\vec{\kappa}_1 - \vec{\kappa}_2|}$.

[9] In a general case one of the waves may include a phase argument, e. g. $\vec{E_2}e^{\imath(\omega t - \vec{\kappa}_1\vec{r} + \varphi)}$. This, however, will not change the following consideration, so one can safely assume here that $\varphi = 0$.

[10] As was noted in Section 2.1.2 and footnote 4, the power density is $I = \left\langle \vec{S} \right\rangle = \frac{cc}{2}E_0^2$. However, the coefficient $\frac{cc}{2}$ will be omitted in all further calculations and a simple relation, $I = E_0^2$, will be used. This will not create any mistake as in all cases two transitions will be made: from the intensity of the individual beams to the electric field and then back to the intensity of the interference pattern.

[11] If $A = ae^{\imath x}$, then (by definition) $A^* = ae^{-\imath x}$. Thus, $A\cdot A^* = ae^{\imath x}ae^{-\imath x} = a^2$.

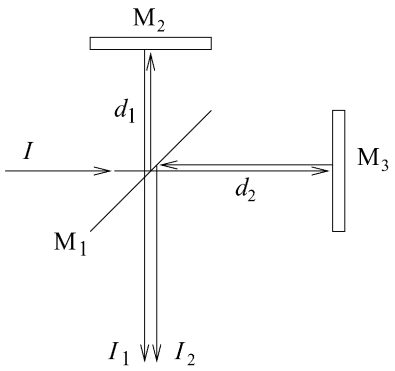

Figure 2.2: Michelson interferometer. M_1 is a semi-transparent mirror, and M_2 and M_3 are 100% reflectors.

Thus, using scalar wavelength $\lambda = \frac{2\pi}{|\vec{\kappa}_1|} = \frac{2\pi}{|\vec{\kappa}_2|}$ instead of wave vectors and introducing the angle between the propagation directions of the waves, α, one can obtain interference period as $L = \frac{\lambda}{2\sin(\alpha/2)}$.

Since $E_1^2 + E_2^2 \geq 2E_1E_2$, the intensity I is never negative. When the waves have the same amplitudes, $E_1 = E_2 = E$, the intensity of the interference pattern changes from 0 to its maximum value of $4E^2$, which is two times greater than the sum of the intensities of the interfering waves, $2E^2$.

Example 2.1: *Interference period of two monochromatic waves.* In order to obtain an interference pattern of two monochromatic plane wave at λ =500 nm (green light) with period of $L = 1$ mm, the angle between the wave propagation directions must be $\alpha = 2\sin^{-1}\left(\frac{\lambda}{2L}\right) \simeq \frac{\lambda}{L} = 0.0005$ radian or $\approx 0.03°$.

There are many optical devices utilizing the phenomenon of wave interference. Such devices are called interferometers. Two types of the interferometers are particularly important for spectroscopy applications and will be discussed here. These are Michelson and Fabry-Perot interferometers.

2.2.1 Michelson interferometer

Michelson interferometer has found numerous applications and was reproduced with multiple modifications. A classic scheme of the interferometer is shown in Fig. 2.2. It consists of three mirrors: a semi-transparent mirror M_1 and two reflectors M_2 and M_3. If incoming beam has intensity I and the mirror M_1 has reflectance R, then the intensity of the reflected beam is RI and the intensity of the transmitted light is $(1 - R)I$, respectively.[12] Let us follow the propagation of the reflected beam first. The mirror M_2, must be adjusted so that

[12]Transmittance of the mirror M_1 is $T = 1 - R$.

the beam reflected by the mirror M_1 is returned back by exactly the same path. Then the reflected beam will hit the semi-transparent mirror M_1 at exactly the same point as incoming beam (I). The intensity of the light, which will cross the mirror M_1, is $I_1 = R(1-R)I$. This is the first beam on the interferometer output. Now let us follow the propagation of the beam which is transmitted by the mirror M_1 at the first incidence of the incoming beam (I). Its intensity after the mirror M_1 is $(1-R)I$. The mirror M_3 must be adjusted so that reflected beam hits the mirror M_1 at exactly the same point as incoming beam (I). Part of the beam will be reflected in the same direction as previously considered beam I_1, and its intensity is $I_2 = R(1-R)I$. Thus, properly adjusted Michelson interferometer splits incoming beam (I) on two beams (I_1 and I_2) of equal intensities and propagating in the same direction.

In order to calculate the resulting output intensity of the interferometer one needs to sum the electric fields created by two beams and find the light intensity for the resulting field. Considering a monochromatic light and taking into account that the beams are propagating in the same direction, e. g. along Z axis, the fields can be written as $E_1(z,t) = E_{out}e^{i(\omega t-\kappa z-\varphi_1)}$ and $E_2(z,t) = E_{out}e^{i(\omega t-\kappa z-\varphi_2)}$, respectively, where E_{out} is the field created by one of the beams (on the interferometer output). The phases φ_1 and φ_2 depend on the propagation distances of the beam from the semi-transparent mirror M_1 to the reflectors (M_2 or M_3, respectively) and back, and can be written as $\varphi_1 = 2\kappa d_1$ and $\varphi_2 = 2\kappa d_2$, where the multiplier 2 is due to the fact that each beam travels twice the distance from the semi-transparent mirror to the corresponding reflector (M_2 or M_3). Thus, for the electric field of the interferometer output one can write

$$\begin{aligned} E &= E_{out}e^{i(\omega t-\kappa z-2\kappa d_1)} + E_{out}e^{i(\omega t-\kappa z-2\kappa d_2)} \\ &= E_{out}e^{i(\omega t-\kappa z)}\left(e^{-2i\kappa d_1} + e^{-2i\kappa d_2}\right) \end{aligned} \tag{2.12}$$

And the light intensity is

$$\begin{aligned} I_{out} &= E\cdot E^* \\ &= E_{out}^2 e^{i(\omega t-\kappa z)}\left(e^{-2i\kappa d_1} + e^{-2i\kappa d_2}\right)e^{-i(\omega t-\kappa z)}\left(e^{2i\kappa d_1} + e^{2i\kappa d_2}\right) \\ &= E_{out}^2\left(2 + e^{-2i\kappa(d_1-d_2)} + e^{2i\kappa(d_1-d_2)}\right) \\ &= 2E_{out}^2\left(1+\cos 2\kappa(d_1-d_2)\right) \end{aligned} \tag{2.13}$$

Converting the wave number κ to the wavelength one obtains

$$I_{out} = 2E_{out}^2\left(1+\cos\frac{4\pi(d_1-d_2)}{\lambda}\right) \tag{2.14}$$

Finally, taking into account the reflectance of the mirror M_1

$$I_{out} = 2I_{in}R(1-R)\left(1+\cos\frac{4\pi(d_1-d_2)}{\lambda}\right) \tag{2.15}$$

Thus, the output intensity depends on relative beam propagation delay $\frac{d_1-d_2}{\lambda}$, and varies from 0 to $4R(1-R)I_{in}$.

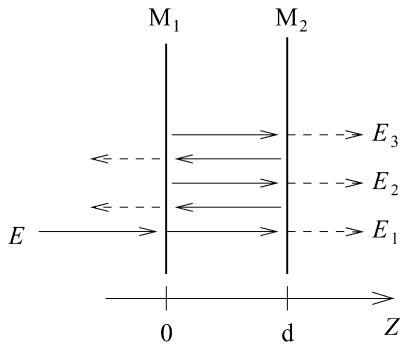

Figure 2.3: Fabry-Perot interferometer

A straightforward application for the Michelson interferometer is direct measurement of the wavelength of monochromatic light. By smooth changing of the distance d_1(or d_2) and counting the interference maxima, which comes as cosine function of the distance, eq. (2.13), one can determine the wave number as number of maxima per unit length,[13] and the wavelength as inverse of the wave number.

A short list of the Michelson interferometer applications in the optical spectroscopy application includes:

- wavelength determination;
- measurements of the light coherence length (the interference pattern can be observed only for coherent beams I_1 and I_2);
- optics diagnostics (an optical component, e. g. a lens, can be inserted between mirrors M_1 and M_2 and any distortions of the wavefront will be seen in distortions of the interference pattern on the interferometer output);
- fine displacement measurements;
- optical correlators (ultra-short pulse width measurements, will be considered in Chapter 4.5.2);
- Fourier transform infrared spectroscopy;

2.2.2 Fabry-Perot interferometer

Fabry-Perot interferometer is formed by a pair of mirrors aligned parallel to each other at some (short) distance, d, as presented in Fig. 2.3. For simplicity we will consider normal

[13]Note that circular frequency was used in eqs. (2.10)-(2.13). In turn, eq. (2.14) was rewritten for "linear units", and corresponding form for the wave number is $I_{out} = 2E_{out}^2 (1 + \cos 4\pi k(d_1 - d_2))$, where k is the wave number, and $\kappa = 2\pi k$. See also footnotes 4 and 5.

incidence of the light and we will suppose that incoming light intensity is 1, i. e. $E_{in} = e^{i(\omega t - \kappa z)}$. For the further simplification, let us suppose that the mirrors have the same reflection coefficient r for the electric field flow, thus the intensity reflection is $R = r^2$. The corresponding transmittance for the electric field component of the wave is $f = \sqrt{1 - r^2}$.

The interference pattern after the interferometer is formed by multiple reflections of the incoming beam between the mirrors. The electric field created by the incoming plane wave right before the mirror M_1 is $E = e^{i(\omega t - \kappa z)} = e^{i\omega t}$ ($z = 0$). Right after the mirror M_1 the electric field of the incident light is $fe^{i\omega t}$, and before the mirror M_2 is $fe^{i(\omega t - \kappa d)}$ ($z = d$). After the mirror M_2 the field is

$$E_1 = ffe^{i(\omega t - \kappa d)} = f^2 e^{i(\omega t - \kappa d)} \tag{2.16}$$

This is the first beam participating in the interference on the interferometer output.

The part of the light, reflected by the mirror M_2, $fre^{i(\omega t - \kappa d)}$, returns back to the mirror M_1, where another portion of the light is reflected in direction to the mirror M_2. The field of the reflected light is $fr^2 e^{i(\omega t - 2\kappa d)}$. Part of this beam will cross mirror M_2 and form the second beam participating in the interference,

$$E_2 = frrfe^{i(\omega t - 3\kappa d)} = f^2 r^2 e^{i(\omega t - 3\kappa d)} = E_1 r^2 e^{-i2\kappa d} \tag{2.17}$$

This re-reflection process will continue again and again giving beams E_3, E_4, ... E_n and so on. It is clear, that for the beam n the electric field is

$$E_n = E_1 r^{2(n-1)} e^{-i2\kappa d(n-1)} \tag{2.18}$$

The resulting electric field after the mirror M_2 is the sum of all the beams

$$\begin{aligned} E_{out} &= \sum_{n=1}^{\infty} E_n = E_1 \sum_{n=1}^{\infty} r^{2(n-1)} e^{-i2\kappa d(n-1)} = E_1 \sum_{n=1}^{\infty} \left(r^2 e^{-i2\kappa d}\right)^{n-1} \\ &= E_1 \sum_{n=0}^{\infty} \left(r^2 e^{-i2\kappa d}\right)^{n} = \frac{E_1}{1 - r^2 e^{-i2\kappa d}} \end{aligned} \tag{2.19}$$

Finally, the intensity of the transmitted beam is[14]

$$I_{out} = \langle E_{out} E_{out}^* \rangle = \frac{(1 - r^2)^2}{(1 - r^2)^2 + 4r^2 \sin^2 \kappa d} \tag{2.20}$$

Converting eq. (2.20) to intensity reflection, $R = r^2$, and wavelength, $\kappa = \frac{2\pi}{\lambda}$, one obtains

$$I_{out} = I_{in} \frac{(1 - R)^2}{(1 - R)^2 + 4R \sin^2 \frac{2\pi d}{\lambda}} \tag{2.21}$$

The transmittance, $T = I_{out}/I_{in}$, of the interferometer in a narrow wavelength range is shown in Fig. 2.4 for $d = 0.01$ mm and $R = 0.5$ and 0.9.

[14] Calculations of the intensity from the electric field eq. (2.19) can be found in e. g. ref. [1] p. 108.

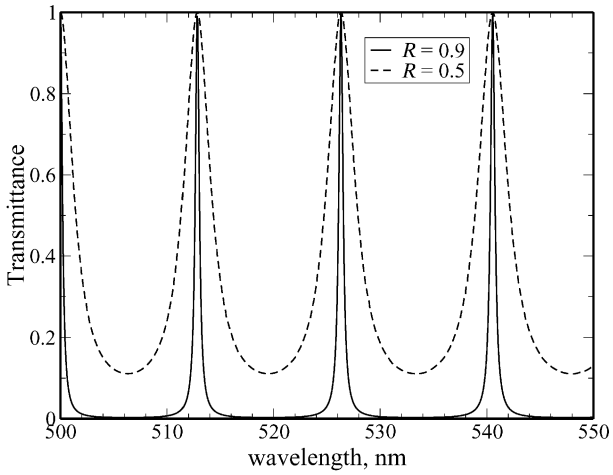

Figure 2.4: Transmittance spectrum of the Fabry-Perot interferometer of thickness $d = 0.01$ mm and formed by mirrors with reflectance $R = 0.5$ (dashed line) and 0.9 (solid line). Equation (2.21) was used for the calculations.

When $\frac{2d}{\lambda} = N$, where N is an integer number (0, 1, 2, ...), $\sin^2 \frac{2\pi d}{\lambda} = 0$ and $I_{out} = I_{in}$, i. e. the light crosses the interferometer without any decrease in the intensity even when the interferometer in formed by two mirrors with high reflectance. Therefore the transmittance spectrum consists of sharp lines at wavelengths satisfying condition $2d = N\lambda$. If there is a transmittance maximum at λ_0, then $N = \frac{2d}{\lambda}$ and the next maximum will be at λ_1 which correspond to $N - 1$. Thus the spectrum distance between maxima is

$$\Delta\lambda_{sp} = \frac{\lambda_0^2}{2d - \lambda_0} \simeq \frac{\lambda_0^2}{2d} \tag{2.22}$$

The spacing between the lines in the wavelength domain decreases as distance between the mirrors increases.

When $\frac{2d}{\lambda} = N + \frac{1}{2}$, i. e. $\sin^2 \frac{2\pi d}{\lambda} = 1$, the transmittance of the interferometer has its minimum value

$$I_{min} = I_{in} \frac{(1-R)^2}{(1+R)^2} \tag{2.23}$$

For example, if $R = 0.5$, then $I_{min} \simeq 0.11 I_{in}$, or the light rejection is higher than what could be expected for two "independent" mirrors, $R^2 = 0.25$.

One of the applications of the Fabry-Perot interferometers in optical spectroscopy is the fine spectrum resolution. Then the value $\Delta\lambda_{sp}$ (eq. (2.22)) can be treated as the spectrum range of the Fabry-Perot interferometer, meaning that if the studied light has wider spectrum the resulting pattern will be overlapped of different spectral parts.

For spectroscopy applications the interferometer is placed on the way of a plane wave front and fine tuning of the interferometer transmittance wavelength is achieved by turning slightly the interferometer. When the light incidence angle, α, is not zero (at normal incidence $\alpha = 0$) eq. (2.21) can be used after substitution $d = h(\cos\alpha - \sin^2\alpha)$ where h is the distance between the mirrors.[15] Thus the transmittance maxima will be at $\lambda = \frac{2h}{N}(\cos\alpha - \sin^2\alpha) = \lambda_0(\cos\alpha - \sin^2\alpha)$. For a small angle α one can use approximations $\cos\alpha \approx 1 - \frac{\alpha^2}{2}$ and $\sin^2\alpha \approx \alpha^2$, so

$$\lambda(\alpha) \approx \lambda_0(1 + \frac{1}{2}\alpha^2) \tag{2.24}$$

At angle $\alpha \simeq \sqrt{\frac{2\Delta\lambda_{sp}}{\lambda_0}}$ the interferometer will be again transparent to light at wavelength λ_0 (at $N + 1$).

For the purpose of spectrum resolution analysis one can introduce a contrast factor $F = \frac{4R}{(1-R)^2}$ and a dimensionless value $\varphi = \frac{2\pi d}{\lambda}$. Then, eq. (2.21) can be rewritten as

$$I_{out} = I_{in}\frac{1}{1 + F\sin^2\varphi} \tag{2.25}$$

The half intensity bandwidth can be determined from condition $I_{out} = \frac{1}{2}I_{in}$, which results in equation

$$1 + F\sin^2\varphi = 2 \tag{2.26}$$

or

$$\sin\varphi = \sqrt{\frac{1}{F}} \tag{2.27}$$

Usually, F is a big value, e. g. $F = 360$ at $R = 0.9$, and $F = 1520$ at $R = 0.95$. Therefore, one can use approximation

$$\varphi \approx \sqrt{\frac{1}{F}} \tag{2.28}$$

which is the equation to be solved in order to evaluate the spectrum resolution of the interferometer.

Interferometers for the fine spectrum resolution are usually constructed so that $d \gg \lambda$. This means that $d \gg 1\mu$ in the optical wavelength range. Then, considering a small deviation of the wavelength from the wavelength of the maximum transmittance, λ_0, one obtains

$$\varphi = \frac{2\pi d}{\lambda_0 - \triangle\lambda} \simeq \frac{2\pi d}{\lambda_0}\left(1 + \frac{\triangle\lambda}{\lambda_0}\right) = \frac{2\pi d}{\lambda_0} + \frac{2\pi d}{\lambda_0^2}\triangle\lambda \tag{2.29}$$

[15] One have to account for the traveling distance between mirrors ($\cos\alpha$) and for the phase shift due to the fact that the wave front is not parallel to the mirror surface ($\sin^2\alpha$).

where $\triangle\lambda$ is the deviation from maximum. Since λ_0 is the wavelength of the transmittance maximum $\frac{2\pi d}{\lambda_0} = \pi N$ ($\sin\frac{2\pi d}{\lambda_0} = 0$) and using eq. (2.28) one can write

$$\frac{2\pi d}{\lambda_0^2}\triangle\lambda \approx \sqrt{\frac{1}{F}} \tag{2.30}$$

Finally, the wavelength resolution is

$$\triangle\lambda \approx \frac{\lambda_0^2}{2\pi d}\sqrt{\frac{1}{F}} \tag{2.31}$$

Example 2.2: *Spectrum resolution of Fabry-Perot interferometer.* Suppose the interferometer is formed by pair of mirrors with reflectance $R = 0.95$ and placed at a distance $d = 1$ mm from each other. The contrast factor of the interferometer is $F = \frac{4R}{(1-R)^2} \approx 1\,520$. At wavelength $\lambda_0 = 500$ nm the spectrum resolution of the interferometer is $\triangle\lambda = \frac{\lambda_0^2}{2\pi d}\sqrt{\frac{1}{F}} \approx 10^{-3}$ nm. The spectrum range is $\Delta\lambda_{sp} \simeq \frac{\lambda_0^2}{2d} = 0.125$ nm.[16] Adjustment angle $\alpha \simeq \sqrt{\frac{2\Delta\lambda_{sp}}{\lambda_0}} \approx 0.02$ radian or $1.3°$. In order to increase the resolution mirrors with higher reflectance can be used, e. g. if $R = 0.98$, then $F = 39\,600$, and $\triangle\lambda \approx 2 \times 10^{-4}$ nm.

2.2.3 Interference filters and mirrors

All of the above considerations can be applied to any combination of parallel light reflecting surfaces. For example, interface between two media with different refractive indexes can be used as the mirror. Although reflection of such interface is not large,[17] one can build a system with multiple reflecting interfaces and place the reflecting surfaces at a distances satisfying the best reflection or transmission conditions at a certain wavelength, thus achieving a high reflectance or high transmittance. One of the applications of this type of structures is the bandpass filters, when the system is designed to be transparent at a certain wavelength range. Alternatively, the structure can be designed to achieve high reflectance in a certain range and can be used as high reflectance mirror. This type of mirrors are commonly called dielectric mirrors. Yet another application is anti-reflecting coating of e. g. lens surfaces.

A big advantage of these systems is that they are made of materials which do not absorb the light – the light is either reflected or transmitted. Therefore they can work at extremely high light power. This is practically important for laser applications where the light peak power or pulse energy can be extremely high (see Chapter 3).

[16] The spectrum width of studied signal must be narrower than $\Delta\lambda_{sp}$, otherwise different parts of the spectrum will overlap each other.

[17] If the refraction index of the medium at one side of the interface is n_1 and at the another is n_2, the reflection from the interface is $R = \frac{(n_1-n_2)^2}{(n_1+n_2)^2}$. Roughly, the refractive indexes of the materials transparent in the visible wavelength range (400-800 nm) vary from 1.35 (cryolite, Na_3AlF_6) to 2.3 (titanium dioxide, TiO_2), thus the maximum reflection from the interface is $R \leq 0.07$.

2.3 Diffraction

Restriction of the plane wave front in space results in distortion of the front. This phenomenon is called diffraction. The diffraction limits the spot size to which the light can be focused. Also the beam presentations can be used only in a limited distance where the change in the beam diameter can be neglected. These are probably two most important implications for optical schemes design.

To explain the diffraction Dutch scientist Christiaan Huygens has proposed a wave theory in 1670. He postulated that each point on a wave front can be treated as a source of a spherical wave called a secondary wave or wavelet. The envelope of those waves, at the same time, is constructed by finding the tangent to the waves. The envelop is assumed to be the new position of the wave [1].

The mathematical treatments of the diffraction were developed later with important contributions by Augustin Fresnel and Joseph Fraunhofer.

2.3.1 Fresnel formulation

Using Huygens principle Fresnel has developed a theory which allows to calculate the electric field amplitudes at any point in space for a wave front defined by some limited surface S. The theory is applied to stationary harmonic waves. The electric field at some point in space pointed by vector $\vec{r}_o$ (observation point) is given by Fresnel integral

$$E(\vec{r}_o) = \int_S C(\vec{r}_s) \frac{E_s(\vec{r}_s)}{|\vec{R}|} e^{i(\vec{\kappa}\cdot\vec{R})} ds \tag{2.32}$$

where $E_s(\vec{r}_s)$ is the electric field at a point on surface S given by vector $\vec{r}_s$, $\vec{R} = \vec{r}_o - \vec{r}_s$, $C(\vec{r}_s)$ is the obliquity factor,[18] and integration is done over the whole surface S. This is a general approach when calculating diffraction of any waves.

2.3.2 Fraunhofer diffraction (far field approximation)

In Fraunhofer or far field approximation the vector $\vec{R}$ is supposed to be much greater than the size of the wave front S. To simplify the problem let us consider plane wave propagating through an aperture, so that the surface S is a plane and it is limited in space. Then, let the plane S be XY plane of the coordinate system. The wave is propagating in almost Z direction and we are interested in the deviation from this direction caused by the diffraction. The deviation is given by function $E = E(\kappa_x, \kappa_y)$, since if the wave propagates in Z direction only (as it takes place at S) then $\kappa_x = 0$ and $\kappa_y = 0$, i. e. $E(\kappa_x, \kappa_y) = 0$ when $\kappa_x \neq 0$ or $\kappa_y \neq 0$, or $E = E_o\delta(\kappa_x, \kappa_y)$. After this simplification, the Fresnel integral (2.32) can be rewritten as

$$E(\kappa_x, \kappa_y) = \int\int E(x, y) e^{i(\kappa_x x + \kappa_y y)} dxdy \tag{2.33}$$

[18] The obliquity factor depends on orientations of the surface S and observation point $\vec{r}_o$. It was later derived by Gustav Kirchhoff, therefore integral (2.32) is also called Fresnel-Kirchhoff integral.

The most simple example of the case is a diffraction of a wave on a slit. If the slit is oriented in Y direction, we can consider only X plane, which gives

$$E(k_x) = \int_{-d}^{+d} e^{i\kappa_x x} dx = -i\frac{e^{i\kappa_x d}}{\kappa_x} \tag{2.34}$$

where $2d$ is the size of the slit and the amplitude of the wave at the slit is supposed to be unity. Taking the real part of the field amplitude (2.34) one obtains

$$E = Re\left(E(\kappa_x)\right) = \frac{\sin(\kappa_x d)}{\kappa_x} \tag{2.35}$$

The projection of the wave vector, κ_x, can be expressed in terms of the observation angle φ, which is deviation from Z axis, and the wave number κ as $\kappa_x = \kappa \sin\varphi = \frac{2\pi}{\lambda}\sin\varphi$. Thus eq. (2.35) gives the wavelength and angular dependence of the amplitude of the diffracted wave

$$E = \frac{\sin\left(\frac{2\pi d \sin\varphi}{\lambda}\right)}{\left(\frac{2\pi \sin\varphi}{\lambda}\right)} \tag{2.36}$$

The light intensity is proportional to the square of the electric field amplitude,

$$I = E^2 = \frac{\sin^2(\kappa_x d)}{\kappa_x^2} = \frac{\sin^2(\frac{2\pi d \sin\varphi}{\lambda})}{\left(\frac{2\pi \sin\varphi}{\lambda}\right)^2} \tag{2.37}$$

Both these functions, eqs. (2.36) and (2.37), are pulse-like functions. The angular dependence of the intensity is shown in Fig. 2.5. The functions have maxima at $\kappa_x = 0$, i. e. $\varphi = 0$, and decrease (with some oscillations) with increase or decrease of κ_x, respectively. The width of the "pulse" can be determined roughly from condition (which gives the first intensity minimum, $I = 0$) $\kappa_x d = \pi$ or $\kappa_x = \frac{\pi}{d}$ or

$$\sin\varphi = \frac{\lambda}{2d} \tag{2.38}$$

This result has an important consequence. A limited in space plane wave cannot propagate as a unidirectional beam. It will have divergence angle given by[19]

$$\sin\varphi \approx \varphi \approx \frac{\lambda}{D} \tag{2.39}$$

where D is the cross-section (aperture) of the diaphragm used to form the beam from "infinite" plane wave and it was assumed that the cross section of the beam is much greater than the wavelength, $D \gg \lambda$. Consequently, even "ideal" lens cannot collect light into a spot smaller than certain size, which is called diffraction limit.

[19] Strictly speaking, the divergence angle of the light after the slit measured as full width at half maximum is $\varphi \simeq \frac{\lambda}{2D}$. However for a round diaphragm the divergence is slightly greater. Therefore relation $\varphi \approx \frac{\lambda}{D}$ is reasonably accurate for a rough estimations of the diffraction effects.

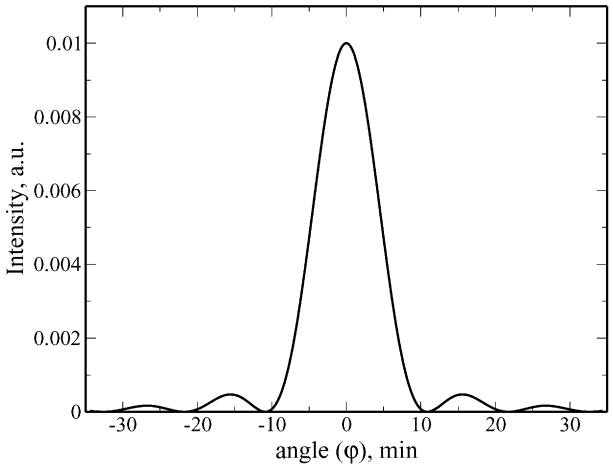

Figure 2.5: Angular distribution of the diffracted light (at $\lambda = 633$ nm) after 0.2 mm slit ($d = 0.1$ mm) calculated using eq. (2.37).

Example 2.3: *Diffraction limit of beam divergence.* Suppose a $D = 0.5$ mm diaphragm is used to form a beam from a plane wave front at the wavelength $\lambda = 500$ nm. The divergence angle of that beam after the diaphragm is $\varphi \approx \frac{\lambda}{D} = 10^{-3}$. At a distance $l = 1$ m from the diaphragm the diffraction will result in the spot size of $l\varphi \approx 1$ mm, which is reasonably close to the size of the diaphragm. However, at 10 m distance the diffraction spot will be 1 cm in diameter, which is essentially larger than the initial beam diameter. In other words, 0.5 mm beam keeps its cross section at distances shorter than 1 m, but at longer distances the diffraction beam spreading becomes essential.

Example 2.4: *Diffraction limit of beam focusing.* Suppose a light at $\lambda = 500$ nm is focused by a lens with focal length $f = 10$ cm and aperture $D = 1$ cm. What is the smallest possible spot size produced by such lens? The diffraction on the lens aperture will be at angle $\varphi \approx \frac{\lambda}{D}$. This divergence gives a spot in focal plane of size $d = f\varphi \approx \lambda\frac{f}{D} = 5\,\mu$. In the best case the spot size will be 10 times greater than the wavelength.

2.3.3 Diffraction grating

Diffraction grating is an optical component that is used to spread light into a spectrum. Typically diffraction grating is a mirror with many thousands of parallel lines, grooves, etched on its surface. The lines must be at one and the same intervals, called grating period. A flat incident wave front is reflected from the grating at a number of angles determined by the

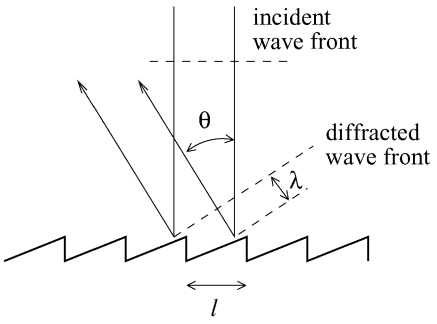

Figure 2.6: Diffraction grating with period l. The first order diffraction angle appears when the delay between reflections from the neighbor grooves is equal to one wavelength, λ.

grating period and the wavelength. The diffraction angles are determined by the condition of phase matching for the reflections form adjacent grooves. A simple illustration in Fig. 2.6 assumes a normal light incidence and one wavelength shift between the reflections. The angular positions of the diffraction maxima are

$$\sin\theta = \frac{m\lambda}{l} \tag{2.40}$$

where l is the grating period and m is an integer number (1, 2, ...), called diffraction order. Naturally, in eq. (2.40) $\frac{m\lambda}{l}$ must be less than one, which means that the maximum possible diffraction order is $m < \frac{l}{\lambda}$.

The diffraction angle can be also expressed in terms of grooves number, which is defined as reciprocal of the grating period, $g = l^{-1}$,

$$\sin\theta = m\lambda g \tag{2.41}$$

The grooves number is the parameter usually found in grating specification, and measured in number of grooves per millimeter, i. e. mm^{-1}.

In spectroscopy applications the gratings are used to spread the light by the wavelength, and thus to measure the light intensity wavelength dependence. Therefore the wavelength resolution is the parameter of interest. In a far field approximation the wavelength resolution is determined by two factors: the wavelength divergence due to diffraction as given by eq. (2.40) and the diffraction divergence due to the limited size of the wave front. The angular resolution is $\triangle\theta \approx \frac{\lambda}{L\cos\theta}$ [1], where L is the length of the illuminated area (i. e. the size of the wave front). Equation (2.40) gives $\triangle\theta\cos\theta = \frac{m\triangle\lambda}{l}$, thus, the wavelength resolution is

$$\frac{\triangle\lambda}{\lambda} \approx \frac{l}{mL} \tag{2.42}$$

In order to improve the wavelength resolution one can increase illuminated area L, which is done by using bigger gratings, or use higher diffraction orders.

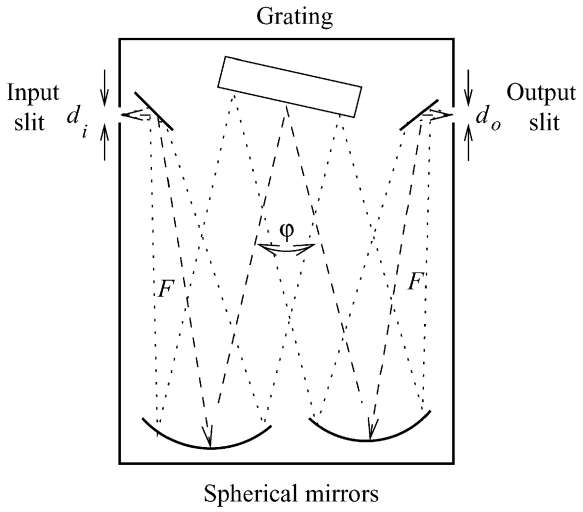

Figure 2.7: Monochromator optical scheme. d_i and d_o are the sizes of input and output slits, respectively, f is the focal length of the mirrors, and φ is the angle between the incident and diffracted beams.

Example 2.5: *Spectrum resolution of diffraction grating.* Typical grooves number for gratings designed for the visible–ultraviolet wavelength range is $g = 1200\,\mathrm{mm}^{-1}$, which corresponds to grating period $l = g^{-1} \approx 0.8\,\mu$. Such gratings work in the first diffraction order. If the size of the grating is $L = 5$ cm, the best possible spectrum resolution of the grating at $\lambda = 500$ nm is $\Delta\lambda \approx \lambda\frac{l}{L} \approx 0.008$ nm.

2.3.4 Monochromator

Monochromator is an optical device which works as narrow band wavelength filter with mechanically adjustable transmission wavelength. A typical optical scheme of monochromator is presented in Fig. 2.7. The incoming light crosses the input slit of size d_i and then it is collected by a spherical mirror of focal length F placed at distance F from the entrance slit. After the mirror a flat wave front is formed and directed to the grating. The diffracted light is collected by the second mirror and focused to the output slit of size d_o. Turning the grating one can change the wavelength which will hit the output slit. In geometry presented in Fig. 2.7, the equation of the light transmission wavelength is

$$l(\sin\alpha + \sin(\varphi + \alpha)) = m\lambda \tag{2.43}$$

where α is the incident angle of the light on the grating and φ is the angle between the incident and diffracted beams (this angle is fixed by the instrument geometry). Using grooves

number the same equation reads

$$\sin\alpha + \sin(\varphi + \alpha) = m\lambda g \tag{2.44}$$

The right side of eq. (2.44) can be simplified if the light incidence angle α is close to 0 (normal incidence of the light on the grating)

$$\sin\alpha + \sin(\varphi + \alpha) = \sin\alpha + \sin\alpha\cos\varphi + \cos\alpha\sin\varphi =$$

$$= \sin\alpha(1 + \cos\varphi) + \cos\alpha\sin\varphi \simeq \alpha(1 + \cos\varphi) + \sin\varphi$$

and

$$\lambda = \alpha\frac{(1 + \cos\varphi)}{mg} + \frac{\sin\varphi}{mg} \tag{2.45}$$

This is the monochromator dispersion equation.[20]

The spectrum resolution of the monochromators is usually determined by the slit sizes, which determines the divergence of the wave formed by the spherical mirror $\triangle\alpha = \frac{d_i}{F}$. In other words, the wave formed after the spherical mirror is not flat and the diffraction limit of the grating spectrum resolution cannot be achieved in most monochromator applications. Then, the spectrum resolution of the monochromator is

$$\triangle\lambda = \frac{d_i(1 + \cos\varphi)}{Fmg} \tag{2.46}$$

As can be seen, to achieve better spectrum resolution one can

1. use smaller slits (decrease d); this is the regular method to change the monochromator resolution, however smaller entrance slit means usually that smaller amount of light will enter the monochromator, and at slits size approaching the wavelength the diffraction on the slit reduces the monochromator efficiency gradually;

2. use grating with higher grooves number (increase g), the wavelength is the limit for grooves number, in the visible part of spectrum the practical limit is 1200 mm^{-1};

3. work at higher diffraction orders (increase m), then there may be overlapping diffraction orders and the diffraction order cannot be greater than $m < (\lambda g)^{-1}$, also with $g = 1200\ \text{mm}^{-1}$ the diffraction order cannot be higher than one in the visible part of the spectrum;

4. use mirrors with longer focal distance (increase F), this increases physical dimensions of the device, i. e. usually bigger monochromators have better spectrum resolution.

[20] Note the linear dependence of the wavelength on the angle α. The linear relation between the grating angle and transmission wavelength explains why the wavelength scale is so common for spectroscopy devices.

An important monochromator parameter is its angular aperture, which is given by the ratio $\frac{R}{F}$, where R is radius of the mirror (mirror size). It determines the maximum deviation of the in-coming beam from the optical axes at the entrance slit. A bigger angular aperture allows to collect more light and is, generally, preferable. However, bigger ratio $\frac{R}{F}$ means bigger aberrations and, therefore, may reduce the wavelength resolution.

Example 2.6: *HR320 monochromator (ISA Inc.).* The HR320 monochromator is an example of a compact monochromator which can be used in many optical spectroscopy applications. The monochromator utilizes 32 cm focal mirrors, and in the visible wavelength range is equipped with $g = 1200$ mm^{-1} grating. It provides the spectrum resolution of 0.05 nm at 0.01 mm slits. The stray light rejection at 8 nm shift from the monochromatic wave is 10^{-5}, which is typical for single grating monochromators. Overall dimensions of the monochromator are $40 \times 34 \times 25$ cm^3.

2.4 Calculations of optical system (matrix formulation)

2.4.1 Geometrical optics approximation

In this section we will consider an approximation of geometrical optics, which deals with essentially flat waves with slowly changing amplitudes, but instead of operating with wave fronts deals with beam and beam trajectories. An example of the device producing a beam can be a laser. Also one can use a lamp with collimator and a diaphragm to obtain a beam.

Three important assumption for transition from the wave light presentation to the geometrical optics are:

1. all dimensions are essentially greater than the wavelength (thus only wave amplitudes are considered),
2. the waves are essentially plane waves,
3. travel distances of the waves are much greater than the wave cross sections.

This allows to replace waves by beams and to discuss the light propagation in terms of beam trajectories. From a point of view of calculations of the light propagation the geometrical optics simplifies the case by omitting the time dependence, so that one needs to consider only the light intensities at different points in space ($I(x, y, z)$). The second simplification is that the wave front is much smaller than the travel distance, so that the only information left from the wave front is the propagation direction, but the size of the wave front itself can be neglected.[21] This means that the problem can be simplified to beam trajectory tracing rather than solving the problem of light intensity distribution in space.

Following the strategy of the beam trajectory tracing one can further simplify the calculations by assuming paraxial approximation, which adds:

[21] The extend to which the size of the wave front can be neglected was discussed in Section 2.3 and Example 2.3.

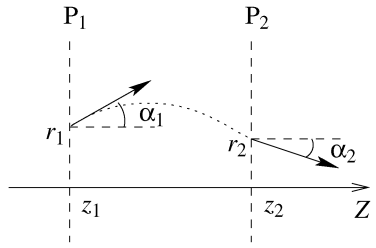

Figure 2.8: Paraxial approximation and beam propagation in optical system.

1. cylindrical symmetry;
2. propagation at sufficiently small angles to the symmetry axis, that it is possible to use approximation $\sin(\alpha) \simeq \alpha$.

In an optical system whose symmetry axis is Z, a paraxial beam at plane z (z =constant) is described by two parameters: the distance from the axis r and the angle it makes with the axis α (see Fig. 2.8). As the result, the whole diversity of the wave theory, as given e. g. by Helmholtz equation (2.9), can be solved in terms of only two parameters, r and α, in geometrical optics paraxial approximation.

2.4.2 Beam transfer matrix

Relation between the parameters at two planes, say planes P_1 ($z = z_1$) and P_2 ($z = z_2$) as shown in Fig. 2.8, is given by a linear system

$$\begin{aligned} r_2 &= Ar_1 + B\alpha_1 \\ \alpha_2 &= Cr_1 + D\alpha_1 \end{aligned} \tag{2.47}$$

or, in the matrix form

$$\begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix} \tag{2.48}$$

or

$$R_2 = MR_1 \tag{2.49}$$

where

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{2.50}$$

is the beam transfer matrix (it is also called ABCD matrix), and $R_1 = \begin{pmatrix} r_1 \\ \alpha_1 \end{pmatrix}$ and $R_2 = \begin{pmatrix} r_2 \\ \alpha_2 \end{pmatrix}$ are the beam parameter vectors.

Determinant of the matrix M is unity, i. e. $\det(M) = AD - CB = 1$, if the media to the left of the input plane and to the right to the output plane have the same refractive indexes. Otherwise $\det(M) = \frac{n_1}{n_2}$, where n_1 is the refractive index to the left of the input plane and n_2 is the refractive index to the right to the output plane.

Simplicity of eq. (2.49) can be extended to an optical system of any complexity. Let us consider a beam propagating in a complex optical system from plane z_1 to plane z_2, then from plane z_2 to plane z_3 and so on to plane z_{n+1}.

$$\begin{array}{ccccccccc} input & & & & & & & & output \\ \longrightarrow & | & \longrightarrow & | & \rightarrow \cdots \rightarrow & | & & \longrightarrow & \\ & z_1 & & z_2 & & z_{n+1} & & & \end{array} \tag{2.51}$$

Let us suppose that the transfer matrices between the neighboring planes are known, i. e. matrices for transfers $R_{i+1} = M_i R_i$ are defined. Then, starting from the last transfer matrix at plane z_n,

$$R_{n+1} = M_n R_n \tag{2.52}$$

one can progress to the next left-side transfer at plane z_{n-1}, which is

$$R_n = M_{n-1} R_{n-1} \tag{2.53}$$

Substituting R_n form eq. (2.53) to eq. (2.52)

$$R_{n+1} = M_n M_{n-1} R_{n-1} \tag{2.54}$$

Applying the same routine consequently one obtains the beam transfer equation from plane z_1 to plane z_{n+1}

$$R_{n+1} = M_n M_{n-1} \ldots M_1 R_1 \tag{2.55}$$

Note the order of matrices in the equation: the indexes increase from the right to the left.

Thus, to calculate beam propagation through an optical system, one needs to divide the propagation path into planes so that the transfer matrices are known for the neighboring planes and to calculate the product of the transfer matrices.

The most often used transfer matrices are collected in Table 2.1. Derivation of the most of these matrices is straightforward and can be done as a home exercise. Naturally, the matrices collected in Table 2.1 can be used to derive transfer matrices of more complex optical systems.

2.4.3 Imaging and magnification

Let us consider a typical imaging system consisting of a single lens. Let the focal length be f and the distance from the object to the lens be d. We need to find the distance from the lens to the object image, which will be denote as x. This optical system can be presented by four principal planes with known transfer matrices as shown in Fig. 2.9. Plane 1 (P_1) is the object plane; this is input plane. Plane 2 (P_2) is placed right behind the lens. Thus the first

Table 2.1: Transfer matrices of simple optical systems.

	matrix
Free space of length d	$\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$
Planar interface between two different media with refractive indexes n_1 and n_2	$\begin{pmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{pmatrix}$
Parallel-sided slab of length d and refractive index n	$\begin{pmatrix} 1 & \frac{d}{n} \\ 0 & 1 \end{pmatrix}$
Thin lens of focal distance f	$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}$
Spherical mirror of radius R	$\begin{pmatrix} 1 & 0 \\ -\frac{2}{R} & 1 \end{pmatrix}$

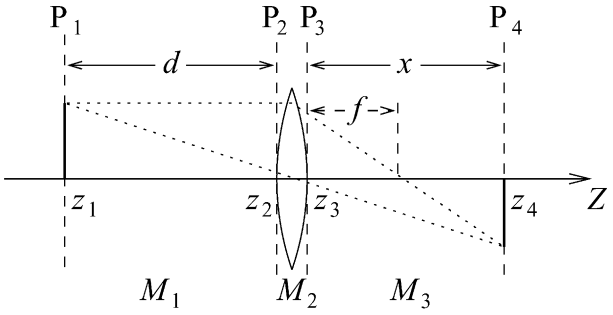

Figure 2.9: Ray tracing to obtain an image (in plane P_4) of an object (in plane P_1) and application of the matrix formalism to calculate the optical system: z_1, z_2, z_3 and z_4 are the principal planes and M_1, $M_2$and M_3 are corresponding transfer matrices.

transfer matrix is that for the free space of length d. Plane 3 (P_3) is the plane right after the lens, and the second transfer matrix is that of the thin lens. The last plane is the image plane or output plane. Consequently, the last transfer matrix is that of the free space of length x. The transfer matrix of the system is the product of three transfer matrices

$$\begin{aligned} M &= M_3 M_2 M_1 = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ -\frac{1}{f} & 1-\frac{d}{f} \end{pmatrix} \\ &= \begin{pmatrix} 1-\frac{x}{f} & d+x-\frac{dx}{f} \\ -\frac{1}{f} & 1-\frac{d}{f} \end{pmatrix} \end{aligned} \tag{2.56}$$

Having the transfer matrix for the optical system we shall answer the question: what does it mean that the plane P_4 is the image plane? The property of the image is that all the beams emitted from one point of the object will reach one and the same point at the image plane. In terms of the transfer matrix formalism this mean that the beam distance from the axis at the image plane, z_4, does not depend on the beam angle at the plane of object, α_1, that is $B = 0$ for matrix M (see eq. (2.50)). Thus, $B = d + x - \frac{dx}{f} = 0$, and solving this equation one obtains a well known formula $x = \left(\frac{1}{f} - \frac{1}{d}\right)^{-1}$. Finally, the transfer matrix from the object to the image plane is

$$M = \begin{pmatrix} \left(1 - \frac{d}{f}\right)^{-1} & 0 \\ -\frac{1}{f} & 1 - \frac{d}{f} \end{pmatrix}$$

The value A of the transfer matrix determines image magnification. This follows from the relation $r_4 = Ar_1 + B\alpha_1 = Ar_1$. Introducing magnification factor

$$m = \frac{r_4}{r_1} = A = \left(1 - \frac{d}{f}\right)^{-1} \tag{2.57}$$

one obtains

$$M = \begin{pmatrix} m & 0 \\ -\frac{1}{f} & \frac{1}{m} \end{pmatrix} \tag{2.58}$$

This is the general form of the transfer matrix of any imaging system.

The angular magnification of the imaging system is given by the element D of the matrix M, i. e. $m_\alpha = \frac{\alpha_4}{\alpha_1} = \frac{1}{m}$. Thus, angular magnification, m_α, is the inverse of the magnification m, thus $m_\alpha m = 1$. This is important and general result.[22]

Let us introduce brightness of an object (or image) as light power emitted by a surface of unit size in an unit angle.[23] If the object brightness is b, then one can calculate the brightness of image. The unit length of the object is converted to the length m of the image. The unit angle of the object is converted to the angle m_α of the image. Thus, the brightness of the image (accounting for the 2D case) is $b_i = m^2 m_\alpha^2 b$, and since $m_\alpha m = 1$, $b_i = b$. In other words, ideal optical imaging system does not change brightness of the object. If the object is magnified, the beams come to the image plane at angles smaller than those when they leave the object. And if the image is smaller than the object, the beams are focused at the image plane at angles greater than those when they leave the object.

In the end of this Chapter Example 2.7 discusses an estimation of the efficiency of the light collection from a lamp to a monochromator entrance slit to obtain a monochromatic light on the output of the system. This is a typical task in optical spectroscopy, e. g. to select the excitation wavelength.

[22] Actually, this comes directly from the fact that the determinant of the transfer matrix is unity, i. e. $AD - CB = 1$. Since $B = 0$, $AD = 1$, and thus $m_\alpha m = 1$.

[23] One can note similarity with the emittance introduced to characterize the black body emission in Section 1.2.1, see eq. (1.21). The spectrum integral of the emittance is equivalent to the brightness.

Example 2.7: *Estimation of the light collection efficiency.* Let us estimate a relative amount of the light emitted by a lamp which could be passed to a monochromator. The light is collected by a lens and focused onto the entrance slit of the monochromator. According to the conclusion made in this Section the brightness does not depend on the magnification of the image on the input slit. However there are two limiting factors: 1) if the linear magnification is too big, the image of the emitting source is bigger than the entrance slit thus a part of the focused light is lost, 2) if the light is focused to a very small spot the angular magnification is high and a part of the light can be lost because of limited angular aperture of the monochromator. Therefore, let us assume that the light is focused in a such way that the image fits inside the slits and the lens diameter is big enough to fill the monochromator angular aperture. Thus the light is collected in solid angle $\Omega \approx \pi \left(\frac{R}{2F}\right)^2$, where F is focal length and R is the radius of the monochromator mirror, see Section 2.3.4 and Fig. 2.7. Since the lamp emits in solid angle 4π the relative amount of the collected light, or the efficiency of coupling lamp with monochromator is $\eta_c \approx \frac{1}{4}\left(\frac{R}{2F}\right)^2 = \frac{1}{16}\left(\frac{R}{F}\right)^2$. If the monochromator angular aperture is $\frac{R}{F} = 0.3$, then $\eta_c \approx \frac{0.09}{16} \approx 0.006 = 0.6\%$.